tvst

Artificial Intelligence

Neurosymbolic AI Framework for Explainable Retinal Disease Classification From OCT Images

Aleksandar Miladinović[1], Alessandro Biscontin[1], Miloš Ajčević[2], Simone Kresevic[2], Agostino Accardo[2], Daniele Tognetto[3], and Leandro Inferrera[3]

[1] Institute for Maternal and Child Health IRCCS "Burlo Garofolo," Trieste, Italy
[2] Department of Engineering and Architecture, University of Trieste, Trieste, Italy
[3] Department of Medicine, Surgery and Health Sciences, Ophthalmology Clinic, University of Trieste, Trieste, Italy

Correspondence: Aleksandar Miladinović, Institute for Maternal and Child Health IRCCS Burlo Garofolo, Italy, Via dell'Istria 65, Trieste 34100, Italy. e-mail: aleksandar.miladinovic@burlo.trieste.it

Received: July 3, 2025
Accepted: December 3, 2025
Published: January 6, 2026

Keywords: neurosymbolic AI; OCT; retinal imaging; artificial intelligence; retinal dystrophies

Citation: Miladinović A, Biscontin A, Ajčević M, Kresevic S, Accardo A, Tognetto D, Inferrera L. Neurosymbolic AI framework for explainable retinal disease classification from OCT images. Transl Vis Sci Technol. 2026;15(1):6, https://doi.org/10.1167/tvst.15.1.6

Purpose: Accurate classification of retinal diseases such as dry age-related macular degeneration, wet AMD, epiretinal membrane, full-thickness macular hole (MH), lamellar MH, and central serous chorioretinopathy (CSC) is essential for effective treatment and clinical decision-making. Traditional deep learning models, however, often struggle with imbalanced datasets and lack interpretability, limiting their translational applicability in ophthalmology.

Methods: We propose a neurosymbolic framework that integrates a convolutional neural network (CNN) with a symbolic reasoning layer based on expert-defined clinical rules. A total of 10,846 optical coherence tomography images were retrospectively collected and categorized into seven diagnostic classes: dry AMD, wet AMD, epiretinal membrane, full-thickness MH, lamellar MH, central serous chorioretinopathy, and healthy retinas.

Results: Our neurosymbolic model achieved macro-precision 0.83, recall 0.82, and F_1 0.81, on internal dataset, having slightly better performance than the CNN (0.64/0.83/0.68). On the external dataset, it retained superior performance, macro-precision 0.85, recall 0.79, F_1 0.78, versus the CNN (0.73/0.64/0.59).

Conclusions: Our hybrid neurosymbolic framework introduces a unified paradigm that couples symbolic reasoning with a conventional CNN, improving diagnostic performance while delivering transparent, clinically interpretable decisions. It is particularly effective for rare and complex conditions that often challenge end-to-end deep learning models.

Translational Relevance: By integrating symbolic clinical logic with visual pattern recognition, the neurosymbolic model fosters trust in artificial intelligence–assisted diagnostics and supports precise, explainable decision-making in retinal care.

Introduction

Optical coherence tomography (OCT) has become one of the most extensively used diagnostic modalities in ophthalmology, owing to its noninvasive nature and its capacity to generate high-resolution cross-sectional images of the retinal structure.[1,2] This imaging technique plays a pivotal role in the assessment and diagnosis of a broad spectrum of retinal disorders, including both the dry and wet forms of age-related macular degeneration (AMD), retinal vein and artery occlusions, central serous chorioretinopathy (CSC), epiretinal membrane (ERM), macular hole (MH), diabetic retinopathy, and glaucoma.[3–16]

A precise differentiation between normal and pathological OCT findings is essential for accurate diagnosis and for the subsequent selection of the most appropriate therapeutic approach.[17,18] Such differentiation is particularly critical in retinal diseases, where OCT features provide key diagnostic clues, each displaying distinctive morphological patterns or signs of retinal abnormalities (further denoted with prefix "s"). Wet AMD (wAMD) is primarily characterized by the macular neovascularization retinal abnormalities sign (sMNV), defined as an aberrant vascular proliferation beneath or within the retina, while subretinal deposits known as drusen sign (sD), may also be present.[19–21] sMNV often results in the formation of intraretinal fluid (sIRF), visualized on OCT as hyporeflective intraretinal spaces, and subretinal fluid (sSRF), which appears as serous exudates accumulating between the neurosensory retina and the retinal pigment epithelial (RPE).[22] Conversely, dry AMD (dAMD) is typically associated with the presence of sD.[23] In advanced stages, it is often accompanied by increased choroidal backscattering (sBS), which represents a stronger reflection of incident light from deeper choroidal structures, and by hypertransmission owing to changes in tissue composition. This finding is indicative of RPE atrophy or loss.[24] Similarly, CSC is characterized by serous detachment of the neurosensory retina, accumulation of sSRF, and characteristic pachychoroid features, including increased choroidal thickness.[25]

Among the structural abnormalities detectable via OCT, the presence of an ERM represents not only a distinct retinal pathology, but also a relevant morphological sign (sERM) of particular clinical significance.[26] Also referred to as macular puckers or cellophane maculopathy, sERMs consist of fibrocellular proliferations on the inner retinal surface, particularly in the macular area, and may be associated with intraretinal distortion in advanced cases.[27] This condition can result in visual distortion and a decrease in central visual acuity. Another important OCT finding is the MH, which may present as a lamellar MH (LMH) or a full-thickness MH (FTMH), denoting partial or complete defects of the central retinal layers, respectively, and is typically associated with significant central visual impairment.[28]

Retinal thickening (sRT) denotes an overall increase in retinal thickness relative to normal values.[29] This finding may be accompanied by retinal distortion (sRDIST), an irregular deformation of the retinal layers resulting from tractional or proliferative processes, and can occur in association with conditions such as CSC, AMD, or LMH.[30]

Finally, RPE detachment is defined as a separation between the retinal pigment epithelium and Bruch's membrane, which may result from serous, drusenoid, or fibrovascular causes and is a hallmark feature in both AMD and CSC.[31,32]

The careful analysis of these characteristic retinal features through OCT imaging facilitates the differentiation of various retinal diseases and supports informed clinical decision-making.[33] However, interpreting OCT scans remains a complex and time-intensive task that often relies on the expertise of highly trained clinicians.

In recent years, artificial intelligence (AI), particularly deep learning methods based on convolutional neural networks (CNNs), has become increasingly prevalent in ophthalmology.[34] Although CNNs have shown impressive accuracy in identifying common conditions, their interpretability and robustness are limited, especially when dealing with nonstandardized or under-represented pathology datasets.[35] These shortcomings hinder clinical adoption, especially because the decisions made by black box models often lack transparency and interpretability.[36]

To overcome these limitations, we adopted a neurosymbolic learning paradigm that integrates deep learning to identify the presence of retinal abnormality signs and symbolic reasoning that guide or constrain the learning process. This model architecture enables clinician-like decision-making and enhances model transparency. Neurosymbolic models can not only recognize patterns, but also provide structured justifications for their classifications,[37,38] making them more suitable for clinical environments where trust and interpretability are paramount.

The aim of this study was to propose and evaluate a neurosymbolic AI framework that tightly integrates CNN-based feature extraction with a symbolic clinical reasoning engine, enabling bidirectional interaction between the two components for the classification of retinal pathologies, specifically dAMD, wAMD, ERM, FTMH, LMH, CSC, and healthy retinas, while assessing its diagnostic accuracy and clinical interpretability, particularly in the context of under-represented conditions.

Materials and Methods

Study Population and Data Acquisition

A total of 10,846 spectral-domain OCT B-scans were retrospectively collected between January 2017 and September 2022 at the University Eye Clinic of Trieste, Italy, using the Spectralis OCT system (Heidelberg Engineering, Heidelberg, Germany). Acquisition parameters included a central wavelength of 815 nm, axial resolution of 3.9 μm/pixel, lateral resolution of 5.7 μm/pixel, and image dimensions of 768 × 496 pixels. Both horizontal and vertical fovea-centered line scans were included.

All images were reviewed independently by two experienced retinal ophthalmologists, who verified diagnostic correctness. Cases with diagnostic disagreement or uncertain findings were jointly reevaluated,

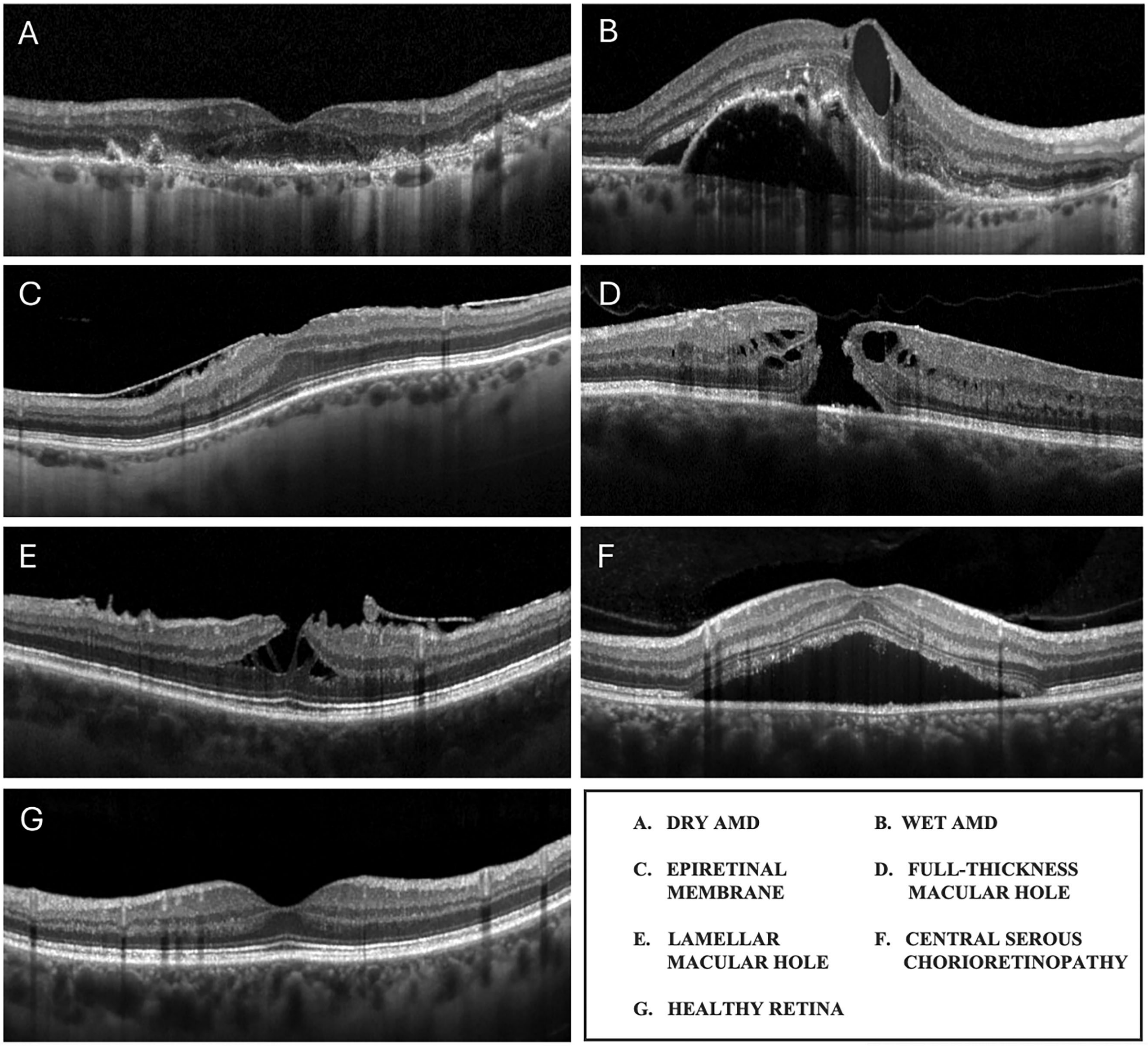

Figure 1. Representative OCT images from a set of seven scans, including six pathological cases and one healthy retina. Panel labels indicate: (**A**) dry age-related macular degeneration (dry AMD), (**B**) neovascular ("wet") age-related macular degeneration (wet AMD), (**C**) epiretinal membrane (ERM), (**D**) full-thickness macular hole (FTMH), (**E**) lamellar macular hole (LMH), (**F**) central serous chorioretinopathy (CSC), and (**G**) healthy retina.

and those lacking consensus were excluded to ensure diagnostic purity and reproducibility of class definitions.

Images were included if they exhibited a clearly identifiable foveal contour with preserved outer retinal layers, a Spectralis Quality Score of 23 or higher, and a characteristic morphology consistent with one of the diagnostic categories.

Images were excluded if they showed severe motion artifacts, segmentation errors, incomplete or off-center foveal coverage, duplicate acquisitions, or concurrent retinal diseases that could confound the primary diagnosis (e.g., retinal vein occlusion and diabetic retinopathy in AMD). No image denoising, flattening, or additional preprocessing was applied before analysis.

This study was conducted in accordance with the Declaration of Helsinki and approved by the Regional Ethics Committee of Friuli Venezia Giulia (CEUR No. 17094/2022). Written informed consent was obtained from all participants for the anonymized use of their imaging data for research.

After quality control, the final dataset comprised seven diagnostic categories: dAMD, wAMD, ERM, FTMH, LMH, CSC, and healthy retina (Fig. 1). Diagnostic definitions followed established clinical

Table 1. Dataset Composition and Overlap of Pathological Features Across Diagnostic Categories

No. of Pathologies	No. of Images
1	3651
2	1824
3	280
4	18
5	3

criteria: dAMD was defined by drusen and/or retinal pigment epithelium irregularities without intraretinal or subretinal fluid; wAMD required macular neovascularization with associated intraretinal or subretinal fluid; ERM corresponded with a hyper-reflective epiretinal layer occasionally associated with intraretinal cystic changes in advanced cases; FTMH and LMH were characterized by full- and partial-thickness defects of the foveal tissue, respectively, resulting in complete or partial disruption of the normal foveal contour; CSC was defined by a localized serous detachment of the neurosensory retina; and healthy retinas exhibited no detectable structural abnormalities. The number of images per category was as follows: 2156 for dAMD, 579 for wAMD, 3574 for ERM, 101 for FTMH, 275 for LMH, 97 for CSC, and 4700 for healthy retinas.

In addition, every image was annotated for the ten retinal signs (sBS, sMNV, sD, sIRF, sSRF, sMH, sERM, sRT, sRDIST, and sRDET) exclusively for evaluation; these labels did not contribute to training the neurosymbolic model, which should infer signs under algorithmic supervision.

The composition of the dataset and the frequency of overlapping pathological features across diagnostic categories are summarized in Table 1.

Neurosymbolic Approach

The overview of the proposed neurosymbolic framework is represented in Figure 2. An input OCT image undergoes bottom-up statistical learning via a CNN, which estimates the probabilities of 10 clinically retinal abnormality signs: (1) sBS, (2) sMNV, (3) sD, (4) sIRF, (5) sSRF, (6) sMH, (7) sERM, (8) sRT, (9) sRDIST, and (10) sRDET. These features are then passed to the symbolic reasoning module, which applies expert-defined logical rules (top–down approach) to infer the likelihoods of specific retinal pathologies. The final output includes the estimated decision tree and probabilities for seven diagnostic categories: dAMD, wAMD, ERM, FTMH, LMH, CSC, and healthy retina. This transparent pipeline integrates visual evidence and expert knowledge to yield interpretable and clinically aligned predictions. In this type of model, the CNN and symbolic inference engine operate in a tightly coupled manner rather than as independent modules. This bidirectional interaction allows the CNN to refine its internal representations based on symbolic feedback, representing a key novelty compared to traditional logic rule-based systems[39,40] that only consume outputs from the CNN without influencing its internal feature learning. Moreover, the backbone CNN was not explicitly trained nor pretrained on the 10 retinal signs; instead, these were learned through algorithmic supervision guided by the symbolic rules, ensuring that sign

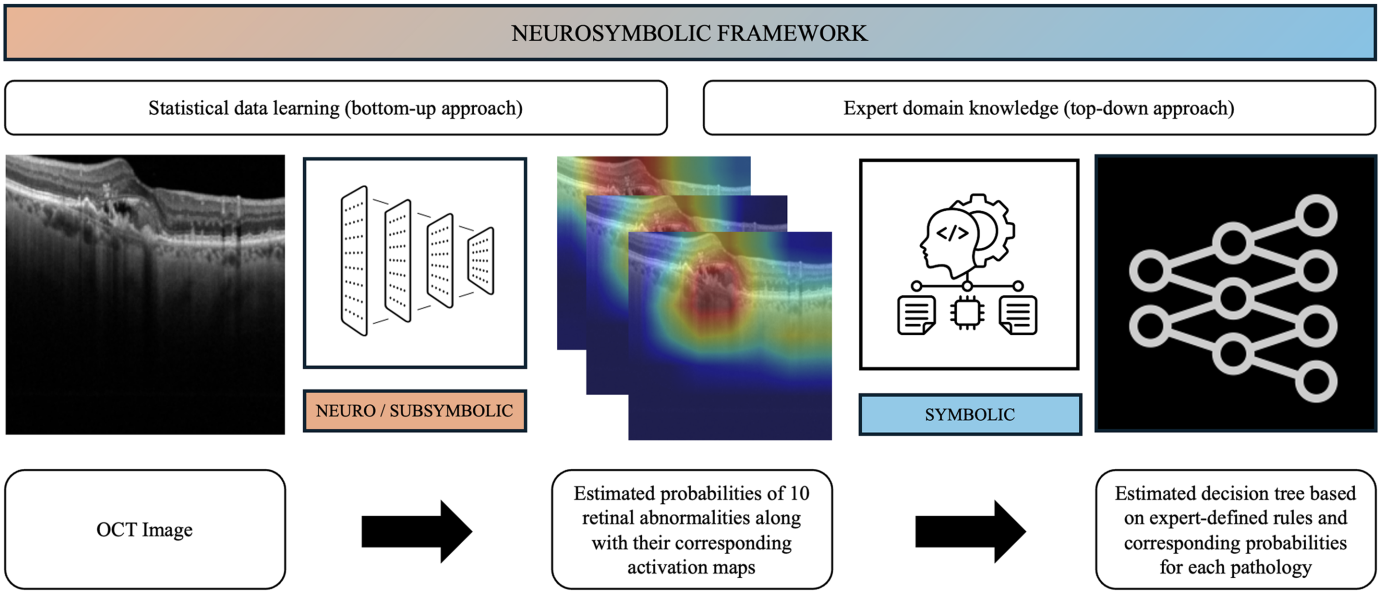

Figure 2. Overview of the proposed neurosymbolic framework for OCT-based retinal disease classification.

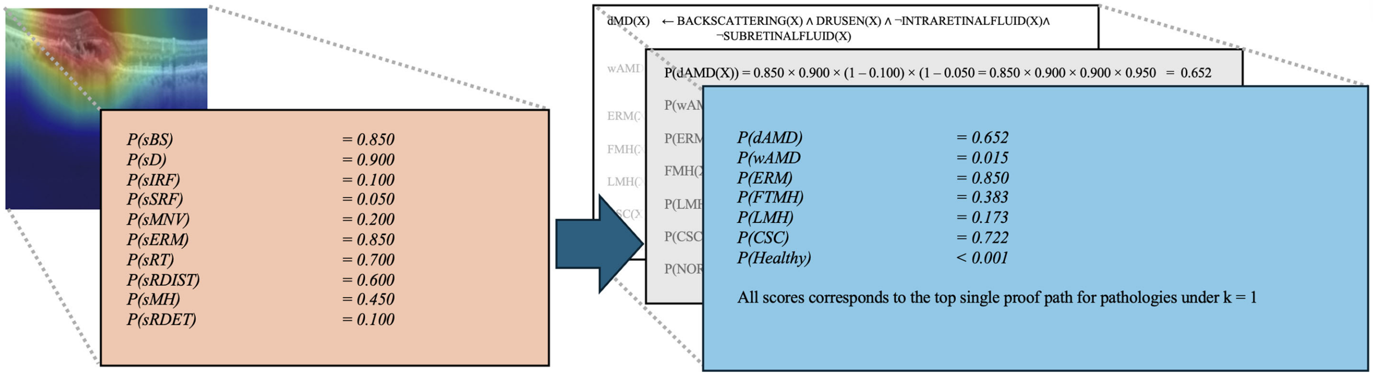

Figure 3. Example of symbolic reasoning via k-top differential provenance.

representations emerged naturally from the reasoning process rather than from direct labeling.

In our framework, the perceptual backbone (a ResNet-50 CNN pretrained on ImageNet) is responsible for extracting a vector of features corresponding with the probability of presence of clinically meaningful 10 retinal abnormality signs. These features are then passed to a symbolic reasoning engine that evaluates them against a set of logical rules codified using first-order logic. Each rule defines the necessary and/or sufficient combination of signs that are indicative of a specific retinal pathology. Importantly, the logical operations do not collapse these inputs into binary true/false outcomes; instead, they propagate and transform uncertainty, yielding probabilistic scores for each pathology. This formulation allows the model to mimic the clinical reasoning process of an ophthalmologist, who would arrive at a diagnosis based on the presence or absence of specific visual features.

At the CNN–symbolic interface, no hard thresholds are applied. The model propagates continuous sign probabilities into the reasoning layer, enabling inference on continuous evidence and preserving diagnostic uncertainty. Outputs are continuous pathology probabilities rather than single categorical assignments (see Fig. 3 for an example). In this framework, retinal sign annotations are used only for evaluation; during training, the system relies on algorithmic supervision to infer both the probabilistic presence and interaction of signs without explicit sign labels. The decision mechanism (Fig. 3) shows how intermediate CNN outputs propagate through symbolic rules to produce clinically interpretable diagnostic probabilities.

To implement this architecture, we used the Scallop neurosymbolic programming language,[41,42] which integrates PyTorch-based CNN deep learning with a symbolic inference engine. Scallop allows the fusion of neural networks and logical reasoning by compiling symbolic rules directly into differentiable computational graphs. Each logical decision in the symbolic engine is accompanied by a probabilistic score, allowing uncertainty-aware inference. Furthermore, we used k-top provenance reasoning,[42] a method for tracing the most probable justifications for each inferred label.

Finally, the neurosymbolic model does not apply a hard decision rule at the pathology level during training. Instead, it outputs continuous probability scores for each diagnostic category, computed by propagating the CNN's soft sign likelihoods through the symbolic reasoning layer. Only during evaluation phase, the threshold for each pathology is set to 0.5. If, for a given pathology, the predicted value exceeds this threshold, the pathology is considered present in the image.

k-Top Differential Provenance and Symbolic Reasoning

This provenance method identifies the top-k most probable minimal proof paths (rule satisfaction chains) that lead to a particular diagnosis. Each path's score is computed as the product of the probabilities of the signs involved in the rule, with negated terms replaced by (1 – probability). The total probability for a pathology is obtained by summing the scores of the top-k highest-weighted proof paths. Thus, unlike traditional rule-based models, no binarization of sign probabilities is performed before reasoning. The symbolic inference engine integrates soft probabilities to compute final pathology likelihoods through weighted proof paths, ensuring consistency between perceptual uncertainty and logical reasoning.

The total probability for a pathology is obtained by summing the scores of the top-k highest-weighted proof paths. The CNN predicts the probability of presence for each intermediate sign, and the symbolic reasoning layer aggregates these using logical rules to infer the final pathology label. In this framework, each rule operates probabilistically, and the final classification probabilities are derived using k-top differential provenance.[41,42]

Symbolic Rules and Clinical Knowledge Representation

These symbolic rules were constructed manually through an interdisciplinary collaboration, guided by widely accepted clinical criteria and diagnostic guidelines for each retinal pathology. Each rule encodes expert-defined logic that links visible OCT abnormalities with specific disease entities, effectively translating implicit diagnostic reasoning into a machine-readable symbolic format. For example, the rule for dAMD requires the presence of sD and sBS (see rule 2), which are hallmark features of early and intermediate dry macular degeneration, while explicitly excluding intraretinal and sSRF, because their presence typically indicates exudative transformation.[4,19–23] In contrast, the rule for wAMD (see rule 3) integrates fluid indicators (sIRF or sSRF) with sMNV, sD, and sBS. This combination reflects the active exudative process, choroidal neovascular complex, and associated tissue disruption that characterizes neovascular AMD.[21] The rule for FTMH (see rule 4) captures the distinct feature of a macular opening through all retinal layers, usually accompanied by structural changes such as backscattering, which reflects photoreceptor loss and disruption.[43] The LMH rule (see rule 5) includes the presence of sMH and sERM, but requires the absence of sRT, to distinguish it from tractional or full-thickness holes.[44] This reflects the partial retinal defect often linked to tangential traction and remodeling, without the elevation typically caused by thickening or edema. The ERM rule (see rule 6) is relatively permissive: it requires either the presence of sERM alone or sERM in conjunction with sIRF.[26] This accounts for the fact that many sERMs are asymptomatic, but may also lead to fluid accumulation and distortion in advanced stages.[26] For CSC (see rule 7), the rule centers on aSRF and explicitly excludes signs commonly associated with AMD such as sD, sMNV, and sIRF. This logic reflects the pathophysiology of CSC, where serous detachment occurs without significant neovascular activity.[25] Finally, the rule for healthy retinas requires the absence of all 10 defined abnormal signs (see rule 8).

To formalize these symbolic rules, let $S = \{s_1, s_2, \ldots, s_n\}$ denote the set of binary signs predicted by the CNN, and $D = \{d_1, d_2, \ldots, d_x\}$ the set of diagnosis labels. Each diagnosis d_E is defined by a rule in the form:

$$dE\,(x) \leftarrow \varphi\,(s_1\,(x)\,,\ s_2\,(x)\,,\ldots,s_n\,(x))\,, \tag{1}$$

where $\boldsymbol{\varphi}$ is a Boolean formula encoding the clinical logic for d_E. Crucially, no hard decision threshold is applied at this interface. The CNN outputs probabilities $p(s_i|x) \in [0, 1]$, and the symbolic layer computes a diagnostic probability $p(d_e/x) = g\_\varphi(p(s_1/x), \ldots, p(s_n/x))$, rather than a binary true/false value. Thus, the absence or presence of any single sign does not strictly rule out a diagnosis; instead, each sign modulates the final probability through the rule structure. In practice, logical AND and OR are realized by differentiable aggregations determined by the provenance semantics of the reasoning engine. In our implementation we use k-Top-Diff proofs provenance (k = 1), which aggregates the top-k most supportive proof paths for φ in a differentiable manner, propagating uncertainty from the sign probabilities to $p(d_e/x)$ and yielding probabilistic, clinically interpretable outputs. Having established the differentiable logical interface, the specific pathology rules used in this study are formalized as follows.

For example, dAMD and wAMD are defined as:

$$\begin{aligned} dAMD\,(x) \leftarrow\ & sBS\,(x) \wedge sD\,(x) \wedge \neg sIRF\,(x) \\ & \wedge\neg sSRF\,(x) \wedge \neg sMNV\,(x) \end{aligned} \tag{2}$$

$$\begin{aligned} wAMD\,(x) \leftarrow\ & sBS\,(x) \wedge sD\,(x) \wedge sMNV\,(x) \\ & \wedge (sIRF\,(x) \vee sSRF\,(x)) \wedge (sRT\,(x) \\ & \vee\ sRDIST\,(x) \vee sRDET\,(x)). \end{aligned} \tag{3}$$

Other pathologies are expressed analogously as conjunctions and/or disjunctions of characteristic signs:

$$FTMH\,(x) \leftarrow sBS\,(x) \wedge sMH\,(x) \tag{4}$$

$$\begin{aligned} LMH\,(x) \leftarrow\ & sMH\,(x) \wedge sERM\,(x) \\ & \wedge \neg sRT\,(x) \wedge sRDIST\,(x) \end{aligned} \tag{5}$$

$$\begin{aligned} ERM\,(x) \leftarrow\ & sERM\,(x) \wedge (sIRF\,(x) \\ & \vee\ sRT\,(x) \vee sRDIST\,(x)) \end{aligned} \tag{6}$$

$$\begin{aligned} CSC\,(x) \leftarrow\ & sSRF\,(x) \wedge \neg sD\,(x) \\ & \wedge \neg sMNV\,(x) \wedge \neg sIRF\,(x)\,. \end{aligned} \tag{7}$$

Finally, healthy retinas are defined by the absence of all pathological signs:

$$\begin{aligned} Healthy\,(x) \leftarrow\ & \neg sBS\,(x) \wedge \neg sMNV\,(x) \wedge \neg sD\,(x) \\ & \wedge \neg sIRF\,(x) \wedge \neg sSRF\,(x) \\ & \wedge \neg sMH\,(x) \wedge \neg sERM\,(x) \neg sRT\,(x) \\ & \wedge \neg sRDIST\,(x) \wedge \neg sRDET\,(x)\,. \end{aligned} \tag{8}$$

Example of the Symbolic Reasoning

The example depicted in Figure 3 highlights the capability of the neurosymbolic model to reason over structured clinical logic while retaining probabilistic sensitivity to intermediate CNN outputs. In this study, we adopted a k-top provenance strategy with $k = 1$, meaning that only the single most probable inference path was used to derive each diagnosis. This choice reflects a clinically intuitive decision-making process, wherein the most likely explanation is typically prioritized for practical interpretation and action. Nonetheless, the underlying framework fully supports $k > 1$, allowing multiple high-probability reasoning paths to be aggregated.

Deep Learning Backbone: Modified ResNet-50 for Multilabel OCT Feature Extraction

To establish a conventional deep-learning baseline, we trained the same ImageNet-pretrained ResNet-50 directly on the pathology labels (dAMD, wAMD, ERM, FTMH, LMH, CSC, and Healthy). The final classification layer was replaced with a 7-unit sigmoid head to model the nonexclusive, multilabel setting of OCT findings. The training protocol (optimizer, loss, early stopping, and data split) was identical to the neurosymbolic model to ensure a fair comparison.

Model Training and Evaluation

Both the baseline and neurosymbolic frameworks used an identical training protocol. The backbone was trained using the Adam optimizer with an initial learning rate of 1×10^{-4}, $\beta_1 = 0.9$, and $\beta_2 = 0.9$. Given the class imbalance typical of clinical cohorts (e.g., fewer CSC/FTMH cases relative to Healthy), we adopted asymmetric focal loss (with parameters $\gamma_+ = 0.4$; $\gamma_- = 0.3$; and clip $= 0.01$) to emphasize rare positives and de-emphasize easy negatives; no oversampling or class-balanced sampling was used, so reported performance reflects realistic prevalence. Training was performed for a maximum of 100 epochs with a batch size of 32. To mitigate overfitting, early stopping was applied based on the validation loss, with training halted if no improvement was observed for 10 consecutive epochs. The model parameters corresponding with the lowest validation loss were retained for final evaluation.

Both the baseline CNN and the neurosymbolic model were trained and evaluated on identical data partitions and preprocessing to ensure a fair comparison. We performed a stratified, patient-level split into training (80%), validation (10%), and test (10%) sets: each subject received a unique identifier, and all images belonging to that subject, including both horizontal and vertical fovea-centered B-scans and, when present, multiple visits, were hierarchically grouped and assigned en bloc to a single partition to prevent leakage across eyes or timepoints while preserving class proportions. All OCT images underwent the same preprocessing (resize to 224 × 224 and intensity normalization).

Neurosymbolic Model

Each logical expression (2)–(8) is compiled into a differentiable program using Scallop, which constructs an inference graph that operates jointly with the CNN backbone. During training, the network receives two learning signals: (1) the primary multilabel loss applied to the final pathology predictions, and (2) gradients propagated through the differentiable logical operators that map sign probabilities to pathology probabilities. Thus, the symbolic component does not function as a post hoc reasoning module, but directly contributes to the optimization objective.

Let p_s denote the predicted probabilities of the ten retinal signs and $p_d = \varphi(p_s)$ the differentiable rule-based mapping to pathology probabilities. The global loss is:

$$L = L_{patholgy}\left(p_d,\, y_d\right),$$

where $\varphi(\cdot)$ denotes Scallop's differentiable inference using top-k provenance. No additional sign-level or consistency penalties are applied; instead, sign representations emerge implicitly, because minimizing the pathology loss requires learning sign patterns consistent with the logical rules. This constitutes algorithmic supervision, meaning that intermediate sign labels are not used during training (only for evaluation).

During each batch, Scallop back-propagates gradients from the rule-based pathology predictions through the logical operators to the CNN sign logits, refining feature representations to satisfy clinical structure. This differentiable coupling distinguishes our method from pipelines where a CNN predicts signs and a fixed rule engine consumes them without influencing feature learning. The CNN produces continuous probabilities for the ten signs (sBS, sMNV, sD, sIRF, sSRF, sMH, sERM, sRT, sRDIST, and sRDET), which are passed to the symbolic layer without thresholding. The symbolic module operates directly on these soft inputs and outputs continuous pathology probabilities, preserving uncertainty. The system thereby combines neural perception and symbolic reasoning into a single end-to-end model. The overall neurosymbolic architecture is illustrated in Figure 4.

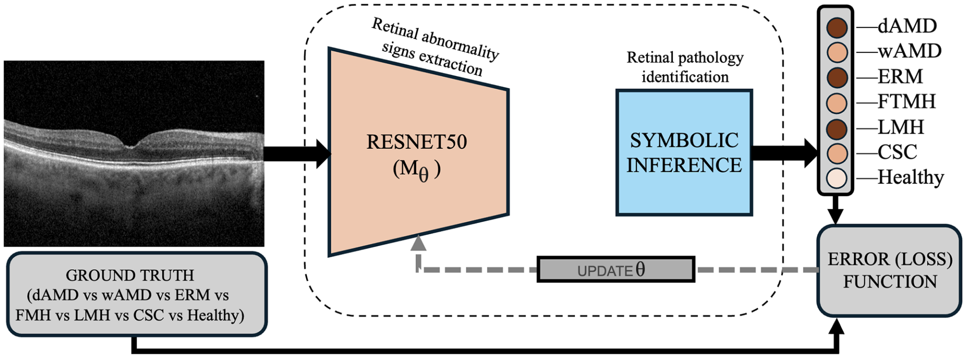

Figure 4. Overview of the implementation of the proposed neurosymbolic architecture for retinal pathology classification. OCT images are first processed by a ResNet50-based CNN ($M\theta$), which predicts the probabilities of retinal abnormality signs (e.g., sIRF, sD, and sERM). These sign probabilities are then passed to a symbolic inference module that applies structured clinical logic to determine the most probable retinal pathology (e.g., dAMD, wAMD, and ERM). The entire system is trained end-to-end by backpropagating the classification error derived from the final diagnosis predictions, enabling algorithmic supervision without requiring intermediate sign-level labels.

Baseline Deep Learning Model

As a reference system, we implemented a direct ResNet-50 multilabel classifier trained to predict final retinal pathologies directly from OCT images. The model's convolutional layers were initialized from ImageNet weights, and the final fully connected layer was replaced with a seven-node sigmoid-activated head, corresponding with the diagnostic categories (dAMD, wAMD, ERM, FTMH, LMH, CSC, and healthy retinas). The remaining architectural, optimization, and early stopping settings were identical to those described elsewhere in this article.

Evaluation Metrics and Procedure

For both models, final pathology predictions were evaluated on the test set using macro-precision, macro-recall, and macro-F_1 score. The macro-F_1 score was used as the primary outcome measure, ensuring equal weighting of all classes despite dataset imbalance. Evaluation procedures were identical for both frameworks, differing only in the additional sign-level analysis performed for the neurosymbolic model. To assess the internal validity of the neurosymbolic inference process, we performed a dedicated sign-level evaluation on the held-out test set. Each image was manually annotated for the ten retinal signs (sBS, sMNV, sD, sIRF, sSRF, sMH, sERM, sRT, sRDIST, and sRDET) solely for evaluation purposes. During training, these labels were not accessed or used for loss computation. For each sign, the model's probabilistic output was binarized using a fixed decision threshold of 0.5, and precision, recall, F_1 score, and classification accuracy were computed.

External Validation

External validation was conducted using the publicly available OCTID dataset (OpenICPSR 108503, version 1),[45] which originates from a different clinical center and imaging device than our internal Heidelberg Spectralis cohort, thereby serving as a true external site validation. OCTID includes four diagnostic categories: AMD, CSC, FTMH, and Healthy. To ensure compatibility with our seven-class internal model, we considered only dAMD predictions as a single AMD category for external evaluation, following the dataset's labeling structure. No other pathology classes in our internal taxonomy are present in OCTID. Both the neurosymbolic and baseline CNN models were evaluated under identical preprocessing and inference procedures. All OCTID scans were resized to 224 × 224 pixels. External testing was limited to final pathology classification to enable a direct comparison of cross-domain generalization between models.

Results

Neurosymbolic Model Performance

The classification performance of the neurosymbolic model across the seven diagnostic categories is summarized in Table 2. The highest F_1 scores were obtained for FTMH (0.92), Healthy (0.92), and ERM (0.85). Moderate performance was observed for dAMD (0.78), LMH (0.84), and CSC (0.72), and the lowest value was recorded for wAMD (0.65). Overall, the model achieved macro-precision = 0.83, macro-recall = 0.82, and macro-F_1 = 0.81 across all categories ($n = 1161$).

Table 2. Classification Performance of the Neurosymbolic Model Across Seven Retinal Disease Categories

Pathology	Precision	Recall	F_1 Score	Support
dAMD	0.92	0.68	0.78	243
wAMD	0.66	0.63	0.65	49
ERM	0.99	0.75	0.85	352
FTMH	0.86	1.00	0.92	12
LMH	0.81	0.87	0.84	30
CSC	0.60	0.90	0.72	10
Healthy	0.96	0.89	0.92	465
Macro AVG	0.83	0.82	0.81	1161

Precision, recall, F_1 score and number of supporting images are reported per class, along with the macro-averaged performance across all classes.

Table 3. Performance of Sign Detection in the Neurosymbolic Model

Retinal Sign	Precision	Recall	F_1 Score	Accuracy
sBS	0.36	0.76	0.49	0.84
sMNV	0.64	0.37	0.47	0.96
sD	0.47	0.55	0.50	0.83
sIRF	0.40	0.92	0.56	0.71
sSRF	0.80	0.65	0.72	0.96
sMH	0.91	0.84	0.88	0.99
sERM	0.92	0.89	0.90	0.95
sRT	0.75	0.79	0.77	0.81
sRDIST	0.54	0.89	0.67	0.62
sRDET	0.22	0.62	0.33	0.67
Macro AVG	–	0.60	0.73	0.63

Performance metrics were computed for each of the ten retinal signs (sBS, sMNV, sD, sIRF, sSRF, sMH, sERM, sRT, sRDIST, and sRDET) on the held-out test set. Values represent precision, recall, F_1 score, and classification accuracy, respectively. Macro-averaged results are reported in the final row.

Intermediate Neurosymbolic Model Sign-Level Performance

The sign-level evaluation of the neurosymbolic model is summarized in Table 3. The model achieved its highest F_1 scores for sERM (0.90), sMH (0.88), and sSRF (0.72), reflecting robust detection of structural alterations involving the vitreoretinal interface and subretinal space. Moderate performance was observed for sRT (0.77), sRDIST (0.67), and sIRF (0.56), while relatively lower scores were obtained for sBS (0.49), sMNV (0.47), sD (0.50), and sRDET (0.33). Overall, the model achieved macro-averaged recall of 0.60, F_1 of 0.73, and accuracy of 0.63.

Table 4. Classification Performance of the CNN Baseline Model Across The Seven Diagnostic Categories (dAMD, wAMD, ERM, FTMH, LMH, CSC, and Healthy)

Pathology	Precision	Recall	F_1 Score	Support
dAMD	0.85	0.79	0.82	243
wAMD	0.44	0.90	0.59	49
ERM	0.95	0.80	0.87	352
FTMH	0.33	1.00	0.50	12
LMH	0.62	0.70	0.66	30
CSC	0.32	0.90	0.47	10
Healthy	0.96	0.73	0.83	465
Macro AVG	0.64	0.83	0.68	1161

Values represent per-class precision, recall, F_1 score, and the total number of test samples (support). Macro-averaged results are reported in the final row.

Table 5. External Validation Results of the Neurosymbolic Model on the OCTID Dataset

Class	Precision	Recall	F_1 Score	Support	Accuracy
AMD	0.59	0.87	0.70	55	0.83
CSC	0.98	0.50	0.66	102	0.83
FTMH	1.00	0.78	0.87	102	0.83
Healthy	0.82	1.00	0.90	206	0.83
Macro AVG	0.85	0.79	0.78	465	0.83

The table reports per-class precision, recall, F_1 score, and accuracy. Macro-averaged values are provided in the final row.

Baseline Deep Learning Performance

The performance of the baseline CNN model across the seven diagnostic categories is summarized in Table 4. The highest F_1 scores were achieved for ERM (0.87), Healthy (0.83), and dAMD (0.82). Moderate results were obtained for LMH (0.66) and wAMD (0.59), whereas CSC (0.47) and FTMH (0.50) displayed the lowest performance, likely reflecting the small number of available samples in these categories. Overall, the model reached a macro-precision of 0.64, macro-recall of 0.83, and macro-F_1 score of 0.68 across all pathologies ($n = 1161$).

Performance Neurosymbolic Model on External Dataset

The performance of the neurosymbolic model on the OCTID external dataset is reported in Table 5. The model achieved F_1 scores for Healthy (0.90) and FTMH (0.87), followed by AMD (0.70) and CSC (0.66). The model maintained balanced sensitivity and specificity, with an overall accuracy of 0.83 and macro-

Table 6. External Validation Results of the Deep Learning Baseline on the OCTID Dataset

Class	Precision	Recall	F_1 Score	Support	Accuracy
AMD	0.63	0.71	0.67	55	0.53
CSC	0.31	0.83	0.45	102	0.53
FTMH	1.00	0.81	0.90	102	0.53
Healthy	1.00	0.20	0.33	206	0.53
Macro AVG	0.73	0.64	0.59	465	0.53

The table reports per-class precision, recall, F_1 score, support, and overall accuracy. Macro-averaged values are provided in the final row.

precision of 0.85, macro-recall of 0.79, and macro-F_1 score of 0.78.

Performance Deep Learning Model on External Dataset

The CNN baseline exhibited uneven cross-domain performance on OCTID, as summarized in Table 6. High precision but low sensitivity for Healthy ($P = 1.00$; $R = 0.20$; $F_1 = 0.33$) indicates poor specificity balance relative to the in-domain setting, whereas CSC showed the converse pattern with higher recall (0.83) but low precision (0.31), yielding an F_1 score of 0.45. Performance for FTMH remained strong ($F_1 = 0.90$) and AMD was moderate ($F_1 = 0.67$). Overall, the model achieved macro-precision of 0.73, macro-recall of 0.64, and macro-F_1 of 0.59, with an accuracy of 0.53 ($n = 465$), underscoring reduced generalization compared with the neurosymbolic approach on the same external dataset.

Comparative Model Performance

A comparative overview of macro-averaged performance metrics between the neurosymbolic model and the CNN baseline is shown in Figure 5. Bars represent macro-precision, macro-recall, and macro-F_1 score for both internal and external (OCTID) evaluations, with error bars indicating 95% confidence intervals computed across classes (internal, 7; external, 4). The neurosymbolic model consistently outperformed the CNN baseline across datasets, showing notably greater precision and higher F_1 scores, indicative of improved calibration and diagnostic reliability, while maintaining comparable recall. Overall, the neurosymbolic framework achieved a more balanced trade-off between sensitivity and specificity and preserved performance on the external cohort, whereas the CNN baseline exhibited a marked decrease in F_1 scores on OCTID.

Discussion

OCT is a cornerstone of modern ophthalmic diagnostics owing to its noninvasive nature and ability to produce high-resolution retinal cross-sections.[1,2] It is routinely used to evaluate a wide range of retinal diseases.[3–16] Classification of retinal diseases such as AMD, ERM, FTMH, LMH, and CSC remains a critical clinical challenge owing to overlapping imaging features and their impact on treatment pathways. Differentiating these conditions using OCT is central to precise diagnosis and therapy

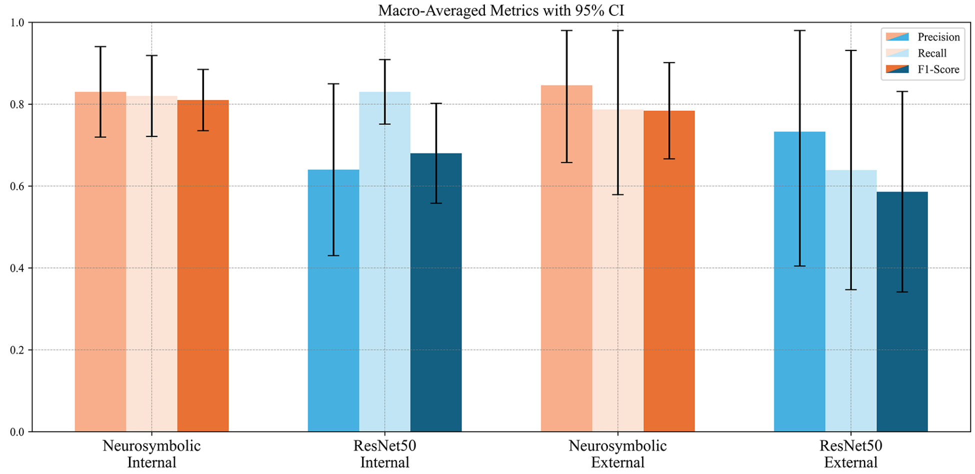

Figure 5. Macro precision, recall, and F_1 scores for the neurosymbolic model and the CNN baseline on the internal test set and the external OCTID set. The neurosymbolic approach shows a more balanced trade-off between precision and recall across both domains. Error bars indicate 95% confidence intervals computed across classes (internal, 7 classes; external, 4 classes), reflecting variability of per-class metrics rather than uncertainty over individual images.

planning.[33] Although deep learning methods have demonstrated strong performance in recent years, concerns over interpretability and generalizability continue to hinder clinical integration.[36,46,47] The neurosymbolic approach is an emerging paradigm in AI that seeks to unify the perceptual strengths of deep learning with the structured, interpretable nature of symbolic reasoning.[34,38,41] In visual reasoning tasks, such as scene understanding, visual question answering, and diagram interpretation, neurosymbolic models have shown improved generalization, interpretability, and robustness to distributional shifts.[34]

In this study, we proposed and evaluated a novel neurosymbolic framework that integrates CNNs with symbolic reasoning, offering an interpretable and modular diagnostic approach. To our knowledge, this application is among the first of a tightly integrated neurosymbolic framework to medical imaging, and specifically to retinal OCT classification. This factor makes our study a novel contribution to both the fields of medical AI and interpretable machine learning. Our results demonstrated that this model performs competitively across a broad range of retinal diseases.

The proposed neurosymbolic model demonstrated consistent and slightly superior diagnostic performance compared with the CNN-only baseline, both in internal evaluation and external validation. On the internal dataset, the neurosymbolic framework achieved macro-precision of 0.83, recall of 0.82, and F_1 score of 0.81, slightly outperforming the baseline CNN (0.64, 0.83, and 0.68, respectively). When tested on the external OCTID dataset, the neurosymbolic model retained stronger generalization ($F_1 = 0.78$, accuracy = 0.83) compared with the CNN baseline ($F_1 = 0.59$, accuracy = 0.53). This sustained performance across datasets highlights the robustness of the neurosymbolic approach to domain variability and unseen imaging conditions. These improvements can be attributed to the integration of explicit symbolic reasoning within the deep learning architecture. By embedding structured ophthalmic knowledge in the form of differentiable logical rules, the model benefits from an additional inductive bias that constrains learning toward physiologically meaningful feature relationships.[48] Unlike conventional CNNs, which depend solely on statistical correlations learned from data, the neurosymbolic framework infers diagnostic categories through the probabilistic interaction of retinal signs (sBS, sMNV, sD, sIRF, sSRF, sMH, sERM, sRT, sRDIST, and sRDET) and encoded logical rules. This strategy yields more interpretable and clinically plausible decision pathways.[34]

Intermediate sign evaluation confirmed that the model accurately inferred key retinal signs with a macro-precision of 0.60, macro-recall of 0.73, macro-F_1 score of 0.63, and accuracy of 0.83, despite the absence of sign-level supervision during training. This supports the principle of algorithmic supervision, whereby the network learns intermediate concepts indirectly via rule-based reasoning rather than through explicit annotation.[49] Such a framework substantially reduces the need for exhaustive manual labeling. For interpretability, a fixed decision threshold ($\tau = 0.5$) was applied only during post hoc evaluation of retinal signs to compute precision, recall, and accuracy. During inference, the symbolic reasoning layer operates directly on the continuous probabilistic outputs of the CNN, preserving diagnostic uncertainty and enabling differentiable reasoning.

Furthermore, the superior generalization observed in the neurosymbolic model, particularly on the external dataset, underscores the benefit of incorporating domain knowledge into model design. The logical rule set introduces an additional layer of interpretability and consistency, acting as a regularizer that discourages implausible co-occurrences of retinal signs. This mechanism enhances resilience to dataset shifts and decreases overfitting to training domain artifacts. The advantage becomes most apparent in smaller or imbalanced datasets, where the added structural knowledge compensates for limited examples, guiding the model toward clinically consistent reasoning.[34]

The neurosymbolic framework, therefore, appears to be especially suited for low-data or heterogeneous clinical scenarios, where symbolic constraints supplement empirical information and stabilize learning. This observation aligns with the broader concept that hybrid reasoning systems combining data-driven and knowledge-driven components can achieve improved robustness and generalization without requiring exponentially larger datasets.[34] Although numerous studies have successfully tackled individual retinal pathologies, most remain restricted to binary or narrow-class classification tasks, limiting direct comparability across frameworks and datasets.[50–53]

From a clinical perspective, one of the principal advantages of the neurosymbolic approach lies in its close alignment with clinical reasoning (see Fig. 2). Rather than relying solely on large-scale data, the model integrates ophthalmic domain knowledge into explicit logical rules that emulate the diagnostic workflow of an ophthalmologist, progressing from structural signs to pathological conclusions, and thus enhances both interpretability and clinical traceability.

Several limitations qualify the interpretation of these findings. The neurosymbolic framework introduces additional architectural complexity relative to conventional CNNs and depends on specialized differentiable-logic frameworks such as Scallop,[41,42] which remain at an early stage of adoption in medical AI research. Although algorithmic supervision mitigates the need for sign-level annotations, rule formulation still requires expert input and careful definition of retinal features. Ambiguities can arise when co-occurring signs (e.g., intraretinal and subretinal fluid) must be differentiated, underscoring the importance of refining and expanding rule sets.

Regarding generalizability, broader validation across multiple institutions and vendors is needed to confirm domain transferability. Furthermore, images with mixed or coexisting pathologies were excluded to ensure unambiguous rule application. Although this restriction facilitated clear rule evaluation, it does not fully represent the complexity of real-world clinical presentations.

Although copathology is common in clinical practice, we excluded mixed diagnoses such as AMD with diabetic retinopathy or retinal vein occlusion in this study. These entities involve distinct pathophysiological markers (including microaneurysms, retinal hemorrhages, venous congestion, and ischemic changes) not annotated in our dataset and not represented in the current symbolic rule set. Incorporating such cases without explicitly modeling their diagnostic signs would risk underspecifying the rules and conflating diagnostic pathways.

Class imbalance, particularly the predominance of Healthy and ERM cases relative to rarer conditions such as CSC and FTMH, may have influenced per-class recall despite balancing strategies. Nonetheless, the improved external generalization of the neurosymbolic system suggests that rule-based reasoning introduces complementary information that enhances learning stability even under such an imbalance.

Future work will aim to extend the current design along several axes. First, symbolic rules could be enriched with spatial or topological descriptors, for example, encoding the position of fluid relative to retinal layers or segmentation-derived anatomical maps.[50–52] Second, advances in large language models offer opportunities to formalize clinical guidelines into consistent, machine-interpretable logic.[53] Large language model–based frameworks could assist in rule synthesis, verification, and expansion while maintaining transparency. Finally, the modular architecture of the proposed system enables independent inspection of the symbolic layer: clinicians can visualize which retinal signs contributed to a decision and generate attention heatmaps over OCT scans. Such transparency supports both clinical validation and iterative refinement when inconsistencies are detected.

Conclusions

This study introduced a neurosymbolic framework for OCT-based retinal disease classification that tightly couples a ResNet-50 backbone with differentiable, rule-based reasoning grounded in ophthalmic knowledge. The model achieved strong internal performance (macro-F_1 score = 0.81; precision = 0.83; recall = 0.82) and retained superior generalization on external validation (OCTID: macro-F_1 score = 0.78; accuracy = 0.83), slightly outperforming a CNN-only baseline (macro-F_1 score = 0.59; accuracy = 0.53). Beyond accuracy, the framework provides transparent inference by tracing how probabilistic retinal signs (sBS, sMNV, sD, sIRF, sSRF, sMH, sERM, sRT, sRDIST, and sRDET) combine through clinically meaningful rules to yield pathology probabilities, preserving diagnostic uncertainty throughout the pipeline.

Embedding structured clinical knowledge as differentiable logic supplies a knowledge-driven inductive bias that improves robustness, especially in small, imbalanced, or heterogeneous datasets, and aligns model behavior with clinician reasoning. The approach reduces dependence on exhaustive sign annotations via algorithmic supervision, yet still enables sign-level interpretability (sign macro-precision/recall/F_1 score/accuracy: 0.60/0.73/0.63/0.83). Its modular design allows the symbolic layer to be inspected, audited, and refined, facilitating adoption in real-world workflows where explainability and traceability are essential.

The neurosymbolic paradigm is broadly applicable to medical imaging tasks where visual features map to established diagnostic logic. Future work will extend to multipathology co-occurrence, multicenter/multivendor validation, and operating-point optimization (e.g., threshold calibration and decision–curve analysis) to better align probabilistic outputs with clinical use.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Disclosure: **A. Miladinović**, None; **A. Biscontin**, None; **M. Ajčević**, None; **S. Kresevic**, None; **A. Accardo**, None; **D. Tognetto**, None; **L. Inferrera**, None

References

1. Eladawi N, Elmogy M, Ghazal M, et al. Classification of retinal diseases based on OCT Images. *Front Biosci (Landmark Ed)*. 2018;23(2):247–264. https://doi.org/10.2741/4589.
2. Özdaş MB, Uysal F, Hardalaç F. Classification of retinal diseases in optical coherence tomography images using artificial intelligence and Firefly algorithm. *Diagnostics*. 2023;13(3):433. https://doi.org/10.3390/diagnostics13030433.
3. Flores R, Carneiro Â, Tenreiro S, Seabra MC. Retinal progression biomarkers of early and intermediate age-related macular degeneration. *Life*. 2021;12(1):36. https://doi.org/10.3390/life12010036.
4. Bressler NM. Age-related macular degeneration is the leading cause of blindness. *JAMA*. 2004;291(15):1900–1901. https://doi.org/10.1001/jama.291.15.1900.
5. Schneider EW, Fowler SC. Optical coherence tomography angiography in the management of age-related macular degeneration. *Curr Opin Ophthalmol*. 2018;29(3):217–225. https://doi.org/10.1097/ICU.0000000000000469.
6. Corvi F, Cozzi M, Invernizzi A, Pace L, Sadda SR, Staurenghi G. Optical coherence tomography angiography for detection of macular neovascularization associated with atrophy in age-related macular degeneration. *Graefes Arch Clin Exp Ophthalmol*. 2021;259(2):291–299. https://doi.org/10.1007/s00417-020-04821-6.
7. Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. *Lancet*. 2010;376(9735):124–136. https://doi.org/10.1016/S0140-6736(09)62124-3.
8. Amoaku WM, Ghanchi F, Bailey C, et al. Diabetic retinopathy and diabetic macular oedema pathways and management: UK Consensus Working Group. *Eye*. 2020;34(S1):1–51. https://doi.org/10.1038/s41433-020-0961-6.
9. Abbas Q, Fondon I, Sarmiento A, Jiménez S, Alemany P. Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features. *Med Biol Engineer Comput*. 2017;55(11):1959–1974. https://doi.org/10.1007/s11517-017-1638-6.
10. Eltanboly A, Ismail M, Shalaby A, et al. A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images. *Med Phys*. 2017;44(3):914–923. https://doi.org/10.1002/MP.12071.
11. Lindtjørn B, Krohn J, Forsaa VA. Optical coherence tomography features and risk of macular hole formation in the fellow eye. *BMC Ophthalmol*. 2021;21(1):351. https://doi.org/10.1186/s12886-021-02111-1.
12. Aoyama Y, Maruko I, Kawano T, et al. Diagnosis of central serous chorioretinopathy by deep learning analysis of en face images of choroidal vasculature: a pilot study. *PLoS One*. 2021;16(6):e0244469. https://doi.org/10.1371/journal.pone.0244469.
13. Xu J, Shen J, Wan C, et al. An Automatic image processing method based on artificial intelligence for locating the key boundary points in the central serous chorioretinopathy lesion area. *Comput Intell Neurosci*. 2023;2023(1):1839387. https://doi.org/10.1155/2023/1839387.
14. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *JAMA*. 2014;311(18):1901–1911. https://doi.org/10.1001/jama.2014.3192.
15. Geevarghese A, Wollstein G, Ishikawa H, Schuman JS. Optical coherence tomography and glaucoma. *Annu Rev Vis Sci*. 2021;7:693–726. https://doi.org/10.1146/annurev-vision-100419-111350.
16. Kim SJ, Cho KJ, Oh S. Development of machine learning models for diagnosis of glaucoma. Liu B, ed. *PLoS One*. 2017;12(5):e0177726. https://doi.org/10.1371/journal.pone.0177726.
17. Akpinar MH, Sengur A, Faust O, Tong L, Molinari F, Acharya UR. Artificial intelligence in retinal screening using OCT images: a review of the last decade (2013-2023). *Comput Methods Programs Biomed*. 2024;254:108253. https://doi.org/10.1016/j.cmpb.2024.108253.
18. Pandya BU, Grinton M, Mandelcorn ED, Felfeli T. Retinal optical coherence tomography imaging biomarkers: a review of the literature. *Retina*. 2024;44(3):369–380. https://doi.org/10.1097/IAE.0000000000003974.
19. Algvere PV, Kvanta A, Seregard S. Drusen maculopathy: a risk factor for visual deterioration. *Acta Ophthalmol*. 2016;94(5):427–433. https://doi.org/10.1111/aos.13011.
20. Ban N, Shinojima A, Negishi K, Kurihara T. Drusen in AMD from the perspective of cholesterol metabolism and hypoxic response. *J Clin Med*. 2024;13(9):2608. https://doi.org/10.3390/jcm13092608.
21. Sheth JU, Stewart MW, Narayanan R, et al. Macular neovascularization. *Surv Ophthalmol*. 2025;70(4):653–675. https://doi.org/10.1016/j.survophthal.2024.08.003.
22. Foss AJE, Almeida D, Cheung CMG, Ogura Y, De Cock E, Empeslidis T. To treat or not to treat? resolving the question of subretinal and intraretinal fluid in age-related macu-

lar degeneration: a narrative review. *Ophthalmol Ther*. 2025;14(3):489–514. https://doi.org/10.1007/s40123-025-01093-3.

23. Karampelas M, Malamos P, Petrou P, Georgalas I, Papaconstantinou D, Brouzas D. Retinal pigment epithelial detachment in age-related macular degeneration. *Ophthalmol Ther*. 2020;9(4):739–756. https://doi.org/10.1007/s40123-020-00291-5.
24. Scheufele TA, Witkin AJ, Schocket LS, et al. Photoreceptor atrophy in acute posterior multifocal placoid pigment epitheliopathy demonstrated by optical coherence tomography. *Retina*. 2005;25(8):1109–1112. https://doi.org/10.1097/00006982-200512000-00027.
25. Daruich A, Matet A, Dirani A, et al. Central serous chorioretinopathy: recent findings and new physiopathology hypothesis. *Prog Retin Eye Res*. 2015;48:82–118. https://doi.org/10.1016/j.preteyeres.2015.05.003.
26. Fung AT, Galvin J, Tran T. Epiretinal membrane: a review. *Clin Exp Ophthalmol*. 2021;49(3):289–308. https://doi.org/10.1111/ceo.13914.
27. Matoba R, Morizane Y. Epiretinal membrane: an overview and update. *Jpn J Ophthalmol*. 2024;68(6):603–613. https://doi.org/10.1007/s10384-024-01127-6.
28. Bikbova G, Oshitari T, Baba T, Yamamoto S, Mori K. Pathogenesis and management of macular hole: review of current advances. *J Ophthalmol*. 2019;2019(1):3467381. https://doi.org/10.1155/2019/3467381.
29. Mino M, Matoba R, Kanzaki Y, et al. Quantitative analyses of retinal traction force and metamorphopsia in lamellar macular hole and related diseases. *Ophthalmol Sci*. 2023;3(3):100305. https://doi.org/10.1016/j.xops.2023.100305.
30. Hecht I, Yeshurun I, Bartov E, Bar A, Burgansky-Eliash Z, Achiron A. Retinal layers thickness changes following epiretinal membrane surgery. *Eye*. 2018;32(3):555–562. https://doi.org/10.1038/eye.2017.233.
31. Spaide RF, Jaffe GJ, Sarraf D, et al. Consensus nomenclature for reporting neovascular age-related macular degeneration data. *Ophthalmology*. 2020;127(5):616–636. https://doi.org/10.1016/j.ophtha.2019.11.004.
32. Kogo T, Muraoka Y, Ishikura M, et al. Pigment epithelial detachment and leak point locations in central serous chorioretinopathy. *Am J Ophthalmol*. 2024;261:19–27. https://doi.org/10.1016/j.ajo.2024.01.012.
33. Leandro I, Lorenzo B, Aleksandar M, et al. OCT-based deep-learning models for the identification of retinal key signs. *Sci Rep*. 2023;13(1):14628. https://doi.org/10.1038/s41598-023-41362-4.
34. Garcez A d'Avila, Lamb LC. Neurosymbolic AI: the 3rd wave. *Artif Intell Rev*. 2023;56(11):12387–12406. https://doi.org/10.1007/s10462-023-10448-w.
35. Yanagihara RT, Lee CS, Ting DSW, Lee AY. Methodological challenges of deep learning in optical coherence tomography for retinal diseases: a review. *Transl Vis Sci Technol*. 2020;9(2):11. https://doi.org/10.1167/tvst.9.2.11.
36. Loyola-Gonzalez O. Black-box vs. white-box: understanding their advantages and weaknesses from a practical point of view. *IEEE Access*. 2019;7:154096–154113. https://doi.org/10.1109/ACCESS.2019.2949286.
37. Chaudhuri S, Ellis K, Polozov O, Singh R, Solar-Lezama A, Yue Y. Neurosymbolic programming. *FNT in Programming Languages*. 2021;7(3):158–243. https://doi.org/10.1561/2500000049.
38. Sheth A, Roy K, Gaur M. Neurosymbolic artificial intelligence (why, what, and how). *IEEE Intell Syst*. 2023;38(3):56–62. https://doi.org/10.1109/MIS.2023.3268724.
39. Ullah N, Guzmán-Aroca F, Martínez-Álvarez F, De Falco I, Sannino G. A novel explainable AI framework for medical image classification integrating statistical, visual, and rule-based methods. *Med Image Anal*. 2025;105:103665. https://doi.org/10.1016/j.media.2025.103665.
40. Musanga V, Viriri S, Chibaya C. A framework for integrating deep learning and symbolic AI towards an explainable hybrid model for the detection of COVID-19 using computerized tomography scans. *Information*. 2025;16(3):208. https://doi.org/10.3390/info16030208.
41. Li Z, Huang J, Naik M. Scallop: a language for neurosymbolic programming. *Proc ACM Program Lang*. 2023;7(PLDI):1463–1487. https://doi.org/10.1145/3591280.
42. Li Z, Huang J, Liu J, Naik M. Neurosymbolic programming in Scallop: principles and practice. *FNT in Programming Languages*. 2024;8(2):118–249. https://doi.org/10.1561/2500000059.
43. Bringmann A, Unterlauft JD, Barth T, Wiedemann R, Rehak M, Wiedemann P. Different modes of full-thickness macular hole formation. *Exp Eye Res*. 2021;202:108393. https://doi.org/10.1016/j.exer.2020.108393.
44. Hubschman JP, Govetto A, Spaide RF, et al. Optical coherence tomography-based consensus definition for lamellar macular hole. *Br J Ophthalmol*. 2020;104(12):1741–1747. https://doi.org/10.1136/bjophthalmol-2019-315432.

45. Gholami P, Roy P, Parthasarathy MK, Lakshminarayanan V. OCTID: optical coherence tomography image database. *arXiv*. Preprint posted online 2018. https://doi.org/10.48550/ARXIV.1812.07056.
46. Tonti E, Tonti S, Mancini F, et al. Artificial intelligence and advanced technology in glaucoma: a review. *J Pers Med*. 2024;14(10):1062. https://doi.org/10.3390/jpm14101062.
47. Rashidisabet H, Sethi A, Jindarak P, et al. Validating the generalizability of ophthalmic artificial intelligence models on real-world clinical data. *Trans Vis Sci Tech*. 2023;12(11):8. https://doi.org/10.1167/tvst.12.11.8.
48. Zhang X, Sheng VS. Neuro-symbolic AI: explainability, challenges, and future trends. *arXiv*. Preprint posted online 2024. https://doi.org/10.48550/ARXIV.2411.04383.
49. Liang B, Wang Y, Tong C. AI Reasoning in deep learning era: from symbolic AI to neural–symbolic AI. *Mathematics*. 2025;13(11):1707. https://doi.org/10.3390/math13111707.
50. Sanchez-Rodriguez G, Lou L, Pardue MT, Feola AJ. RetOCTNet: deep learning–based segmentation of OCT images following retinal ganglion cell injury. *Trans Vis Sci Tech*. 2025;14(2):4. https://doi.org/10.1167/tvst.14.2.4.
51. Burke J, Engelmann J, Hamid C, et al. An open-source deep learning algorithm for efficient and fully automatic analysis of the choroid in optical coherence tomography. *Transl Vis Sci Tech*. 2023;12(11):27. https://doi.org/10.1167/tvst.12.11.27.
52. Shamsi F, Liu R, Owsley C, Kwon M. Identifying the retinal layers linked to human contrast sensitivity via deep learning. *Invest Ophthalmol Vis Sci*. 2022;63(2):27. https://doi.org/10.1167/iovs.63.2.27.
53. Billi M, Pisano G, Sanchi M. Fighting the knowledge representation bottleneck with large language models. In: Savelka J, Harasta J, Novotna T, Misek J, eds. *Frontiers in Artificial Intelligence and Applications*. Amsterdam, the Netherlands: IOS Press; 2024. https://doi.org/10.3233/FAIA241230.