Received: 22 July 2024 | Revised: 8 November 2024 | Accepted: 10 November 2024

DOI: 10.1002/mp.17695

RESEARCH ARTICLE

MEDICAL PHYSICS

Stable and discriminating OCT-derived radiomics features for predicting anti-VEGF treatment response in diabetic macular edema

Sudeshna Sil Kar[1] | **Hasan Cetin**[2] | **Sunil K. Srivastava**[2,3] | **Anant Madabhushi**[1,4] | **Justis P. Ehlers**[2,5]

[1]Department of Biomedical Engineering, Emory University, Atlanta, Georgia, USA

[2]The Tony and Leona Campane Center for Excellence in Image-Guided Surgery and Advanced Imaging Research, Cole Eye Institute, Cleveland Clinic, Cleveland, Ohio, USA

[3]Novartis Pharmaceuticals, East Hanover, New Jersey, USA

[4]Atlanta VA Medical Center, Atlanta, Georgia, USA

[5]Vitreoretinal Service, Cole Eye Institute, Cleveland Clinic, Cleveland, Ohio, USA

Correspondence
Justis P. Ehlers, Cole Eye Institute, Cleveland Clinic, 9500 Euclid Avenue/Desk i32, Cleveland, OH 44195, USA.
Email: ehlersj@ccf.org

Anant Madabhushi, Wallace H Coulter Department of Biomedical Engineering, Emory University and Georgia Institute of Technology, 1750 Haygood Drive, Suite 647, Research Career Scientist, Atlanta VA Medical Center, Atlanta, Georgia 30322, USA.
Email: anantm@emory.edu

Anant Madabhushi and Justis P. Ehlers are co-senior authors.

Funding information
NIH-NEI P30, Grant/Award Number: IP30EY025585; The Research to Prevent Blindness, Inc; Cleveland Eye Bank Foundation awarded to the Cole Eye Institute, Grant/Award Number: K23-EY022947-01A1; National Cancer Institute, Grant/Award Numbers: R01CA249992, R01CA202752, R01CA208236, R01CA216579, R01CA220581, R01CA257612, R01CA268207A1, U01CA239055, U01CA248226, U54CA254566; National Heart, Lung and Blood Institute, Grant/Award Numbers: R01HL151277, R01HL158071; National Institute of Biomedical Imaging and Bioengineering, Grant/Award Number:

Abstract

Background: Radiomics-based characterization of fluid and retinal tissue compartments of spectral-domain optical coherence tomography (SD-OCT) scans has shown promise to predict anti-VEGF therapy treatment response in diabetic macular edema (DME). Radiomics features are sensitive to different image acquisition parameters of OCT scanners such as axial resolution, A-scan rate, and voxel size; consequently, the predictive capability of the radiomics features might be impacted by inter-site and inter-scanner variations.

Purpose: The main objective of this study was (1) to develop a more generalized classifier by identifying the OCT-derived texture-based radiomics features that are both stable (across multiple scanners) as well as discriminative of therapeutic response in DME and (2) to identify the relative stability of individual radiomic features that are associated with specific spatial compartments (e/g. fluid or tissue) within the eye.

Methods: A combination of 151 optimal responders and rebounders of anti-VEGF therapy in DME were included from the PERMEATE (imaged using Cirrus HD-OCT scanner) and VISTA clinical trials (imaged using Cirrus HD-OCT and Spectralis scanners). For each patient within the study, a set of 494 texture-based radiomics features were extracted from the fluid and the retinal tissue compartment of OCT images. The training set (S_t) included 76 patients and the independent test set (S_v) comprised of 75 patients. Features were ranked based on (i) only discriminability criteria, that is, maximizing area under the receiver operating characteristic curve (AUC) and (ii) both stability and discriminability criteria. The subset of radiomic features for which the feature expression remained relatively consistent between the two datasets, as assessed by Wilcoxon rank-sum test, were considered to be stable. Different

SD-OCT, spectral domain-optical coherence tomography; DME, diabetic macular edema; ILM, inner limiting membrane; RPE, retinal pigment epithelium.

R43EB028736; National Center for Research Resources, Grant/Award Number: C06 RR12463-01; VA Merit Review, Grant/Award Number: IBX004121A; United States Department of Veterans Affairs Biomedical Laboratory Research and Development Service; Breast Cancer Research Program, Grant/Award Number: W81XWH-19-1-0668; Prostate Cancer Research Program, Grant/Award Numbers: W81XWH-15-1-0558, W81XWH-20-1-0851; Lung Cancer Research Program, Grant/Award Numbers: W81XWH-18-1-0440, W81XWH-20-1-0595; Peer Reviewed Cancer Research Program, Grant/Award Numbers: W81XWH-18-1-0404, W81XWH-21-1-0345; Kidney Precision Medicine Project; Bristol Myers-Squibb, Boehringer-Ingelheim, Eli-Lilly; Astrazeneca

machine learning (ML) classifiers (such as k-nearest neighbors, Random Forest, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Support Vector Machine using linear and radial basis kernel, Naive Bayes) were trained using the features selected based on both the stability and discriminability criteria on S_t and then subsequently validated on S_v. The ML classifier (M_g) that yielded maximum AUC on S_v was considered to be more generalized and stable for distinguishing anti-VEGF therapy treatment response as well as less sensitive to the effect of inter-site and inter-scanner variability.

Results: The model M_g (based on both stability and discriminability criteria) achieved higher AUC compared to the criteria based off feature discrimination alone on S_v (maximum AUCs of 0.9 versus 0.81; p-value = 0.048). The texture-based radiomic features pertaining to the retinal tissue compartment were found to be more stable compared to the fluid related features across the two datasets.

Conclusions: Our study suggests that incorporating both stable and discriminatory texture-based radiomic features extracted from fluid and retinal tissue compartments of OCT scans, a more generalized radiomic classifier can be developed to predict therapeutic response in DME. Also, the feature stability was found to be a function of the spatial location within the eye from where the features were extracted.

KEYWORDS
anti-VEGF therapy, diabetic macular edema, discriminability, radiomics, spectral domain-optical coherence tomography, stability

1 | INTRODUCTION

Radiomics-based characterization of spectral-domain optical coherence tomography (SD-OCT) scans provide unique opportunities in biomarker discovery, early detection and monitoring of disease progressions, predicting therapeutic response and treatment decision making for multiple retinal diseases.[1–10] OCT is the dominant non-invasive imaging modality that provides high resolution cross-sectional tissue imaging of the retina and has been mostly used by the clinicians for quantitative assessment of retinal thickness to determine the severity level of different ocular diseases including diabetic macular edema (DME).[11] Radiomics offers extraction of sub-visual imaging attributes characterizing the heterogeneity within different OCT subcompartments.[12,13] Quantification of the subtle variation within the texture from SD-OCT scans allows for monitoring disease manifestation and predicting therapeutic response in DME. The role of OCT-derived texture-based radiomics features from different spatial compartments (fluid and retinal tissue compartment) in predicting anti-VEGF therapy treatment response in DME patients are well investigated.[12,13] In a recent work from our group,[12] the texture based radiomics features pertaining to the fluid compartments of OCT images were found to be most associated with therapeutic response for DME patients.

For diagnosis, predicting treatment response and management of diseases, the radiomics features should be stable and robust across multiple sites and institutions irrespective of inter-site and inter-scanner variation. However, the radiomics features are known to be sensitive to scanner specific image acquisition parameters that influence the brightness, contrast, resolution and overall image quality.[14–16] Recently a number of studies[17–21] in literature have focused on identifying stable and discriminatory features across multi-site data in the context of prostate[17–19] and lung cancer[20] and glioblastoma tumors.[21] These studies investigated the sensitivity of radiomics features to magnetic resonance (MR)[21] and computed tomography (CT)[20]-based image acquisition parameters (such as slice thicknesses, image contrast and voxel resolutions) as well as for digitized radical prostatectomy specimens[17] and quantitatively linked to variability in radiomics signatures. The SD-OCT image acquisition parameters such as axial resolution, sensitivity, A-scan rate, wavelength, voxel size also vary across different scanners and have potential impact on overall image acquisition.[22–24] For example, higher axial resolution of OCT scanners yield better image quality followed by detailed visualization of retinal structures and explicit differentiation of retinal layers.[23] OCT systems with higher sensitivity are capable of providing higher contrast images.[24] Considering the impact of CT and MRI image acquisition parameters on overall image quality,[17]–[21] it is likely to be true that the variability in OCT image appearance (as a function of acquisition related parameters) might also have impact on the extracted radiomic features and their predictive capability. Since the OCT scanners used for retinal image acquisition vary across multiple sites and institutions, radiomics model developed on single-institution data might not be generalized across multiple-sites; hence the stability and reproducibility of

OCT-derived radiomics features to build a more generalized and robust model to predict therapeutic response in DME across multi-sites needs thorough investigation.

Most of the radiomic studies[25,26] have focused on just using traditional feature selection methods to build models, these selection strategies aim to prioritize features that maximize discrimination. However, given the sensitivity of radiomics to acquisition parameters, there is an opportunity to include additional criteria like stability to create more robust and generalized models that explicitly invoke both stability and discriminability. Leo et al.[17] introduced the concept of Preparation Induced (PI) instability measure to quantify feature instability which is defined in terms of the frequency at which any feature is found to be significantly different in distribution among multiple sites and presented that incorporating both the stability (based off PI score) and discriminability (based off maximizing area under receiver operating characteristics curve (AUC)) criteria, highly accurate and highly stable features could be identified for prostate cancer detection.

Additionally, even though prior work from our group[12] has suggested that spatial location (fluid and retinal tissue compartment) within the eye can allow for unique radiomic signatures to predict treatment response in DME, it is still unknown whether spatial location can result in more or less stable radiomic features. Therefore, the association between feature stability and the spatial compartment for treatment response prediction also needs further exploration. In the present study we explored two scientific premises: (1) Incorporating both the stability and discriminability criteria we sought to develop a more generalized classifier that are more robust to variations in image acquisition parameters of the OCT scanners across multiple sites and simultaneously favorably distinguishes between the optimal responders and rebounders of anti-VEGF therapy in DME and (2) the stability of the features associated with different spatial locations/compartments within retina.

2 | MATERIALS AND METHODS

2.1 | Patient cohort

The present study included a combination of 151 optimal responders and rebounders of anti-VEGF therapy in DME from the PERMEATE[27] and VISTA[28] clinical trials. PERMEATE was a 12-month prospective open-label study for treatment-naïve eyes with foveal-involving retinal edema secondary to DME and retinal vein occlusion (RVO). The study protocol was approved by the Cleveland Clinic Investigational Review Board (IRB) and the study was adhered to the tenets of the Declaration of Helsinki and US Code 21 of Federal Regulations. The study included 31 subjects. Inclusion criteria included ≥18 years of age, foveal-involving retinal edema secondary to DME or RVO based on SD-OCT, Early Treatment Diabetic Retinopathy Study (ETDRS) best-corrected visual acuity (BCVA) of 20/25 to hand motion (HM), that is, 0 to 83 ETDRS letters in the study eye.[27] Three eyes from 31 subjects were excluded due to poor image quality, insufficient follow-up, or patient drop-out during the study. Patients were given intravitreal aflibercept injection (IAI) in two phases over 12 months period of time: 2 mg of IAI every 4 weeks (q4 dosing) for the first 6 months followed by bimonthly dosing (q8 dosing) for the last 6 months.[29] For the stability analysis of the texture based radiomics features in predicting therapeutic response on DME patients, we excluded all the RVO subjects ($N = 15$) and considered only DME subjects ($N = 13$) from the PERMEATE study. Based on the treatment response, eyes were categorized as optimal responders (that maintained/improved BCVA following the first 8-week therapeutic challenge, $N = 6$) and rebounders (that showed at least one letter worsening in BVCA following the first 8-week challenge $N = 7$).

VISTA was a double-masked randomized phase III clinical trial that investigated the efficacy and safety of intravitreal aflibercept injection (IAI) in eyes with center involving DME.[28,30,31] The study was conducted in 54 sites across the United States and the principles of Health Insurance Portability and Accountability Act, the Declaration of Helsinki, and the International Conference on Harmonization were followed. Local IRB approval was obtained at each participating institution, and all subjects provided written informed consent prior to study enrollment. The primary inclusion criteria were adults with type 1 or type 2 diabetes presenting center-involving DME with BCVA in ETDRS letter score between 24 and 73 letters in the study eye. Only 1 eye per subject was enrolled. This post-hoc analysis included 138 subjects in the q8 dosing group that underwent SD-OCT with the Cirrus HD-OCT device (Zeiss, Oberkochen, Germany) or Spectralis OCT (Heidelberg Engineering, Heidelberg, Germany). Subjects were treated with 2 mg IAI every 4 week (q4 dosing) followed by 2 mg IAI every 8 week (q8 dosing) after the loading phase of five consecutive monthly injections. Based on treatment response at week 20–24, eyes were categorized as rebounders ($N = 115$) and optimal responders ($N = 23$). The final cohort thus comprised of 151 DME patients (13 from PERMEATE and 138 from VISTA).

2.2 | OCT image acquisition

SD-OCT images for each patient in the PERMEATE clinical trial were captured using the Cirrus HD-OCT device. The subjects of VISTA clinical trial underwent SD-OCT with either the Cirrus HD-OCT device or the Spectralis OCT. These devices generated macular cube scans with 512×128 A-scans covering a nominal 6×6 mm scan area. Additional details regarding data preprocessing are provided in the Supplementary Material Section I.

2.3 | Region of interest/image segmentation

OCTViewer (Cleveland Clinic, Cleveland, OH)[2] is a machine learning (ML) based retinal layer segmentation platform that performs semi-automated segmentation of the internal limiting membrane (ILM), ellipsoid zone (EZ), retinal pigment epithelium (RPE) band, intraretinal fluid (IRF) and subretinal fluid (SRF) boundary lines. Macular cube scans (512 × 128 A-scans) were imported to OCTViewer to generate the retinal segmentation lines. The software generated segmentation lines were reviewed by two trained experts/readers and necessary corrections were done, if required. Multiple certified readers were involved in this assessment as it was performed within a large reading center. All readers undergo over 250 h of focused training on advanced OCT segmentation. Senior readers have more than 1.5 years of experience with advanced segmentation in retinal diseases. Their professional role is as an advanced image analyst for the Cole Reading Center.

2.4 | Feature extraction

From the image I, we represented a sub-volume (of I) corresponding to the segmentation of fluid and the retinal tissue compartment between inner limiting membrane (ILM) to retinal pigment epithelium (RPE) by I_f and I_{rtc}, respectively. A total of 494 3D texture-based radiomics features (F_{oct}) were extracted from the I_f and I_{rtc} sub-volumes, on a MATLAB platform (version 2021b; Mathworks, Natick, Mass). The radiomics feature set included 65 Haralick features[32] (capturing texture heterogeneity), 152 Laws energy[33] (capturing presence of spots, edges, waves, and ripples in an image), 225 Gabor wavelet[34] (capturing structural details at different orientations and scales) and 52 CoLIAGe[35] (capturing anisotropic tensor gradient differences across similar appearing pathologies in an image) features on a per-voxel basis. Statistics of mean, median, standard deviation, skewness, and kurtosis were then calculated from the feature responses of all voxels within the region of interest and a total of 2470 statistical features were obtained. Following feature extraction, z score feature normalization (mean of 0 and standard deviation of 1) was applied to ensure that radiomic features extracted from different sites lie within a comparable range of values. Redundant features were removed by computing the Spearman correlation coefficient (SCC) between every possible feature pair followed by pruning the feature with a higher Wilcoxon rank sum p-value from every feature pair with an SCC > 0.8. This resulted in a total of 1418 uncorrelated features. A detailed description of the texture-based OCT features is provided in Supplementary material section II.

2.5 | Feature stability assessment

In the present study, we used the concept of PI[17] to assess feature stability. Through cross-data set comparisons PI score was computed by counting the frequency at which a feature was found to be differentially expressed across the images captured by different scanners (Supplementary material section III). A high PI score represents a feature that is frequently different between scanners, therefore, likely to be affected by variations in OCT scanners and acquisition parameters and was considered to be unstable, whereas a low PI score reflects that the corresponding feature is less likely to be affected by the scanner-specific variation and was therefore more stable. In the present study, F_{oct} within S_t was ranked based on their PI score. PI > 0.1 indicates that the corresponding radiomic feature was found to be significantly different for more than 10% of comparisons across multiple scanners and was considered to be highly unstable. A detailed description of the feature stability is presented in Supplementary material section III.

2.6 | Feature discriminability assessment

The discriminability criteria was based on maximizing AUC. To assess the discriminability of features, each feature within S_t was used separately to train seven different ML classifiers, such as k-nearest neighbors (KNN), Random Forest (RF), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Support Vector Machine (SVM) using linear and radial basis kernel, Naive Bayes (NB) in a 3-fold cross-validated setting over 500 runs. The average of AUC values yielded by each of the classifiers was considered as the final AUC for each feature within S_t. A PI score of 0.1 suggests that a radiomic feature associated with the OCT sub compartment was significantly different between different sites for 10% of comparisons. Therefore PI < 0.1 is indicative of the fact that for 90% of cases the feature value is similar between the different scanners, therefore could be considered as stable features. On the other hand, our previous work[12] on radiomics based characterization of OCT sub compartments for predicting anti-VEGF therapy treatment response distinguished between optimal responders and rebounders of anti-VEHG therapy with minimum AUC of 0.65 using the most discriminating features. Therefore, in the present work features with PI < 0.1 and AUC > 0.65 were considered to be the most stable and highly discriminating features.

2.7 | Statistical analysis

In the present study we performed two experiments. In the first experiment, we developed a more generalized ML classifier to distinguish between the anti-VEGF therapy treatment response groups considering the inter-scanner variability in the present study. The final cohort of 151 DME patients was randomly split into the training set (S_t, $N = 76$) that consisted of 61 optimal responders and 15 rebounders and the test set (S_v, $N = 75$) that comprised of 61 optimal responders and 14 rebounders ensuring the balance in the number of optimal responders and rebounders within S_t and S_v. To avoid the curse of dimensionality and the risk of overfitting, the top nine features were selected independently from the stable and discriminating feature pool (within S_t) using three different feature selection methods including Wilcoxon rank-sum, minimum Redundancy maximum Relevance (mRmR), and *t*-test (Detailed Description in Supplementary material section IV). The top selected features were then used to train seven different cross-platform ML classifiers in a 3-fold cross-validated setting involving over 1000 iterations. Finally, the performance of the seven ML classifiers was evaluated on S_v. Among these seven classifiers, the classifier (M_g) that yielded maximum AUC on S_v was considered to be the most stable and generalizable.

In the second experiment, we assessed radiomic feature stability across I_f and I_{rtc} compartments for individual feature families (such as Gabor, Haralick, Laws and CoLIAGe). We evaluated feature resilience as a function of the geographic or spatial location within the retina from where the features were extracted.

3 | EXPERIMENTAL RESULTS

3.1 | Experiment 1: Developing robust machine learning model to differentiate between optimal responders and rebounders of anti-VEGF therapy using the most stable and discriminating features by evaluating the effect of scanner variability

The PI-AUC plot of the 1418 features within S_t for 151 patients are presented in Figure 1 where the *x*- and *y*-axis represents PI score and average AUC of each feature, respectively. Separate color-coding scheme is used to represent each feature family. Each feature is represented by a circle, the radius of the circle being the standard deviation of the AUC value from the average AUC value of the seven different classifiers. The rectangular box (in black) represents the set of features that follow both the stability (PI < 0.1) and the discriminability criteria (AUC > 0.65). In the present study, we identified 201 features to be stable and discriminating, out of them 77 (53 Laws, seven Gabor and 17 CoLIAGe) features were found to correspond to the fluid compartment and 124 (72 Gabor and 52 Laws) features belonged to the retinal tissue compartment. Detailed description of the stable and most discriminating features is presented in Supplementary material section V. The average PI score for F_f and F_{rtc} were 0.04 and 0.035, respectively.

The AUC values generated from a pair-wise combination of three feature selection methods along with seven ML classifiers is presented in Table 1. Out of these seven classifiers, the RF classifier yielded the highest AUC on S_v, hence this classifier was considered to be the most stable and generalizable classifier (M_g).

The "kurtosis-Laws W5E5W5" feature from the fluid compartment was found to be the most stable and discriminating feature across the PERMEATE and VISTA datasets. The Laws W5E5W5 captures patterns of waves in horizontal and diagonal directions and edges in vertical using a $5 \times 5 \times 5$ convolutional kernel. The feature map for the Laws W5E5W5 is illustrated in Figure 2 for one case of optimal responder and one case of rebounder. The feature value is highly expressed for the rebounders which is reflective of higher heterogeneity within the fluid compartment of the rebounders.

Additionally, from the 1418 features within S_t, we identified 761 features (596 fluid features and 165 retinal tissue features) based on the discrimination criteria alone. The AUC values generated by different feature selection and classifier combination is presented in Table 1. The classifiers built using both the stability and discriminability criteria achieved higher AUC compared to the classifiers built using only on the discriminability criteria (maximum AUC of 0.9 vs. maximum AUC of 0.81 for LDA; p-value $= 0.048$).

3.2 | Experiment 2: assessing feature stability as a function of geographical location

The PI-AUC plot for the individual feature family for F_f and F_{rtc} is illustrated in Figure 3(a)–(d) and the Boxplot of PI scores for different feature families is shown in Figure 3(e). A total of 110 Laws, 24 CoLIAGe and 32 Gabor features within I_f and 132 Laws, 82 Gabor, 11 CoLIAGe features within I_{rtc} were found to be stable with a PI score < 0.1. The average PI scores obtained for F_f and F_{rtc} are 0.14 and 0.094, respectively for Gabor and 0.11 and 0.05, respectively for CoLIAGe feature family. This reflects that both Gabor and CoLIAGe features are more stable for I_{rtc} compared to I_f. With PI score $= 0.23$ for both the compartments, Laws features were found to have intermediate stability. The average PI score for the Haralick feature was found to be 0.65 for both I_f and I_{rtc}, Haralick features were identified as among the least stable features.

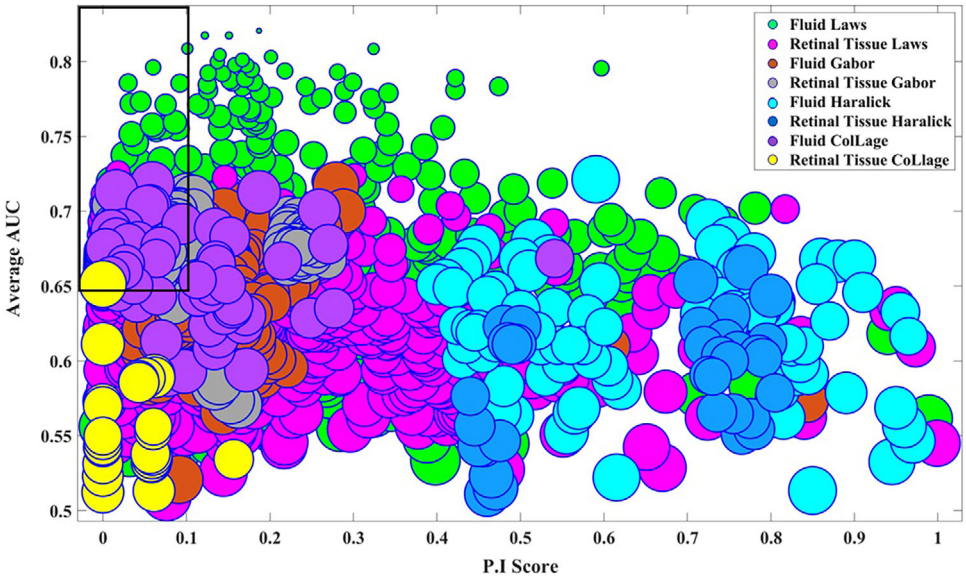

FIGURE 1 The PI-AUC plot for 1418 features for 151 patients within the training set across PERMEATE and VISTA datasets. Each feature is represented by a circle and each feature family is represented by a separate color. The size of each circle is the standard deviation of the seven different classifiers used for calculating average AUC. On the X-axis is the PI value for each feature associated with the Optimal responders across the two datasets. The average AUC value corresponding to each feature is plotted on the Y-axis. The AUC values were averaged across 1000 iterations of three-fold cross-validation across all 151 patients from the two datasets. The rectangular black box represents the most stable and discriminating feature pool across the two data sets. AUC, area under receiver operating characteristics curve; PI, preparation-induced instability score.

TABLE 1 AUC values yielded by different classifier and feature selection combination in distinguishing optimal responders and rebounders on S_v ($N = 75$).

Criteria	Feature selection	KNN	RF	QDA	LDA	NB	SVM (linear)	SVM (RBF)
Both stability-discriminability (PI < 0.1, AUC > 0.65)	Wilcoxon	0.65	0.84	0.57	0.73	0.62	0.57	0.72
	t Test	0.67	0.65	0.56	0.62	0.6	0.57	0.64
	mRmR	0.7	**0.9**	0.58	0.73	0.62	0.56	0.68
Discriminability alone (AUC > 0.65)	Wilcoxon	0.64	0.78	0.55	0.71	0.61	0.6	0.56
	t Test	0.66	0.47	0.55	0.72	0.56	0.58	0.55
	mRmR	0.69	**0.81**	0.56	0.69	0.64	0.61	0.64

Note: Bold value represents the highest AUC yielded based on each of the criteria.

4 | DISCUSSION

Radiomics offers extraction and quantification of high-throughput sub-visual imaging attributes from medical images obtained using computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI) or optical coherence tomography (OCT).[36–38] Radiomics-based signatures are widely used for developing machine learning (ML) models for predicting treatment response and risk progression for different retinal disorders.[2–4] There is an interest in developing radiomics-based diagnostic and clinical decision-support tools for individualized treatment management for multiple retinal disorders. Although radiomics hold promise in clinical applications, the main challenge in translating radiomics signature into clinical decision-support tool is the repeatability and reproducibility of radiomics features across multiple sites and their associated variability.[39]

Repeatability and reproducibility represent the extent of variability of the measured radiomics-based feature values of a particular region of interest (ROI) for repeated measurements and whether the feature values are consistent and stable for the same disease subtype across multiple sites.[39] Multiple sites/reading centers use multiple scanners with different image acquisition parameters and their effect on the variability of the extracted radiomics features needs to be investigated. The effect of inter-site/inter-scanner variability on radiomics signals are well studied in the context of CT and MRI images.[17–20] There are certain OCT image acquisition parameters such as axial resolution, scanning rate and voxel size that influence overall image quality and might have impact on predictive capability

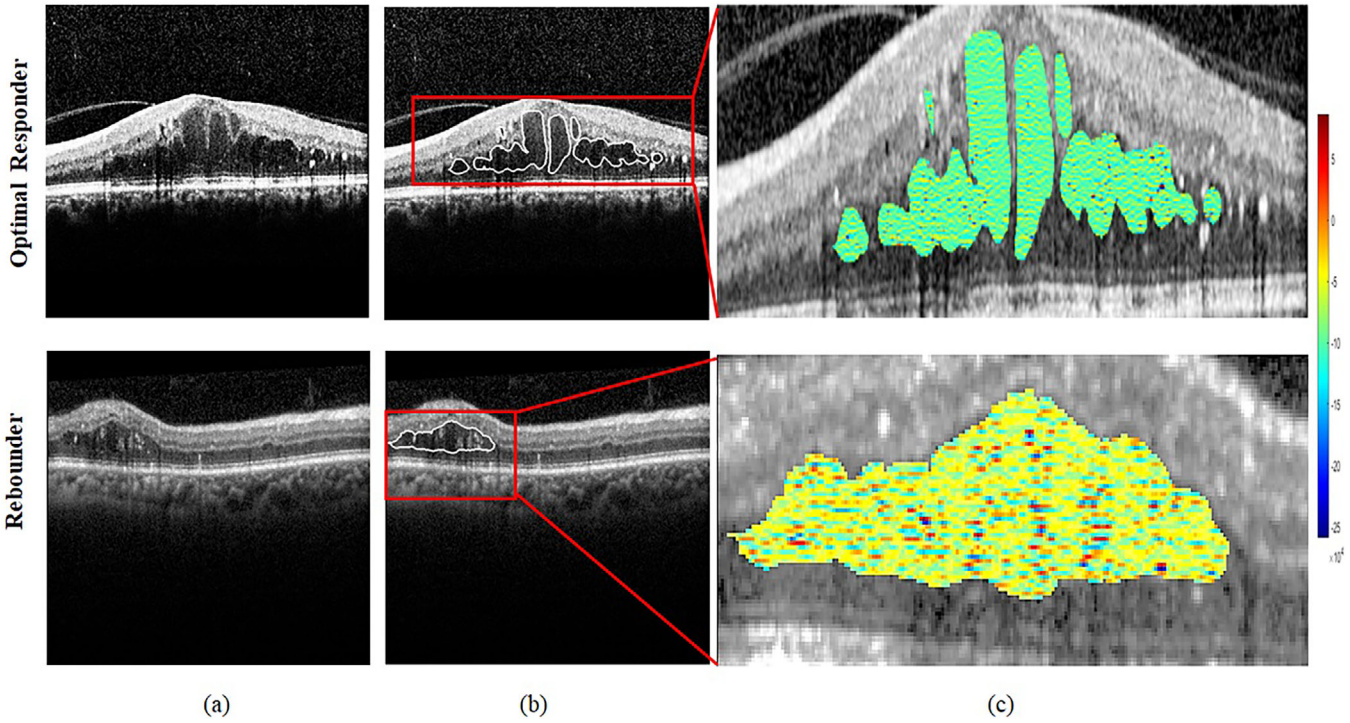

FIGURE 2 Feature map of the most stable and discriminating Laws W5E5W5 feature (PI = 0.06, AUC = 0.8), for one case of optimal responder and one case of rebounder. (a) Original SD-OCT image, (b) ROI corresponding to the fluid compartment on SD-OCT scan, (c) Feature heatmap showing higher feature expression reflected by the warmer color tone for the rebounder.

of OCT-derived radiomics features.[22–24] While identification of reproducible radiomics signatures is critical for developing computational imaging biomarkers, to the best of our knowledge, to date no studies have investigated the role of inter-scanner variability on predictive capability of radiomics features extracted from spectral domain (SD)-OCT scans or the feasibility of having a model that generalizes across scanner devices. A few prior studies have assessed the effect of signal strength and image quality fluctuations on the reproducibility of retinal nerve fiber layer thickness measurement using time-domain OCT and SD-OCT in the peripapillary area for glaucoma diagnosis.[37,38] Yang et al.[38] demonstrated the effect of image quality fluctuations on the reproducibility of swept source (SS)-OCT measurement of peripapillary retinal nerve fiber (PP-RNFL) and ganglion cell-inner plexiform (GC-IPL) layers. However, still there is a lack of knowledge on the degree of sensitivity of the radiomics features to inter-scanner variability and heterogeneous image acquisition parameter settings that likely to impact their predictive capability. Prior work from our group[12] investigated that the predictive capability of radiomics features to anti-VEGF therapy treatment response in DME patients are associated with the geographic/spatial location of the features. Particularly the fluid features were found to be most implicated in therapeutic response. In this study for the first time, we rigorously evaluated stability of features as a function of the spatial and geographic location within the eye from where they were extracted.

In the present study, we performed two experiments. The main objective of the first experiment was to evaluate the role of inter-scanner variability on radiomics features extracted from SD-OCT images, captured by different SD-OCT scanners/devices. We identified the most stable and discriminating texture-based radiomics features across PERMEATE (imaged using Cirrus device) and VISTA (imaged using Cirrus and Spectralis) clinical trials and developed a generalized cross-platform ML model (M_g) to distinguish between the optimal responders and rebounders of anti-VEGF therapy for diabetic macular edema (DME). The Laws texture feature pertaining to the fluid compartment were found to be stable as well as most discriminatory in predicting therapeutic response. Predominant overexpression of Laws feature values were identified for the rebounders which is reflective of higher order of textural heterogeneity within fluid compartment. This supports our previous finding where Laws texture features within the fluid compartment were found to be most associated with anti-VEGF therapy treatment response for DME patients.[12]

In the second experiment we evaluated feature resilience as a function of geographic/spatial location within the eye. Our results revealed that Haralick features are most unstable for both the fluid and retinal

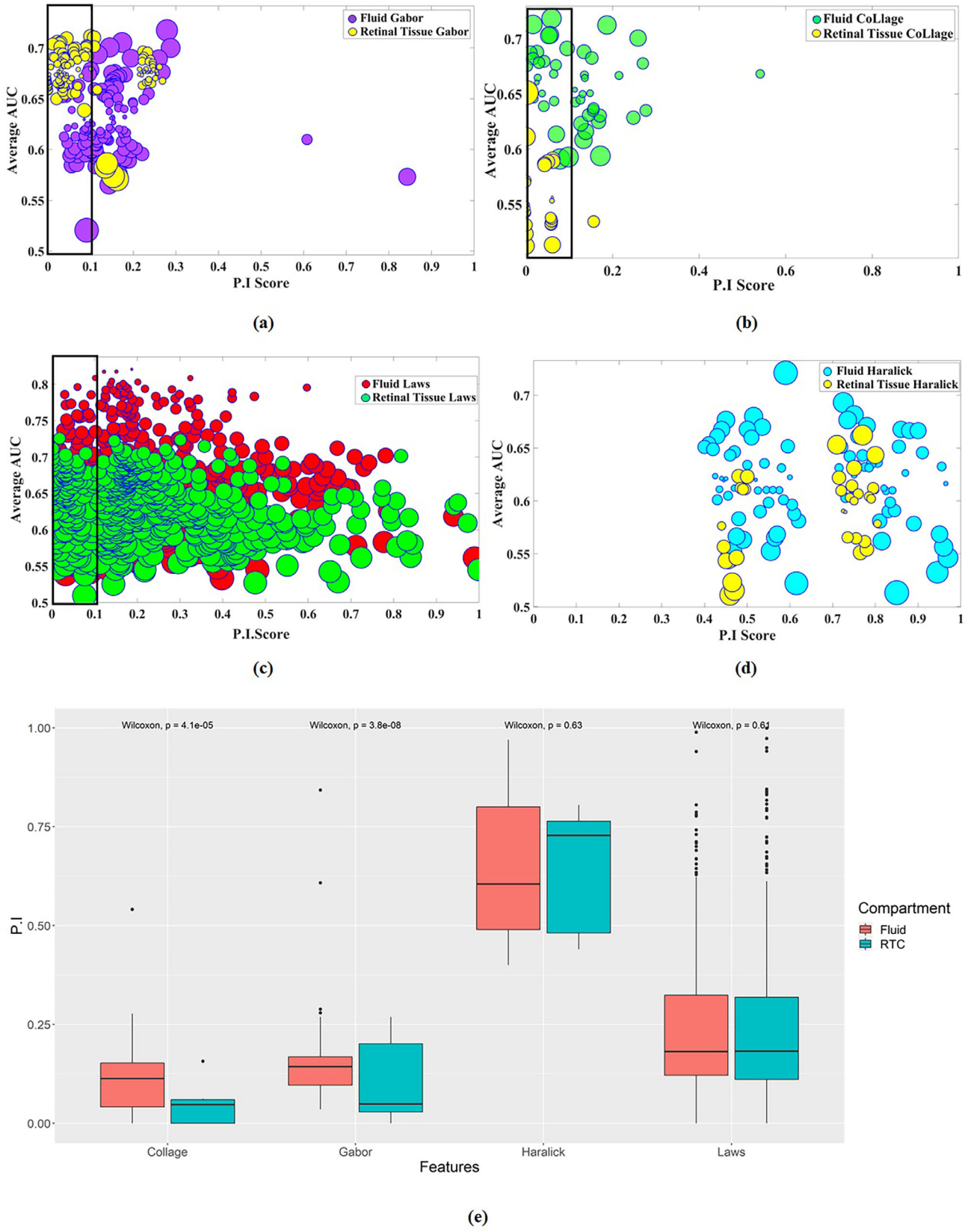

FIGURE 3 The PI-AUC plot for (a) Gabor, (b) CoLIAGe, (c) Laws and (d) Haralick feature family for fluid and retinal tissue compartments within the training set across PERMEATE and VISTA datasets. The size of each circle represents the standard deviation of the seven different classifiers used for calculating average AUC. The black rectangular box for each subfigure shows the subset of most stable features with PI < 0.1 for each of the feature family. Smaller circle size represents higher stability. (e) Boxplot of PI scores for different feature families between fluid and retinal tissue compartment. AUC, area under receiver operating characteristics curve; PI, preparation-induced instability score.

tissue compartments. Haralick features are extracted using higher-order derivatives and joint statistics from OCT images. Higher-order derivatives are known to be noisy and might be the reason for instability of this feature family.[39,40] Laws features are intermediate stable for both the compartments. Both Gabor and CoLIAGe features are more stable for retinal tissue compartment than the fluid compartment, however retinal tissue features are less discriminatory than the fluid features.

We acknowledge that our study did have limitations. The PERMEATE dataset was overall quite small and provided increased numbers of Cirrus scans. Since variance in datasets of limited size impact the ability of classifiers to learn, the generalizability of the classifier in this study must be investigated on larger datasets and additional validation is needed on large multi-site datasets in order for these features to be implemented in clinal decision support tools. A major limitation is that we did not evaluate the impact of individual acquisition related parameters on radiomics in this study as all acquisition parameters were standardized across the clinical sites. We did not analyze how the individual image acquisition parameters impacted the predictive capability of the radiomics features. The accuracy of the segmentation software was also not assessed. Since the voxel size depends on the retinal thickness map, measured by the segmentation software; segmentation error may have impacted voxel size. This could in turn affect the discriminability and reproducibility of the corresponding radiomics features. Despite these limitations, we believe that the findings presented in the study could allow for developing more generalized classifiers in treatment response prediction for DME patients on SD-OCT images.

5 | CONCLUSIONS

Our findings suggest that considering the most stable and discriminating features across multiple institutions/sites, it is possible to develop radiomics model less sensitive to inter-scanner and inter-site variation to discriminate between the favorable optimal responders and rebounders of anti-VEGF therapy in DME. Additionally, feature stability was found to be associated with geographical/spatial location within eye. Retinal tissue features were identified as most stable, however less discriminatory.

ACKNOWLEDGMENTS

Research reported in this publication was supported by NIH-NEI P30 Core Grant (IP30EY025585) (Cole Eye Institute), Unrestricted Grants from The Research to Prevent Blindness, Inc (Cole Eye Institute), Cleveland Eye Bank Foundation awarded to the Cole Eye Institute (Cole Eye), K23-EY022947-01A1 (JPE), the National Cancer Institute under award numbers R01CA249992, R01CA202752, R01CA208236, R01CA216579, R01CA220581, R01CA257612, R01CA268207A1, U01CA239055, U01CA248226, U54CA254566, National Heart, Lung and Blood Institute R01HL151277, R01HL158071, National Institute of Biomedical Imaging and Bioengineering R43EB028736, National Center for Research Resources under award number C06 RR12463-01, VA Merit Review Award IBX004121A from the United States Department of Veterans Affairs Biomedical Laboratory Research and Development Service, the Office of the Assistant Secretary of Defense for Health Affairs, through the Breast Cancer Research Program (W81XWH-19-1-0668), the Prostate Cancer Research Program (W81XWH-15-1-0558, W81XWH-20-1-0851), the Lung Cancer Research Program (W81XWH-18-1-0440, W81XWH-20-1-0595), the Peer Reviewed Cancer Research Program (W81XWH-18-1-0404, W81XWH-21-1-0345), the Kidney Precision Medicine Project (KPMP) Glue Grant and sponsored research agreements from Bristol Myers-Squibb, Boehringer-Ingelheim, Eli-Lilly and Astrazeneca.

This content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the U.S. Department of Veterans Affairs, the Department of Defense, or the United States Government.

CONFLICT OF INTEREST STATEMENT

S.K.S. has research support from Regeneron, Allergan, and Gilead; is a consultant for Bausch and Lomb, Novartis, and Regeneron. A.M.: Research Funding: Astrazeneca, Bristol Myers-Squibb, Boehringer-Ingelheim, Eli-Lilly. Equity: Picture Health Inc, Inspirata Inc., Elucid Bioimaging, Consultant: Aiforia, Picture Health, Simbiosys. J.P.E has research support from the following: Aerpio, Alcon, Thrombogenics/Oxurion, Regeneron, Genentech, Novartis, Allergan; is a consultant for the following: Aerpio, Adverum, Alcon, Allegro, Allergan, Genentech/Roche, Stealth, Novartis, Thrombogenics/Oxurion, Leica, Zeiss, Regeneron, Santen; and holds a patent with Leica. SSK and HC have no financial disclosures to report.

DATA AVAILABILITY STATEMENT

The data and code are not available for public access because of patient privacy concerns but are available from the corresponding authors if there is a reasonable request and approval from the institutional review boards of the affiliated institutions.

REFERENCES

1. Chen TC, Hoguet A, Junk AK, et al. Spectral-Domain OCT: helping the clinician diagnose glaucoma: a report by the American Academy of Ophthalmology. *Ophthalmology*. 2018;125(11):1817-1827. 10.1016/j.ophtha.2018.05.008. Epub 2018 Jul 7. Erratum in: Ophthalmology. 2019 Jun;126(6):915.

2. Ehlers JP, Zahid R, Kaiser PK, et al. Longitudinal assessment of ellipsoid zone integrity, subretinal hyperreflective material, and subretinal pigment epithelium disease in neovascular age-related macular degeneration. *Ophthalmol Retina*. 2021;5(12):1204-1213. 10.1016/j.oret.2021.02.012. Epub 2021 Feb 26.
3. Ehlers JP, Uchida A, Hu M, et al. Higher-order assessment of OCT in diabetic macular edema from the VISTA Study: ellipsoid zone dynamics and the retinal fluid index. *Ophthalmol Retina*. 2019;3(12):1056-1066. 10.1016/j.oret.2019.06.010. Epub 2019 Jul 6.
4. Kalra G, Talcott KE, Kaiser S, et al. Machine learning–based automated detection of hydroxychloroquine toxicity and prediction of future toxicity using higher-order OCT biomarkers. *Ophthalmol Retina*. 2022;6(12):1241-1252. 10.1016/j.oret.2022.05.031
5. Korobelnik JF, Do DV, Schmidt-Erfurth U, et al. Intravitreal aflibercept for diabetic macular edema. *Ophthalmology*. 2014;121(11):2247-2254. 10.1016/j.ophtha.2014.05.006. Epub 2014 Jul 8.
6. Maheshwary AS, Oster SF, Yuson RM, Cheng L, Mojana F, Freeman WR. The association between percent disruption of the photoreceptor inner segment-outer segment junction and visual acuity in diabetic macular edema. *Am J Ophthalmol*. 2010;150:63–67 e1.
7. A Alasil T, Keane PA, Updike JF, et al. Relationship between optical coherence tomography retinal parameters and visual acuity in diabetic macular edema. *Ophthalmology*. 2010;117(12):2379-2386. 10.1016/j.ophtha.2010.03.051. Epub 2010 Jun 18.
8. Ugwuegbu O, Uchida A, Singh RP, et al. Quantitative assessment of outer retinal layers and ellipsoid zone mapping in hydroxychloroquine retinopathy. *Br J Ophthalmol*. 2019;103(1):3-7. 10.1136/bjophthalmol-2018-312363. Epub 2018 Sep 6.
9. Banaee T, Singh RP, Champ K, et al. Ellipsoid zone mapping parameters in retinal venous occlusive disease with associated macular edema. *Ophthalmol Retina*. 2018;2(8):836-841. 10.1016/j.oret.2017.11.009. Epub 2018 Jan 6.
10. Itoh Y, Ehlers JP. Ellipsoid zone mapping and outer retinal characterization after intravitreal ocriplasmin. *Retina*. 2016;36(12):2290-2296. 10.1097/IAE.0000000000001110
11. Pe'er J, Folberg R, Itin A, Gnessin H, Hemo I, Keshet E. Upregulated expression of vascular endothelial growth factor in proliferative diabetic retinopathy. *Br J Ophthalmol*. 1996;80(3):241-245.
12. Kar SS, Sevgi DD, Dong V, Srivastava SK, Madabhushi A, Ehlers JP. Multi-Compartment spatially-derived radiomics from optical coherence tomography predict anti-VEGF treatment durability in macular edema secondary to retinal vascular disease: preliminary findings. *IEEE J Transl Eng Health Med*. 10.1109/JTEHM.2021.3096378
13. Kar SS, Cetin H, Lunasco L, et al. OCT-derived radiomic features predict anti-VEGF response and durability in neovascular age-related macular degeneration. *Ophthalmol Sci*. 2022;2(4):100171. 10.1016/j.xops.2022.100171
14. Styner MA, Charles HC, Park J, et al. Multisite validation of image analysis methods: assessing intra- and intersite variability. *Proc SPIE Int Soc Opt Eng*. 2002;46841:278-286.
15. Um H, Tixier F, Bermudez D, et al. Impact of image preprocessing on the scanner dependence of multi-parametric MRI radiomic features and covariate shift in multi-institutional glioblastoma datasets. *Phys Med Biol*. 2019;64:165011. Medline.
16. Zhao B, Tan Y, Tsai WY, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep*. 2016;6:23428. Medline.
17. Leo P, Lee G, Shih NC, et al. Evaluating stability of histomorphometric features across scanner and staining variations: prostate cancer diagnosis from whole slide images. *J Med Imaging (Bellingham)*. 2016;3:047502. Medlin.
18. Hiremath A, Shiradkar R, Merisaari H, et al. Test-retest repeatability of a deep learning architecture in detecting and segmenting clinically significant prostate cancer on apparent diffusion coefficient (ADC) maps. *Eur Radiol*. 2021;31:379-391. Medline.
19. Merisaari H, Taimen P, Shiradkar R, et al. Repeatability of radiomics and machine learning for DWI: short-term repeatability study of 112 patients with prostate cancer. *Magn Reson Med* 2020;83:2293-2309. Medline14.
20. Khorrami M, Bera K, Thawani R, et al. Distinguishing granulomas from adenocarcinomas by integrating stable and discriminating radiomic features on non-contrast computed tomography scans. *Eur J Cancer*. 2021;148:146-158. 10.1016/j.ejca.2021.02.008. Epub 2021 Mar 17.
21. Verma R, Hill VB, Statsevych V, et al. Stable and discriminatory radiomic features from the tumor and its habitat associated with progression-free survival in glioblastoma: a multi-institutional study. *AJNR Am J Neuroradiol*. 2022;43(8):1115-1123. 10.3174/ajnr.A7591
22. Leung CK, Ye C, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography a study on diagnostic agreement with Heidelberg Retinal Tomograph. *Ophthalmology*. 2010;117:267-274.
23. Araie M. Test-retest variability in structural parameters measured with glaucoma imaging devices. *Jpn J Ophthalmol*. 2013;57:1-24.
24. Wang J, Shousha MA, Perez VL. Ultra-high resolution optical coherence tomography for imaging the anterior segment of the eye. *Ophthalmic Surg Lasers Imaging*. 2011;42:15-27. Suppl.
25. Min X, Li M, Dong D, et al. Multi-parametric MRI-based radiomics signature for discriminating between clinically significant and insignificant prostate cancer: cross-validation of a machine learning method. *Eur J Radiol*. 2019;115:16-21. 10.1016/j.ejrad.2019.03.010. Epub 2019 Mar 15.
26. Cuocolo R, Stanzione A, Ponsiglione A, et al. Clinically significant prostate cancer detection on MRI: a radiomic shape features study. *Eur J Radiol*. 2019;116:144-149. 10.1016/j.ejrad.2019.05.006. Epub 2019 May 7.
27. Figueiredo N, Srivastava SK, Singh RP, et al. Longitudinal panretinal leakage and ischemic indices in retinal vascular disease after aflibercept therapy: the PERMEATE Study. *Ophthalmol Retina*. 2020;4:154-163.
28. Heier JS, Korobelnik JF, Brown DM, et al. Intravitreal aflibercept for diabetic macular edema: 148-week results from the VISTA and VIVID studies. *Ophthalmology*. 2016;123(11):2376-2385.
29. Ehlers JP, Peripheral and macular retinal vascular perfusion and leakage in DME and RVO (PERMEATE). [Online] Available: https://clinicaltrials.gov/ct2/show/NCT02503540
30. Korobelnik JF, Do DV, Schmidt-Erfurth U, et al. Intravitreal aflibercept for diabetic macular edema. *Ophthalmology*. 2014;121(11):2247-2254.
31. Brown DM, Schmidt-Erfurth U, Do DV, et al. Intravitreal aflibercept for diabetic macular edema: 100-week results from the VISTA and VIVID studies. *Ophthalmology*. 2015;122(10):2044-2052.
32. RMHaralick KS, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern*. 1973;SMC-3(6):610-621. doi:10.1109/TSMC.1973.4309314
33. Laws KI, Textured image segmentation. University of Southern California Los Angeles Image Processing Inst, 1980. Report No.: USCIPI-940. Accessed November 24, 2016 http://www.dtic.mil/docs/citations/ADA083283
34. Jain AK, Farrokhnia F, "Unsupervised texture segmentation using Gabor filters," 1990 IEEE International Conference on Systems, Man, and Cybernetics Conference Proceedings, Los Angeles, CA, USA, 1990, pp. 14-19, 10.1109/ICSMC.1990.142050
35. Prasanna P, et al. Co-occurrence of local anisotropic gradient orientations (CoLIAGe): a new radiomics descriptor. *Sci Rep*. 2016;6:37241. 10.1038/srep37241
36. Leite MT, Rao HL, Zangwill LM, et al. Comparison of the diagnostic accuracies of the Spectralis, Cirrus, and RTVue optical coherence tomography devices in glaucoma. *Ophthalmology*. 2011;118:1334e1339.

37. Lee EJ, Lee KM, Kim H, Kim TW. Glaucoma diagnostic ability of the new circumpapillary retinal nerve fiber layer thickness analysis based on Bruch's membrane opening. *Invest Ophthalmol Vis Sci*. 2016;57:4194e4204.
38. Yang H, Lee HS, Bae HW, et al. Effect of image quality fluctuations on the repeatability of thickness measurements in swept-source optical coherence tomography. *Sci Rep*. 2020;10:13897. 10.1038/s41598-020-70852-y
39. Prenosil GA, Weitzel T, Fu¨rstner M, Hentschel M, Krause T, Cumming P, et al. Towards guidelines to harmonize textural features in PET: haralick textural features vary with image noise, but exposure-invariant domains enable comparable PET radiomics. *PLoS One*. 2020;15(3):e0229560. 10.1371/journal.pone.0229560
40. Caramella C, Allorant A, Orlhac F, et al. Can we trust thecalculation of texture indices of CT images? A phantom study. *Med Phys*. 2018;45(4):1529e36. 10.1002/mp.12809

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kar SS, Cetin H, Srivastava SK, Madabhushi A, Ehlers JP. Stable and discriminating OCT-derived radiomics features for predicting anti-VEGF treatment response in diabetic macular edema. *Med Phys*. 2025;52:2762–2772.
https://doi.org/10.1002/mp.17695