Anomaly Detection and Biomarkers Localization in Retinal OCT Scans

Liran Tiosano  

tiosano@hadassah.org.il

Hadassah-Hebrew University Medical Center, Hebrew University – Hadassah School of Medicine

Ron Abutbul  

Hebrew University of Jerusalem

Rivkah Lender  

Hadassah Medical Center

Yahel Shwartz  

Hadassah Medical Center

Itay Chowers  

Hadassah-Hebrew University Medical Center, Hebrew University – Hadassah School of Medicine

Yedid Hoshen  

Hebrew University of Jerusalem

Jaime Levy  

Hadassah-Hebrew University Medical Center, Hebrew University – Hadassah School of Medicine

Article

Keywords: Age-related macular degeneration, deep learning, anomaly detection, optical coherence tomography angiography

Posted Date: September 21st, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3330479/v1

License: ☕️  This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.
Abstract

Anomaly detection combined with localization in retinal scans can help identify retinal anomalies scans and localize pathologies that might otherwise be difficult to detect. We designed a novel approach for detecting anomalies and localization by applying AI-based tools to optical coherence tomography (OCT) scans in retinal disease. High-resolution OCT-scans from the public and a local dataset were used in four state-of-the-art self-supervised frameworks. The backbone for these frameworks was a pre-trained convolutional neural network, which allowed us to extract meaningful features from the OCT-images. Anomalous images included choroidal neovascularization, diabetic macular edema, and drusen. The resulting anomaly detectors were then evaluated using area under the receiver operating characteristic curve (ROC AUC) scores, F1 scores, and accuracy. Approximately 30000 OCT-images were used. The best-performing anomaly detectors had an ROC AUC score of 0.99, and all frameworks achieved high performance and generalized well to various retinal diseases. Using pre-trained feature extractors, the frameworks tested here can be generalized to retinal OCT-scans, yielding high image-level ROC AUC scores. The localization results obtained using these frameworks can successfully capture areas indicating the presence of retinal pathology; moreover, these frameworks may also reveal new biomarkers. Finally, these frameworks can be integrated into clinical decision-making and automated screening systems, thereby facilitating treatment management.

Introduction

Retinal diseases affect millions of people worldwide and often result in reduced visual acuity and even blindness. Currently, clinical diagnoses and follow-up exams rely extensively on optical coherence tomography (OCT) scans, which provide high-resolution, cross-sectional images of the retinal layers. Obtaining a timely and accurate diagnosis is essential for optimal treatment and outcome; however, in the earlier stages of retinal disease, imaging biomarkers and lesions are typically difficult to detect. Furthermore, the relatively subjective nature of describing the features visible in a retinal scan often leads to disagreement, even among experts in the field. Therefore, there is currently a need for an automated clinical decision-making support system to help ophthalmologists detect biomarkers for predicting disease development and progression as early as possible.

Anomaly detection (AD) is a subclass of machine learning–based methods based on a binary classification between normal and anomalous classes. Although anomaly detection in ophthalmology is highly valuable, it also poses significant challenges. For example, a major challenge is that it is not always possible to train a model with full supervision for a given task, given the wide variety of visual anomalies and the current lack of classified anomalous data. Thus, although retinal datasets often contain annotations in the disease stage, they often lack annotations for biomarkers and the level of the lesions. The field of image anomaly detection has advanced considerably in recent years due in large part to the emergence of deep networks. The strength of these networks lies in their ability to learn meaningful representations of images. Similar semantic images are typically assigned a closely associated numerical representation; in contrast, raw image pixels often provide a much weaker
representation. The latest state-of-the-art anomaly detection methods use a relatively simple three-step paradigm. First, a numerical representation is computed for each image in the normal dataset using a deep neural network. Next, the normal data distribution is estimated using a probabilistic model. Finally, the likelihood of every test sample is calculated using the probabilistic model. If the model considers an image to be unlikely, it is labeled anomalous. Despite being trained on non-medical datasets, these methods can perform well on retinal images; for example, Burlina et al. recently showed that such methods can achieve state-of-the-art performance when applied to retinal fundus images.

Several retinal pathologies such as choroidal neovascularization (CNV), diabetic macular edema (DME), and drusen often lead to a diverse set of anomalous images. Specifically, CNV severely disrupts the retinal structure, DME results in swelling of the retinal layers, and drusen appear as localized elevations of the retina. Thus, the aim of this study was to identify these common retinal pathologies using a novel self-supervised deep learning–based anomaly detection method.

**Methods**

In this retrospective study, we used the Kaggle dataset, a public dataset containing 84,495 OCT slabs obtained from 44,346 individuals, and the Hadassah Imaging Dataset, a local real-world dataset containing 348,657 OCT scans obtained from 4663 patients diagnosed with age-related macular degeneration (AMD); all scans in both datasets were obtained using a Spectralis OCT device (Heidelberg Engineering GmbH, Heidelberg, Germany). Four categories were used to train the model: (i) normal retinal structure, (ii) CNV, (iii) DME, and (iv) early or intermediate AMD with the presence of drusen. The study adhered to the tenets established by the Declaration of Helsinki, and the Hadassah Medical Center Institutional Review Board (IRB) approved the study (0382–19). Because this was a retrospective study involving data collection, informed consent was waived.

The combined dataset was divided randomly into training, validation, and testing sets. We used testing and validation sets to evaluate the algorithm. In addition, we used a validation set to apply Sparsity-constrained Generative Adversarial Network (Sparse-GAN) as recommended by Akcay et al. This validation set was comprised of 4000 normal and 1000 anomalous image slabs that were randomly chosen from each of the four categories described above. Importantly, this dataset presents a challenge for anomaly detection, as the retinal scans are not aligned, not centered, and contain significant noise. In addition—and perhaps most challenging—the anomalies have diverse characteristics and locations through the various layers of the retina.

Our novel approach used the latest state-of-the-art (SOTA) methods paradigm and utilized local regions represented in scans of healthy retinas. We first applied the pre-trained neural network using the database of healthy retinal scans in order to obtain a local representation bank. We then fit a probabilistic model to the local representations. This probabilistic model was then applied to a new set of OCT scans in order to estimate the likelihood of their local regions. Local regions with low likelihood scores were defined as abnormal regions. Finally, retinal scans that contained one or more abnormal regions were classified as...
anomalous. In addition, we used the regions’ likelihood scores to localize anomalous patches. An initial self-supervised fine-tuning step was also performed on the pre-trained neural network, in which we adapted the neural network to the normal retinal dataset to improve the method’s performance. We also performed center loss adaptation using PANDA (pre-trained anomaly detection adaptation) in order to fine-tune the pre-trained neural network and achieve better results.²

We compared four SOTA frameworks for anomaly detection and segmentation in the images. All four frameworks used a pre-trained convolutional neural network (CNN) for feature extraction and embedding. The first SOTA framework was PANDA;² this framework consists of two stages: i) feature extraction using a pre-trained deep neural network and adaptation, and ii) retrieval of the k-nearest neighbors (kNN) to the target and calculation of anomaly scoring. The second SOTA framework, SPADE (Semantic Pyramid Anomaly Detection),⁵ consists of three stages: i) extraction of embedded images using a pre-trained deep neural network, ii) retrieval of kNN to the target, and iii) calculation of dense pixel level correlation between the target and normal images. The third framework, PaDiM (Patch Distribution Modeling),⁶ consisted of three stages: i) extraction of embedded images using a pre-trained deep neural network, ii) fitting a multivariate Gaussian distribution to every patch using the embedded normal images, and iii) estimation of the plausibility of every patch in the target image using the Mahalanobis distance measured between the target patches and the corresponding multivariate Gaussian distribution. Finally, the fourth framework, PatchCore framework,⁷ consists of three stages: i) creating embedded patches for every image in the training set, ii) sub-sampling the embedded patches using the coreset mechanism, and iii) calculating segmentation maps and image-level anomaly scores for each sample in the test set using the subset of embedded patches in the normal images.

Implementation details

The initial preprocessing stage included removal of black and white margins using classic image-processing algorithms. Specifically, a contour detector was applied to identify the white and black margins. In cases in which the corners were part of a contour, we checked to see whether each row or column contained more than 25% white or black pixels; if so, we removed that row or column. The images were then resized to 224x224 pixels. ResNet152 pre-trained on ImageNet-1k was then used as a deep neural network to extract the representation.⁸ We evaluated both the local region and the image-level representation methods. We represented each local region by combining the features of the second and third residual blocks in the ResNet152 model. In addition, the features obtained after pooling the global averages were used as image-level representations. The kNN algorithm was used to estimate density with k = 5.⁹ For the local region approach, the density estimation used a global approach. A local region density was estimated against all local regions within the nominal bank, regardless of their spatial location. To reduce the size of the local region bank and lower the computational cost, we used a coreset generated using greedy approximation. The nominal bank retains 1% of the healthy local regions for the open source and 10% of the healthy local regions for the clinical dataset. In the fine-tuning step, we used PANDA to adapt the features over the training dataset, with 40 epochs.²
After training the model on normal-only OCT scans, we extracted a nominal local bank. The model was then operated to score each of the OCT scans in the test set using kNN. The area under the receiver operating characteristic curve (ROC AUC) was then used as a parameter-free criterion for scoring the model. Each experiment was conducted using 3 different random seeds, and all studies were performed on NVIDIA RTX 2080 GPU cards using the machine learning framework PyTorch.

Anomaly detection of local retinal regions

Data representation: Numerical representations of local regions were extracted from healthy retinal scans using a deep neural network. The neural network was pre-trained on non-medical images, but has been shown to learn to effectively represent medical images, including color fundus photographs.\textsuperscript{10–12} The representations were required to satisfy the following criteria: regions with similar properties must be mapped to numerically similar representations, while dissimilar local retinal regions must be mapped to dissimilar numerical representations. For example, regions in an OCT scan that contain drusen should map to the same deep representation and should differ from fluid presented in the OCT representation (Fig. 1).

Density estimation

After extracting the representations, we estimated the distribution of nominal retinal regions. To represent the normal data, we trained a probabilistic model. Such models typically fall under two categories, namely parametric models (which follow a simple, mathematically specified function) and non-parametric models. Unfortunately, the appearance of local retinal regions is relatively complex and highly variable and therefore does not fit a simple mathematical model. Therefore, we opted to use a well-known non-parametric model, the kNN algorithm.\textsuperscript{9} Using a target image, this model determines the likelihood of a region by first retrieving the k nominal local region with the most similar representation using the nominal region bank. We then recorded the average distance between the numerical representation of the target local region and each of the representations of the k nearest local regions. The kNN distance provides a measure of the likelihood that the local region is anomalous. Regions with a large kNN distance are considered highly atypical (i.e., their representation is unlike the representation of any previously observed normal region); these regions have a very low likelihood and are therefore classified as anomalous. Conversely, representations of regions with a small kNN distance are similar to those observed as normal; thus, these regions are expected to be normal. Note that kNN distance is not a calculation of probability.\textsuperscript{13,14} However, their results are highly correlated to kNN results and the kNN is less computational starving.

Efficient nearest neighbors

To make the similarity search more efficient, we reduced the number of healthy local regions in the nominal bank. The computation of kNN can be relatively slow for large datasets due to neighbor retrieval, and this was the case in our retinal localization, in which each training image contained many local
retinal regions. Moreover, the training set contained thousands of normal scans, greatly increasing the number of local retinal regions. Because likelihood is estimated over local regions, the calculation can be computation-intensive. To speed up the computation process of the kNN algorithm, most approaches attempt to reduce the number of potential matches without removing the likely matches. We used this approach to choose a reduced number of prototypical regions, thus providing a compressed representation of all of the nominal regions observed during training. The most commonly used approaches for performing this compression are K means and coreset methods; in our study, we used PatchCore to run an approximation algorithm of the coreset in order to compress the normal region dataset.\(^7,9\)

**Fine-tuning the representation using retinal data**

Neural networks pre-trained on external non-medical datasets can extract powerful representations, even for ophthalmic data. The representation can be further improved by performing another training stage using the healthy retinal scan data. We therefore used the deep network and extracted image-level representations for the healthy training set. We then computed the center of the numerical representations using their average. A center loss was then used to fine-tune the network to the healthy scans. This approach requires that the representations of normal scans be close to the calculated center, therefore, our goal was for the normal scan representations to be closer to the center than the anomalous scans. To minimize the risk of catastrophic interference (also known as catastrophic forgetting), we used PANDA to include an elastic weight consolidation regularization loss in addition to a representation of center loss as described previously.\(^2\)

**Full retinal scan anomaly detection**

After extracting the anomaly scores for each local region of a retinal scan, we then classified each scan as either normal or abnormal. Our approach was to define the anomaly score as the largest anomaly score measured in all local regions within the retinal scan. The underlying logic was that an anomaly would be of interest even if it was in a small region of the retina. Using this approach allowed us to detect even small lesions and/or changes in the retinal structure.

**Results**

This study was designed to answer several questions. First, should anomaly detection be performed at the local region level, or the image level? Second, are representations that were learned using non-medical images effective for anomaly detection using OCT retinal scans? Third, given the effectiveness of pre-trained features, are there benefits to self-supervised training using normal data? Fourth, can inter-scan alignment assumptions be used in the nearest neighbor procedure? And finally, are the local region anomaly scores informative for use in patients?

Using SOTA methods, we found that the local region approach had generally good accuracy, sensitivity, specificity, and F1 scores, averaging 94.8, 94.4, 95.9, and 0.95, respectively. Table 1 summarizes the
results obtained using the various methods.

Table 1
Quantitative comparison with the state of the art methods. AUC- area under the curve, GAN- Generative Adversarial Networks

<table>
<thead>
<tr>
<th></th>
<th>Validation set</th>
<th>Local region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>Sensitivity</td>
</tr>
<tr>
<td>Auto-Encoder</td>
<td>72.9</td>
<td>78.3</td>
</tr>
<tr>
<td>AnoGAN</td>
<td>81.5</td>
<td>84.6</td>
</tr>
<tr>
<td>F-AnoGAN</td>
<td>84.9</td>
<td>88.2</td>
</tr>
<tr>
<td>Pix2pix</td>
<td>83.7</td>
<td>87.4</td>
</tr>
<tr>
<td>Sparse-GAN</td>
<td>88.5</td>
<td>92.5</td>
</tr>
<tr>
<td>Local region</td>
<td>98.6</td>
<td>99.7</td>
</tr>
</tbody>
</table>

Image-level representation vs. local region representation

In their recent study, Burlina et al.\(^3\) performed anomaly scoring using image-level features, in which a representation is computed for the entire retinal scan and an anomaly score is assigned to the image based on the extracted representation. Here, we opted to use a local approach, which scores regions rather than the entire retinal scan. Using this approach, we computed local region representations using a deep network followed by a region anomaly scoring stage. The set of region scores was then used to compute an image-level anomaly score. These two approaches (local versus image-level) are compared in Table 2.

Table 2
Comparison of local region and image-level anomaly detection methods (ROC AUC%)

<table>
<thead>
<tr>
<th></th>
<th>Validation set</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Image-level</td>
<td>Local-level</td>
</tr>
<tr>
<td>CNV</td>
<td>98.7</td>
<td>99.9</td>
</tr>
<tr>
<td>DME</td>
<td>95.4</td>
<td>99.1</td>
</tr>
<tr>
<td>Drusen</td>
<td>92.1</td>
<td>96.7</td>
</tr>
<tr>
<td>Average</td>
<td>95.4</td>
<td>98.6</td>
</tr>
</tbody>
</table>

AUC- area under the curve, CNV-choroidal neovascularization, DME-diabetic macular edema
Self-supervised vs. auxiliary pre-trained representations of features

Next, we compared two approaches for learning representations. The first approach is self-supervised representation learning, which uses only the training data consisting of healthy OCT retinal scans. The second approach uses a feature representation that was trained on ImageNet, an auxiliary, non-medical dataset. Note that every image in the ImageNet dataset contains a label indicating its object category (e.g., dog, cat, car, etc.). ImageNet images differ considerably from OCT scans; in addition to the obvious fact that the ImageNet images do not contain a retina, they are also color images and are relatively noise-free. Table 3 summarizes the differences between the pre-trained and self-supervised approaches, showing that the ImageNet-pre-trained representations outperformed the self-supervised representation. This is an important finding, as the model was not trained on retinal OCT scans.

<table>
<thead>
<tr>
<th></th>
<th>Validation set</th>
<th></th>
<th>Test set</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self-supervised</td>
<td>Pre-trained</td>
<td>Self-supervised</td>
<td>Pre-trained</td>
</tr>
<tr>
<td>CNV</td>
<td>96.7</td>
<td>99.8</td>
<td>97.4</td>
<td>100</td>
</tr>
<tr>
<td>DME</td>
<td>96.7</td>
<td>98.9</td>
<td>98.6</td>
<td>100</td>
</tr>
<tr>
<td>Drusen</td>
<td>81.9</td>
<td>97</td>
<td>82.1</td>
<td>99.1</td>
</tr>
<tr>
<td>Average</td>
<td>91.8</td>
<td>98.6</td>
<td>92.7</td>
<td>99.7</td>
</tr>
</tbody>
</table>

Fine-tuning the representations

Although representations that are pre-trained on non-retinal images are clearly effective for anomaly detection using retinal OCT scans, this begs the question of whether adding retinal scans to the training images will improve the results even further. To address this question, we fine-tuned the ImageNet–pre-trained representation on a training set that also contained healthy OCT scans using PANDA. As shown in Table 4, the adaptation stage improved both the image-level and local-level anomaly detection approaches. The increase in performance was particularly large for the OCT scans containing drusen. Thus, training on both auxiliary non-retinal data and normal OCT scans is more effective than training on only the auxiliary pre-training dataset.
Table 4
An evaluation of feature adaptation (ROC AUC%) CNV-choroidal neovascularization, DME-diabetic macular edema, AUC- area under the curve

<table>
<thead>
<tr>
<th></th>
<th>Image-level</th>
<th></th>
<th>Local-level</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-trained</td>
<td>Adapted</td>
<td>Pre-trained</td>
<td>Adapted</td>
</tr>
<tr>
<td>CNV</td>
<td>98.7</td>
<td>99.1</td>
<td>99.8</td>
<td>100</td>
</tr>
<tr>
<td>DME</td>
<td>95.4</td>
<td>95.9</td>
<td>98.9</td>
<td>99.9</td>
</tr>
<tr>
<td>Drusen</td>
<td>92.1</td>
<td>93</td>
<td>97</td>
<td>99.4</td>
</tr>
<tr>
<td>Average</td>
<td>95.4</td>
<td>96</td>
<td>98.6</td>
<td>99.8</td>
</tr>
</tbody>
</table>

Slice alignment assumptions

Previous approaches for local region anomaly detection (e.g., PaDiM) suggest that the anomaly score for each local region can be computed by comparing its representation to the representation of other regions in the image with the same spatial location. In other words, a region centered around the \((x, y)\) pixel location can be compared with other regions in other slices and scans that are also centered around the \((x, y)\) pixel. However, this assumes that different scans are aligned to each other. Indeed, this approach has several benefits when the scan are aligned, including faster runtime and higher accuracy, but these advantages are only realized when the scans are aligned well. We therefore compared this approach to the standard, global approach of comparing against all local regions regardless of their location relative to the center. As summarized in Table 5, we found that the alignment assumption is less effective with non-aligned retinal scans and results in reduced performance compared to the global search.

Table 5
Evaluation of the alignment assumption ROC AUC% CNV-choroidal neovascularization, DME-diabetic macular edema

<table>
<thead>
<tr>
<th></th>
<th>Aligned</th>
<th>Global</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNV</td>
<td>84.3</td>
<td>98.3</td>
</tr>
<tr>
<td>DME</td>
<td>85.1</td>
<td>93.7</td>
</tr>
<tr>
<td>Drusen</td>
<td>74.2</td>
<td>86.7</td>
</tr>
<tr>
<td>Average</td>
<td>81.2</td>
<td>92.9</td>
</tr>
</tbody>
</table>

Discussion

Here, we present a novel approach for anomaly detection in retinal OCT scans. This approach is based on local region anomaly detection and localization. First, we computed a numerical representation for each
local region of the retinal scan. We then fit a model for the local representations, thereby estimating the likelihood that each local region in the retinal scan is normal. We then used this model to localize the normal and anomalous regions in the retinal scans. Finally, retinal scans that contained one or more anomalous region were classified as anomalous. Importantly, our method represents data using a neural network that was pre-trained on an external, non-medical dataset. Moreover, we found that this method performed well on two datasets, namely a widely used public dataset and a locally obtained real-world clinical dataset.

Anomaly detection and localization are vital in visual machine learning, and the methods used in the field can be categorized as reconstruction-based, distribution-based, and classification-based.

Reconstruction-based methods learn to reconstruct the normal training data using a set of basic functions. During the reconstruction process, these methods use a bottleneck to create a compressed representation, which is then used to reconstruct the input sample. The main idea in this approach is that normal samples should reconstruct accurately, while anomalous data will not. Therefore, reconstruction loss is used as the anomaly criterion. Classic reconstruction-based methods include k-nearest neighbors (kNN), K-means, and principal component analysis (PCA). In recent years, the use of deep learning methods has expanded to include the field of reconstruction-based methods. Deep learning architectures such as autoencoders (AEs), variational autoencoders (VAEs), and generative adversarial networks (GANs) are all trained to reconstruct normal images. The common reconstruction loss functions include Euclidean distance, L1, and the structural similarity index measure (SSIM). Deep learning architectures also use deep perceptual loss and discriminators as adversarial loss functions.

Distribution-based methods attempt to learn the probability density function (PDF) of the distribution of normal samples. The anomaly score for a given image in the test set is calculated using the PDF. The expectation is that normal images will be mapped to dense areas using the PDF so that their likelihood will be relatively low compared to anomalous images. Therefore, in cases in which the anomaly score of an image is larger than a predetermined threshold, the image is deemed to be anomalous. Various distribution-based methods differ with respect to the distributional assumptions that they use. These methods include parametric models (e.g., multivariate Gaussian and Gaussian mixture models) and non-parametric models such as kernel density estimation, robust kernel density estimation, and kNN (which also performs density estimation).

Due to the use of PDF estimation, achieving good results using this approach can be difficult when handling high-dimensional data. To mitigate this problem, deep learning methods can be used to project the high-dimensional data to lower-dimensional space. Tools such as PANDA, SPADE, and PatchCore use pre-trained convolutional neural networks (CNNs) to extract embedding and use kNN to evaluate the anomaly score for a sample in the test set. Other approaches use AE or GAN architectures such as the deep autoencoding Gaussian mixture model (DAGMM), which uses a combination of distribution and reconstruction loss.
Finally, classification-based methods find a separate manifold between the normal data and the anomalous data in a given representation space. These methods include the one-class support vector machine (OC-SVM) and its variants such as support vector data description (SVDD), which obtains a spherically shaped boundary around a given dataset. Recently, deep learning methods were used to improve upon these classic approaches, resulting in methods such as deep SVDD and deep OC-SVM. An additional set of deep learning methods uses the self-supervised training approach, in which models attempt to solve an auxiliary task using either pre-existing data or easily obtained data. For example, Golan and El-Yaniv devised the Rot-Net-based method, a self-supervised method that uses rotation transformations to classify test samples\(^{20}\). This method was later improved and expanded.\(^7\)

Another unique challenge presented by visual anomaly detection is localization of the anomaly. For every detected anomalous scan, the anomaly detection algorithm segments all of the areas that were identified as anomalous; this setting is important for the algorithm’s explainability (Fig. 2), as it sheds light on the algorithm’s decision. In addition, and relevant to the medical imaging field, anomaly segmentation can help detect disease biomarkers and lesions in the scans. These algorithms even have the potential to identify new biomarkers and can therefore improve diagnosis. Anomaly segmentation has received relatively less attention compared to image-level anomaly detection. Early studies proposed patch-based localization using classic approaches,\(^{21,22}\) while more recent studies expanded these patch-based methods and used pre-trained CNNs as a feature extractor for a K-means classifier.\(^{23}\) For example, Bergmann et al. introduced the MVTec anomaly detection dataset and then evaluated this dataset using GAN, AE, and classic methods, followed by a student-teacher approach using a pre-trained CNN.\(^{24}\) Venkataramanan et al. introduced an attention-guided anomaly localization framework that uses VAE and other methods such as GradCAM and adversarial loss.\(^{25}\) Recently, a number of distribution-based frameworks, including PANDA, SPADE, PaDiM, and PatchCore, achieved SOTA performance in the task of anomaly localization. These frameworks use pre-trained CNNs for feature extraction and embedding creation and use density estimations for anomaly scoring.

Here, we adapted four SOTA frameworks for visual anomaly detection to the domain of retinal OCT scans. Anomaly detection and localization frameworks have the potential to detect disease biomarkers and lesions, providing the ability to diagnose retinal pathologies using OCT scans. We found that by using pre-trained feature extractors, these frameworks can generalize to the domain of retinal OCT scans and achieve high image-level ROC AUC scores. In addition, adapting the feature extractor to OCT scans can allow the framework to outperform the "raw" baseline. The localization results obtained using these frameworks are promising, as they can successfully capture areas that indicate the presence of retinal pathology. Importantly, we evaluated OCT scans of eyes affected by AMD and DME, the most common cause of blindness in the elderly and in working-age populations, respectively. However, these two conditions are not the only sight-threatening retinal pathologies, and the fact that our unsupervised anomaly detection frameworks achieved clinical-level performance suggests that other retinal pathologies may also be readily identified in OCT scans. Such frameworks for anomaly detection and
localization can even be integrated into systems that support clinical decision-making, thereby helping ophthalmologists with patient screening, diagnosis, follow-up, and treatment design.

A caveat of our study is that it included a relatively low number of scans and differed in image quality parameters such as the number of scans in each B-scan volume, differences in automatic real time (ART) values, and differences in image alignment. However, all of the OCT images in our dataset were obtained using a commonly used OCT device that produces high-quality sections. In addition, we used an unsupervised approach, which reduced the risk of bias generated by training.

Although preliminary, this study establishes a solid basis for the further development of automated anomaly detection frameworks for clinical use. Future studies should focus on exploiting the information that exists beyond a given OCT scan, including the patient’s diagnosis and/or medical history, thereby improving the algorithm’s clinical performance.

**Declarations**

Contributors: LT and RA contributed equally to this work.

YH and JL contributed equally to this work.

LT, JL, RL, and IC contributed to the study concept and design, acquisition, analysis and/or interpretation of the data, drafting of the manuscript and critical revision of the manuscript.

YS contributed to the acquisition, analysis, and/or interpretation of the data.

RA and YH contributed to the analysis and/or interpretation of the data, drafting of the manuscript, and critical revision of the manuscript.

Funding: This work was supported in part by the Israel Science Foundation (grant 3485-19).


Patient consent for publication: Waiver for informed consent was approved by Hadassah Medical Center Institutional Review Board (IRB) for the study (0382-19).

Provenance and peer review: Not commissioned; externally peer reviewed.

Availability of Data and Materials: The data collected in this study are available upon request. Please contact the corresponding author tiosano@hadassah.org.il

**References**

Figures
Figure 1

Overview of the workflow for local region anomaly detection. Top, the non-medical pre-trained neural network was used to learn local region representations. Middle, regions with similar properties were mapped to numerically similar representations, while dissimilar local retinal regions were mapped to dissimilar representations. Bottom, a test scan of a retina with choroidal neovascularization, in which the fluid representation differs from the presence of pigment epithelial detachment and drusen.
Figure 2

Heat maps of anomalous retinal OCT scans containing choroidal neovascularization (A), diabetic macular edema (B), and drusen (C). The original OCT scans are shown at the left, and the heat maps were extracted by local region using aligned density estimation (column 2) and local region with global density estimation (column 3).