scientific reports

Check for updates

OPEN

A deep learning approach to hard exudates detection and disorganization of retinal inner layers identification on OCT images

Lisa Toto[1], Anna Romano[1]✉, Marco Pavan[2], Dante Degl'Innocenti[2], Valentina Olivotto[2], Federico Formenti[1], Pasquale Viggiano[3], Edoardo Midena[4,5] & Rodolfo Mastropasqua[6]

The purpose of the study was to detect Hard Exudates (HE) and classify Disorganization of Retinal Inner Layers (DRIL) implementing a Deep Learning (DL) system on optical coherence tomography (OCT) images of eyes with diabetic macular edema (DME). We collected a dataset composed of 442 OCT images on which we annotated 6847 HE and the presence of DRIL. A complex operational pipeline was defined to implement data cleaning and image transformations, and train two DL models. The state-of-the-art neural network architectures (Yolov7, ConvNeXt, RegNetX) and advanced techniques were exploited to aggregate the results (Ensemble learning, Edge detection) and obtain a final model. The DL approach reached good performance in detecting HE and classifying DRIL. Regarding HE detection the model got an AP@0.5 score equal to 34.4% with Precision of 48.7% and Recall of 43.1%; while for DRIL classification an Accuracy of 91.1% with Sensitivity and Specificity both of 91.1% and AUC and AUPR values equal to 91% were obtained. The P-value was lower than 0.05 and the Kappa coefficient was 0.82. The DL models proved to be able to identify HE and DRIL in eyes with DME with a very good accuracy and all the metrics calculated confirmed the system performance. Our DL approach demonstrated to be a good candidate as a supporting tool for ophthalmologists in OCT images analysis.

Diabetes mellitus (DM) is a metabolic disorder frequently complicated by diabetic retinopathy (DR) and diabetic macular edema (DME). Diabetic retinopathy is the major sight threatening eye disease in working-age populations with an estimated prevalence of any form of DR of 34.6% and proliferative DR (PDR) of 6.96% in the diabetic patients.[1] Diabetic macular edema, which occurs in any stage of DR, is the most common cause of visual loss in DR. DME consists of retinal thickening involving or approaching the fovea related to capillary leakage and retinal fluid accumulation in the macula.[2]

Vascular, neurodegenerative, and inflammatory components have been implicated in DME occurrence.[3–6]

Nowadays, optical coherence tomography (OCT) a well-established retinal imaging technique allows to evaluate the status and the progression of different retinal pathologies. OCT is widely used for DME detection and classification based on the type of retinal edema.[7]

In addition, OCT is helpful in imaging biomarkers detection such as retinal fluid distribution, the status of the inner retina and in particular the disorganization of retinal inner layers (DRIL), the status of the outer retina and particularly of the photoreceptor layer, the presence of hard exudates (HE), and the identification of hyperreflective foci (HF).[8–11]

Several studies investigated OCT biomarkers in DME and found them to be predictive of treatment outcomes after intravitreal anti-VEGF injection and intravitreal steroid implant and to be related to the final visual outcome.[12–17]

[1]Ophthalmology Clinic, Department of Medicine and Ageing Science, University "G. D'Annunzio" of Chieti-Pescara, Via Dei Vestini Snc, 66100 Chieti, Italy. [2]Datamantix S.R.L. Artificial Intelligence Company, Via Paolo Sarpi, 14/15, 33100 Udine, Italy. [3]Ophthalmology Clinic, Department of Translational Biomedicine Neuroscience, University of Bari "Aldo Moro", Bari, Italy. [4]Department of Ophthalmology, University of Padova, 35128 Padova, Italy. [5]IRCCS- Fondazione Bietti, 00198 Roma, Italy. [6]Ophthalmology Clinic, Department of Neuroscience, Imaging and Clinical Science, "G. D'Annunzio" University of Chieti-Pescara, Via Dei Vestini Snc, 66100 Chieti, Italy. ✉email: romano-anna@hotmail.it

In the last years, the high demand for screening and monitoring patients with DR prompted the development of Artificial Intelligence (AI) in the ophthalmology field in order to improve access of patients to DR screening and to improve diagnostic accuracy, by means of retinal fundus images. AI and particularly deep learning (DL) models were applied to fundus images for automated detection of fundus lesions, such as microaneurysms, hard exudates, hemorrhages, and soft exudates.[18–21]

Moreover, recently the need to diagnose correctly DME features at baseline avoiding subjectivity of OCT biomarkers interpretation and to monitor DME response after intravitreal therapy favored the application of AI to structural OCT.

Some authors applied Machine Learning (ML) to OCT to identify macular edema of various etiologies including DME, exploiting the classification capability of AI models.[22,25] In particular, ML is a branch of AI in which programs learn from raw data to develop an automated classification of medical images.

Sandhu et al. showed high accuracy in the diagnosis of non-proliferative DR, applying ML techniques to multiple data such as OCT images alone or combined with retinal imaging techniques (such as OCT angiography) and clinical and demographic patient data. When OCT images were analyzed alone by the system demonstrated an accuracy of diagnosis of 76%, sensitivity of 85% and specificity of 87% in diagnosing DR.[26]

Recently other authors used DL models to detect imaging biomarkers in OCT images such as intraretinal and subretinal fluid, status of the external retina, and HF showing high accuracy of automating quantification for the different OCT biomarkers.[27–30]

Singh et al. also developed a DL-based algorithm for DRIL detection in DR patients using OCT images showing high accuracy, specificity and sensitivity in identifying DRIL.[31]

The aim of this study was to assess the accuracy of AI implementing a DL model to identify and localize two different OCT biomarkers related to DR and DME. The biomarkers chosen for this purpose are HE, an early and classic sign of DR related to the breakdown of blood retinal barrier (BRB), and the DRIL, a biomarker of inner retinal status frequently related to the worst visual acuity in DR patients, despite good response to intravitreal treatment.

Two different DL approaches were developed based on specific architectures with HE identification with details of their position and area and DRIL identification by running a whole image classification. Given that this paper has two primary objectives (identifying HE and DRIL), and the techniques for implementing them are of different natures, it is necessary to use the appropriate metrics for each of them. It will be the responsibility of the specialist to consider the two predictions provided by the two AI systems, aggregate the results, and draw conclusions. The novelties of our study are: (i) the use of a recent and complex DL approach combined with ensemble methods to exploit the best from different architectures and improve DL model classification performance applied to OCT images in DR and DME; (ii) identification/classification of two different OCT biomarkers of different nature using different AI methodologies and aggregating the results. Briefly, differently from the other works in the literature, we designed a more complex software architecture to combine several modules and take advantage of both the AI object detection approach and the image classification one.

Methods

Patients

This was a retrospective, cross-sectional, single center Italian study. It was carried out at the University of Chieti-Pescara for patients' enrollment, and in collaboration with Datamantix S.r.l. Artificial Intelligence Company (Udine, Italy) for data analysis.

This study was conducted in accordance with the Declaration of Helsinki and was approved by the institutional review board of the Department of Medicine and Ageing Science of the University 'G. d'Annunzio' of Chieti-Pescara (Chieti, Italy). All patients provided written informed consent prior to enrollment. Data were collected between March 2022 and September 2022 and analyzed from October 2022 to December 2022.

Inclusion criteria were age > 18 years and a diagnosis of type 2 diabetes mellitus with non-proliferative DR (NPDR) according to the simplified version of the ETDRS classification complicated by center-involved DME. Exclusion criteria were: a diagnosis of a retinal disease other than DR (retinal vascular occlusions, retinal dystrophies, age-related macular degeneration, macular holes etc.), vitreomacular interface (VMI) alterations such as vitreomacular traction; media opacity precluding high-quality OCT imaging, and myopia > 6 diopters.

Three hundred patients (442 eyes) were enrolled in the study, 155 females and 145 males, with a mean ± SD age of 65.5 ± 10.5 years (range 48 and 78 years), a mean diabetes duration of 15.2 ± 8.2 years and mean ± SD HbA1c (%) of 7.2 ± 0.6.

One hundred and fifty-eight eyes of the 300 patients enrolled in the study were not considered because not fulfilling the inclusion and exclusion criteria and in detail 48 eyes were not considered due to absence of center-invoved DME and/or presence of proliferative DR; 42 eyes were excluded for coexistence of vitreomacular interface (VMI) alterations; 68 eyes were excluded for media opacity precluding high-quality OCT imaging.

All patients had a diagnosis of DM type 2 with NPDR complicated by center-involved DME.

All patients underwent spectral domain OCT (SDOCT) using Spectralis ® HRA + OCT (Heidelberg Engineering; Heidelberg, Germany). The acquisition protocol for SDOCT included a macular horizontal and vertical B-scans centered on the fovea. Eye tracking and automatic real-time (ART) averaging techniques were used to improve image quality. A High-Resolution mode, a mean ART > 90, and a Quality index > 30 were used to ensure image quality. All selected OCT images were exported in PNG format for analysis.

OCT biomarker identification

At OCT images hard exudates are recognized as previously described as hyperreflective spots (HRS) mostly localized in the outer retinal layers with size > 30 µm, reflectivity similar to retinal pigment epithelium (RPE)-Bruch membrane and back shadowing (Fig. 1A).[32,33]

DRIL as previously described are defined as the loss of boundaries between ganglion cells, Inner plexiform layer, inner nuclear layer, and outer plexiform layer measured as the transverse extent (Fig. 1B).[34]

The deep learning system: design and development

This work requires simultaneous object detection for HEs and DRIL classification. To achieve this dual task effectively, a workflow that includes two different DL approaches was designed. The proposed pipeline is shown in Fig. 2 and it consists of a common phase of data collection and annotation followed by two main branches for DRIL classification and HEs detection, executed separately due to the different techniques involved.

Since the classification and the detection algorithms focus on various image features, two different data preparation strategies were carried out. Furthermore, the state-of-the-art to choose the best architectures was examined, and multiple experiments to compare the performance and the robustness of the models were developed. In the end, the results of the two branches have been combined. In Fig. 3 is depicted how the AI system provide the results.

All the experiments were carried out using an AI workstation with an NVIDIA v100 GPU with 32 GB of VRAM memory. The implementations were based on Python v3.8.10 and PyTorch v.1.13.0. Python libraries such as OpenCV v4.6.0 and Pillow v9.3.0 were also used.

In the following sections, a comprehensive description of the annotation process, the two key AI techniques implemented in our study, and the evaluation methodologies is provided.

Image annotation and analysis

Image data annotation refers to the process of labeling specific features or regions of interest within images, providing valuable information to train and evaluate DL models.

Each image was analyzed by three experienced retinal specialists (LT, RM, PV). In particular, to identify all the HEs in the pictures, two specialists separately analyzed the samples using a specific tool to draw bounding boxes around the biomarkers. The third specialist reviewed the annotations by merging and/or cleaning boxes.

On the other hand, the three specialists independently annotated the DRIL by indicating for each image its presence or absence. The majority voting has been applied to assign the proper annotation to each image. We also computed the inter-grader agreement and the Kappa coefficient in order to provide details about the

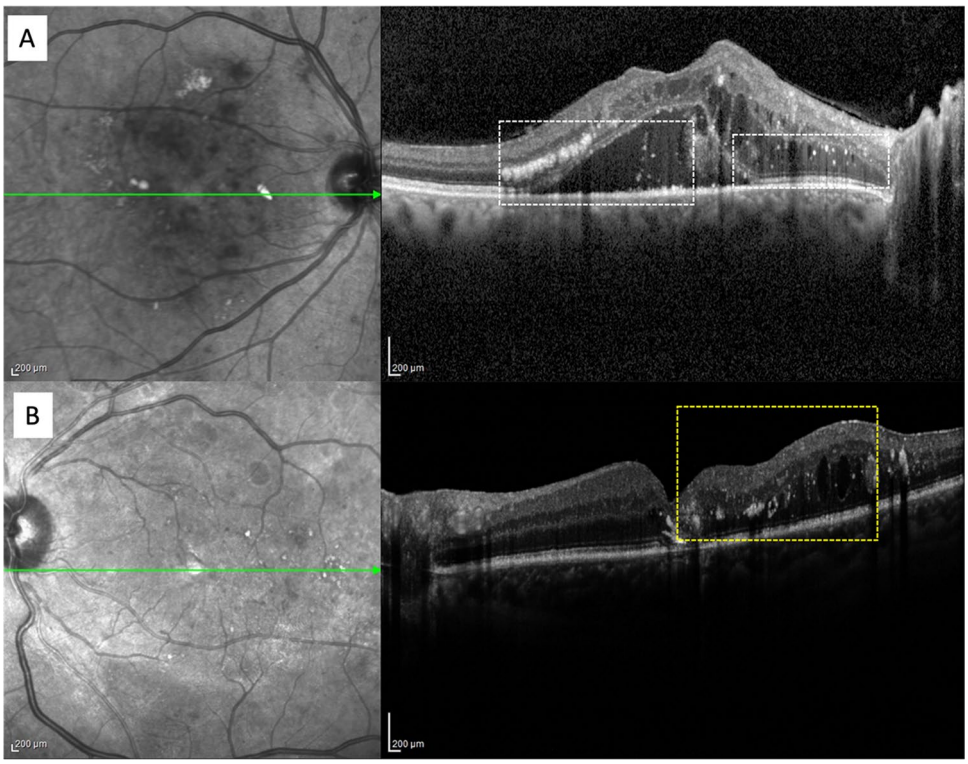

Figure 1. **A** Infrared (IR) image and horizontal optical coherence tomography (OCT) scan of a patient with diabetic macular edema showing hard exudates (dashed white rectangle). **B** IR image and horizontal passing through the fovea OCT scan passing through the fovea showing diabetic macular edema and disorganization of retinal inner layers (dashed yellow rectangle).

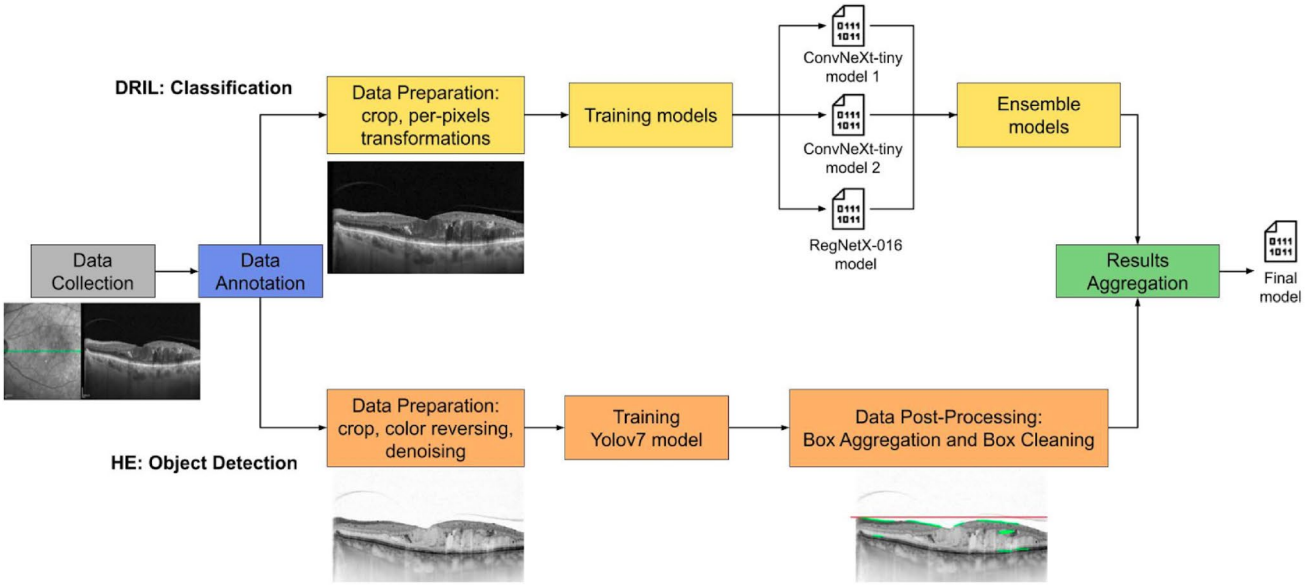

Figure 2. The operational workflow of our AI system. Due to the complexity of the task we designed and implemented a custom pipeline. It involves a shared phase of data collection and annotation and it is followed by two distinct branches, one for DRIL classification and the other one for HE detection. By combining these approaches, we are able to achieve the final goal with a model that fulfills our purpose.

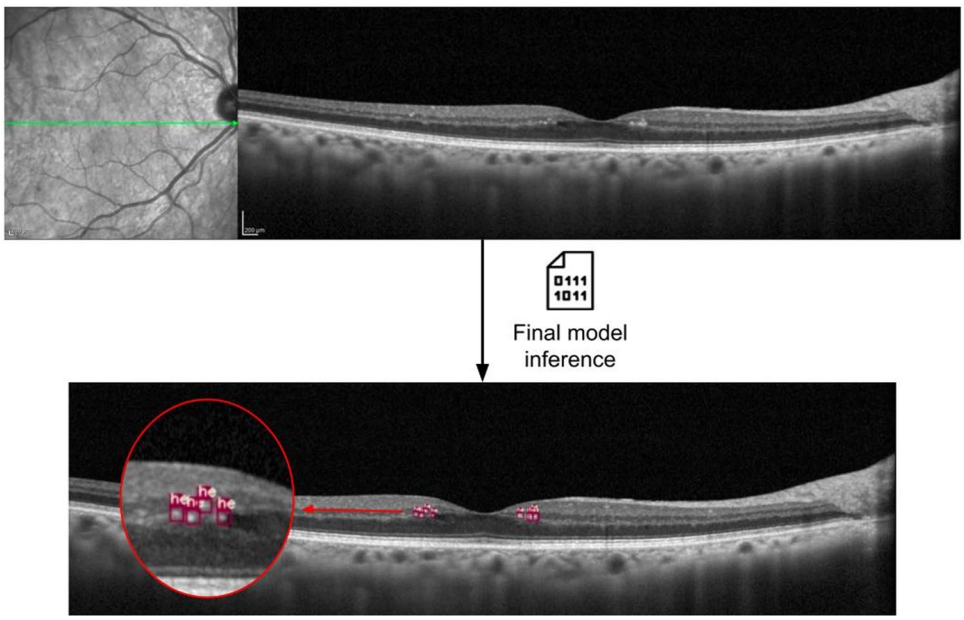

Figure 3. An example of input image and its corresponding output after the AI model inference. The upper part of the figure reports a typical OCT image that can be used as input data of our system to analyze the benchmarks presence. After the inference of the final model the image is labeled with a class ("*DRIL*" or "*NO DRIL*") that highlights if a DRIL is present or not, and all the predicted HEs are marked with bounding boxes. Regarding the specific sample, the predicted class is "*NO_DRIL*", while the HEs are identified by red boxes as emphasized in the zoom area.

agreement level among the different pairs of specialists. In general, the overall inter-grader agreement reaches a score equal to 83.26% (90% on DRILs and 76.58% on NO_DRILs images). In the following Table 1 we report all the computed scores for each specialists' pair.

A total of 442 OCT images were collected, among which 220 displaying DRILs (49.8%) and a total of 6847 HEs have been annotated.

Specialists' pair	% Agreement	Kappa coefficient
Rater 1–Rater 2	89.14	0.66
Rater 1–Rater 3	88.91	0.78
Rater 2–Rater 3	88.46	0.77

Table 1. Annotation agreement scores for each specialists' pair.

HE detection using AI object detection approach

Object detection technique allows us to simultaneously predict the class of an object and its bounding box coordinates in images. To maximize the performance and train a stable model, we implemented a set of image transformations and model configurations. We cropped the images to emphasize the region of interest and we applied the median filter technique to reduce the speckle noise issue. Moreover, we reversed the colors of collected images to highlight the objects of interest, in order to emphasize HEs against the surrounding tissue.

To proceed with the AI training phase, we chose the YOLO algorithm, because of its accuracy and ability to generalize well to new objects.[35] At the time of the experiments done in this study, version YOLOv7 surpassed all known object detectors, and for this reason, it was selected to maximize the experiments' performance.[36] First of all, we split the dataset randomly into a training set (60%, 264 images), a validation set (20%, 88 images), and a test set (20%, 90 images), dividing at the patient level. During the training phase, each input image was resized to 1024×1024 pixels, since high resolution improves the performance of object detection, especially for small objects such as HEs. A batch size of 16 images, a momentum of 0.937, and a learning rate in the range [0.01, 0.1] were set as the principal hyperparameters of the training step. Moreover, we analyzed and checked the validation loss curve to avoid overfitting.

After running the trained YOLO model on the test set we applied two post-processing steps to improve the precision of the results: merging boxes that overlap over a certain threshold and cleaning up predictions outside the region of interest. We exploited the Intersection Over Union metric (IOU) to measure the ratio of the boxes' intersection over the union of their areas and we merged boxes if their IOU is greater than 50%. Moreover, to address false positives outside the region in which HEs can be located, we detected the main edges using the Canny algorithm followed by the Probabilistic Hough Line Transform.[37,38] Finally, we removed the bounding boxes predicted above these edges.

DRIL Identification using AI image classification

DRIL were identified as present or not and the task was addressed implementing a whole image classification based on Convolutional Neural Networks.

In order to prepare data for the training process, all the images were cropped to keep only the part of the interest of the raw pictures. We also created copies of the original dataset to obtain different versions by applying various pixel transformations to the images. This involved adjusting the contrast and brightness of each picture and applying colorization effects. As a result, we obtained a total of six datasets (including the original one) which were used in the training phase. All of them were split randomly into three subsets (train, validation, and test set), following the same percentages of the object detection task and maintaining the same proportion of the target classes in each set, that makes the evaluation of the model more rigorous.

As previously stated, we examined the state-of-the-art Deep Neural Network architectures to select the most suitable ones that align with our primary objective, and we chose to adopt ResNet, ConvNeXt, and RegNetX.[39,40] We designed and implemented several experiments using the selected network architectures and the versions of the dataset created to compare the different combinations and highlight the best ones.

In Table 2 the top three models that reach the best performance are shown.

We combined the strengths of these models to identify the DRILs on images implementing an ensemble phase. We carried out an approach based on soft voting, where the three models have been considered with equal weight, and their predictions were combined to make the final decision.

Architecture	Dataset transformation	Epochs	Learning rate	Batch size	Image input size	Resize method
ConvNeXt tiny	Color substitution: black → red white → yellow midtone → blue	50	0.01	16	434×1024	Squish
ConvNeXt tiny	Color substitution: black → blue white → yellow midtone → red	100	0.001	16	434×1024	Squish
RegNetX_016	Brightness beta = -30 and contrast alpha = 1.5	50	0.001	16	434×1024	Squish

Table 2. Top three classification models and their main hyperparameters.

Evaluation and statistical analysis

In this section, we discuss the evaluation metrics used to assess the models' performance, to provide insights on their strengths and limitations, and to evaluate their potential for clinical applicability. The selected metrics to evaluate the DL system depend on the task considered since object detection and image classification problems have different evaluation processes.

The typical metrics calculated to evaluate an object detector are mean Average Precision (mAP), Precision, and Recall, that can be computed using different IOU thresholds. Usually, an IOU score greater than 0.5 is used to define a correct prediction, therefore we chose to adopt it. In our case, we do not need the standard mAP metric, which computes the mean (m) of the Average Precision (AP) over the different object classes, since there is only one class to predict. Therefore, we calculated the AP@0.5 as the Area Under the Precision-Recall Curve evaluated at the IOU threshold equal to 0.5. Regarding Precision and Recall, we followed the standard calculation. Moreover, each box is predicted by our model with a certain confidence level. In our specific case, since the objects are very small on many occasions and the images do not always have a high quality, we selected a confidence level of 10%. In this way, our AI system ensures to detect the maximum number of HEs possible, acting as a good supporting tool and without missing any crucial objects of interest in the image.

On the other hand, with regard to the DRIL classification, the principal metrics that we analyzed are Accuracy, Sensitivity, Specificity, Area Under the ROC Curve (AUC), and Area Under the Precision-Recall Curve (AUPR). Taking into consideration the confusion matrix data, we calculated two important statistical measures called p-value and Kappa coefficient. In this context, the p-value is computed considering a one-sided binomial test to determine if the accuracy surpasses the no-information rate. Moreover, the Kappa coefficient is a measure of inter-rater reliability, which assesses the level of agreement between two or more raters in evaluating a set of items. In the context of DL model evaluation, the Kappa coefficient is used to evaluate the performance of a model in a classification task by calculating the agreement between the predicted labels and the true labels.

Results

This section presents the results obtained calculating the described metrics over the HE detection and DRIL classification on the respective test set.

With reference to the HEs detection task, the AP@0.5 reaches a value of 34.4%, with the corresponding Precision-Recall curve shown in Fig. 4. In particular, the Precision results at 48.7%, and the Recall value is equal to 43.1% as summarized in Table 3.

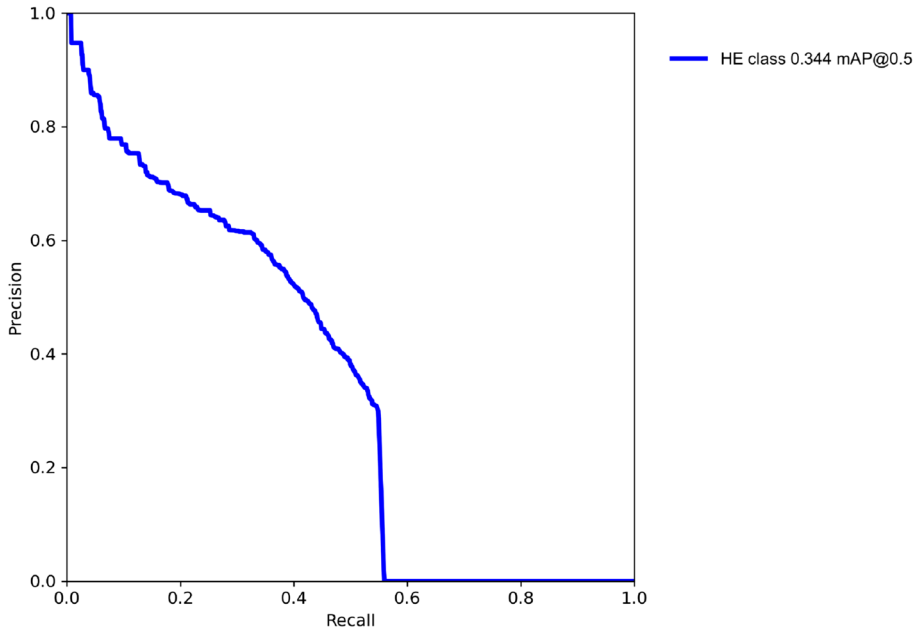

Figure 4. The results of HE detection on test set images. Precision-Recall curve of HE detection model considering IOU (Intersection Over Union) threshold to 0.5.

Object detection model	AP@0.5	Precision	Recall
Yolov7	34.4%	48.7%	43.1%

Table 3. The results of HE detection on test set images. *AP* average precision.

Finally, we computed the Pearson correlation between the number of HEs detected by the AI model and the number of HEs annotated by experts per each image, in order to evaluate the system under the HEs counting point of view. We obtained a score of 0.804 with a two-sided p-value lower than 0.05.

Considering the DRIL classification, we reached an overall accuracy score of 91.1% (a total of 82 correct predictions out of 90) with a 95% confidence interval of [0.852 ,0.969]. Both classes obtained the "Excellent" grade on the AUC interpretation rating scale.[41] The results indicate that the model effectively differentiates between positive and negative samples, with a 91.1% as score for both sensitivity and specificity metrics, as it is possible to note in the in Fig. 5. Table 4 depicts the principal metrics for the three selected models and the resulting ensembled final model.

The AUC and the AUPR values, both equal to 91%, confirm the remarkable performance of the classification model. The Kappa coefficient (the inter agreement between AI and the human majority voting) equal to 0.82 and the P-value lower than 0.05 indicate that the final model predictions are statistically significant. The obtained Kappa value is considered "Excellent" according to Cicchetti's and Fleiss' benchmarks and "Almost Perfect" according to Landis & Koch's benchmark.[42,43] In order to measure also the agreement between AI and each rater we computer the percentage of agreement and the Kappa coefficient for each "AI-rater" pair.[40] In Table 5 we illustrate these scores.

Finally, a more in-depth and thorough qualitative analysis of the final model predictions was conducted on each image. Expert ophthalmologists confirmed the true effectiveness of our AI system as a supporting tool for ophthalmologists.

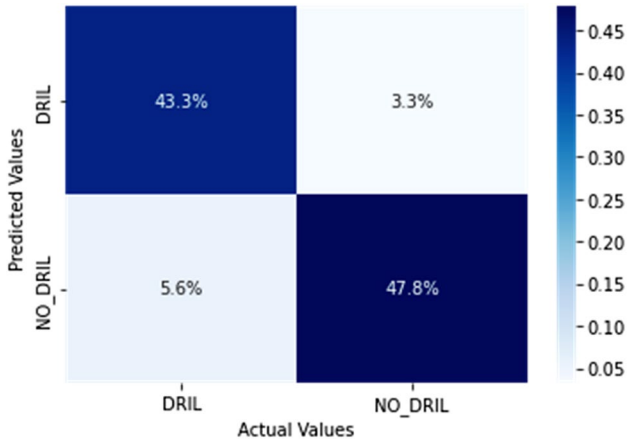

Figure 5. Confusion matrix of DRIL classification results on the test set images. It presents the percentage result of the predictions made by the model, compared to the actual values assigned by expert ophthalmologists during the annotation process.

Classification model	Accuracy	Accuracy 95% CI	Sensitivity	Specificity
ConvNeXt tiny model 1	83.3	[75.6%, 91.0%]	OVERALL: 83.3% DRIL: 83.7%, NO_DRIL: 83.0%	OVERALL: 83.3% DRIL: 83.0%, NO_DRIL: 83.7%
ConvNeXt tiny model 2	76.7	[67.9%, 85.4%]	OVERALL: 77.0% DRIL: 79.5%, NO_DRIL: 74.5%	OVERALL: 77.0% DRIL: 74.5%, NO_DRIL: 79.5%
RegNetX_016	73.3	[64.2%, 82.5%]	OVERALL: 73.3% DRIL: 72.7% NO_ DRIL: 73.9%	OVERALL: 73.3% DRIL: 73.9%, NO_DRIL:72.7%
Final Ensamble Classification	91.1	[85.2%, 96.9%]	OVERALL: 91.1% DRIL: 88.6%, NO_ DRIL: 93.5%	OVERALL: 91.1% DRIL: 93.5%, NO_ DRIL: 88.6%

Table 4. The results of DRIL classification on test set images. *CI* confidence interval.

"AI–rater" pair	% Agreement	Kappa coefficient
AI–Rater 2	91.11	0.82
AI–Rater 3	88.89	0.78
AI–Rater 3	88.89	0.78
AI-Majority Voting	91.11	0.82

Table 5. Agreement scores for each "AI-rater" pair.

Discussion

RD hallmark signs and particularly OCT biomarkers of DME have been investigated in several studies and post hoc analyses to assess their correlation to visual function and their prediction concerning the potential response to intravitreal pharmacotherapy.[9,11,12,44–49]

Hard exudates in the retina are one of the early signs of DR and their presence in the macula is related to diabetic maculopathy.

The quantification of HE is an important aspect due to the relationship between severity of HE in the macular region and visual loss. In addition, their assessment and monitoring during time is fundamental for the management of DME as their sedimentation in the foveal region can compromise visual recovery.[47,48]

Moreover, the regression of HE from the foveal region has been found to differ with different intravitreal treatments being greater after dexamethasone intravitreal implant compared to intravitreal anti VEGF.[49]

DRIL has been proposed as a biomarker strongly associated with visual acuity in patients with DME, particularly with worse VA and with DR severity.[50,51]

In addition, DRIL is a potential biomarker of final post-treatment VA in eyes with DME and is associated to non-responders to intravitreal anti-VEGF and steroids in DME.[34]

In the past years identification and quantification of DR signs has been performed exclusively by human grading mainly on fundus images and more recently on OCT images.

Semiautomated quantification of RD biomarkers including HE in color fundus photographs and OCT images has been proposed and has been shown to be reliable and accurate, nevertheless, manual corrections had to be applied to obtain accurate quantification.[52–56]

Deep learning models have been applied for the classification, segmentation, and detection of lesions in DR including HE using public DDR datasets of fundus images by several authors demonstrating good performance in disease classification and lower performance for microlesion recognition (detection and segmentation).[18–20]

A new approach based on the YOLOv5 deep neural network model implemented through the open-source machine learning library PyTorch has been applied for fundus lesions detections such as microaneurysms, HE, hemorrhages, and soft exudates in diabetic retinopathy.[21] In this study, a public dataset for diabetic retinopathy (DDR) of 757 fundus images was used and demonstrated the superiority of this deep learning-based approach for lesion detection compared to other studies.

Deep learning has also been used to detect different imaging biomarkers in OCT images such as intraretinal and subretinal fluid, status of the external retina, and HF.[27–30] Midena et al. using a DL model showed an accuracy of the automatic quantification of retinal fluids and retinal status ranging between 94.7 and 95.7%.

Tejero et al. using a DL approach demonstrated high accuracy in automatic detection of IRF and SRF in Early Treatment Diabetic Retinopathy Study (ETDRS) rings of OCT scans.[28]

Suciu et al. investigated the usefulness of AI-assisted diagnosis of DME using AI neural networks (EfficientNetV2 and ConvNeXt) for detection of intraretinal cystoid spaces (i.e. DME), HF, and neurosensory detachment (ND) and demonstrated high sensitivity and specifity. In detail DME detection showed a sensitivity of 0.92 and a specificity of 0.98, ND a sensitivity of 0.82 and a specificity of 0.97, HF a sensitivity of 0.97 and a specificity of 0.78. 30.

Vahadane et al. used a patch-based approach to classify OCT frames as being indicative of DME detecting and classifying patches corresponding to HE and fluid using image processing and deep learning.[57]

Singh et al. developed a Deep Learning-based algorithm for DRIL detection in a cohort of patients with DR using OCT images and demonstrated high accuracy of 88.3%, specificity of 90% and sensitivity of 82.9% in identifying DRIL.[31]

In contrast to their methodology, we employed a more performant CNN and utilized ensemble methods to exploit the best from different architectures. Thanks to this approach, we achieved an advancement in the DRIL classification task, reaching metrics that demonstrate a high level of performance in classification capability.

The identification and quantification of HE and DRIL could be challenging in clinical practice even if performed by trained retinal experts and could be time-consuming and need OCT images of high quality.

This study uses DL for automated detection of OCT biomarkers in DME patients such as HE and DRIL with the aim of facilitating their identification and quantification and making possible correlations with response to intravitreal treatment and visual function.

As highlighted by all the metrics that were considered, the AI system designed and developed reached good results in both the two main tasks. In more detail, the performance on HE detection measured using AP@0.5, Precision, and Recall, confirmed that the DL model was able to correctly identify the right regions on the images. Moreover, the obtained Pearson score indicates a statistically significant correlation between the number of HEs manually annotated and the ones predicted by the AI system, enforced by a very low P-value. The model has been positively evaluated also by expert ophthalmologists, from a qualitative point of view. In fact, despite the complexity of this task due to the small size of HEs and the possible presence of confusing elements, the predicted boxes pointed out the regions of interest. These positive results indicate a possible future clinical usefulness.

Regarding the DRIL classification, looking at the overall accuracy, it is possible to validate the remarkable results of our AI system. In particular, high values on sensitivity and specificity confirm the robustness of the model that shows no difficulties in the target classes. This fact is also supported by statistical indicators such as p-value and Kappa coefficient. Moreover, the Kappa scores per each "AI-rater" pair highlight how the AI proposed model reaches comparable performance as the pool of three specialists.

All these results were achieved thanks to the design of our pipeline, which exploits both object detection and classification, to build a complex AI system. Moreover, we also paid attention to computational costs, and we optimized the inference time of our model, getting the image analyzed in less than one second.

As previously stated, this is one of the first studies on this topic and represents a preliminary investigation of AI in retinal diseases such as diabetic retinopathy that could be implemented enriching the dataset and analyzing other biomarkers. Moreover, we plan to run more experiments on different network architectures and more complex solutions in order to further investigate on the criticalities of the studied tasks. In particular, we will focus future works on dataset improvements, additional data augmentation analysis, and more sophisticated algorithms.

In conclusion, this study demonstrated the capability of AI in identify and/or quantify OCT biomarkers in DME. Application of AI and in particular deep learning models on OCT images could fasten OCT biomarkers assessment at baseline allowing correlation with visual function and prediction of treatment.

Data availability

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Received: 11 January 2024; Accepted: 3 June 2024
Published online: 19 July 2024

References

1. Yau, J. W. *et al.* Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care* **35**(3), 556–564. https://doi.org/10.2337/dc11-1909 (2012).
2. Varma, R. *et al.* Prevalence of and risk factors for diabetic macular edema in the United States. *JAMA Ophthalmol.* **132**(11), 1334–1340. https://doi.org/10.1001/jamaophthalmol.2014.2854 (2014).
3. Starace, V. *et al.* The role of inflammation and neurodegeneration in diabetic macular edema. *Ther Adv Ophthalmol.* **13**, 25158414211055964. https://doi.org/10.1177/25158414211055963 (2021).
4. Capitão, M. & Soares, R. Angiogenesis and inflammation crosstalk in diabetic retinopathy. *J Cell Biochem.* **117**(11), 2443–2453. https://doi.org/10.1002/jcb.25575 (2016).
5. Miyamoto, K. *et al.* Prevention of leukostasis and vascular leakage in streptozotocin-induced diabetic retinopathy via intercellular adhesion molecule-1 inhibition. *Proc Natl Acad Sci U S A.* **96**(19), 10836–10841. https://doi.org/10.1073/pnas.96.19.10836 (1999).
6. Lee, H., Jang, H., Choi, Y. A., Kim, H. C. & Chung, H. Association between soluble CD14 in the aqueous humor and hyperreflective foci on optical coherence tomography in patients with diabetic macular edema. *Invest Ophthalmol Vis Sci.* **59**(2), 715–721. https://doi.org/10.1167/iovs.17-23042 (2018).
7. Virgili, G. *et al.* Optical coherence tomography (OCT) for detection of macular oedema in patients with diabetic retinopathy. *Cochrane Database Syst Rev.* **1**, CD008081. https://doi.org/10.1002/14651858.CD008081.pub3 (2015).
8. Huang, H., Jansonius, N. M., Chen, H. & Los, L. I. Hyperreflective dots on OCT as a predictor of treatment outcome in diabetic macular edema: A systematic review. *Ophthalmol Retina.* **6**(9), 814–827. https://doi.org/10.1016/j.oret.2022.03.020 (2022).
9. Sorour, O. A. *et al.* Persistent diabetic macular edema: Definition, incidence, biomarkers, and treatment methods. *Surv Ophthalmol.* **68**(2), 147–174. https://doi.org/10.1016/j.survophthal.2022.11.008 (2023).
10. Zhang, J. *et al.* Diabetic macular edema: Current understanding, molecular mechanisms and therapeutic implications. *Cells.* **11**(21), 3362. https://doi.org/10.3390/cells11213362 (2022).
11. Kwan, C. C. & Fawzi, A. A. Imaging and biomarkers in diabetic macular edema and diabetic retinopathy. *Curr Diab Rep.* **19**(10), 95. https://doi.org/10.1007/s11892-019-1226-2 (2019).
12. Chou, H. D. *et al.* Optical coherence tomography and imaging biomarkers as outcome predictors in diabetic macular edema treated with dexamethasone implant. *Sci Rep.* **12**(1), 3872. https://doi.org/10.1038/s41598-022-07604-7 (2022).
13. Diabetic Retinopathy Clinical Research Network et al. Relationship between optical coherence tomography-measured central retinal thickness and visual acuity in diabetic macular edema. *Ophthalmology*. 2007;114(3):525–536. https://doi.org/10.1016/j.ophtha.2006.06.052
14. Fickweiler, W. *et al.* Predictive value of optical coherence tomographic features in the Bevacizumab and Ranibizumab in patients with diabetic macular edema (BRDME) study. *Retina.* **38**(4), 812–819. https://doi.org/10.1097/IAE.0000000000001626 (2018).
15. Giocanti-Aurégan, A. *et al.* Functional and anatomical outcomes in patients with serous retinal detachment in diabetic macular edema treated with ranibizumab. *Invest. Ophthalmol. Vis. Sci.* **58**(2), 797–800. https://doi.org/10.1167/iovs.16-20855 (2017).
16. Seo, K. H., Yu, S. Y., Kim, M. & Kwak, H. W. Visual and morphologic outcomes of intravitreal Ranibizumab for diabetic macular edema based on optical coherence tomography patterns. *Retina.* **36**(3), 588–595. https://doi.org/10.1097/IAE.0000000000000770 (2016).
17. Pelosini, L. *et al.* Optical coherence tomography may be used to predict visual acuity in patients with macular edema. *Invest Ophthalmol Vis Sci.* **52**(5), 2741–2748. https://doi.org/10.1167/iovs.09-4493 (2011).
18. Mateen, M., Wen, J., Nasrullah, N., Sun, S. & Hayat, S. Exudate detection for diabetic retinopathy using pretrained convolutional neural networks. *Complexity.* https://doi.org/10.1155/2020/5801870 (2020).
19. Alyoubi, W. L., Abulkhair, M. F. & Shalash, W. M. Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors.* **21**(11), 3704. https://doi.org/10.3390/s21113704 (2021).
20. Tsiknakis, N. *et al.* Deep learning for diabetic retinopathy detection and classifiction based on fundus images: A review. *Comput Biol Med.* **135**, 104599. https://doi.org/10.1016/j.compbiomed.2021.104599 (2021).
21. Santos, C., Aguiar, M., Welfer, D. & Belloni, B. A new approach for detecting fundus lesions using image processing and deep neural network architecture based on YOLO model. *Sensors.* **22**(17), 6441. https://doi.org/10.3390/s22176441 (2022).
22. Samagaio, G. *et al.* Automatic macular edema identification and characterization using OCT images. *Comput Methods Programs Biomed.* **163**, 47–63. https://doi.org/10.1016/j.cmpb.2018.05.033 (2018).
23. Alsaih, K. *et al.* Machine learning techniques for diabetic macular edema (DME) classification on SD-OCT images. *BioMed Eng OnLine* **16**(1), 68. https://doi.org/10.1186/s12938-017-0352-9 (2017).
24. Syed, A. M., Hassan, T., Akram, M. U., Naz, S. & Khalid, S. Automated diagnosis of macular edema and central serous retinopathy through robust reconstruction of 3D retinal surfaces. *Comput Methods Programs Biomed.* **137**, 1–10. https://doi.org/10.1016/j.cmpb.2016.09.004 (2016).
25. Li, F., Chen, H., Liu, Z., Zhang, X. & Wu, Z. Fully automated detection of retinal disorders by image-based deep learning. *Graefes Arch Clin Exp Ophthalmol.* **257**(3), 495–505. https://doi.org/10.1007/s00417-018-04224-8 (2019).
26. Sandhu, H. S. *et al.* Automated diagnosis of diabetic retinopathy using clinical biomarkers, optical coherence tomography, and optical coherence tomography angiography. *Am J Ophthalmol.* **216**, 201–206. https://doi.org/10.1016/j.ajo.2020.01.016 (2020).
27. Midena, E. *et al.* Validation of an automated artificial intelligence algorithm for the quantification of major OCT parameters in diabetic macular Edema. *Clin Med.* **12**(6), 2134. https://doi.org/10.3390/jcm12062134 (2023).

28. Tejero, J. G. *et al.* Predicting OCT biological marker localization from weak annotations. *Sci Rep.* **13**(1), 19667. https://doi.org/10.1038/s41598-023-47019-6 (2023).
29. Wang, X. *et al.* Automated evaluation of retinal hyperreflective foci changes in diabetic macular edema patients before and after intravitreal injection. *Front Med (Lausanne).* **6**(10), 1280714. https://doi.org/10.3389/fmed.2023.1280714 (2023).
30. Suciu, C. I., Marginean, A., Suciu, V. I., Muntean, G. A. & Nicoară, S. D. Diabetic macular Edema optical coherence tomography biomarkers detected with EfficientNetV2B1 and ConvNeXt. *Diagnostics (Basel).* **14**(1), 76. https://doi.org/10.3390/diagnostics14010076 (2023).
31. Singh, R. *et al.* Deep learning algorithm detects presence of disorganization of retinal inner layers (DRIL)—An early imaging biomarker in diabetic retinopathy. *Transl Vis Sci Technol.* **12**(7), 6. https://doi.org/10.1167/tvst.12.7.6 (2023).
32. Vujosevic, S. *et al.* Hyperreflective retinal spots in normal and diabetic eyes: B-Scan and En face spectral domain optical coherence tomography evaluation. *Retina* **37**(6), 1092–1103. https://doi.org/10.1097/IAE.0000000000001304 (2017).
33. Fragiotta, S. *et al.* Significance of hyperreflective foci as an optical coherence tomography biomarker in retinal diseases: Characterization and clinical implications. *J Ophthalmol Mol* **17**(2021), 6096017. https://doi.org/10.1155/2021/6096017 (2021).
34. Sun, J. K. *et al.* Disorganization of the retinal inner layers as a predictor of visual acuity in eyes with center-involved diabetic macular edema. *JAMA Ophthalmol.* **132**(11), 1309–1316. https://doi.org/10.1001/jamaophthalmol.2014.2350 (2014).
35. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. You only look once: Unified, real-time object detection. (2015).
36. Wang, C.Y., Bochkovskiy, A., & Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. (2022); https://doi.org/10.48550/arXiv.2207.02696.
37. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **8**(6), 679–698. https://doi.org/10.1109/TPAMI.1986.4767851 (1986).
38. Stephens, R. S. Probabilistic approach to the Hough transform. *Image and Vision Computing* **9**(1), 66–71. https://doi.org/10.1016/0262-8856(91)90051-P (1991).
39. He, K., Zhang, X., Ren, S., Sun, J. Deep residual learning for image recognition. In *CVPR* 770–778 (2016).
40. Liu, Z. et al. A ConvNet for the 2020s. In *CVPR*, 11966–11976 (2022) https://doi.org/10.1109/CVPR52688.2022.01167.
41. Radosavovic, I., Kosaraju, R.P., Girshick, R., He, K., Dollár, P. Designing network design spaces. In *CVPR*. 10425–10433 (2020). https://doi.org/10.1109/CVPR42600.2020.01044.
42. Bekkar, M., Djemaa, K. H. & Alitouche, T. A. Evaluation measures for models assessment over imbalanced data sets. *J. Inf. Eng. Appl.* **3**(10), 2013 (2013).
43. Fleiss, J. L. Measuring nominal scale agreement among many raters. *Psychol. Bull.* **76**(5), 378–382 (1971).
44. Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics.* **33**(1), 159–174 (1977).
45. Ashraf, M., Souka, A. & Adelman, R. Predicting outcomes to anti-vascular endothelial growth factor (VEGF) therapy in diabetic macular oedema: A review of the literature. *Br J Ophthalmol.* **100**(12), 1596–1604. https://doi.org/10.1136/bjophthalmol-2016-308388 (2016).
46. Parravano, M., Costanzo, E. & Querques, G. Profile of non-responder and late responder patients treated for diabetic macular Edema: Systemic and ocular factors. *Acta Diabetol.* **57**(8), 911–921. https://doi.org/10.1007/s00592-020-01496-7 (2020).
47. Cusick, M. *et al.* Histopathology and regression of retinal hard exudates in diabetic retinopathy after reduction of elevated serum lipid levels. *Ophthalmology.* **110**(11), 2126–2133. https://doi.org/10.1016/j.ophtha.2003.01.001 (2003).
48. Møller, F. & Bek, T. The relation between visual acuity, fixation stability, and the size and location of foveal hard exudates after photocoagulation for diabetic maculopathy: A 1-year follow-up study. *Graefes Arch Clin Exp Ophthalmol.* **241**(6), 458–462. https://doi.org/10.1007/s00417-003-0661-5 (2003).
49. Mehta, H. *et al.* Efficacy of dexamethasone versus bevacizumab on regression of hard exudates in diabetic maculopathy: Data from the BEVORDEX randomised clinical trial. *Br J Ophthalmol.* **100**(7), 1000–1004. https://doi.org/10.1136/bjophthalmol-2015-307797 (2016).
50. Singuri, S. Clinical utility of spectral-domain optical coherence tomography marker disorganization of retinal inner layers in diabetic retinopathy. *Ophthalmic Surg Lasers Imaging Retina.* **54**(12), 692–700. https://doi.org/10.3928/23258160-20231031-02 (2023).
51. Das, R., Spence, G., Hogg, R. E., Stevenson, M. & Chakravarthy, U. Disorganization of inner retina and outer retinal morphology in diabetic macular Edema. *JAMA Ophthalmol.* **136**(2), 202–208. https://doi.org/10.1001/jamaophthalmol.2017.6256 (2018).
52. Marupally, A. G. *et al.* Semi-automated quantification of hard exudates in colour fundus photographs diagnosed with diabetic retinopathy. *BMC Ophthalmo.* https://doi.org/10.1186/s12886-017-0563-7 (2017).
53. Sasaki, M. *et al.* Quantitative measurement of hard exudates in patients with diabetes and their associations with serum lipid levels. *Invest Ophthalmol Vis Sci.* **54**(8), 5544–5550. https://doi.org/10.1167/iovs.13-11849 (2013).
54. Kessler, L. J., Bagautdinov, D., Łabuz, G., Auffarth, G. U. & Khoramnia, R. Semi-automated quantification of retinal and choroidal biomarkers in retinal vascular diseases: Agreement of spectral-domain optical coherence tomography with and without enhanced depth imaging mode. *Diagnostics (Basel)* **12**(2), 333. https://doi.org/10.3390/diagnostics12020333 (2022).
55. Tao, L. *et al.* Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. *Inf. Sci.* **2019**(501C), 511–522. https://doi.org/10.1016/j.ins.2019.06.011 (2019).
56. Porwal, P. *et al.* IDRiD: Diabetic retinopathy—segmentation and grading challenge. *Med Image Anal.* **59**, 101561. https://doi.org/10.1016/j.media.2019.101561 (2020).
57. Vahadane, A., Joshi, A., Madan, K., & Dastidar, T.R. Detection of diabetic macular edema in optical coherence tomography scans using patch based deep learning. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 1427–1430 (Washington, DC, USA, 2018) https://doi.org/10.1109/ISBI.2018.8363840.

Author contributions

L.T. and R.M. contributed to the conception of the work; A.R. and F.F. carried out the acquisition of data; M.P., D.D.I., V.O and P.V. designed and carried out analysis and interpretation of data; L.T wrote the main manuscript text; R.M. and E.M. reviewed the paper;

Competing interests

Lisa Toto, Anna Romano, Marco Pavan, Dante Degl'Innocenti, Valentina Olivotto, Federico Formenti, Pasquale Viggiano, Edoardo Midena and Rodolfo Mastropasqua declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.