

Systematic Review

Weakly Supervised Deep Learning for Ocular Image Segmentation: A Systematic Review of Fundus and OCT Methods

Pedro Penedo ¹, Jorge Machado ² , Rita Anjos ³ , Ana Marta ^{4,5} , Aristófaes Corrêa Silva ⁶ 
and António Cunha ^{2,7,*} 

¹ Department of Sciences and Technologies, Universidade Aberta (UAb), 1269-001 Lisboa, Portugal; pmapenedo@gmail.com

² School Sciences and Technologies, Universidade de Trás-os-Montes e Alto Douro (UTAD), 5000-801 Vila Real, Portugal; miguelj1992@outlook.com

³ Unidade Local de Saúde de São José, 1150-199 Lisboa, Portugal; rita.s.anjos@gmail.com

⁴ Unidade Local de Saúde de Santo António, Department of Ophthalmology, 4099-001 Porto, Portugal; analuisamarta2@gmail.com

⁵ Unit for Multidisciplinary Research in Biomedicine, Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, 4050-346 Porto, Portugal

⁶ Applied Computer Group Department of Electrical Engineering, Federal University of Maranhão, São Luís 65085-580, Brazil; ac.silva@ufma.br

⁷ Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), 4200-465 Porto, Portugal

* Correspondence: acunha@utad.pt

Abstract

Eye diseases, such as glaucoma, diabetic retinopathy, and age-related macular degeneration, drive the growing need for reliable and scalable analyses of fundus and optical coherence tomography (OCT) images. Deep learning performs strongly in ocular structure segmentation. However, it typically relies on dense pixel-wise annotations, which are costly and difficult to obtain at scale. Weakly supervised learning (WSL) can reduce this burden by leveraging coarse labels, limited strong annotations, and unlabeled data. This systematic umbrella review synthesizes survey and review articles on weakly supervised deep learning for image segmentation, with a focus on ocular imaging (fundus and OCT/OCTA). After analyzing twenty-one secondary studies, the main finding reveals an “empty intersection”: WSL-focused segmentation surveys are often modality-agnostic. Conversely, ocular reviews are predominantly fully supervised and seldom offer quantitative evidence on annotation-effort savings or direct comparisons between weak and fully supervised methods on identical datasets. Across the included reviews, label-efficient strategies cluster around CAM/MIL formulations, sparse supervision (points/scribbles/boxes), pseudo-labelling/self-training, and semi-/self-supervised learning, implemented mainly with U-Net/DeepLab families and increasingly Transformer or hybrid backbones. These results provide a structured map of available WSL mechanisms and, critically, identify reproducible reporting gaps that currently prevent fair benchmarking in ocular segmentation. Therefore, this review supports the development of ocular-specific benchmarks and minimum reporting practices that link segmentation performance to annotation effort.

Keywords: weakly supervised learning; label-efficient deep learning; medical image segmentation; ocular imaging; fundus photography; optical coherence tomography; optic disc and cup segmentation; multiple-instance learning; class activation maps; pseudo-labelling



Academic Editors: Martina Iammarino, Chiara Verdona and Stefano Galantucci

Received: 23 January 2026

Revised: 15 February 2026

Accepted: 22 February 2026

Published: 26 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Retinal diseases comprise a heterogeneous group of conditions that can lead to progressive and often irreversible vision loss. Among them, diabetic retinopathy and age-related macular degeneration are highly prevalent and account for a substantial proportion of visual impairment and blindness worldwide [1–5]. In clinical practice, colour fundus photography and optical coherence tomography (OCT) are routinely used to document retinal structure, quantify disease progression and support treatment decisions [1,4,6]. In these modalities, reliable segmentation of the optic nerve head structures (optic disc and cup) in fundus images and retinal layers and vascular networks in OCT/OCTA is essential for deriving quantitative biomarkers, automating diagnosis and monitoring longitudinal changes [1,4,6].

Deep learning has shown great potential in ophthalmology, mirroring advances already seen in radiology, pathology and dermatology [1–7]. Convolutional and Transformer-based networks can learn complex representations from large collections of imaging data and have achieved state-of-the-art performance in multiple retinal tasks, including disease detection, grading and structure segmentation [1–10]. When trained on high-quality annotated datasets, these models can help reduce diagnostic errors, support personalized management and reveal subtle image patterns that may not be easily captured by human observers. Additional recent ocular-focused surveys have reported deep learning segmentation approaches for retinal vessels, the optic disc/cup, diabetic retinopathy lesions, and retinal OCT layers, providing a domain-specific context for the present review [11–21].

Despite these advances, developing deep learning segmentation models in ophthalmology is limited by the requirement for dense pixel-wise annotations [1–4,8–10,22–24]. Creating reference masks for the optic disc and cup boundaries, thin retinal layers, or small-calibre vessels requires experienced graders, specialized software, and iterative quality control. This process is expensive and time-consuming and is subject to inter- and intra-observer variability, which can be particularly pronounced in ophthalmology for structures or lesions with unclear boundaries. Consequently, although many widely used ocular segmentation benchmarks provide dense annotations, in practice, labels are often incomplete, limited to small subsets, noisy/variable across graders, or available only at coarse levels (e.g., image-level diagnosis), which makes weakly supervised and label-efficient learning attractive for leveraging large-scale data while reducing the annotation burden [8,23–25].

Traditional strategies to mitigate data scarcity include patch-based learning, conventional data augmentation and transfer learning from large natural image collections [8–10]. Although these approaches can improve performance, they do not fundamentally address the dependence on strong pixel-wise labels. In the context of segmentation, more flexible learning paradigms are needed that can leverage incomplete, inexact or noisy supervision and systematically use large unlabeled repositories.

Weakly supervised learning (WSL) has emerged as a promising solution to this challenge [8,22–33]. In WSL, models are trained using weaker annotations than those required at the test time (e.g., image-level tags, bounding boxes, scribbles, points, or coarse masks) or using only a limited subset of fully annotated images. These settings are often combined with semi- and self-supervised learning, in which unlabeled data contribute additional training signals, and with domain-specific priors, such as anatomical constraints, shape regularization, or topology-aware losses, to encourage biologically plausible predictions [8,24–33]. These approaches are particularly for retinal imaging, where anatomy is relatively well understood and disease manifestations follow characteristic spatial patterns [1–7].

While methodological progress has led to several surveys reviewing weak supervision, label-efficient learning, and deep segmentation in general computer vision and medical imaging [8,22–34], reviews specifically in ophthalmology have focused on retinal image analysis but typically emphasize fully supervised classification or segmentation and do not systematically

address weak labelling regimes [1–7]. Consequently, the current evidence on weakly supervised deep learning for ocular image segmentation is scattered across generic WSL surveys, organ- or modality-specific reviews, and a small number of ophthalmic overviews [1–8,22–34].

To capture this fragmented evidence base, review-type literature (2020–2026) was identified primarily via Google Scholar using Publish or Perish (PoP) and complemented by targeted searches in PubMed and Scopus. Inclusion was restricted to English full-text surveys/reviews meeting predefined eligibility criteria.

This systematic umbrella review of surveys and review articles on weakly supervised deep learning segmentation, reported according to PRISMA 2020 [35], places particular emphasis on ocular imaging (fundus and OCT). This review consolidates and critically appraises this fragmented literature from an ocular perspective by examining how weakly supervised deep learning approaches for image segmentation are presented in survey and review articles, identifying which strategies and architectures are reported for fundus and OCT segmentation, and summarizing what is known about their performance relative to fully supervised baselines and their impact on annotation effort. By organizing existing knowledge in this way, this review provides a structured reference for researchers and clinicians interested in label-efficient retinal segmentation and highlights priority areas for future work, including the development of ocular-specific benchmarks and more systematic evaluations of annotation–efficiency trade-offs.

A preliminary search of the PROSPERO database [36] did not identify registered umbrella reviews specifically addressing weakly supervised deep learning for fundus and OCT image segmentation. Accordingly, this umbrella review makes three contributions: (i) maps weakly supervised and label-efficient strategies reported across review literature for fundus and OCT/OCTA segmentation; (ii) identifies gaps in annotation-effort reporting and in explicit weak-versus-fully supervised comparisons; and (iii) appraises included reviews using AMSTAR-2 and proposes minimum reporting elements to support reproducible benchmarking.

A conceptual overview of the suitability weak-label and label-efficient for fundus versus OCT/OCTA segmentation is provided in Figure 1.

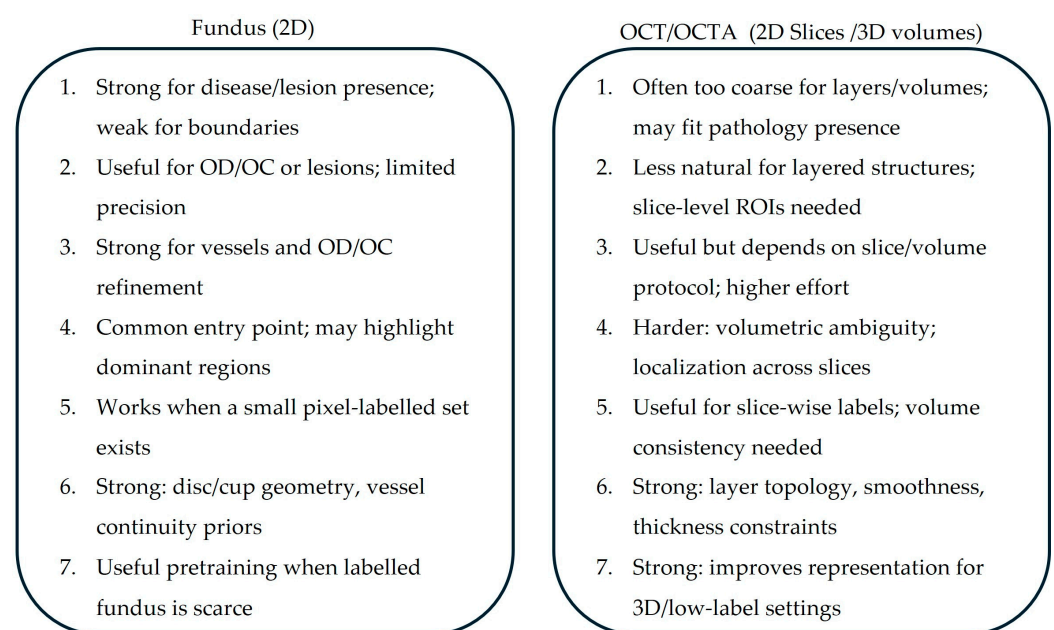


Figure 1. Conceptual suitability of weak labels for fundus vs. OCT/OCTA segmentation.

2. Materials and Methods

This study was conducted as a systematic umbrella review in which the unit of analysis comprised secondary research articles rather than individual primary studies. This study aimed to synthesize the literature on weakly supervised deep learning approaches for medical image segmentation, with specific emphasis on ocular imaging modalities, namely fundus photography and optical coherence tomography (OCT). The review was conducted and reported in accordance with the PRISMA 2020 guidelines for systematic reviews. The PRISMA 2020 checklist (Supplementary Materials) guided the reporting of this umbrella review, and the completed checklist is included in the Supplementary Materials. The review protocol was registered in the Open Science Framework (OSF) with registration ID wd8gu (<https://osf.io/wd8gu/>, accessed on 4 February 2026) (embargoed). Prior to the literature search, the scope of the review was defined, eligibility criteria were prespecified, relevant information sources were identified, and the data items to be extracted, together with the planned synthesis approach, were determined a priori to enhance transparency and reproducibility of the study. The following subsections briefly outline the search and data extraction procedures used.

2.1. Related Work

Weakly supervised learning (WSL) has emerged as a key strategy for reducing the prohibitive cost of dense annotations in segmentation tasks. Zhou formalized weak supervision in terms of inexact, incomplete, and inaccurate labels and linked these categories to typical learning set-ups such as multiple-instance learning (MIL), class-activation maps (CAMs), and noisy labels, thereby providing much of the vocabulary used today [22]. In medical imaging, Cheplygina et al. reviewed semi-supervised learning, MIL and transfer learning and argued that realistic clinical workflows often rely on hybrid supervision rather than a single “pure” regime [23]. Method-centric surveys then moved closer to segmentation. Zhang et al. summarized semi- and weakly supervised semantic segmentation pipelines that convert weak signals into pixel-wise masks via CAM seeding, refinement with priors, partial-pixel supervision from points, scribbles and boxes, and self-training with pseudo-labels and consistency regularization [8]. Ouassit et al. catalogued losses, architectures, and post-processing schemes for weakly supervised semantic segmentation (WSSS), with a strong focus on CAM-based methods [34]. More recently, Shen et al. proposed the notion of “label-efficient deep image segmentation,” organizing methods along a spectrum from fully supervised to weakly and unsupervised and systematizing practical strategies such as pseudo-label curricula, contrastive pre-training, and shape/topology constraints [24]. Qu et al. complemented these segmentation-oriented views with a survey of label-efficient automatic diagnosis and analysis that unifies weak, semi-supervised and self-supervised techniques in clinical pipelines [25].

Beyond supervision regimes, several reviews have emphasized representation learning and architectural trends. Green reviewed representation learning with imperfect labels and incomplete views, highlighting noise-robust objectives, sample re-weighting, and multi-view consistency as generic tools for WSL perception systems [26]. Zhang et al. aggregated non-fully supervised medical image segmentation methods, including weak, semi-supervised, and self-supervised approaches, and proposed a unified terminology and reporting recommendations [27]. Bohlender et al. focused on shape-constrained deep learning for medical image segmentation, surveying anatomy-aware losses, topology constraints and statistical shape priors that can be combined with both convolutional and Transformer backbones [9]. In parallel, Gao et al. provided a comprehensive review of deep learning-based medical image segmentation, tracing the evolution from encoder–decoder CNNs (U-Net, DeepLab) towards Transformer and hybrid architectures and discussing

datasets, evaluation metrics, and reproducibility issues [10]. Together, these studies offer rich taxonomies of weak labels (CAM, MIL, sparse annotations), learning strategies (pseudo-labels, self-training, anatomy rules), and backbone families; however, they remain largely modality-agnostic.

A growing body of surveys analyze WSL in domains outside medical imaging, underlining the generality of the core ideas. In remote sensing, Fasana et al. reviewed weakly supervised object detection with image-level labels or points for localizing small targets in very large aerial and satellite images [28]. In industrial settings, Martínez-Heredia and Ventura surveyed weak supervision for predictive maintenance, and Zhao et al. reviewed bearing fault diagnosis methods based on WSL, where expert labels are scarce and failures are rare [29,30]. These studies emphasize MIL formulations, anomaly detection, and self-training on unlabeled or weakly labelled sensor data. Similar themes appear in medical-adjacent surveys, such as Hassan et al., who reviewed supervised and weakly supervised deep learning models for COVID-19 CT diagnosis and documented how CAM-based localization, pseudo-labelled lung/lesion masks, and MIL are deployed when full annotations are unavailable [31]. Modality-specific segmentation reviews, for example, on ultrasound and lung CT, analyze how noise characteristics, data scarcity, and architecture choice affect segmentation quality and denoising strategies and occasionally touch on non-fully supervised training [32,33].

Within retinal imaging, most existing surveys focus on disease detection or fully supervised segmentation pipelines. Alawad et al. reviewed machine learning and deep learning techniques for optic disc and cup segmentation, documenting datasets, evaluation metrics and the dominance of U-Net-style encoder–decoder architectures under full supervision, with limited attention to weak labels or pseudo-label loops [1]. Goutam et al. provided a comprehensive review of deep learning strategies for retinal disease diagnosis from fundus photographs, covering tasks such as lesion detection, vessel extraction, and optic disc/cup analysis, again mostly under dense pixel-wise supervision [2]. Anusuya and Masoodhu Banu surveyed glaucoma detection from fundus images, focusing on classification and screening models built on convolutional neural networks and transfer learning rather than on label-efficient segmentation [3]. Zedan et al. reviewed automated glaucoma screening and diagnosis from fundus photographs, discussing both structural biomarkers derived from disc/cup segmentation and end-to-end classification pipelines, but without a systematic analysis of weakly supervised training regimes [4].

More recent ophthalmic reviews have broadened the scope of the modalities and processing pipelines. Rizvana and Narayanan reviewed deep learning for ocular disease detection from both fundus and OCT images, providing an architectural overview and covering tasks ranging from diabetic retinopathy grading and glaucoma detection to retinal vessel and disc segmentation, while briefly touching on weak supervision (mainly in the context of CAM-based explainability) [5]. Xue et al. analyzed atherosclerotic plaque assessment using intravascular OCT and described deep learning pipelines for plaque component segmentation and classification under full supervision [7]. Chen et al. presented a comprehensive review of OCT angiography (OCTA) denoising, segmentation and volumetric rendering, focusing on network designs and pre- and post-processing for 3D data; however, they did not provide a dedicated taxonomy of weak labels or label-efficient training strategies [6]. Beyond ophthalmology, modality-centric segmentation reviews for ultrasound [32] and lung CT [33] further illustrate that most domain-specific syntheses still assume dense ground-truth labels and use WSL only peripherally, for example, as a data augmentation or pre-training tool.

In summary, the existing literature offers (i) generic weakly supervised and label-efficient learning taxonomies that are largely modality-agnostic [8–10,22–27,29–31,34] and

(ii) ophthalmic reviews that focus mainly on diagnostic classification or fully supervised segmentation in fundus and OCT imaging [1–7]. However, there is still no systematic, modality-specific synthesis of weakly supervised deep learning segmentation for ocular imaging. Current surveys do not consistently distinguish fundus tasks (e.g., optic disc/cup) from OCT tasks (e.g., retinal layers, vessels and lesions), map concrete weak-label types (CAM-style image/volume-level tags, MIL bags, sparse scribbles/points/boxes) to segmentation pipelines, or compare backbone families (U-Net, DeepLab, Transformer/hybrid) and 2D/3D/hybrid designs under weak supervision. The present review addresses this gap by focusing on weakly supervised ocular segmentation in fundus and OCT imaging, mapping the landscape since 2020, examining how annotation–efficiency trade-offs are reported, and providing a structured comparison of weak-label types, learning strategies, and architectures. To make this fragmentation explicit, Table 1 summarizes the included surveys and reviews and contrasts their coverage of ocular modalities (fundus/OCT), weak-label types (CAM, MIL, sparse), label-efficient strategies (auto-/pseudo-labelling and anatomy-aware constraints), and backbone families (U-Net, DeepLab, Transformer/hybrid).

As indicated in Table 1, generic surveys provide detailed weak supervision taxonomies but are largely modality-agnostic, whereas ophthalmic reviews cover fundus/OCT pipelines but rarely systematize weak labels for segmentation. This motivates the modality-specific synthesis of weakly supervised ocular segmentation in the present review.

2.2. Research Questions

This review is structured around three research questions that aim to clarify what has been done, how weakly supervised methods compare with full supervision, and which model designs are the most promising for ocular image segmentation. In this review, these questions were investigated through existing surveys and review articles, rather than by reanalyzing primary segmentation studies.

- RQ1: Which weakly supervised deep learning approaches have been proposed for ocular image segmentation in fundus and OCT images?
- RQ2: For optic disc and cup segmentation in fundus images, how do weakly supervised deep learning methods compare to fully supervised baselines in terms of segmentation performance and annotation effort?
- RQ3: For retinal layer and vessel segmentation in OCT, how do 2D, 3D, and hybrid 2D + 3D deep learning designs trained under weakly supervised regimes perform compared with fully supervised baselines?

Together, these questions define the scope of the review: RQ1 maps the landscape of weakly supervised ocular segmentation methods as reported in surveys and reviews, whereas RQ2 and RQ3 focus on two clinically important settings to examine what current review articles reveal about potential gains in label efficiency and architectural choices.

2.3. Search Strategy

This umbrella review targeted review-type literature (surveys, systematic/scoping reviews, and tutorials) on weakly supervised deep learning for medical image segmentation, with an emphasis on ocular imaging (fundus and OCT). Searches were conducted between 2020 and 2026 and restricted to English-language full texts.

The primary search was conducted in Google Scholar using Publish or Perish (PoP) to ensure broad cross-publisher coverage, capturing journal and conference surveys, as well as relevant preprints. To improve the coverage of biomedical and interdisciplinary indexing, complementary targeted searches were performed in PubMed and Scopus using database-adapted Boolean strategies. The full search strings, PoP settings, execution dates, and applied filters are reported in Appendix A.

Table 1. Overview of the surveys and reviews included in this study and their coverage of weakly supervised segmentation aspects.

Review Articles	Description	Medical Imaging		Weak Labels			Learning Strategies			Backbones	
		OCT	Fundus	CAM	MIL	Sparse	Auto-Labels	Anatomy Rules	U-Net Family	DeepLab Family	Transformer/Hybrid
Rizvana & Narayanan (2024) [5]	Deep learning of fundus images and optical coherence tomography images for ocular disease detection—A review	✓	✓						✓		
Xue et al. (2024) [7]	Analysis of Atherosclerotic Plaques Using OCT Images Based on Deep Learning: A Comprehensive Review	✓							✓	✓	
Chen et al. (2025) [6]	Denosing, segmentation and volumetric rendering of optical coherence tomography angiography (OCTA) images using deep learning techniques: A comprehensive review	✓							✓		
Alawad et al. (2022) [1]	Machine Learning and Deep Learning Techniques for Optic Disc and Cup Segmentation—A Review		✓								
Goutam et al. (2022) [2]	A Comprehensive Review of Deep Learning Strategies in Retinal Disease Diagnosis Using Fundus Images		✓	✓					✓	✓	
Anusuya & Masoodhu Banu (2023) [3]	A Comprehensive Review of Glaucoma Detection from Fundus Images using Deep Learning		✓								
Zedan et al. (2023) [4]	Automated Glaucoma Screening and Diagnosis Based on Retinal Fundus Images Using Deep Learning Approaches		✓	✓					✓	✓	
Zhang et al. (2020) [8]	A survey of semi- and weakly supervised semantic segmentation of images			✓	✓	✓	✓	✓	✓	✓	✓
Ouassit et al. (2022) [34]	A Brief Survey on Weakly Supervised Semantic Segmentation			✓	✓	✓	✓		✓	✓	
Hassan et al. (2022) [31]	Supervised and weakly supervised deep learning models for COVID-19 CT diagnosis: A systematic review				✓		✓		✓	✓	✓
Qu et al. (2022) [25]	Towards label-efficient automatic diagnosis and analysis: a comprehensive survey of advanced deep learning-based weakly supervised, semi-supervised and self ...				✓		✓		✓	✓	
Shen et al. (2023) [24]	A Survey on Label-Efficient Deep Image Segmentation: Bridging the Gap Between Weak Supervision and Dense Prediction			✓	✓	✓	✓	✓	✓	✓	✓
Bohlender et al. (2023) [9]	A Survey on Shape-Constraint Deep Learning for Medical Image Segmentation							✓	✓		
Zhang et al. (2025) [27]	A Review of Non-Fully Supervised Deep Learning for Medical Image Segmentation			✓	✓	✓	✓	✓	✓	✓	✓
Gao et al. (2025) [10]	Medical Image Segmentation: A Comprehensive Review of Deep Learning-Based Methods					✓	✓	✓	✓		
Green (2025) [26]	Learning with Imperfect Labels and Incomplete Views: A Review of Representation Methods for Weakly Supervised Perception						✓				✓
This review (PRISMA)	Weakly Supervised Deep Learning for Ocular Image Segmentation: A Systematic Umbrella Review of Surveys and Reviews on Fundus and OCT Methods	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

For each article, the table indicates whether OCT and/or fundus imaging are discussed, which weak-label types are considered (CAM-based, multi-instance learning, sparse annotations), which label-efficient learning strategies are addressed (automatic or pseudo-labelling, anatomy/shape rules), and which segmentation backbone families are reported (U-Net, DeepLab and Transformer/hybrid architectures). Check marks denote an explicit discussion of the corresponding modality, label type, strategy or backbone. The last row summarizes the scope of the present review.

For the PoP component, results were restricted at the search stage to survey-type publications by requiring “review” OR “survey” in the title. Three PoP title/keyword profiles were used (executed in January 2026): (1) weak supervision-focused reviews, (2) ocular modality-focused reviews (fundus/OCT) with weak supervision keywords, and (3) segmentation-focused reviews with ocular + weak supervision keywords (Appendix A).

For PubMed and Scopus, Boolean strategies combined (i) weak supervision/weak-label terminology, (ii) segmentation terms, (iii) ocular imaging/anatomical terms (fundus/OCT/retina/ocular), and (iv) review-type terms (Appendix A). The PubMed search retrieved four records, and the Scopus search retrieved six records (2020–2026). These were merged with the PoP set and subjected to deduplication and screening; none met the inclusion criteria during the full-text assessment (predominantly due to not meeting IC2/IC3). These complementary searches are reported to document the breadth of sources considered and support reproducibility.

In addition to the PRISMA-counted identification process, a targeted PoP query was used to identify recent ocular image segmentation reviews (fundus/OCT) to strengthen the background coverage. These records were used for narrative context only and were not counted as included studies unless they met IC2 (explicitly addressing weak supervision). Accordingly, citations [11–21] are referenced as contextual background and did not alter the PRISMA flow diagram.

Screening was conducted in two stages by a single reviewer, as follows: First, titles, abstracts, and keywords were screened to confirm review-type status and topical relevance to weak supervision and segmentation/related perception tasks. Second, the full texts were assessed against predefined inclusion and exclusion criteria (Tables 2 and 3). The identification, screening, and eligibility workflows followed the PRISMA 2020 [35] guidelines. Two key pre-2020 tutorial papers were used for the conceptual background only and were not part of the PRISMA count.

Table 2. Inclusion criteria.

Identifier	Inclusion Criteria
IC1	Review-type articles published in journals or conferences; preprints were also eligible.
IC2	Uses deep learning and addresses weak supervision.
IC3	Discusses methods or applications relevant to weak supervision and/or segmentation or related perception tasks (e.g., segmentation, detection/localization, saliency, temporal localization), including medical and ocular imaging.
IC4	The full text is available in English.
IC5	Published between 2020 and 2026.

Table 3. Exclusion criteria.

Identifier	Exclusion Criteria
EC1	Not a review-type article.
EC2	Not involving deep learning methods.
EC3	Not about weak supervision and not relevant to segmentation/related perception tasks.
EC4	Full text not obtainable (or not available in English).
EC5	Duplicate record of another included review.

2.4. Data Extraction

For each of the 21 included reviews, data were extracted into a structured spreadsheet using a pre-defined extraction form. Data extraction was performed by a single reviewer.

To mitigate potential errors from single-reviewer procedures, independent verification was performed by a second reviewer on a random subset of the included reviews. Full-text eligibility decisions and key extracted fields were rechecked against the source articles, and discrepancies were resolved by consensus. The extracted items included the following: (i) bibliographic metadata (first author, year, publication venue); (ii) review scope (domain focus and imaging modalities covered); and (iii) relevance to ocular segmentation (whether segmentation and fundus/OCT imaging were primary topics or secondary components). Imaging modalities were coded as reported and grouped into fundus-based and OCT-based categories, as previously described. When specified, fundus imaging was recorded as colour fundus photography/retinography, FA, or FAF, and OCT was coded as structural OCT or OCTA. Generic mentions of “fundus” without acquisition details were recorded as unspecified.

To support the synthesis aligned with RQ1–RQ3, the following methodological descriptors were additionally coded: weak-label categories (e.g., CAM-based, MIL-based, and sparse supervision), label-efficient learning strategies (e.g., pseudo-/auto-labelling, self-training, semi-/self-supervision, and shape/anatomy-constrained learning), and model families/backbones (e.g., U-Net variants, DeepLab-like architectures, and Transformer/hybrid designs). For RQ2 and RQ3, the reporting of (a) segmentation metrics, (b) annotation-effort indicators, and (c) explicit comparisons to fully supervised baselines were also recorded. These data were used to populate summary tables and support the qualitative synthesis presented in the Section 3.

No effect measures were calculated because no quantitative synthesis or meta-analysis was performed. Reporting bias was not assessed for the same reasons. Certainty (confidence) in the body of evidence was not assessed because this review provided a descriptive synthesis of secondary studies.

2.5. Methodological Quality Assessment of Included Reviews

The methodological quality was appraised using AMSTAR-2. For each included review, items were rated as Yes (Y), Partial Yes (PY), No (N), Unclear (U), or Not applicable (NA) based on what was explicitly reported; unreported items were coded as U. Item-level ratings are reported in Table S1 (Supplementary Materials), and overall confidence ratings are summarized in Table 4. Overall confidence (High, Moderate, Low, or Critically Low) followed the AMSTAR-2 guidance, considering key domains such as protocol registration, search adequacy, justification of exclusions, and consideration of risk of bias in interpretation. Findings were interpreted with greater emphasis on high/moderate reviews, while Low/Critically Low surveys were used primarily for taxonomy mapping and were not treated as equally robust evidence. The protocol registration in Table 4 refers to the registration reported by the included reviews; the umbrella review protocol is registered in OSF (registration ID wd8gu, embargoed).

2.6. Synthesis and Quality Appraisal

A descriptive qualitative synthesis of secondary studies (surveys/reviews) was performed, grouping reviews by domain focus and imaging modality (fundus vs. OCT) and summarizing each research question using Tables 4–8 and narrative synthesis. No statistical synthesis/meta-analysis was planned or performed; therefore, heterogeneity exploration and sensitivity analyses were not performed. Because the unit of analysis comprised secondary studies, no primary study risk-of-bias assessment was performed, which is acknowledged as a limitation. However, the methodological quality of the reviews included was appraised using AMSTAR-2 (Section 2.5), with item-level ratings reported

in Table S1 and overall confidence summarized in Table 4. These ratings were considered when interpreting the findings.

The overlap of primary studies across the included reviews was not formally quantified (e.g., corrected covered area) because the unit of analysis was secondary studies, and many included records were narrative surveys that did not report complete lists of included primary studies. However, potential overlap was considered qualitatively during the synthesis by noting the repeated citation of benchmark datasets and widely used architecture across reviews. This limitation is acknowledged in the Section 4.

Table 4. AMSTAR-2 summary appraisal of included reviews.

ID	Included Review	Review Type	Protocol Registered?	Adequate Search?	Selection/Extraction in Duplicate?	RoB Assessed?	Overall Confidence
R1	Zhang (2020) [8]	Narrative survey	N	N	Y/N	N	Critically low
R2	Ouassit (2022) [34]	Narrative survey	N	N	Y/N	N	Critically low
R3	Shen (2023) [24]	Narrative survey	N	PY	Y/N	N	Critically low
R4	Qu (2022) [25]	Narrative survey	N	PY	Y/N	N	Critically low
R5	Green (2025) [26]	Narrative survey	N	N	N/N	N	Critically low
R6	Zhang (2025) [27]	Narrative survey	N	PY	Y/N	N	Critically low
R7	Bohlender (2023) [9]	Narrative survey	N	PY	Y/N	N	Critically low
R8	Gao (2025) [10]	Narrative survey	N	PY	Y/Y	N	Critically low
R9	Fasana (2022) [28]	Narrative survey	N	PY	N/N	N	Critically low
R10	Martínez-Heredia (2025) [29]	Narrative survey	N	Y	Y/N	Y	Critically low
R11	Zhao (2023) [30]	Narrative survey	N	PY	N/N	N	Critically low
R12	Hassan (2022) [31]	Systematic review	N	Y	Y/N	N	Critically low
R13	Liu (2023) [32]	Narrative review	N	PY	Y/N	N	Critically low
R14	Mehrnia (2025) [33]	Narrative review	N	Y	Y/Y	N	Critically low
R15	Alawad (2022) [1]	Narrative review	N	PY	Y/N	N	Critically low
R16	Goutam (2022) [2]	Narrative review	N	PY	Y/N	N	Critically low
R17	Anusuya (2023) [3]	Narrative review	N	PY	Y/N	N	Critically low
R18	Zedan (2023) [4]	Narrative review	N	Y	Y/N	N	Critically low
R19	Rizvana & Narayanan (2024) [5]	Narrative review	N	N	Y/N	N	Critically low
R20	Xue (2024) [7]	Narrative review	N	PY	Y/N	N	Critically low
R21	Chen (2025) [6]	Narrative review	N	PY	Y/N	N	Critically low

Note: AMSTAR-2 was applied based on what was explicitly reported in the included studies. Most of the included articles were narrative surveys/reviews, for which items such as duplicate screening/extraction and protocol registration are often not reported; these were coded as Unclear (U), which lowers the overall confidence.

3. Results

The structured citation search performed with Harzing’s Publish or Perish (PoP) identified 28 records, and PubMed ($n = 4$) and Scopus ($n = 6$) added 10 additional records for a total of 38 (Figure 2). After screening the titles and abstracts, 38 records were retained for full-text assessment. One report could not be retrieved despite comprehensive searches using institutional access and open-access sources. Of the remaining 37 records, 15 were excluded at the full-text stage for not meeting IC2 and/or IC3, and one was excluded because the full text was not available in English (IC4). Overall, 21 review articles were included in the qualitative synthesis (Figure 2), and the selection process was reported in accordance with PRISMA 2020 [35]. Two foundational weak-supervision papers published before 2020 were added as conceptual backgrounds [22,23] but were not counted in the PRISMA flow.

In the following sections, each “study” refers to an included survey or review article (R1–R21, Table 5). Primary weakly supervised segmentation methods are discussed only through their coverage in secondary sources. The main characteristics of the 21 reviews are summarized in Table 5. Publications span 2020–2026, with a marked increase after 2022. The corpus can be grouped into three categories: (i) generic weakly supervised learning and segmentation surveys [8,24–34], (ii) organ- or modality-specific medical segmentation reviews (e.g., ultrasound, CT) [9,10,32,33], and (iii) ocular and OCT-focused reviews [1–7].

Generic WSL segmentation surveys provide broad taxonomies of weak labels, label-efficiency strategies and architectures across computer vision and medical imaging [8,24–27,34], while organ-specific reviews characterize segmentation architectures and practical constraints for particular modalities [9,10,32,33]. Ocular reviews concentrate on fundus-based

glaucoma and retinal disease analysis and on structural OCT pipelines [1–7]. A simple timeline (Figure 3) shows that generic WSL and segmentation surveys preceded a more recent wave of modality-focused and ocular-related reviews, although ocular segmentation remains a minority topic in label-efficient literature.

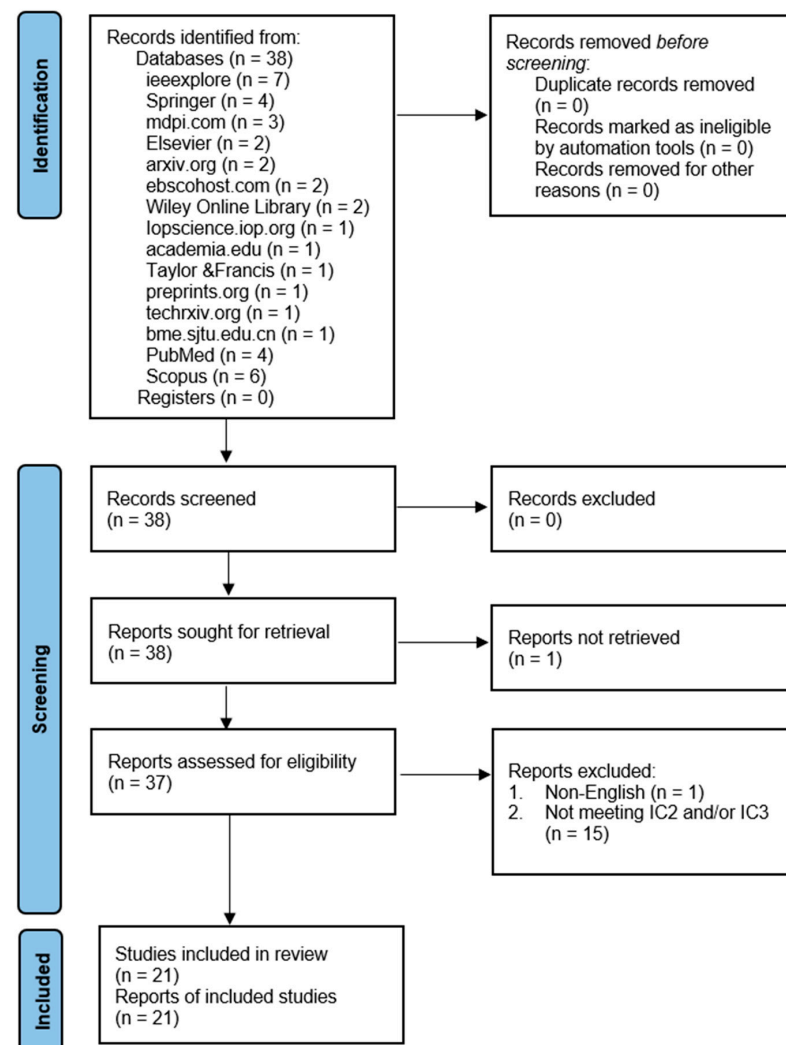


Figure 2. PRISMA flow diagram.

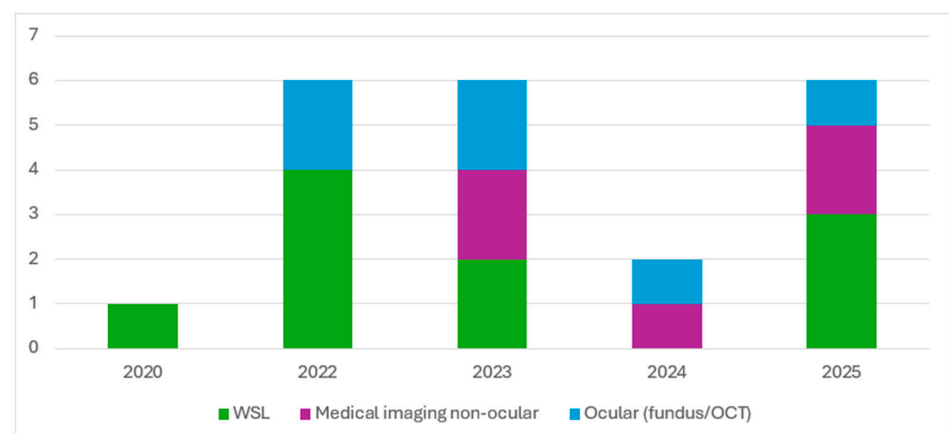


Figure 3. Timeline of included reviews by year and category.

Table 5. Characteristics of included reviews (global overview).

ID	First Author (Year)	Domain Focus	Segmentation as a Main Topic?	Ocular Imaging Covered?
R1	Zhang (2020) [8]	Generic WSL/segmentation	Yes	No
R2	Ouassit (2022) [34]	Generic WSL/segmentation	Yes	No
R3	Shen (2023) [24]	Generic WSL/segmentation	Yes	No
R4	Qu (2022) [25]	Generic WSL/segmentation	Partly	No
R5	Green (2025) [26]	Generic WSL	No	No
R6	Zhang (2025) [27]	Medical segmentation	Yes	No
R7	Bohlender (2023) [9]	Medical segmentation	Yes	Occasional
R8	Gao (2025) [10]	Medical segmentation	Yes	Occasional
R9	Fasana (2022) [28]	Other domain	No	No
R10	Martínez-Heredia (2025) [29]	Other domain	No	No
R11	Zhao (2023) [30]	Other domain	No	No
R12	Hassan (2022) [31]	Medical imaging	Partly	No
R13	Liu (2023) [32]	Medical segmentation	Yes	No
R14	Mehrnia (2025) [33]	Medical segmentation	Yes	No
R15	Alawad (2022) [1]	Ocular	Yes	Fundus
R16	Goutam (2022) [2]	Ocular	Partly	Fundus
R17	Anusuya (2023) [3]	Ocular	Partly	Fundus
R18	Zedan (2023) [4]	Ocular	Partly	Fundus
R19	Rizvana & Narayanan (2024) [5]	Ocular	Partly	Fundus + OCT
R20	Xue (2024) [7]	OCT	Yes	OCT
R21	Chen (2025) [6]	Ocular	Yes	OCTA

Methodological quality was appraised using AMSTAR-2 (Section 2.5). Overall confidence ratings are summarized in Table 4, with item-level assessments in Table S1 (Supplementary Materials). Most included records were narrative surveys/reviews with limited reporting of protocol registration, duplicate screening/extraction, and risk-of-bias assessment of primary studies; therefore, many items were coded as Unclear, yielding predominantly Low/Critically Low confidence. These ratings were used to contextualize the strength of evidence during synthesis.

The depth of weak-label and label-efficiency coverage is uneven (Table 6). Six generic surveys (Zhang et al. (2020), Ouassit et al., Shen et al., Qu et al., Green, and Zhang et al. (2025)) detail weak labels (e.g., CAM, MIL, sparse points/scrabbles/boxes) and strategies such as pseudo-/auto-labelling, self-training, consistency regularization, and self-supervised pretraining [8,24–27,34]. Recent work increasingly combines weak supervision with self-supervised/contrastive pretraining and Transformer-based backbones (e.g., ViT-style or CNN–Transformer hybrids), but the ocular review literature rarely discusses these SSL/CL and Transformer-oriented approaches in a task-specific manner for fundus or OCT/OCTA segmentation, limiting reproducible weak-vs-fully supervised comparisons under modern architectures [24,25,27]. Shape-constraint and architectural reviews emphasize anatomy-aware losses and the shift from U-Net/DeepLab to Transformer/hybrid designs but largely assume dense supervision [9,10]. Ocular and OCT reviews mention weak supervision only sporadically (often CAM visualization or transfer learning) and do not systematically categorize weak-label types or label-efficient strategies [1–7]. Consequently, no included article simultaneously has weak supervision as the primary topic, segmentation as the main task, and ocular imaging as the central domain (Figure 4).

For RQ1, namely which weakly supervised deep learning approaches have been proposed for ocular image segmentation, the included ocular reviews indicate that deep learning-based segmentation of optic disc/cup (OD/OC), vessels, lesions and retinal layers is now routine but is almost always presented as a fully supervised task. Alawad et al. systematically review OD/OC segmentation in fundus photographs and report that U-Net-style encoder–decoder architectures dominate this literature [1]. Goutam et al. describes retinal-

disease pipelines in which segmentation of vessels, lesions and OD/OC regions supports the classification of diabetic retinopathy, glaucoma and age-related macular degeneration [2]. Glaucoma-specific surveys similarly focus on supervised OD/OC segmentation as a source of structural biomarkers feeding downstream classifiers [3,4]. Rizvana and Narayanan cover both fundus and OCT, giving architectural context for segmentation and diagnosis; weak supervision appears mainly in the form of CAM-/attention-based visual explanations rather than as a primary training paradigm [5]. For OCT, Xue et al. discuss 2D and 3D CNNs for intravascular OCT plaque component segmentation [7], and Chen et al. review OCTA denoising, segmentation and volumetric rendering using 2D, 3D and hybrid architectures [6]. When contrasted with the broader weakly supervised segmentation toolbox summarized in generic WSL surveys (image-level labels, MIL, sparse annotations, pseudo-labelling, semi-/self-supervision, anatomy-aware losses), these ocular reviews rarely frame segmentation as a weakly supervised problem and do not assemble ocular-specific WSL design patterns. Overall, deep learning segmentation for ocular structures is well represented, but weakly supervised ocular segmentation is not systematically addressed in existing reviews.

Table 6. Weak labels, strategies and backbones.

ID	First Author (Year)	Weakly Supervised Learning Coverage	Weakly Supervised Label Types	Label-Efficient Strategies
R1	Zhang (2020) [8]	Detailed	CAM, MIL, sparse (points/scribbles/boxes)	Pseudo-/self-training, consistency, semi-supervision
R2	Ouassit (2022) [34]	Detailed	CAM, seeds, sparse, some MIL	Seed expansion, losses, post-processing, some pseudo-labelling
R3	Shen (2023) [24]	Detailed	Weak labels broadly	Pseudo-label curricula, consistency, SSL/contrastive, semi-/self-supervision
R4	Qu (2022) [25]	Detailed	Weak, semi-, self-supervision	Pseudo-labels, self-training, SSL, multi-task/curriculum
R5	Green (2025) [26]	Conceptual	Inexact/inaccurate labels, partial views	Robust objectives, sample re-weighting, multi-view consistency
R6	Zhang (2025) [27]	Detailed	Weak, semi-, self-supervision; sparse labels	Pseudo-/auto-labelling, SSL, teacher–student, hybrid supervision
R7	Bohlender (2023) [9]	Conceptual	Not main focus; mostly assumes full labels	Shape/topology priors, anatomy-aware losses, CRF
R8	Gao (2025) [10]	Conceptual	Mostly fully supervised; brief WSL/SSL mentions	Transfer learning, occasional semi-/self-supervision
R9	Fasana (2022) [28]	Detailed	Image-level labels, MIL, proposal boxes	MIL training, pseudo-labelling, OICR-style refinement
R10	Martínez-Heredia (2025) [29]	Detailed	Weak labels in time-series (incomplete/inexact)	Self-training, EM-style refinement, heuristics
R11	Zhao (2023) [30]	Detailed	Incomplete/inaccurate labels, weak supervision	Domain adaptation, pseudo-labels, MIL, SSL
R12	Hassan (2022) [31]	Conceptual	Weak labels mainly for diagnosis (COVID-19 CT)	Pseudo-labelling, semi-supervision (classification-heavy)
R13	Liu (2023) [32]	Conceptual	Weak labels mentioned only tangentially	Focus on denoising + supervised segmentation; some transfer learning
R14	Mehrnia (2025) [33]	Conceptual	Non-fully supervised approaches briefly noted	Mainly full supervision; occasional SSL/semi-supervision
R15	Alawad (2022) [1]	Minimal	WSL rarely/only briefly mentioned	Transfer learning, some CAM-based visualization
R16	Goutam (2022) [2]	Conceptual	WSL/label efficiency mentioned at high level	Transfer learning, data augmentation, some CAM
R17	Anusuya (2023) [3]	None	–	Focus on supervised glaucoma classification and features
R18	Zedan (2023) [4]	Minimal	CAM/heatmaps occasionally for explainability	Supervised pipelines; transfer learning
R19	Rizvana & Narayanan (2024) [5]	Conceptual	CAM/attention for interpretability	Transfer learning, some attention mechanisms
R20	Xue (2024) [7]	None	–	Supervised training, transfer learning
R21	Chen (2025) [6]	None	–	Supervised training; some multi-task/denoising

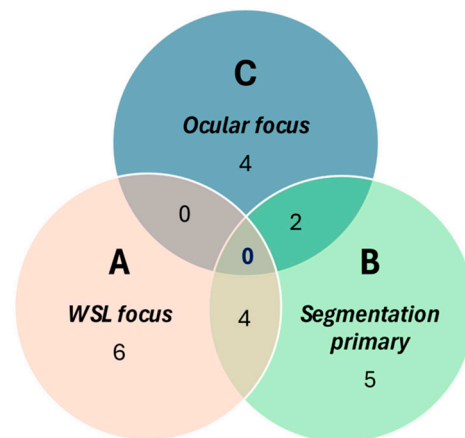


Figure 4. Overlap between WSL focus, segmentation focus and ocular focus.

To contextualize RQ1 in terms of data, representative ocular segmentation datasets recurring across the included reviews and their cited primary studies are summarized in Table 7 [1–7]. For fundus imaging, vessel segmentation relies primarily on DRIVE [37], STARE [38], CHASE_DB1 [39], HRF [40], DR HAGIS [41], RITE [42] and IOSTAR [43], which provide pixel-wise vessel masks in small to moderate cohorts [2,5]. OD/OC and glaucoma-related segmentation studies predominantly use ORIGA(-light) [44], DRISHTI-GS1 [45], DRIONS-DB [46], RIM-ONE DL [47], RIGA [48], REFUGE/REFUGE2 [49,50] and G1020 [51], which supply disc and cup boundaries, often with glaucoma labels [1,3–5]. Diabetic retinopathy lesion segmentation is typically based on IDRiD [52], e-optha (MA/EX) [53] and SUSTech-SYSU EX [2,5,54]. For OCT and OCTA, commonly cited datasets include Duke DME/SD-OCT [55] and RETOUCH [56] for fluid and layer segmentation, and OCTA-500 [57] and ROSE [58] for vessel and foveal avascular zone segmentation [5–7]. Across these datasets, labels are almost exclusively dense pixel-wise or contour annotations; weak labels (image-level tags, scribbles, boxes) rarely exist as native annotations [8,10,24,25,27,34]. In practice, weak supervision in ocular segmentation is therefore usually induced by the training strategy rather than by dataset design, and this distinction is not yet made explicit in current ocular reviews [1–8,24,25,34].

For RQ2, Table 8 summarizes the five fundus-based reviews that include OD/OC segmentation [1–5]. This task differs from retinal segmentation and is often more subjective between graders, which makes ground-truth harder to define. All five identify manual disc and cup delineation as labour-intensive and variable, and several report segmentation metrics for fully supervised OD/OC methods. None of these articles, however, report quantitative measures of annotation effort and none presents head-to-head comparisons between weakly supervised and fully supervised OD/OC approaches on the same datasets. Mentions of weak or label-efficient ideas, where they appear at all, are confined to generic transfer learning, CAM-based visualizations or coarse region proposals, rather than dedicated weakly supervised training schemes. As a result, the existing review literature does not yet quantify how much manual effort weakly supervised OD/OC pipelines can save, nor how their segmentation performance compares systematically with fully supervised baselines.

For OCT/OCTA, reported weak supervision evidence was heterogeneous across tasks (fluid, layer surfaces, vessels/FAZ) and data structures (2D B-scans vs. 3D volumes); therefore, findings are summarized by task family rather than pooled across all OCT studies. For RQ3, Table 9 summarizes the reviews that discuss volumetric deep learning designs for medical image segmentation [6,7,10,27,32,33]. Multi-modal segmentation surveys describe 2D slice-based, 3D volumetric and hybrid 2D + 3D architectures across CT, MRI and

ultrasound, typically assuming fully supervised training [10,27,32,33]. Xue et al. compare 2D slice-wise and 3D volumetric CNNs for intravascular OCT plaque segmentation [7], and Chen et al. survey 2D, 3D and multi-view hybrid networks for OCTA vessel and plexus segmentation and volumetric rendering [6]. These works qualitatively characterize trade-offs between 2D, 3D and hybrid designs in terms of 3D context, memory footprint and computational burden, and they report method-specific segmentation performance. However, none of the reviews focus on weakly supervised training, and none examine how 2D, 3D or hybrid OCT designs for retinal layer or vessel segmentation behave under weak labels relative to fully supervised baselines, or how performance and annotation cost trade off across these architectural choices.

In summary, the PRISMA flow (Figure 2) and the synthesis across Tables 4–8 indicate that the existing survey and review literature provides a solid conceptual and architectural background for weakly supervised learning, as well as for ocular image segmentation. However, three consistent gaps remain: (i) no review is dedicated specifically to weakly supervised ocular segmentation; (ii) no review quantifies the label-efficiency advantages of weakly supervised OD/OC segmentation in fundus images; and (iii) comparative analyses of 2D, 3D, and hybrid OCT architectures trained under weak supervision against fully supervised baselines are largely absent. Together, these gaps motivate a follow-on systematic review of primary studies on weakly supervised ocular segmentation.

Table 7. Datasets used for ocular image segmentation.

Dataset	Modality	Dim.	Primary Segmentation Tasks	Size (Approx.)	Label Type	Weak-Label Availability (Native vs. Constructed)	Public
DRIVE [37]	Fundus	2D	Retinal vessels	40 images (20 train, 20 test)	Pixel-wise vessel masks (2 annotators for test)	Dense only	Yes
STARE [38]	Fundus	2D	Retinal vessels	20 images	Pixel-wise vessel masks (2 annotations)	Dense only	Yes
CHASE_DB1 [39]	Fundus	2D	Retinal vessels	28 images	Pixel-wise vessel masks	Dense only	Yes
HRF [40]	Fundus	2D	Retinal vessels	45 images (15 + 15 + 15)	Pixel-wise vessel masks	Dense only	Yes
DR HAGIS [41]	Fundus	2D	Retinal vessels	39 images	Pixel-wise vessel masks	Dense only	Yes
RITE [42]	Fundus	2D	Vessels + artery/vein (A/V) labels	40 images	Vessel masks + A/V labels	Dense only	Yes
IOSTAR [43]	Fundus	2D	Retinal vessels, OD, A/V ratio	30 images (1024 × 1024)	Pixel-wise vessels, OD/A/V labels	Dense only	Yes
ORIGA (-light) [44]	Fundus	2D	OD/OC	650 images (482 healthy, 168 glaucoma)	OD & OC boundaries	Dense only	Partly
DRISHTI-GS1 [45]	Fundus	2D	OD/OC	101 images (train/test split)	OD & OC masks (multi-expert)	Dense only	Yes
DRIONS-DB [46]	Fundus	2D	OD	110 images	OD contours (2 annotations)	Dense only	Yes
RIM-ONE DL [47]	Fundus	2D	OD/OC	485 images (313 normal, 172 glaucoma)	OD/OC masks	Dense only	Yes
RIGA [48]	Fundus	2D	OD/OC	750 images from 3 sources	OD/OC boundaries (multi-expert)	Dense only	Yes
REFUGE/REFUGE2 [49,50]	Fundus	2D	OD/OC (+glaucoma labels)	1200 images	OD/OC masks + diagnosis labels	Native image-level + dense	Yes
G1020 [51]	Fundus	2D	OD/OC	1020 images (724 healthy, 296 glaucoma)	OD/OC masks + labels	Native image-level + dense	Yes

Table 7. Cont.

Dataset	Modality	Dim.	Primary Segmentation Tasks	Size (Approx.)	Label Type	Weak-Label Availability (Native vs. Constructed)	Public
IDRiD [52]	Fundus	2D	Lesion segmentation (MA, EX, HE, soft EX) + OD/FAZ in some splits	81 images with pixel-wise lesion labels (+extra images for grading)	Pixel-wise lesion masks (+ DR grades)	Native image-level + dense	Yes
e-optha (MA/EX) [53]	Fundus	2D	Lesion segmentation (MAs, exudates)	MA: 148 lesion + 233 normal; EX: 47 lesion + 35 normal	Pixel-wise lesion masks	Dense only	Yes
SUSTech-SYSU EX [54]	Fundus	2D	Exudate segmentation	1400+ images (various subsets)	Pixel-wise exudate masks	Dense only	Yes
Duke DME/Duke SD-OCT [55]	OCT	2D	Fluid regions ± layers	~110 B-scans from 10 eyes (canonical dataset)	Pixel-wise fluid/layer labels	Dense only	Yes
RETOUCH [56]	OCT	3D	Fluid segmentation (IRF, SRF, PED)	>70 training volumes + test from 3 vendors	Voxel-wise fluid labels	Dense only	Yes
OCTA-500 [57]	OCT + OCTA	3D + 2D	Vessels (large/capillary), arteries/veins, FAZ (2D/3D), retinal layers	500 subjects, multi-FOV	Multi-label (vessels, FAZ, layers)	Dense only	Yes
ROSE [58]	OCTA	2D	Vessel segmentation, FAZ	229 images in 3 subsets	Centerline + pixel-wise vessel/FAZ labels	Native sparse + dense	Yes
OCT layer challenge datasets [59,60]	OCT	2D + 3D	Retinal layers	Dozens–hundreds of volumes depending on dataset	Layer boundary/surface annotations	Mixed/unclear	Mixed

Weak-label availability was coded from the dataset’s reported label types. “Native image-level” denotes diagnosis/grades provided by the dataset. “Native sparse” denotes sparse annotations such as centerlines; otherwise, datasets were coded as “Dense only”.

Table 8. OD/OC label-efficiency evidence in fundus reviews (RQ2).

ID	First Author (Year)	OD/OC Segmentation Role	WSL Methods for OD/OC?	Annotation-Effort Metrics Reported?	WSL vs. FS OD/OC Comparison?	OD/OC Metrics Reported
R15	Alawad (2022) [1]	Main focus	No systematic WSL; occasional mention of CAM/feature extraction only	No	No	Yes
R16	Goutam (2022) [2]	Component	Only high-level/brief references to weak or label-efficient ideas	No	No	Yes
R17	Anusuya (2023) [3]	Minimal/indirect	None for segmentation; focus on supervised classification and hand-crafted or learned features	No	No	Yes
R18	Zedan (2023) [4]	Component	Minimal; weak supervision not analyzed	No	No	Yes
R19	Rizvana & Narayanan (2024) [5]	Component	Conceptual mentions of CAM/attention as interpretability; no explicit WSL OD/OC segmentation analysis	No	No	Yes

Table 9. 2D vs. 3D vs. hybrid designs for volumetric segmentation (focus on OCT; RQ3).

ID	First Author (Year)	Modality Focus (Volumetric)	Main Segmentation Tasks	Designs Discussed (2D/3D/Hybrid)	Supervision Regime for These Designs	Quantitative Comparison Between Designs?	Example Datasets Mentioned
R6	Zhang (2025) [27]	Multi-modal medical imaging	Organ/tumour and structure segmentation across CT, MRI, etc.	2D, 3D, hybrid	Mix of full, weak, semi- and self-supervision; WSL examples mostly non-ocular	Qualitative only	Various CT/MRI datasets
R8	Gao (2025) [10]	Multi-modal medical imaging	General organ and lesion segmentation (CT, MRI, US, OCT as one of many)	2D, 3D, hybrid	FS training	Qualitative comparison of context vs. computation	Common multi-organ CT/MRI benchmarks; OCT only as minor example
R13	Liu (2023) [32]	Ultrasound	Organ/lesion segmentation with denoising emphasis	2D, 3D, hybrid	FS training	Qualitative remarks on 2D vs. 3D performance	Several US benchmarks (non-ocular)
R14	Mehrnia (2025) [33]	Lung CT	Lung tumour/lesion segmentation	2D only	FS training	No	Lung CT datasets
R20	Xue (2024) [7]	Intravascular OCT	Plaque component segmentation	2D, 3D; hybrid	FS training	Qualitative comparison of 2D vs. 3D	Intravascular OCT datasets from specific centres
R21	Chen (2025) [6]	OCTA	Vessels, plexus segmentation	2D, 3D, hybrid	FS training	Qualitative discussion of 2D vs. 3D vs. hybrid designs	OCTA-500, ROSE and other OCTA datasets

4. Discussion

This review aimed to understand how weakly supervised deep learning strategies are currently reflected in the review literature on ocular image segmentation. Using an umbrella-review perspective and structuring the 21 included articles along three axes, weakly supervised learning focus, segmentation focus and ocular focus (Tables 4 and 5, Figure 4), the intention was to identify not only what is already known but also where the conversation has not yet reached.

With respect to RQ1, the main message is one of imbalance. On the one hand, generic weak-supervision segmentation surveys form a mature conceptual backbone. They define types of weak labels, connect them to learning formulations such as CAM, MIL and sparse annotation, and describe rich strategy toolboxes that include pseudo-labelling, self-training, consistency regularization and self-supervised pretraining [8,22–34]. Shape-constraint and architecture-oriented reviews add further depth by organizing anatomy-aware losses and by tracing the evolution from U-Net and DeepLab families to Transformer-based and hybrid backbones [9,10]. On the other hand, when ocular imaging is considered specifically, existing reviews remain almost entirely framed around fully supervised pipelines. OD/OC segmentation in fundus images is comprehensively catalogued [1], segmentation of vessels and lesions is described as a component of broader retinal disease pipelines [2], and glaucoma surveys mostly treat OD/OC masks as structural biomarkers that feed into supervised classifiers [3,4]. OCT reviews document denoising, segmentation and volumetric rendering strategies, as well as 2D, 3D and hybrid architectures [6,7], but weak supervision is only mentioned in passing, typically in the form of CAM-based visualization or transfer learning [5]. In other words, the pieces needed to build weakly supervised ocular segmentation pipelines exist in generic literature, but they have not yet been assembled in an ocular-specific way.

Recent label-efficient segmentation literature increasingly combines weak supervision with self-supervised/contrastive pretraining and Transformer-based backbones (e.g., ViT style encoders or CNN–Transformer hybrids), which can reduce reliance on dense masks by strengthening representations from unlabeled data and improving robustness to sparse,

noisy, or coarse labels. However, within the ocular review literature included here, SSL/CL and Transformer-oriented strategies are rarely discussed in a task-specific manner for fundus or OCT/OCTA segmentation, limiting reproducible weak-vs-fully supervised comparisons under modern architectures [24,25,27].

For RQ2, the gap is even more concrete. All five fundus-based OD/OC or glaucoma reviews acknowledge that manual delineation of disc and cup is labour-intensive and subject to reader variability [1–5]. They report segmentation performance in terms of Dice, IoU or cup-to-disc ratio error, and they repeatedly rely on a shared set of well-established datasets such as ORIGA(-light) [44], DRISHTI-GS1 [45], RIM-ONE DL [47], RIGA [48], REFUGE/REFUGE2 [49,50] and G1020 [51] (Table 7). However, none of these reviews provide quantitative evidence about annotation effort, and none contrast weakly supervised OD/OC methods with fully supervised baselines in a systematic way. Measures such as the proportion of images with dense masks, clicks or scribbles per image, or annotation time per case simply do not appear. Weakly supervised learning, when mentioned, is limited to generic transfer learning, CAM visualization or coarse region proposals, without explicit accounting of how much annotation is saved. From a practical perspective, particularly for large-scale glaucoma screening, this means that clinicians and engineers are still left without a clear answer to the following question: “How much accuracy is lost, and how much labelling time is saved, if full supervision is replaced by a weakly supervised OD/OC pipeline?”

The picture for RQ3 is similarly incomplete, but in a different way. Several included reviews discuss architectural choices for volumetric data. Generic segmentation surveys and lung- or ultrasound-focused reviews differentiate 2D slice-wise, 3D volumetric and hybrid designs, and they describe trade-offs between spatial context and computational cost [10,27,32,33]. Intravascular OCT reviews go further by presenting concrete 2D and 3D networks and multi-view hybrids for plaque or vessel segmentation and volumetric rendering [6,7]. However, all these comparisons are carried out under full supervision. None of the reviews ask explicitly how 2D, 3D and hybrid architectures behave when labels are weak, sparse or noisy, and none contrast weakly supervised OCT segmentation against fully supervised baselines on a common footing. Considering that the OCT datasets most often cited such as Duke DME/SD-OCT [55], RETOUCH [56], OCTA-500 [57] and ROSE [58] are relatively modest in size and may come with heterogeneous labels (Table 7), this represents a non-trivial gap for real-world deployment.

To make the implications of RQ1–RQ3 actionable, a minimal procedure for future ocular weak-supervision studies and reviews is suggested. For RQ1 (what exists), methods should be categorized jointly by weak-label type (e.g., CAM/MIL; points/scribbles/boxes), training strategy (e.g., pseudo-labelling/self-training/SSL), and backbone family (CNN vs. Transformer/hybrid). For RQ2 (label efficiency in fundus OD/OC), studies should report a matched fully supervised baseline (same backbone, split, and budget) alongside explicit annotation-cost measures (e.g., time per case, clicks/scribbles, % densely labelled images) and the corresponding performance change. For RQ3 (OCT/OCTA), evidence should be summarized by task family (fluid, layers/surfaces, vessels/FAZ) and data structure (2D B-scans vs. 3D volumes), and comparisons should again include matched fully supervised baselines to isolate the effect of weak supervision.

Taken together, the evidence points to two practical priorities for future work. First (datasets), almost all commonly used ocular datasets provide dense pixel-wise or contour labels, so weak supervision is usually introduced by the training strategy rather than by the annotations themselves (Table 7). Future fundus and OCT datasets could deliberately include multi-level supervision, for example, a small subset with dense labels and a larger pool with coarse or weak annotations and report annotation protocols and labelling effort,

as encouraged in generic WSL segmentation surveys [8,24,25,34]. Second (architectures), architectural surveys highlight a shift from classical CNNs (U-Net, DeepLab) towards Transformer-based and hybrid models [10,27], but ocular segmentation practice still relies largely on U-Net-style encoder–decoder networks for fundus and OCT tasks [1,2,6,7]. Whether some backbones, for example, attention-based Transformers or CNN–Transformer hybrids, are more robust to weak, noisy or incomplete labels in ocular applications remains an open question that existing reviews do not address.

To support reproducible benchmarking in ocular weakly supervised segmentation, a minimum reporting set is recommended (Table 10). The current review literature rarely quantifies annotation effort, specifies how weak labels are obtained/constructed, or provides matched fully supervised baselines under identical data splits and backbones [8,24,25,34]. Standardized reporting of label source, annotation cost, training protocol, and evaluation setup would enable meaningful comparisons across modalities (fundus vs. OCT/OCTA), tasks (OD/OC, vessels, fluid, layers, FAZ), and data structures (2D vs. 3D). In addition, whenever weak and fully supervised models are compared, the same dataset splits, pre-processing, backbone capacity, and evaluation metrics should be used to isolate the effect of label supervision. Where applicable, studies should report uncertainty (e.g., repeated runs/CI) and statistical testing (e.g., Wilcoxon/Friedman) when comparing methods; such reporting is rarely described in the included ocular surveys and limits the strength of comparative conclusions.

Table 10. Recommended reporting template for weakly supervised ocular image segmentation studies.

Category	Item to Report	Minimum Details
Task & data	Modality/task/dimension	Fundus OD/OC (2D); OCT fluid (3D); OCTA vessels + FAZ (2D/3D).
Dataset	Dataset name + split	Dataset version; train/val/test split; cross-validation if used.
Label source	Label type	Image-level, points, scribbles, boxes, partial masks, temporal/clinical labels.
Weak-label availability	Native vs. constructed	Native weak labels or constructed from dense masks.
Annotation effort	Cost metric	Images labelled; clicks/scribbles; minutes per image; annotator expertise; tools used.
Baselines	Matched fully supervised baseline	Same backbone, same split, same preprocessing; training budget matched.
Method	Weak supervision mechanism	CAM/MIL, region growing/propagation, graph constraints, shape priors, pseudo-labelling, SSL/CL pretraining.
Training details	Core settings	Input resolution, augmentations, loss terms, optimizer/LR, epochs, early stopping.
Evaluation	Metrics + protocol	Dice/IoU, boundary metrics, AUPR (vessels), FAZ error; volume-wise vs. slice-wise rules.
Statistics	Uncertainty/testing	Confidence intervals; repeated runs; Wilcoxon/Friedman if applicable.
Reproducibility	Code + model availability	Code link, pretrained weights, seed control, hardware.
Bias & limitations	Failure modes	Domain shift, class imbalance, label noise, leakage risks.

Annotation burden directly limits the scalability of screening and monitoring workflows because dense delineations are often impractical in high-volume settings. Consistent reporting of annotation effort and matched weak-vs-fully supervised baselines is therefore necessary to judge whether label-efficient methods deliver clinically meaningful savings without unacceptable performance loss.

This review has several limitations. Study selection and data extraction were primarily conducted by a single reviewer, which may increase the risk of selection bias or extraction errors; however, an independent second reviewer verified full-text eligibility decisions and key extracted fields for a random subset, with discrepancies resolved by consensus. Future updates could further strengthen reliability by implementing full dual screening and dual extraction with adjudication. The search relied on Harzing’s Publish or Perish

with Google Scholar as the indexed source, supplemented by reference-list screening rather than parallel searches across multiple bibliographic databases, so some relevant reviews may have been missed. Because the unit of analysis was the review article, no formal risk-of-bias assessment at primary-study level was performed, and nuanced uses of weak supervision in individual experimental papers may therefore remain hidden. In addition, overlap of underlying primary studies across the included reviews was not quantified, so some conclusions may reflect repeated evidence from a small set of commonly cited primary sources. Finally, the categorization into “WSL focus”, “segmentation primary” and “ocular focus” simplifies heterogeneous article scopes, and some borderline cases required judgement based on the dominant narrative of each article.

Despite these limitations, several strengths should be noted. This appears to be the first systematic umbrella review that explicitly links generic weakly supervised and label-efficient segmentation surveys [8,22–34] with ocular segmentation reviews [1–7]. By mapping each article onto three focus axes (Figure 4) and organizing the evidence into concise tables for global characteristics, weak-label coverage, OD/OC segmentation, volumetric OCT design and datasets (Tables 4–8), this review offers a transparent view of where the literature is dense and where it is thin. In particular, it highlights that, although conceptual and architectural guidance for weakly supervised segmentation is ample, existing reviews do not yet provide the ocular-specific, quantitative evidence needed to determine how much label effort can be saved for OD/OC (RQ2), or which 2D/3D/hybrid designs are preferable for weakly supervised OCT segmentation (RQ3), and that no review is dedicated specifically to weakly supervised ocular segmentation (RQ1).

In summary, the field appears to be at an interesting crossroads. Deep learning segmentation is now standard for fundus and OCT imaging, and the broader weak-supervision community has developed powerful tools for learning from imperfect labels. The next step is to bring these strands together in a targeted way: through primary ocular studies that report not only accuracy but also annotation cost; through dataset designs that intentionally include weak labels and multi-level supervision; and through future ocular-focused WSL reviews or benchmarks that fill the empty intersection in Figure 4. Doing so will be essential if weakly supervised segmentation is to move from a conceptual promise to a practical tool for glaucoma screening, retinal disease monitoring and quantitative OCT analysis in everyday clinical practice.

5. Conclusions

This umbrella review shows that the review literature has not yet consolidated weak supervision into an ocular-segmentation evidence base: ocular surveys remain predominantly fully supervised, while weak-supervision surveys are largely modality-agnostic, leaving a persistent “empty intersection” for fundus and OCT/OCTA segmentation. To move beyond conceptual taxonomies, future work should prioritize (i) ocular datasets and benchmarks that include multi-level supervision (dense + weak labels), (ii) routine reporting of annotation effort alongside segmentation performance, and (iii) matched weak-versus-fully supervised comparisons under identical splits, backbones, and evaluation protocols. Closing these gaps is essential to judge whether label-efficient segmentation can deliver scalable benefits for glaucoma screening, retinal disease monitoring, and quantitative OCT analysis.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app16052241/s1>, Table S1: AMSTAR-2 item-level appraisal of included reviews.

Author Contributions: Conceptualization, P.P.; methodology, P.P.; validation, P.P. and A.C.; formal analysis, A.C.; investigation, P.P.; writing—original draft preparation, P.P.; writing—review and editing, A.C., A.M., R.A., J.M. and A.C.S.; supervision, A.C.; project administration, A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work is financed by National Funds through the Portuguese funding agency, FCT—Fundação para a Ciência e a Tecnologia, within project UID/50014/2025 (DOI: 10.54499/UID/50014/2025), and is associated with the clinical study registered at ClinicalTrials.gov under the identifier NCT06839443.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed at the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CL	Contrastive Learning
CNN	Convolutional Neural Network
DL	Deep Learning
FA	Fluorescein Angiography
FAF	Fundus Autofluorescence
FS	Fully Supervised
MIL	Multiple-Instance Learning
OCT	Optical Coherence Tomography
OCTA	Optical Coherence Tomography Angiography
OD/OC	Optic Disc/Optic Cup
PoP	Publish or Perish
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RoB	Risk of Bias
SSL	Self-Supervised Learning
ViT	Vision Transformer
WSL	Weakly Supervised Learning
WSSS	Weakly Supervised Semantic Segmentation

Appendix A

Full Search Strategies and Search Yields

Table A1. Information, sources, dates, filters and settings.

Source	Interface	Years	Language	Fields Searched	Key Settings
Google Scholar	Publish or Perish (PoP) v8.19	2020–2026	English	Title + keywords	Max results per query: 200; include citations: No; include patents: No; only review articles: Yes
PubMed	PubMed (web)	2020–2026	English	Title/ Abstract (tiab)	Filters: year range; English
Scopus	Scopus (Advanced Search)	2020–2026	English	TITLE-ABS-KEY	Filters: year range; English

Table A2. Reproducible search strategy (full queries, settings, and hits) for PoP/Google Scholar, PubMed, and Scopus.

Google Scholar (Publish or Perish v8.19)

Settings: Years 2020–2026; Language English; Fields searched Title + Keywords; Max results 200; exclude patents; exclude citations.

Query 1 (weak-supervision surveys):

Title: (“review” OR “survey”) AND (“weakly supervised” OR “weak supervision”)

Hits: 14

Query 2 (ocular modalities + weak supervision):

Title: (“review” OR “survey”) AND “deep learning” AND (“optical coherence tomography” OR “OCT” OR “fundus”); Keywords: (“weak supervision” OR “weakly supervised”)

Hits: 6

Query 3 (segmentation + ocular + weak supervision):

Title: (“review” OR “survey”) AND (“deep learning segmentation”); Keywords: (“optical coherence tomography” OR “OCT” OR “fundus images”) AND (“weak supervision” OR “weakly supervised”)

Hits: 8

PubMed (executed: 2020–2026)

Query:

((“weak supervision”[tiab] OR “weakly supervised”[tiab] OR “weakly-supervised”[tiab] OR “weak label*”[tiab] OR “inexact label*”[tiab] OR “sparse annotation*”[tiab] OR scribble*[tiab] OR “point annotation*”[tiab] OR “bounding box*”[tiab] OR “box annotation*”[tiab] OR “pseudo-label*”[tiab] OR “pseudo label*”[tiab] OR “multiple instance”[tiab] OR MIL[tiab]) AND (fundus[tiab] OR OCT[tiab] OR “optical coherence tomography”[tiab] OR retina*[tiab] OR ocular[tiab] OR ophthalm*[tiab]) AND (review[tiab] OR survey[tiab] OR “systematic review”[tiab] OR “scoping review”[tiab] OR overview[tiab] OR tutorial[tiab]))

Hits: 4

Scopus (executed: 2020–2026)

Query:

TITLE-ABS-KEY((“weakly supervised” OR “weak supervision” OR “weakly-supervised” OR “weak label*” OR “inexact label*” OR “sparse annotation*” OR scribble* OR “point annotation*” OR “bounding box*” OR “box annotation*” OR “pseudo-label*” OR “pseudo label*” OR “multiple instance” OR MIL) AND (segment* OR segmentation OR “image segmentation” OR “semantic segmentation”) AND (fundus OR OCT OR “optical coherence tomography” OR retina* OR retinal OR ocular OR ophthalm* OR “optic disc” OR “optic nerve head”) AND (review OR survey OR “systematic review” OR “scoping review” OR overview OR tutorial))

Hits: 6

References

1. Alawad, M.; Aljouie, A.; Alamri, S.; Alghamdi, M.; Alabdulkader, B.; Alkanhal, N.; Almazroa, A. Machine Learning and Deep Learning Techniques for Optic Disc and Cup Segmentation—A Review. *Clin. Ophthalmol.* **2022**, *2022*, 747–764. [\[CrossRef\]](#)
2. Goutam, B.; Hashmi, M.F.; Geem, Z.W.; Bokde, N.D. A Comprehensive Review of Deep Learning Strategies in Retinal Disease Diagnosis Using Fundus Images. *IEEE Access* **2022**, *10*, 57796–57823. [\[CrossRef\]](#)
3. Anusuya, S.; Masoodhu Banu, N.M. A Comprehensive Review of Glaucoma Detection from Fundus Images Using Deep Learning. In Proceedings of the 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 23–25 August 2023; pp. 1–6.
4. Zedan, M.J.M.; Zulkifley, M.A.; Ibrahim, A.A.; Moubark, A.M.; Kamari, N.A.M.; Abdani, S.R. Automated Glaucoma Screening and Diagnosis Based on Retinal Fundus Images Using Deep Learning Approaches: A Comprehensive Review. *Diagnostics* **2023**, *13*, 2180. [\[CrossRef\]](#)
5. Rizvana, M.; Narayanan, S. Deep Learning of Fundus Images and Optical Coherence Tomography Images for Ocular Disease Detection—A Review. *Multimed. Tools Appl.* **2024**, *83*, 88745–88789.
6. Chen, K.; Gao, G.; Yang, X.; Wang, W.; Na, J. Denoising, Segmentation and Volumetric Rendering of Optical Coherence Tomography Angiography (OCTA) Images Using Deep Learning Techniques: A Comprehensive Review. *arXiv* **2025**, arXiv:2502.14935.

7. Xue, H.; Hamed, H.N.B.A.; Isyaku, B.; Qichen, S.; Xin, D. Analysis of Atherosclerotic Plaques Using OCT Images Based on Deep Learning: A Comprehensive Review. *KSII Trans. Internet Inf. Syst. TIIS* **2024**, *18*, 3256–3277.
8. Zhang, M.; Zhou, Y.; Zhao, J.; Man, Y.; Liu, B.; Yao, R. A Survey of Semi- and Weakly Supervised Semantic Segmentation of Images. *Artif. Intell. Rev.* **2020**, *53*, 4259–4288. [[CrossRef](#)]
9. Bohlender, S.; Oksuz, I.; Mukhopadhyay, A. A Survey on Shape-Constraint Deep Learning for Medical Image Segmentation. *IEEE Rev. Biomed. Eng.* **2021**, *16*, 225–240. [[CrossRef](#)]
10. Gao, Y.; Jiang, Y.; Peng, Y.; Yuan, F.; Zhang, X.; Wang, J. Medical Image Segmentation: A Comprehensive Review of Deep Learning-Based Methods. *Tomography* **2025**, *11*, 52. [[CrossRef](#)] [[PubMed](#)]
11. Ilesanmi, A.E.; Ilesanmi, T.; Gbotoso, G.A. A systematic review of retinal fundus image segmentation and classification methods using convolutional neural networks. *Healthc. Anal.* **2023**, *4*, 100261. [[CrossRef](#)]
12. Qin, Q.; Chen, Y. A review of retinal vessel segmentation for fundus image analysis. *Eng. Appl. Artif. Intell.* **2024**, *128*, 107454. [[CrossRef](#)]
13. Abdulsahib, A.A.; Mahmoud, M.A.; Mohammed, M.A.; Rasheed, H.H.; Mostafa, S.A.; Maashi, M.S. Comprehensive review of retinal blood vessel segmentation and classification techniques: Intelligent solutions for green computing in medical images. *Netw. Model. Anal. Health Inform. Bioinform.* **2021**, *10*, 20.
14. Abdi, S.; Abdulazeez, A.M. A comprehensive review of deep learning in OCT image segmentation and classification. *Med. Novel Technol. Devices* **2025**, *28*, 100396.
15. Rajarajeshwari, G.; Selvi, G.C. Application of artificial intelligence for classification, segmentation, early detection, early diagnosis, and grading of diabetic retinopathy from fundus retinal images: A comprehensive review. *IEEE Access* **2024**, *12*, 172499–172536. [[CrossRef](#)]
16. Zazueta, L.J.G.; Covarrubias, B.L.L.; Cota, C.X.N.; Briseño, M.V.; Hipólito, J.I.N.; Rodríguez, G.J.A. Segmentation Algorithms in Fundus Images: A Review of Digital Image Analysis Techniques. *Appl. Sci.* **2025**, *15*, 11324. [[CrossRef](#)]
17. Ma, X.; Cao, G.; Chen, Y. A review of optic disc and optic cup segmentation based on fundus images. *IET Image Process.* **2024**, *18*, 2521–2539. [[CrossRef](#)]
18. Veena, H.N.; Muruganandham, A.; Kumaran, T.S. A review on the optic disc and optic cup segmentation and classification approaches over retinal fundus images for detection of glaucoma. *SN Appl. Sci.* **2020**, *2*, 1476. [[CrossRef](#)]
19. Gautam, A.; Shanker, R. Diabetic retinopathy detection from fundus images: A wide survey from grading to segmentation of lesions. *Comput. Biol. Med.* **2025**, *196*, 110715. [[CrossRef](#)]
20. Ilesanmi, A.; Ilesanmi, T.; Idowu, O.P.; Torigian, D.A.; Udupa, J.K. Segmentation and Classification of Retinal Fundus Images using Convolutional Neural Networks: A systematic review of methods and latest trends. *Res. Sq.* **2022**. [[CrossRef](#)]
21. Zhang, H.; Yang, B.; Li, S.; Zhang, X.; Li, X.; Liu, T.; Higashita, R.; Liu, J. Retinal OCT image segmentation with deep learning: A review of advances, datasets, and evaluation metrics. *Comput. Med. Imaging Graph.* **2025**, *123*, 102539. [[CrossRef](#)] [[PubMed](#)]
22. Zhou, Z.-H. A Brief Introduction to Weakly Supervised Learning. *Natl. Sci. Rev.* **2018**, *5*, 44–53.
23. Cheplygina, V.; de Bruijne, M.; Pluim, J.P.W. Not-so-Supervised: A Survey of Semi-Supervised, Multiple-Instance, and Transfer Learning in Medical Image Analysis. *Med. Image Anal.* **2019**, *54*, 280–296.
24. Shen, W.; Peng, Z.; Wang, X.; Wang, H.; Cen, J.; Jiang, D.; Xie, L.; Yang, X.; Tian, Q. A Survey on Label-Efficient Deep Image Segmentation: Bridging the Gap Between Weak Supervision and Dense Prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 9284–9305. [[CrossRef](#)] [[PubMed](#)]
25. Qu, L.; Liu, S.; Liu, X.; Wang, M.; Song, Z. Towards Label-Efficient Automatic Diagnosis and Analysis: A Comprehensive Survey of Advanced Deep Learning-Based Weakly Supervised, Semi-Supervised and Self-Supervised Techniques. *Med. Image Anal.* **2022**, *79*, 102475.
26. Green, R. Learning with Imperfect Labels and Incomplete Views: A Review of Representation Methods for Weakly Supervised Perception. *TechRxiv* **2025**. [[CrossRef](#)]
27. Zhang, X.; Wang, J.; Wei, J.; Yuan, X.; Wu, M. A Review of Non-Fully Supervised Deep Learning for Medical Image Segmentation. *Information* **2025**, *16*, 433. [[CrossRef](#)]
28. Fasana, C.; Pasini, S.; Milani, F.; Fraternali, P. Weakly Supervised Object Detection for Remote Sensing Images: A Survey. *Remote. Sens.* **2022**, *14*, 5362. [[CrossRef](#)]
29. Martínez-Heredia, A.M.; Ventura, S. Weak Supervision: A Survey on Predictive Maintenance. *WIREs Data Min. Knowl. Discov.* **2025**, *15*, e70022. [[CrossRef](#)]
30. Zhao, B.; Cen, J.; Si, W.; Chen, H. A Review of Bearing Fault Diagnosis Based on Weakly Supervised Learning. *Machines* **2023**, *11*, 112.
31. Hassan, H.; Ren, Z.; Zhou, C.; Khan, M.A.; Pan, Y.; Zhao, J.; Huang, B. Supervised and Weakly Supervised Deep Learning Models for COVID-19 CT Diagnosis: A Systematic Review. *Comput. Methods Programs Biomed.* **2022**, *218*, 106731. [[CrossRef](#)]
32. Wan, X.C.; Chen, L. Is Denoising Necessary for Ultrasound Image Segmentation Deep Learning: Review and Benchmark. *Learning* **2023**, *25*, 45.

33. Mehrnia, S.S.; Safahi, Z.; Mousavi, A.; Panahandeh, F.; Farmani, A.; Yuan, R.; Rahmim, A.; Salmanpour, M.R. Landscape of 2D Deep Learning Segmentation Networks Applied to CT Scan from Lung Cancer Patients: A Systematic Review. *J. Imaging Inform. Med.* **2025**, *38*, 3711–3740. [\[CrossRef\]](#)
34. Ouassit, Y.; Ghoumid, K.; Damoiseaux, J.-L.; Benmoussa, N. A Brief Survey on Weakly Supervised Semantic Segmentation. *IEEE Access* **2022**, *10*, 81268–81289. [\[CrossRef\]](#)
35. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [\[CrossRef\]](#)
36. PROSPERO. International Prospective Register of Systematic Reviews. Centre for Reviews and Dissemination, University of York. Available online: <https://www.crd.york.ac.uk/prosperto/> (accessed on 11 January 2026).
37. Staal, J.; Abramoff, M.D.; Niemeijer, M.; Viergever, M.A.; van Ginneken, B. Ridge-Based Vessel Segmentation in Color Images of the Retina. *IEEE Trans. Med. Imaging* **2004**, *23*, 501–509. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Hoover, A.; Kouznetsova, V.; Goldbaum, M. Locating Blood Vessels in Retinal Images by Piecewise Threshold Probing of a Matched Filter Response. *IEEE Trans. Med. Imaging* **2000**, *19*, 203–210. [\[CrossRef\]](#)
39. Fraz, M.M.; Remagnino, P.; Hoppe, A.; Uyyanonvara, B.; Rudnicka, A.R.; Owen, C.G.; Barman, S.A. An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 2538–2548. [\[CrossRef\]](#)
40. Odstrcilik, J.; Kolar, R.; Budai, A.; Hornegger, J.; Jan, J.; Gazarek, J.; Kubena, T.; Cernosek, P.; Svoboda, O.; Angelopoulou, E. Retinal Vessel Segmentation by Improved Matched Filtering: Evaluation on a New High-Resolution Fundus Image Database. *IET Image Process.* **2013**, *7*, 373–383. [\[CrossRef\]](#)
41. Holm, S.; Russell, G.; Nourrit, V.; McLoughlin, N. DR HAGIS—A Fundus Image Database for the Automatic Extraction of Retinal Surface Vessels from Diabetic Patients. *J. Med. Imaging* **2017**, *4*, 014503. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Hu, Q.; Abramoff, M.D.; Garvin, M.K. Retinal Images vessel Tree Extraction (RITE) Dataset. *Med. Image Comput. Comput. Assist. Interv.* **2013**, *16*, 436–443. [\[PubMed\]](#)
43. Zhang, J.; Dashtbozorg, B.; Bekkers, E.; Pluim, J.P.W.; Duits, R.; ter Haar Romeny, B.M. Robust Retinal Vessel Segmentation via Locally Adaptive Derivative Frames in Orientation Scores. *IEEE Trans. Med. Imaging* **2016**, *35*, 2631–2644. [\[CrossRef\]](#)
44. Zhang, Z.; Yin, F.S.; Liu, J.; Wong, W.K.; Tan, N.M.; Lee, B.H.; Cheng, J.; Wong, T.Y. ORIGA(-light): An Online Retinal Fundus Image Database for Glaucoma Analysis and Research. In Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina, 31 August–4 September 2010; pp. 3065–3068.
45. Sivaswamy, J.; Krishnadas, S.R.; Joshi, G.D.; Jain, M.; Tabish, A.U. Drishti-GS: Retinal Image Dataset for Optic Nerve Head (ONH) Segmentation. In Proceedings of the 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), Beijing, China, 29 April–2 May 2014; pp. 53–56.
46. Carmona, E.J.; Rincón, M.; García-Feijó, J.; Martínez-de-la-Casa, J.M. Identification of the Optic Nerve Head with Genetic Algorithms. *Artif. Intell. Med.* **2008**, *43*, 243–259. [\[CrossRef\]](#)
47. Batista, A.F.; Costa, A.E.; Bernardino, A. RIM-ONE DL: A Unified Retinal Image Database for Assessing Glaucoma Using Deep Learning. *Image Anal. Stereol.* **2020**, *39*, 161–167.
48. Almazroa, A.A.; Alodhayb, S.; Osman, E.; Ramadan, E.; Hummadi, M.; Dlaim, M.; Alkatee, M.; Raahemifar, K.; Lakshminarayanan, V. Retinal Fundus Images for Glaucoma Analysis: The RIGA Dataset. In Proceedings of the SPIE MEDICAL IMAGING, Houston, TX, USA, 10–15 February 2018; Volume 10579, pp. 55–62.
49. Orlando, J.I.; Fu, H.; Breda, J.B.; van Keer, K.; Bathula, D.R.; Diaz-Pinto, A.; Fang, R.; Heng, P.-A.; Kim, J.; Lee, J.; et al. REFUGE Challenge: A Unified Framework for Evaluating Automated Methods for Glaucoma Assessment from Fundus Photographs. *Med. Image Anal.* **2020**, *59*, 101570. [\[CrossRef\]](#)
50. Fang, H.; Li, F.; Wu, J.; Fu, H.; Sun, X.; Son, J.; Yu, S.; Zhang, M.; Yuan, C.; Bian, C.; et al. REFUGE2 Challenge: A Treasure Trove for Multi-Dimension Analysis and Evaluation in Glaucoma Screening. *arXiv* **2022**, arXiv:2202.08994. [\[CrossRef\]](#)
51. Bajwa, M.N.; Singh, G.A.P.; Neumeier, W.; Malik, M.I.; Dengel, A.; Ahmed, S. G1020: A Benchmark Retinal Fundus Image Dataset for Computer-Aided Glaucoma Detection. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020.
52. Porwal, P.; Pachade, S.; Kamble, R.; Kokare, M.; Deshmukh, G.; Sahasrabuddhe, V.; Meriaudeau, F. Indian Diabetic Retinopathy Image Dataset (IDRiD): A Database for Diabetic Retinopathy Screening Research. *Data* **2018**, *3*, 25. [\[CrossRef\]](#)
53. Decencière, E.; Zhang, X.; Cazuguel, G.; Lay, B.; Cochener, B.; Trone, C.; Gain, P.; Ordonez, R.; Massin, P.; Erginay, A.; et al. Feedback on a Publicly Distributed Image Database: The Messidor Database. *Image Anal. Ster.* **2014**, *33*, 231. [\[CrossRef\]](#)
54. Lin, L.; Li, M.; Huang, Y.; Cheng, P.; Xia, H.; Wang, K.; Tang, X. The SUSTech-SYSU Dataset for Automated Exudate Detection and Diabetic Retinopathy Grading. *Sci. Data* **2020**, *7*, 409. [\[CrossRef\]](#)
55. Chiu, S.J.; Li, X.T.; Nicholas, P.; Toth, C.A.; Izatt, J.A.; Farsiu, S. Kernel Regression Based Segmentation of Optical Coherence Tomography Images with Diabetic Macular Edema. *Biomed. Opt. Express* **2015**, *6*, 1172–1194. [\[CrossRef\]](#) [\[PubMed\]](#)

56. Bogunovic, H.; Venhuizen, F.; Klimscha, S.; Apostolopoulos, S.; Bab-Hadiashar, A.; Bagci, U.; Beg, M.F.; Bekalo, L.; Chen, Q.; Ciller, C.; et al. RETOUCH: The Retinal OCT Fluid Detection and Segmentation Benchmark and Challenge. *IEEE Trans. Med. Imaging* **2019**, *38*, 1858–1874.
57. Li, M.; Huang, K.; Xu, Q.; Yang, J.; Zhang, Y.; Ji, Z.; Xie, K.; Yuan, S.; Liu, Q.; Chen, Q. OCTA-500: A Retinal OCTA Dataset for Multi-Scale Vascular Analysis. *Med. Image Anal.* **2024**, *93*, 103092. [[CrossRef](#)] [[PubMed](#)]
58. Ma, Y.; Hao, H.; Xie, J.; Fu, H.; Zhang, J.; Yang, J.; Liu, J.; Zheng, Y.; Wang, Y. ROSE: A Retinal OCT-Angiography Vessel Segmentation Dataset and Algorithmic Benchmark. *Med. Image Anal.* **2021**, *73*, 102162.
59. Fang, H.; Li, F.; Fu, H.; Wu, J.; Zhang, X.; Xu, Y. Dataset and Evaluation Algorithm Design for GOALS Challenge. *arXiv* **2022**, arXiv:2207.14447. [[CrossRef](#)]
60. MICCAI Registered Challenges. *Glaucoma OCT Analysis and Layer Segmentation (GOALS)*; Zenodo: Geneva, Switzerland, 2022.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.