APPLICATIONS OF
MODELLING AND SIMULATION
http://arqiipubl.com/ams eISSN 2600-8084 VOL 10, 2026, 196-212

Deep Learning Applications in OCT and OCT Angiography for Retinal Disease Diagnosis: A Review

Jalaluddin Muhammad-Sukki[1], Audrey Huong[1,*], Ser Lee Loh[2], Firdaus Muhammad-Sukki[3] and Xavier Ngu[1]

[1]Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia
[2]Centre for Robotics and Industrial Automation, Fakulti Teknologi dan Kejuruteraan Elektrik, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[3]School of Computing, Engineering & the Built Environment, Edinburgh Napier University, Edinburgh EH10 5DT United Kingdom

[*]Corresponding author: audrey@uthm.edu.my

Submitted 04 January 2026; Revised 21 June 2026; Accepted 09 July 2026; Available online 25 July 2026.

Abstract: Retinal diseases are among the most pressing global health challenges for which novel diagnostic approaches are urgently needed. Optical Coherence Tomography (OCT) and OCT Angiography (OCTA) are novel imaging techniques that provide high-resolution visualization of retinal microvascular abnormalities. Nonetheless, interpreting complex OCT/OCTA images requires extensive knowledge and experience, prompting the need for automated analysis using deep learning technology. This review provides a critical overview of recent progress in deep learning applications for various OCT and OCTA data analysis, including the model training paradigms, optimization techniques, and image processing applications. Three hundred seventy-three articles were found based on the keywords "retinal disease", "OCT" or "OCTA", and "deep learning" in literature searches. Following title, abstract, and keyword screening, 170 articles were reviewed in full text for eligibility, with 27 studies included in the final review. Selected articles were reviewed to determine the dataset used, the employed deep learning and image preprocessing methods, and the evaluated performance metrics. This paper summarizes the applications of deep learning in improving the performance of retinal disease classification and identifies future directions for enhancing the reliability of AI-based models. Finally, this review outlines the challenges associated with current research and existing strategies, and suggests future directions to enhance the medical relevance of deep-learning diagnostic systems.

Keywords: Classification; Deep learning; Optical coherence tomography; Retinal disease.

1. INTRODUCTION

Retinal diseases are progressively becoming dominant public health problems in the world and considered as a major threat to public health [1], [2]. Age-related macular degeneration (AMD), glaucoma, and diabetic retinopathy (DR) are among the most common retinal diseases and the leading causes of blindness and visual impairment worldwide. Earlier reports showed that there are approximately 8 million people living worldwide with AMD, while the number of patients with glaucoma and DR conditions are estimated as 7.7 million and 3.9 million, respectively [3], and the health and economic burdens associated with these conditions are projected to increase substantially by 2045 [4]. These statistics call for an urgent need to review current advances in technology to efficiently address issues related to early disease diagnosis and the prevention of vision problems.

Optical coherence tomography (OCT) is one of the advanced imaging techniques popular for detecting retinal diseases [5]-[7] due to its ability to produce high-quality cross-sectional images, enabling ophthalmologists to identify subtle morphological patterns that aid in the diagnosis of retinal diseases and the assessment of treatment outcomes [8], [9]. This optical imaging technology uses low-coherence interferometry for depth-resolved visualization of deep structures [10], [11], distinguishing it from other imaging modalities, such as fundus photography, fluorescein angiography (FA), and indocyanine green angiography (ICGA) [12], [13]. Both FA and ICGA are invasive imaging techniques that require intravenous injection of a contrast agent [13]–[15], fundus photography is limited in its ability to capture the peripheral retina, where clinically significant abnormalities may occur, compromising early retinal disease detection [16]. Optical coherence tomography angiography (OCTA) is a new addition to the complementary technology of OCT that enables visualization of microvascular blood flow without the need for dye injection, thus improving the detection of early vascular changes, especially those associated with DR and AMD [17], [18].

Artificial intelligence (AI), particularly deep learning, has been applied to numerous real-world problems across various fields, including medicine [19], engineering [20], and telecommunication [21]. Deep learning algorithms, especially convolutional neural networks (CNNs), can be applied to increasingly complex retinal imaging data to facilitate more accurate disease classification [21]. AI resolves the difficulties in human pattern recognition, leading to improved diagnostic accuracy and

increased robustness and efficiency in clinical workflows [19]. Various pretrained models and task-specific architectures have been studied to evaluate the intensity and accuracy of detection in OCT and OCTA images, to achieve higher-performance metrics in terms of accuracy (ACC), precision (PREC), sensitivity (SENS), specificity (SPEC), and F1-score [19]–[21].

Although numerous systematic reviews and meta-analyses on the application of AI in retinal imaging have been published [22]–[27], most have focused on either specific retinal diseases or a particular imaging modality. For instance, the use of OCT and fundus imaging to detect DR [24] or combined ultra-widefield color fundus photography and OCTA with a fused pretrained model that enhanced the classification of DR [28]. However, very few exist that analyze the application of classification-based deep learning across various retinal diseases using OCT or OCTA alone. Additionally, there is a lack of direct comparison between different training paradigms, such as pretrained versus task-specific models, preprocessing techniques, and diagnostic frameworks for retinal diseases.

For this purpose, this paper reviews the application of deep learning techniques for diagnosing retinal diseases using pretrained and task-specific models. The paper further discusses the recent efforts for effective data enhancement, image preprocessing, and optimization methods used in these systems. This review also includes existing methods for handling and managing retinal data, and a comparison of the diagnostic performance of these strategies. The challenges of deep learning-based systems and future work are also discussed. The following summarizes this article's contributions to the existing literature:

a) A systematic search and review of articles published between January 2020 and June 2026 related to deep learning for diagnosing eye disease using clinical imaging samples from OCT and OCTA.
b) Highlighting the contributions of the existing studies by identifying the key study elements, such as the data source used, data processing techniques, proposed architectures, and performance evaluation metrics.
c) Identifying and discussing the challenges and current advancements in deep learning-based systems for diagnosing eye disease.

Section 2 describes the types of retinal diseases. Section 3 describes the paper selection activity, i.e., the inclusion and exclusion criteria. Section 4 discusses recent diagnosis models that use different training paradigms. Section 5 explains different data enhancement methods for retinal disease detection. Model optimization strategies are summarized in Section 6, while different image preprocessing techniques are discussed in Section 7. Critical insights into the existing challenges and future work are included in Section 8. Finally, the paper is concluded in Section 9.

2. TYPES OF RETINAL DISEASES, COMMERCIAL SYSTEMS, AND CURRENT LIMITATIONS

Retinal diseases are a group of conditions that can cause vision loss and blindness. This paper focused on the most prevalent retinal conditions, specifically DR, AMD, choroidal neovascularization (CNV), diabetic macular edema (DME), and drusen. Each of these disorders has a distinct pathogenesis, diagnostic methods, and risk of complications.

DR is a diabetes complication associated with damage to blood vessels in the retina [29]. DR is divided into two main stages: non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). Clinical features include micro-aneurysms, retinal hemorrhages, and exudate at the earlier stage, i.e., NPDR, and neovascularization at the later stage, i.e., PDR [30]. Complications of DR are a significant loss of vision from macular edema (ME) or retinal detachment [29]. DME is a manifestation of DR, characterized by fluid accumulation in the macula due to increased vascular permeability, and is a key pathological feature of diabetes mellitus [31]. Hypertension and dyslipidemia have a significant impact on the development of DME due to their capability to intensify diabetes associated microvascular damage [32]. Complications that can arise as a result of DME are continued vision loss, retinal detachment [31], further progression of the condition may lead to sight-threatening stages of DR [33].

Meanwhile, AMD, which is a result of macular degeneration, is the leading cause of vision loss among older people. This disease can be classified into two types: dry (non-neovascular) and wet (neovascular) [34]. Wet AMD can be distinguished by invasion of vessels from the choroidal capillaries into the retina, also referred to as CNV, while dry AMD involves geographic atrophy of the retinal pigment epithelium and the overlying macular photoreceptors [35]. An end-stage AMD can be characterized by the loss of retinal pigment epithelium (RPE) function, which is important for supporting photoreceptors and the outer blood-retinal barrier [36]. CNV is a common complication in AMD, characterized by the growth of new blood vessels from the choroid into the subretinal space, resulting in fluid leakage and retinal scarring. Even so, this disease may also happen in other pathological conditions, e.g., in myopia and inflammatory disorders [36], [37]. The diagnosis is typically based on the FA and OCT results, which enable the observation of neovascular membranes and the extent of retinal damage [36], [37].

Drusen are yellowish-white frothy deposits beneath the RPE, i.e., a hallmark feature of AMD [29]. They are categorized according to their dimensions and appearance. Hard drusen are small, well-defined, whereas soft drusen are larger and more diffuse; the latter is often associated with a higher risk of progression to advanced AMD and severe vision loss [28], [38]. The key characteristics of drusen, such as size, number, and distribution, can be evaluated using fundus photography and OCT [39], [40].

Several commercial AI-based screening platforms have been developed for automated retinal disease detection, including IDx-DR, Medios AI, EyeArt®, and Retmarker [41], [42]. The performance of these systems, which are primarily based on CNNs, have been evaluated against manual consensus grading, i.e., the clinical reference standard. IDx-DR, Medios AI, and EyeArt have reported sensitivities and specificities of 88–100%, while Retmarker achieved a sensitivity of approximately 85% [41]. However, most of these commercially available systems are designed specifically for DR screening. Pegasus-OCT is one of the few deep learning solutions designed for automated retinal layer segmentation and detection of macular abnormalities using OCT technology. This software supports the classification of AMD and DME, achieving Area Under the Receiver Operating Characteristic (AUROC) values in the range of 0.98–0.99 [43]. Instead of fully automated diagnosis, Zeiss Cirrus and Heidelberg Spectralis are commercial OCT imaging platforms that provide clinician-assisted diagnostic and monitoring capabilities for retinal vascular visualization and quantitative assessment [44]. Despite their clinical utility and regulatory

approval from medical regulatory bodies, these systems have several limitations. Many of these commercially available systems are trained on benchmark datasets such as EyePACS and Messidor [45], which may limit generalizability across different populations, imaging devices, and acquisition protocols. Their performance may also be affected by poor-quality images, resulting in higher referral rates [45]. Furthermore, many systems focus on single-disease detection, provide limited capability for disease progression prediction, and rely on black-box decision-making processes, which may reduce clinician trust. These limitations have driven ongoing research into domain adaptation, interpretable AI, image enhancement, multi-disease classification, and disease progression prediction.

3. METHODOLOGY

This section describes the procedure for choosing articles for this study. The searches of the articles were conducted from Jan. 2 to Jun. 10, 2026, on three databases: Springer Nature Link, ScienceDirect, and IEEE Xplore, using the keywords “retinal disease” and “OCT” and “deep learning”, and “retinal Disease” and “OCTA” and “deep learning”. English-language, peer-reviewed articles published between January 2020 and June 2026 were eligible for inclusion; 373 articles met the inclusion criteria after duplicates were eliminated. It must be mentioned that PubMed is an alternative citation database well-known for its comprehensive coverage in biomedical disciplines, but this review did not consider this search engine due to its limited search results, therefore requiring users to refine their queries to locate relevant studies and potentially affecting search reproducibility [46].

Titles, abstracts, and article types (restricted to open-access journals and conference proceedings, which carry higher prestige and academic standards) were manually screened to ensure relevance to this study objective. The division into words or frequencies was not performed, as the focus remained on inclusion and exclusion criteria. Two hundred and three articles that failed to meet the inclusion criteria were taken out. The inclusion criteria required that:

a) Articles must be published in English.
b) Articles that were published between January 2020 and June 2026.
c) Articles focused on deep learning and retinal disease classification using OCT/OCTA images.

The inclusion and exclusion criteria are further elaborated in Table 1. The search reduced the list to 170 articles that would be assessed.

Table 1. Inclusion and exclusion criteria.

Criteria	Inclusion	Exclusion
Publisher	Springer Nature Link, ScienceDirect, and IEEE Xplore	Others
Source Type	Journal articles and conference proceedings	Book, patents, review articles, abstracts, case reports, magazines, e-book, comments, theses and dissertations, commentaries, pre-prints, preliminary notes, or other non-peer-reviewed documents
Timeframe	1 Jan 2020 –1 June 2026	Published in years other than the timeframe
Dataset	Optical coherence tomography/ angiography (OCT/ OCTA) images	Fluorescein angiography (FA), fundus photography, indocyanine green angiography (ICGA) images, and other radiographic/imaging images
Deep Learning Model	Classification method only	Segmentation method, segmentation-classification method
Retinal Condition	Choroidal neovascularization (CNV), diabetic macular edema (DME), drusen, age-related macular degeneration (AMD), diabetic retinopathy (DR), normal	Other diseases
Reporting	English language	Except English or without translation

At the eligibility stage, 170 were assessed for relevance in terms of eligibility criteria, and 67 articles were chosen for detailed evaluation. A total of 103 studies were excluded for the following reasons: (1) Sixty studies focused on vascularization, biomarkers, disease progression, or segmentation tasks rather than classification. (2) Twenty-one studies combining vessel segmentation and retinal disease classification tasks within a single framework. (3) Five studies did not use deep learning methods, instead relying on classical machine learning approaches, whilst 17 studies used imaging modalities outside the scope of this review, including fundus photography, FA, ICGA, or multimodal imaging approaches combining OCT/OCTA with other imaging techniques. Forty papers were further excluded at the eligibility screening stage as they investigated retinal diseases not covered in this review (e.g., retinitis pigmentosa and hypertensive retinopathy). These exclusions were necessary to maintain the focus of the present review. This procedure resulted in the selection of 27 articles for inclusion in the review. Figure 1 presents an overview of the study selection and screening workflow.

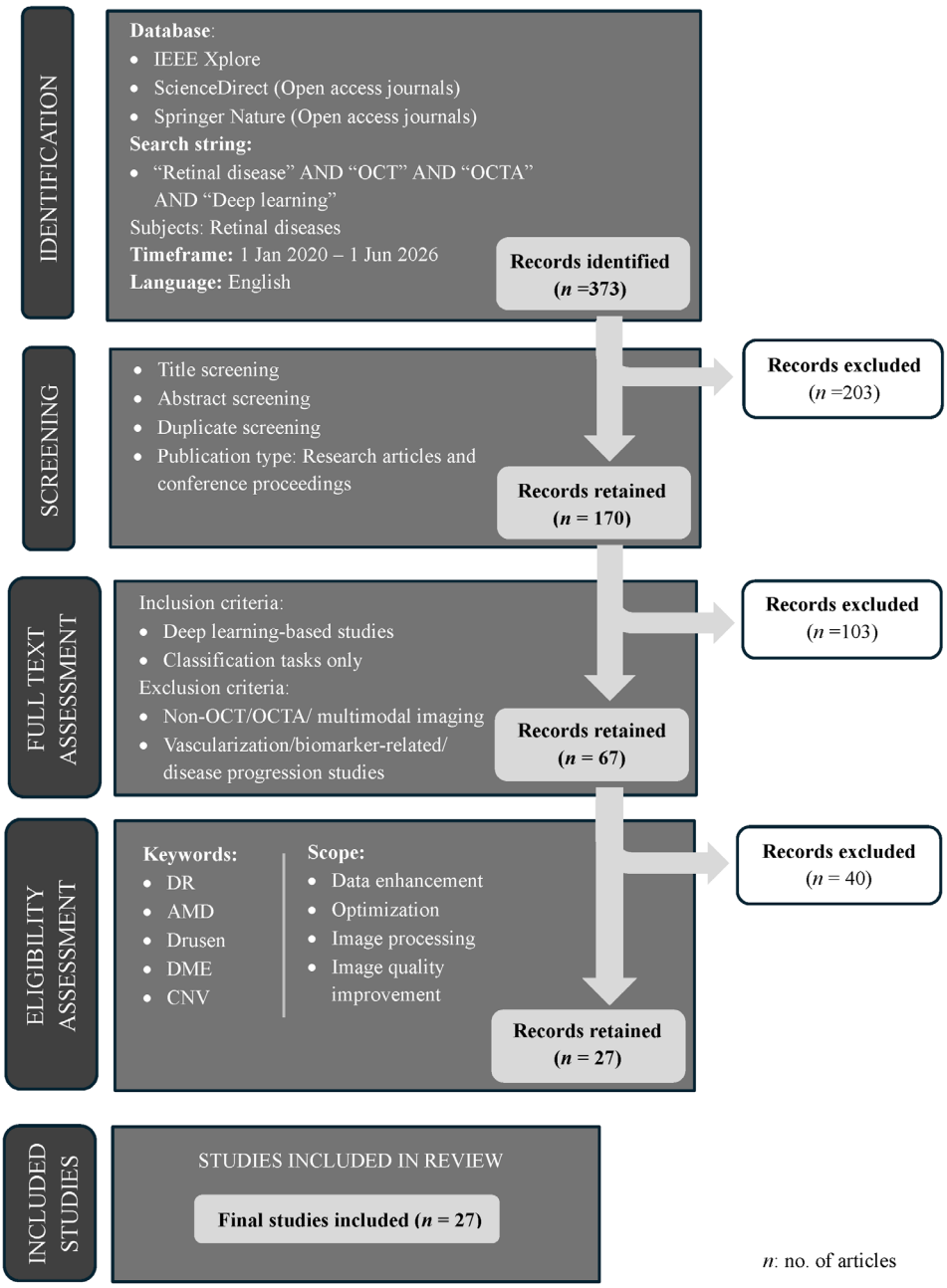

Figure 1. Flow diagram of the article selection process from IEEE Xplore, ScienceDirect, and Springer Nature articles collection.

In this study, the article selection process was conducted manually by a single reviewer, who applied the criteria outlined in Table 1, while a second reviewer independently cross-checked the selections to mitigate subjectivity and minimize potential bias. Rather than relying on automated screening tools in [22]–[27], each article was carefully assessed by reading its full text to ensure accuracy and relevance to the topic investigated. Subsequently, the selected studies were manually labeled according to their subject matter and contents. Among the data collected were the dataset sources, preprocessing methods, target applications, deep learning models (including pretrained and task-specific models), and the reported performances. The results of the research were analyzed through narrative synthesis, as the chosen articles varied in terms of the datasets, model types, and performance reporting.

In this review, models are categorized as pretrained when they are trained through transfer learning (i.e., transferring features from external datasets), whereas task-specific models are architecture designed and trained specifically for retinal disease classification without relying on pretrained backbones. Figure 2 reveals that 19 of the 27 studies used pretrained deep learning architectures to classify retinal disease, while the remaining employed task-specific architectures. Among these, one study used both pretrained and task-specific architectures. These papers were also compared in terms of their methods for data enhancement, optimization, and preprocessing associated with the image, and quality improvement, as summarized in Figure 2. Since this is a narrative review, the selection of included studies was guided by their relevance to the topic, and no formal or risk-of-bias assessment was conducted. These included studies were discussed in the context of their design, methodological strengths, and limitations as reported in their original publications.

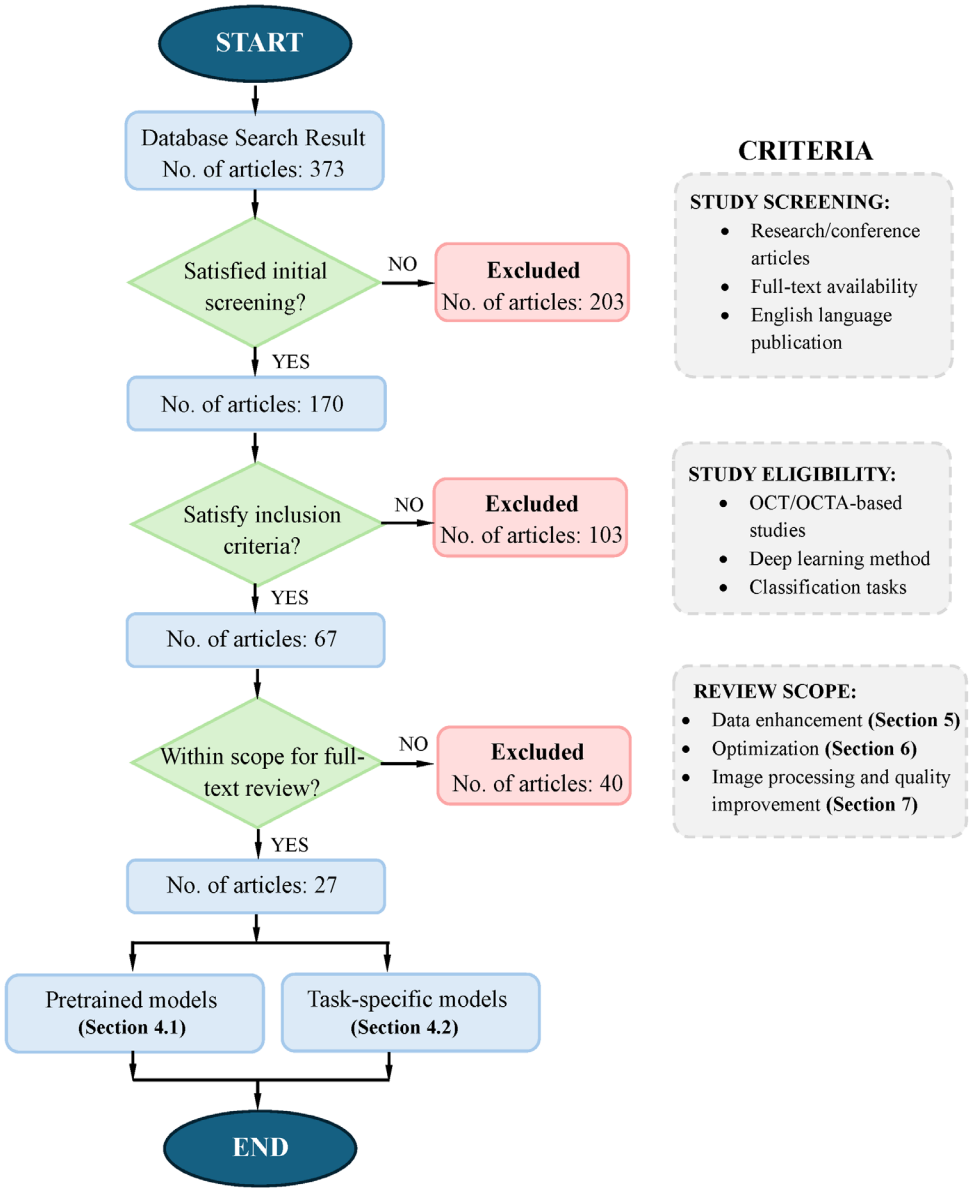

Figure 2. Flow diagram showing the number of selected articles reviewed following the screening and analysis process. The reviewed articles are categorized by model training paradigm and evaluation criteria. The diagram also indicates the paper subsections corresponding to each content area.

4. DEEP LEARNING NETWORK

Deep learning has transformed the landscape of various domains, particularly in medical imaging, natural language processing, and computer vision, primarily using pretrained models. Pretrained models have the advantage of better generalization despite being trained on small datasets. This approach works by fine-tuning the weights of the models and was reported to produce high accuracy in disease detection [47], [48]. However, the practical application of this transfer-training strategy is often hindered by the prevalence of model overfitting to the training data and poor interpretability of these "black-box" models [47]. Therefore, task-specific models were designed to overcome these limitations, enhancing the classification robustness and increasing model interpretability [47]. The existing pretrained and task-specific deep learning models used for classifying retinal disease of interest are compared and discussed in the following subsections.

4.1 Pretrained Model

In 2020, Shih and Patel [48] demonstrated the use of VGG16 for classifying OCT images into CNV, DME, drusen, and normal. The model was reported to achieve a training accuracy of 99.48%, while the testing precision and recall were found to be 99.6% and 95.1%, respectively, implying its reliability for real-world applications. The high F1-score of 99% furher indicated a well-balanced precision and sensitivity. A similar study by [49] used the same VGG16 architecture on a simpler task, i.e., classifying DR against normal cases using OCTA images. The VGG16 was fine-tuned on the final nine layers using 177 OCTA images, achieving 87.27% accuracy, and sensitivity and specificity of 83.76% and 90.83%, respectively. The smaller dataset size was reported as the key factor that contributed to the reduced performance. Other studies in [48]-[49] addressed this issue using data augmentation and RGB color mapping. In 2023, Khalaf *et al.* [50] fine-tuned VGG16 to classify between CNV,

DME, drusen, and normal, and obtained an accuracy of 99%, a precision of 97%, a recall or sensitivity rate of 98%, and an F1-score of 98%. VGG19 is another variant of VGG architecture employed by Wang *et al.* [51] to classify OCT images into AMD, drusen, and normal categories. The model achieved 93.14% accuracy, 95.65% precision, and 91.67% sensitivity through tuning of the last two convolutional and fully connected layers. Alternative architectures to VGG were also explored. For example, Khalil *et al.* [52] proposed OCTNet, which combines InceptionV3 with a dual-attention system, i.e., Channel Attention Modules (CAM) and Modified Spatial Attention Modules (SAM), to effectively capture both spatial and inter-channel dependencies. The hybrid system produced a good accuracy of 99.50% and 99.65%, respectively, on OCT2017 and OCTID datasets, a precision of 99.51%, and an F1-score of 99.50%. Gencer *et al.* [53] used EfficientNet-B0 for the binary classification of DR and no DR classes using a relatively small and imbalanced dataset. Although the study reported high performance across all evaluated metrics (~90-100%), the number of positive DR misclassification was considerably high, reflecting insufficient model generalization.

Asia *et al.* [54] conducted a comprehensive comparison of pretrained CNN architectures, namely EfficientNet-B0, ResNet50, DenseNet161, and MobileNetV2, for the classification of AMD, DR, drusen, and normal retinal images. In addition to the architectural comparison, Walrus Optimization Algorithm (WaOA) was employed to determine suitable training configurations for the models. The optimization process reduced inference time by approximately 12% while improving the classification accuracy by 1.6–3% compared with the non-optimized counterparts. Among the evaluated models, EfficientNet-B0 achieved the best overall performance, with its accuracy increasing from 95.88% to 97.26% following the WaOA optimization. Unlike studies that focus on multiclass retinal disease classification, the earlier work in [55] investigated the use of ResNet34 as a baseline model for DR grading. Although the network achieved 93.1% accuracy, 91.5% F1 score, and an AUC of 95.2%, its performance remained inferior to that of the more specialized architecture, i.e., DRCNN, designed for retinal disease analysis. This specialized network is incorporated with a lesion proxy module, which acts as an auxiliary feature enhancement mechanism within an end-to-end classification framework using ResNet34. This hybrid model integrates features from pretrained ResNet34 with lesion-aware representations, enabling the capture of both contextual and fine-grained details without requiring extensive computational resources. The good performance of this approach across datasets was confirmed with relatively consistent accuracy ranging from 96% to 98%, precision and sensitivity between 93% and 96% and 92% to 94%, respectively, and specificity between 94% and 97%. This consistency is further supported by a 5-fold cross-validation on the composite dataset. The deep learning-based retinal disease diagnosis systems using OCT/OCTA and pretrained deep learning models are summarized in Table 2. These earlier studies suggested that simple pretrained models can provide promising support for clinical decision-making.

Apart from the conventional pretrained models, many previous studies have proposed ensemble models to further improve the classification performance. These include earlier studies by [56]-[57], who used the Mendeley dataset published in [58], while [59]-[60] used the Kaggle OCT dataset [61] in their demonstrations. The study in [56] suggested an ensemble model comprising ResNet50 and EfficientNetB0 for residual learning and squeeze-and-excitation operations. The model achieved a high testing accuracy of 97.50%, with precision, sensitivity, and F1 scores of 98.10%, 97.23%, and 98.21%, respectively, and an area under the curve (AUC) of 99.83%; these results suggest that combining complementary feature extraction capacities can enhance classification. Similarly, an Ensemble Transfer Learning (ETL) model proposed by [59] employed DenseNet169 and InceptionV3 for classifying macular disease. The ensemble model achieved a consistently high accuracy of 99.8%, a sensitivity of 99.8%, and 100% specificity, demonstrating its ability to learn complex retinal features and achieve diagnostic accuracy approaching human-level performance.

Conventional CNN-based models often exhibit limited generalization when trained on small datasets. Transformer-based models are a new machine learning approach proposed for addressing diverse problems in image classification and analysis, utilizing self-attention to preserve long-range interactions. A hybrid Transformer-CNN Retina (HTC-Retina) was developed by [62] based on convolutional patch token embedding, residual depthwise-pointwise convolution, and Vision Transformer (ViT) encoder to capture both local and global retinal features. The model generalizability and translation invariance were enhanced by global average pooling, achieving an accuracy of 99.40% and a sensitivity of 99.41%. Yang *et al.* [63] proposed a hybrid network, HRS-Net, which combines a residual network (ResNet50) and a Swin-Transformer Network. In their study, the model was trained on the Mendeley dataset to extract local features and capture global contextual information associated with CNV, DME, drusen, and normal classes by integrating both branches at a fully connected layer prior to the final classification decision. Their findings demonstrated the effectiveness of this approach, achieving a prediction accuracy of 97.16%, with precision, recall, and F1-scores of 96.05%, 93.77%, and 94.90%, respectively, while the relatively lower performance in drusen detection was attributed to dataset imbalance. Meanwhile, MultiModalNet proposed by [64] is a dual-branch hybrid architecture that leverages vascular information from OCTA encoded with a ResNet101 encoder, and structural features from OCT B-scans, extracted with a ViT-Large model, for the classification of AMD, DR, and normal cases. This joint feature strategy, which integrates vascular and structural cues, achieved 94.59% accuracy, 95.06% precision, 94.59% sensitivity, and 94.49% F1-score under a 5-fold cross-validation in the three-class classification task, outperforming single-modality baselines by approximately 5%.

Another hybrid model integrating MobileNetV1 and a temporal-based Long Short-Term Memory (LSTM) model was proposed in [65] for AMD classification. In this architecture, MobileNetV1 was used as the feature extractor, while LSTM captures sequential dependencies for wet and dry AMD classification. The model achieved 98.85% accuracy, 99.09% sensitivity, and 99.10% specificity. The performance of this model was further improved through advanced preprocessing strategies, including weighted peak signal-to-noise ratio (WPSNR), L2 regularization, and early stopping, to minimize overfitting.

In addition to the fusion of different deep learning models, other innovative strategies have also been reported for enhancing existing pretrained models, such as a multiscale transfer learning algorithm (MTLA) introduced by Zhang *et al.*

[57] to enhance image clarity, reduce noise, and improve the interpretability of OCT-based diagnoses. This model comprises a self-enhancement module that learns multiscale features extracted from fine-tuning the pretrained InceptionV3 for disease classification. The proposed approach achieved a prediction accuracy of 94.50%, with sensitivity, specificity, and precision values of 97.70%, 97%, and 97.02%, respectively. Another study by Kim and Tran [66] investigated the effectiveness of the bagging ensemble strategy across six pretrained models: VGG16, VGG19, ResNet50, ResNet152, DenseNet121, and InceptionV3. Among these models, ResNet152 demonstrated superior performance, and was attributed to its deep residual architecture and the use of a large dataset.

Since the model's convergence relies on its capability in extracting important features, Adel *et al.* [60] compared the performance of InceptionV3 and Xception networks, which have different trainable layers, for classification. These models were ensembled with Support Vector Machines (SVMs) and evaluated using a 10-fold cross-validation. The Xception that employs depth-wise separable convolutions to improve feature extraction was shown to outperform its counterpart with an accuracy of 98%. A similar effort by [67] recommended a Squeeze-and-Excitation ResNet (SeResNet), which incorporated channel-wise attention to emphasize informative features, reported a near-perfect accuracy of 99%, indicating improved feature representation and classification reliability.

Table 2. A summary of pretrained models for classifying retinal diseases using OCT/OCTA images.

Ref	Dataset Sources	No of Images	Retinal Class	Dataset Partition	Model	Performance Metrics
[48]	1. Heidelberg Engineering, Germany 2. Shiley Eye Institute of the University of California, San Diego 3. California Retinal Research Foundation, 4. Medical Center Ophthalmology Associates 5. Shanghai First People's Hospital 6. Beijing Tongren Eye Center	84,452	CNV, DME, drusen, normal	Training: 80% Validation: 10% Testing: 10%	VGG16	ACC: 99.48% PREC: 99.60% SENS: 95.10% F1: 99%
[49]	1. University of Illinois, Chicago 2. National Taiwan University Hospital	177	DR, No DR	Training: 80% Testing: 20%	VGG16	ACC: 87.27% SENS: 83.76% SPEC: 90.83% AUC: 96%
[50]	1. Shiley Eye Institute of the University of California San Diego 2. California Retinal Research Foundation	1,336	CNV, DME, drusen, normal	Training: 70% Testing: 30%	VGG16	ACC: 99% PREC: 97% SENS: 98% F1: 98%
[51]	Northwestern Memorial Hospital	497	AMD, drusen, normal	Training: 80% Testing: 20%	VGG19	ACC: 93.14% PREC: 95.65% SENS: 91.67%
[52]	1. OCT2017 2. OCTID	108,885	CNV, DME, drusen, normal	Training: 80% Testing: 20%	OCT-Net	ACC: 99.50% PREC: 99.51% SENS: 99.50% F1: 99.50%
[53]	OCT image	4,274	DR and No DR	Training: 65% Validation: 15% Testing: 20%	EfficientNet-B0	ACC: 99% PREC: 99-100% SENS: 90-99% F1: 95-100% AUC: 100%
[54]	Kaggle Retinal OCT-C8	19,700	AMD, DR, drusen, normal	Training: 75% Validation: 12.5% Testing: 12.5%	MobileNetV2, ResNet50, EfficientNetB0*, DenseNet161	ACC: 96-98%

[55]	1. EyePACS 2. Messidor-2 3. APTOS 2019 4. DDR 5. DIARETDB1 6. IDRiD	12,982	Abnormal DR grades and normal	Training: 85% Validation: 10% Testing: 5%	ResNet34 DRCNN-Lesion Proxy*	ACC: 96-98% PREC: 93-96% SENS: 92-94% SPEC: 94-97% F1: 95-97% AUC: 96-98%
[56]	Mendeley	84,452	Drusen, normal, CNV, DME	Training: 70% Validation: 15% Testing: 15%	Hybrid (ResNet50 and EfficientNetB0)	ACC: 97.50% PREC: 98.10% SENS: 97.23% F1: 98.21% AUC: 99.83%
[57]	Mendeley	38,057	CNV, DME, drusen, normal	Training: 90% Testing: 10%	MTLA and InceptionV3	ACC: 94.50% PREC: 97.02% SENS: 97.70% SPEC: 97%
[59]	Kaggle	84,484	CNV, DME, drusen, normal	Training: 70% Validation: 15% Testing: 15%	Hybrid (DenseNet169 and InceptionV3)	ACC: 99.8% PREC: 100% SENS: 99.8% SPEC: 100% F1: 100% AUC: 100%
[60]	Kaggle	84,495	CNV, DME, drusen, normal	Training: 80% Validation: 10% Testing: 10%	InceptionV3 Xception*	ACC: 98% PREC: 97% SENS: 95% F1: 96%
[62]	1. OCT-C8 2. OCT-2014 3. OCT-2017	75,805	CNV, DME, drusen, normal	Training: 70% Validation: 15% Testing: 15%	HTC-Retina	ACC: 99.40% PREC: 99.39% SENS: 99.41% SPEC: 99.70% F1: 99.40%
[63]	Mendeley	109,309	CNV, DME, drusen, normal	Training: 80% Validation: 10% Testing: 10%	HRS-Net (ResNet50 and Swim-Transformer)	ACC: 97.16% PREC: 96.05% SENS: 93.77% F1: 94.90%
[64]	OCTA-500	185	AMD, DR, normal	Training: 70% Validation: 15% Testing: 15%	MultiModalNet	ACC: 94.56% PREC: 95.06% SENS: 94.59% F1: 94.49%
[65]	Kaggle	84,495	dry AMD, wet AMD	Training: 70% Validation: 15% Testing: 15%	Hybrid (MobileNetV1 and LSTM)	ACC: 98.85% PREC: 98.5% SENS: 99.09% SPEC: 99.10% F1: 98.34%
[66]	1. Shiley Eye Institute, University of California San Diego 2. California Retinal Research Foundation 3. Medical Center Ophthalmology Associates 4. Shanghai First People's Hospital Beijing Tongren Eye Center	109,309	CNV, DME, drusen, normal	Training: 90% Testing: 10%	VGG16 VGG19 ResNet50 ResNet152* DenseNet121 InceptionV3	ACC: 98.90% SENS: 98.90% SPEC: 99.60%
[67]	1. Spectralis OCT system 2. Shiley Eye Institute California	84,495	CNV, DME, drusen, normal	Training: 70% Validation: 15% Testing: 15%	SeResNet	ACC: 99% PREC: 99% SENS: 99% F1: 99%

[68]	Mendeley	84,568	CNV, DME, drusen, normal	Training: 70% Validation: 15% Testing: 15%	VGG19	ACC: 99.17% PREC: 99% SENS: 99% F1: 99% AUC: 99.99%

*Indicate the best techniques and achieve the best performance metrics.
nAMD: Neovascular age-related macular degeneration; CNV: Choroidal neovascularization; DR: Diabetic retinopathy; DME: Diabetic macular edema; No DR: Diabetic without retinopathy; AUC: Area under curve

Based on the summary of the review studies in Table 2, it can be observed that most researchers fine-tuned different variants of VGG for the classification task of interest. This is likely owing to their simple design, high capacity for feature extraction, and versatility in usage. In addition, the small 3×3 convolutional filters enable the development of deeper network architectures with relatively low computational complexity, allowing for the extraction of fine-grained image features with reasonable success [69]. Similarly, ResNet152 (i.e., with 152 layers) has a larger representational capacity compared to ResNet50 and is well-known for complex classification problems [66]. This was demonstrated in [56], where ResNet152 outperformed the hybrid ResNet50-EfficientNetB0 model. Hybrid architecture is also gaining popularity due to its flexibility and performance benefits. Most researchers used CNNs with transformers to extract both local and global features, such as HTC-Retina [62]. SeResNet is an example of innovative models that can be further integrated in the network to improve feature extraction through squeeze-and-excitation operations [67]. The differences in performance among various models reported in the table depend largely on the complexity of the tasks, the characteristics of the dataset, and the preprocessing techniques.

4.2 Task-specific Models

Apart from the pretrained models, different deep learning models were designed specifically for eye disease classification problems. These studies are summarized in Table 3.

One of the earlier works in this domain was by [70], who proposed DcardNet, a dense and continuous connection network with adaptive rate dropout, to automatically classify the different stages of DR using OCT and OCTA images of the Casey Eye Institute at Oregon Health and Science University. The dataset exhibits class imbalance, with representations of No DR, mild/moderate NPDR, severe NPDR, and PDR, the latter being the least represented class, comprising only 50 scans. This model comprises 36 layers (35 convolutional and one fully connected) with 9.26 million parameters; all layers are fully connected to improve feature spreading and minimizing overfitting. Adaptive dropout rates, which vary with the position of the blocks within the bottleneck, were also adopted. The model produced a considerably poor overall accuracy of 71%, with sensitivity and specificity of 87.8% and 87.1%, respectively. Although the model was shown to perform superiorly in determining the stages of DR, its sensitivity to NPDR detection was poor (12%), likely due to class imbalance.

Another study by [71] developed RetNet, i.e., a lightweight CNN, to classify CNV, DME, drusen, and normal using the Kaggle dataset published in [61]. The architecture consists of 22 layers, including six convolutional, batch normalization, max pooling, and three dense layers, resulting in a total of 517,188 trainable parameters. RetNet yielded an accuracy of 97.85% and a good recall of 97.96%, indicating that it is a competitive network and computationally efficient. These findings make RetNet a competent alternative to pretrained networks. In 2022, [72] proposed DeepOCT, a compact CNN that was created to classify DME and normal images. The model consisted of three convolutional layers with filter sizes of 3×3 and 13×13, containing approximately 7.9 million parameters. Model overfitting was reduced by using a dropout rate of 0.5, with the final classification being performed by a softmax layer. The model achieves an accuracy of 99.20%, a sensitivity and specificity of 100% and 98.40%, respectively. Although DeepOCT achieved strong performance and provided interpretable activation maps, its external validity may be limited because variations in imaging equipment and preprocessing procedures, such as flattening and cropping, may remove or obscure subtle diagnostic patterns.

To overcome the limitations of CNNs in modeling long-range dependencies, [54] proposed a task-specific model that combines a BiLSTM, Additive Attention Mechanism (AM), and Deep & Cross Network (DCN) within a unified framework for OCT retinal disease classification. This design leverages the BiLSTM to capture long-range spatial dependencies, additive attention to emphasize relevant retinal regions, and DCN to model feature interactions. Their ablation study confirms that each component contributes to the overall performance improvement. The proposed hybrid model achieved superior results across all metrics, with an accuracy of 98.84%, 96.96% precision, 95.74% recall, and F1-score of 98.94%, based on 5-fold cross-validation. The model is also notably lighter (~3.28 million parameters) than other pretrained networks, which contain between 3.51 and 28.7 million parameters.

An Automated Optical Coherence Tomography Network (AOCT-Net) was proposed by [73] to classify CNV, DME, drusen, and normal retinal tissue. Its architecture consisted of 19 layers, which were lighter than those of the standard deep models, allowing it to train much faster with lower computational costs. Although the AOCT-Net was shown to work well without extensive preprocessing, the workers suggested dataset expansion and the use of denoising and preprocessing methods to further enhance the model's robustness. Their recommendations were adopted by [74], who employed contrast-limited adaptive histogram equalization capsule network (CLAHE-CapsNet) for OCT classification. The architecture contained three convolutional layers, a primary capsule layer, and a power squash activation to maintain spatial relationships between pixels and control the magnitudes of activations. The model achieved an accuracy of 97.90%, higher than that of its earlier study in [73]. Global Attention Block Network (GABNet) is a customized model proposed by Huang *et al.* [75], consisting of a 20-layer convolutional, pooling, and global attention block. This design is intended to minimize model complexity and enhance feature differentiation in data-scarce conditions. Although the results revealed inferior diagnostic performance as compared to

earlier studies [72]-[74], with an accuracy of 80.47%, specificity of 93.32%, and an AUC of 96.54%, they confirmed the model's robustness even under dataset imbalance

Table 3. A summary of task-specific models for classifying retinal diseases using OCT/OCTA images.

Ref	Dataset Sources	No of Images	Retinal Class	Dataset Partition	Model	Performance Metrics
[70]	Casey Eye Institute, Oregon Health & Science University	596	DR, PDR, NPDR, normal	Training: 80% Testing: 20%	DcardNet	ACC: 71.0% SENS: 87.8% SPEC: 87.1%
[71]	Kaggle	30,904	CNV, DME, drusen, normal	Training: 70% Validation: 10% Testing: 20%	RetNet	ACC:97.85% PREC: 97.93% SENS:97.96%
[72]	1. Zhang Lab 2. SERI (Singapore Eye Research Institute)	62,521	DME, normal	Training: 80% Testing: 20%	DeepOCT	ACC: 99.20% SENS: 100% SPEC: 98.40%
[54]	Kaggle (Retinal OCTC8)	19,700	AMD, DR, drusen, normal	Training: 75% Validation: 12.5% Testing: 12.5%	BiLSTM-AM-DCN	ACC: 98.84%, PREC: 96.96%, SENS: 95.74%, F1: 98.94%
[73]	1. Zhang Lab at the University of California at San Diego (UCSD) 2. Farsiu Ophthalmology 2013 AMD	136,187	AMD, CNV, DME, drusen, normal	Training: 70% Testing: 30%	AOCT-Net	ACC: 97.12% SENS: 97.12% SPEC: 99.28%
[74]	Kaggle	84,495	CNV, DME, drusen, normal	Training: 80% Validation: 10% Testing: 10%	CLAHE-CapsNet	ACC: 97.90% PREC: 99.30% SENS: 99.50%
[75]	1. Shiley Eye Institute of the University of California San Diego 2. California Retinal Research Foundation 3. Medical Center Ophthalmology Associates 4. Shanghai First People's Hospital 5. Beijing Tongren Eye Center	108,312	CNV, DME, drusen, normal	Training: 70% Validation: 15% Testing: 15%	GABNet	ACC: 80.47% SENS: 80.47% SPEC: 93.32% F1: 79.88% AUC: 96.54%
[76]	Chinese University of Hong Kong Eye Center	4,644	DME, normal	Training: 60% Validation: 20% Testing: 20%	S-MIL	ACC: 89.88% SENS: 90.25% SPEC: 90.23% F1: 77.49% AUC: 95.51%
[77]	1. Heidelberg-DME 2. Triton-DME 3. Duke-AMD	1,509	DME, AMD	Training: 60% Validation: 20% Testing: 20%	UD-MIL	ACC: 93.30% F1: 91.70% AUC: 98.70%

DR: Diabetic retinopathy; NPDR: Non-proliferative DR; PDR: Proliferative DR; DME: Diabetic macular edema; AUC: Area under curve; DR: Referable DR; vtDR: Vision-threatening DR; CNV: Choroidal neovascularization; AMD: Age-related macular degeneration

Since unlabeled datasets are common in deep learning, Wang *et al.* [76] proposed semi-supervised multiple instance learning (S-MIL) to identify DME with labeled and unlabeled OCT images, and dealt with noisy labels through confidence-based filtering and dynamic weighting. The method grouped cases into bags and used both instance-level and bag-level classifiers to refine predictions. The high accuracies, recall, specificity, and AUC of 89.88%, 90.25%, 90.23%, and 95.51%, respectively, indicate that the approach is less reliant on large, labeled datasets. The same strategy was adopted by [77], who used an uncertainty-driven multiple instance learning (UD-MIL) model to capture global contextual dependencies between image instances, thereby enhancing uncertainty estimation and robustness.

One of the similarities most models in Table 3 share is their use of CNNs to extract features from retinal images. Most of these studies incorporated either a dropout layer to prevent overfitting or an attention mechanism to enhance feature representation. Among the developed models, DcardNet [70], RetNet [71], and DeepOCT [72] are some of the popular examples adopting a dropout strategy in their operations, whereas GABNet [75] employed varying attention mechanisms for the same purpose. There exist differences in the complexity and design decisions of these models. The DcardNet [70] uses a dense connection design, whereas RetNet [71] exhibits an optimal architecture for classification. Some models were designed to overcome the hardware and computational challenges of deep learning, e.g., the lightweight CNN architecture of DeepOCT [72] and AOCT-Net [73], with limited layers and fewer trainable parameters to enhance training efficiency. The CLAHE-CapsNet [74] that combines sophisticated attention mechanisms is another model designed to reduce pooling operations, thus improving computational efficiency. These differences in architectural design and the incorporation of various improvement strategies reflect a thorough exploration of approaches to the classification problem.

5. DATA ENHANCEMENT

Dataset enhancement is one of the most adopted methods to improve model performance by artificially increasing the size and diversity of training datasets. This process is crucial in scenarios where labeled data is scarce or imbalanced, as it helps mitigate model overfitting to the majority classes and enhances the model's generalization capabilities.

Some of the data augmentation studies worth highlighting include the study by [48], who used random reflection and translation, and grayscale-to-RGB translation, to adjust OCT images for classification using VGG16. A related study in [49] adopted random rotations, flips, and zooming to create a more diverse dataset for DR detection, whereas [56] used a 90-degree rotation and mirroring method to fine-tune a hybrid ResNet50-EfficientNetB0 model. Ismail *et al.* [59] proposed a simple yet effective brightness normalization model, followed by random shifts, rotations, and flips, creating a balanced dataset and enhancing the model's robustness. Other studies in [50] and [71] employed richer variations of augmentation techniques, i.e., rescaling, random rotations, width and height shifts, shearing, zooming, and flipping, to increase model generalization. There were also reports of different rotation angles used for data augmentation, [66] used random sampling and 25-degree rotations, Zang *et al.* [70] applied 90-degree rotations, and [74] used 20-degree random rotations, horizontal flipping, and spatial shifts, while also prioritizing the reliability of annotations through expert validation. Asia *et al.* [54] recommended the use of random horizontal flipping, rotation, and color jittering to enlarge the training dataset and improve model generalizability. Similarly, [55] adopted data augmentation together with dropout regularization and early stopping to mitigate overfitting. However, the authors did not provide details regarding specific augmentation techniques or the extent of dataset expansion. The results demonstrated that incorporating augmentation strategies achieved better generalization, with a validation accuracy of 97.1%, compared with 92.5% without augmentation. Beyond conventional augmentation techniques, such as rotation, flipping, zooming, brightness adjustment, contrast modification, and Gaussian noise addition, Gencer *et al.* [53] further employed a dual-stage data enhancement strategy. This approach utilized Generative Adversarial Networks (GANs) to improve class distribution and enhance generalization, while an autoencoder was applied to improve image quality for classification tasks.

Additional preprocessing strategies were explored to improve robustness and consistency. Yang *et al.* [63] used small-angle rotation and nearest-neighbor padding to deal with the missing pixel area, whereas [51] used vertical flipping, random rotation, and undersampling of the dominant classes to deal with class imbalance. To overcome low-quality OCT images, [57] proposed a self-enhancement method based on edge detection to improve the visibility of features. Wang *et al.* [77] used large B-scan datasets to improve class balance and model generalization. Furthermore, more advanced augmentation techniques, including random grid shuffling, cropping, gamma correction, rescaling, and flipping, were employed in [62] to enhance robustness against illumination and orientation variations while preserving diagnostically relevant regions.

Generally, the augmentation methods employed in most research include rotation, flipping, translation, rescaling, and brightness to enhance training data and alleviate overfitting. Methods, such as grayscale-to-RGB mapping, brightness normalization, and geometric transformations [48], [49], [66], are effective to increase the features diversity, other methods, e.g., rescaling, shearing, zooming, and flipping strategies, are applied in [50], [65], [68] to enhance the model robustness and generalization under different imaging conditions.

6. MODEL OPTIMIZATION AND HYPERPARAMETER TUNING

Model optimization is another important strategy for enhancing model convergence during training. The process often requires tuning training hyperparameters, selecting the appropriate algorithms, and employing iterative optimization procedures to enhance the model's performance on new, unseen data.

The study by [51] used conventional Gradient Descent with a learning rate of 0.0001 and a momentum of 0.9 to enhance generalization. Stochastic Gradient Descent with Momentum (*SGDM*) is a variant of the Gradient Descent optimizer, implemented by Shih and Patel [48] on VGG16, to achieve convergence stability and effectively minimize the error. Another study that used an *SGD* optimizer was by [66], which employed a decay rate of $1e^{-6}$ and a Nesterov momentum of 0.9 to achieve faster convergence. In the paper, several CNNs, i.e., VGG16, VGG19, ResNet50, ResNet152, DenseNet121, and

InceptionV3, were trained to increase feature transferability. Zang *et al.* [70] used *SGD* with a Nesterov momentum (0.9), a learning rate of 0.01, and a cosine decay, and employed 10-fold cross-validation as a measure of performance assessment. Khalil *et al.* [52] applied *SGD* with a momentum of 0.9, a learning rate of 0.0001, a batch size of 24, and 50 training epochs, optimized through extensive experimentation.

The *Adam* optimizer has also been widely used in the field due to its fast convergence and adaptive learning capabilities. Le *et al.* [49] applied *Adam* optimizer, early stopping, and fivefold cross-validation to enhance the generalization of DR detection. The hybrid ResNet50 and EfficientNetB0 model was trained using the chosen settings, which were recognized as the best hyperparameters in [56]. Adel *et al.* [60] trained InceptionV3 and Xception architectures using the *Adam* optimizer with a learning rate of 0.00025, 100 repeated epochs, and batch sizes of 64 and 32 for training and validation, respectively. *Adam* with cross-entropy loss was also used to train HRS-Net and VGG19 models, respectively, in [63] and [68]. The adaptive moment estimation used by this optimizer adjusts learning rates based on first- and second-order gradient statistics, and this strategy has been demonstrated in [48] to allow a more rapid training convergence. The same optimizer and loss function were used in [54] to train the task-specific LSTM-AM-DCN model for 50 epochs, with a learning rate of 0.001 and a mini-batch size of 64. The *Adam* optimizer was also used in [53] across the GAN generator, autoencoder, and EfficientNet to effectively minimize their loss functions; the EfficientNet was trained using the same learning rate as in [54] to expedite convergence, with 20 epochs and a batch size of 32 to balance computational efficiency and generalization. The study also experimented with a dropout rate of 0.5, which randomly drops 50% of neurons, to reduce overfitting during training. In [55], different hyperparameter settings were explored while using the *Adam* optimizer to optimize inference performance. Although multiple learning rates (ranging from 0.0001 to 0.01) and mini-batch sizes (4 to 32) were tested, the results revealed minor variations, and the best prediction performance was obtained with a learning rate of 0.001 and a mini-batch size of 16.

A study by [65] incorporated L2 regularization and early stopping into the training process and used the *Adam* optimizer to minimize overfitting. Others employed the *Adam* optimizer to achieve effective transfer learning [59]. The workers used a learning rate of 0.0001, 100 epochs, and a batch size of 64 in their demonstration, while [71] chose the same optimizer due to its computational efficiency. Meanwhile, *AdamW* used in [62] is a variant of *Adam* popular for enhancing generalization and minimizing overfitting. Likewise, [64] combines *AdamW* with a focal loss function to address class imbalance in the proposed MultiModalNet model, which was trained with 50 epochs, a learning rate of $1e^{-5}$, and a mini-batch size of 32.

Findings from earlier studies showed that important hyperparameters, such as the learning rate, batch size, regularization, and training time, should be optimally tuned to improve the model's performance for retinal image analysis. While most studies tend to explore learning rate within a relatively narrow range of $1e^{-5}$ to $1e^{-3}$, greater variability is observed in the choice of training strategies, including early stopping, cross-validation, and learning rate scheduling, which vary more widely and have been extensively investigated to enhance the training process. Among the investigated optimizers, the *Adam* optimizer is the most popular in the field, largely due to its flexibility and efficiency [49], [59], [60]. The *AdamW* variant proposed in [62] further enhances model generalization by weight decay. Even so, *SGD* and its variants in [48], [51], [67] remain reliable for providing convergence through controlled updates. Unlike the relatively consistent initial learning rate values (typically $1e^{-3}$, $1e^{-4}$, and $1e^{-5}$), there was a large variation in the value of other significant hyperparameters, such as early stopping, cross-validation, and learning rate scheduling, chosen and extensively explored to improve the training process.

Instead of relying on brute-force hyperparameter tuning, [54] is among the earlier studies to integrate an optimization-based approach to improve model convergence speed and stability in the complex, non-convex optimization problems. Specifically, the study employed heuristic WaOA to identify optimal configurations for training pretrained models. The same method was also used for feature selection, extracting the most informative features from pretrained model outputs before feeding them into a task-specific BiLSTM-AM-DCN classifier.

7. IMAGE PROCESSING AND QUALITY IMPROVEMENT

Image processing is a multidisciplinary field that encompasses all domains of the arts and sciences concerning methods that enhance, interpret, and analyze images. The steps are typically image acquisition, preprocessing, feature extraction, and classification.

Different studies adopted different image resizing strategies to ensure compatibility with deep learning models. The most common image sizes chosen for the task are 227×227 pixels for training VGG16 and hybrid ResNet50-EfficientNetB0 models [48], [49], [56], and 224×224 pixels was commonly used for ResNet, EfficientNet, and MobileNet architectures [59], [62], [65], [67], [70]–[72], [76]. Alternative resolutions were also explored, which include 150×150 pixels for lightweight VGG19 in [68], 48×48 pixels for CLAHE-CapsNet [74], 384×384 pixels for HRS-Net [63], and 256×256 pixels for VGG-based and task-specific AOCT-Net models [50], [51], [73]. Meanwhile, the 299×299 pixel resolution was commonly used to match the input requirements of the InceptionV3 and Xception models in [52] and [60].

Preprocessing methods can also be used to reduce noise and enhance pathological details, thus improving image quality. Gaussian filtering and histogram equalization are two image preprocessing methods that improve image contrast and illumination distribution [59]. These enhancement techniques were used by Hamid *et al.* [65], who evaluated and compared the image quality using WPSNR. Since OCT images are often subject to noise, intensity variations, background artifacts, and retinal misalignment, [66] employed a fully connected network (FCN) to eliminate salt-and-pepper noise. The workers also performed retinal region segmentation and corrected retinal tilt using projection-based alignment to improve clinical decision support. Other preprocessing techniques proposed in the field also include an en-face projection strategy to convert the volumetric OCT/OCTA data into a 2D representation in [70], CLAHE to enhance contrast [71], and noise reduction techniques, i.e, BM3D filtering with retinal flattening [72], and adaptive wavelet thresholding [73]. These methods were shown to enhance overall image clarity, standardize the quality of input data, and increase the robustness of the systems for retinal disease analysis

[77]. Another noise reduction approach reported in [53] utilized a denoising autoencoder to reduce noise in both real and GAN-generated OCT images, thereby improving image quality and dataset reliability.

8. DISCUSSION, CHALLENGES, AND FUTURE WORK

Although substantial efforts have been made in the automatic detection of retinal diseases using OCT/OCTA and deep learning approaches, existing studies remain heterogeneous and inconsistency, hindering direct comparison and limit clinical translation. This article reviews recent advancements in deep learning-based classification of five important retinal disorders, i.e., DR, CNV, DME, drusen, and AMD, which were selected due to the availability of extensive datasets associated with each disease. Earlier studies in [48], [49], [56], [57], [59]–[68], [70]–[77] reported superior disease detection and classification accuracy, sensitivity, and specificity, independent of the training paradigms. Most pretrained models, e.g., VGG19 and VGG16 in [48], [49], [56], [57], [59]–[68], were found to generalize well even on a small dataset, demonstrating the efficiency of the transfer learning strategy. Nonetheless, this strategy poses a risk of overfitting the training data and the majority class. This is evident from the near-perfect performance of the model reported in [53], which may reflect dataset-specific learning rather than robust and generalizable disease detection. In addition, the evaluation was limited to a simplified binary classification task rather than a clinically relevant multistage DR grading framework. This concern is further supported by the relatively high number of false negatives observed during testing.

By comparison, the task-specific models introduced in [54], [70]–[77] can potentially surpass the performance of pretrained architectures, addressing challenges such as overfitting and interpretability issues. Based on the reviewed studies, several methodology patterns can be identified and applied in future research, which include using preprocessing operations, such as normalization, augmentation, and resizing, to enhance model robustness and reduce computational complexity. While domain-specific preprocessing methods designed for OCT/OCTA data would provide further improvements, the generation of a more diverse dataset through augmentation helps mitigate biases [26], [27]. Data resizing and normalization are preprocessing steps often used to reduce data variability and bias in imaging, thereby optimizing computational efficiency. Ensemble or hybrid models that integrate the strengths of several models, such as that proposed in [56], [57], [59], [62], [63], [65], have the potential to improve the efficiency of feature extraction, and hence classification performance, rendering them important solutions to overfitting problems.

Performance metrics, such as accuracy, sensitivity, and specificity, are commonly used for evaluating the efficiency and robustness of an architecture design. Earlier results revealed that although hybrid or customized architectures are effective at mitigating model overfitting, pretrained models generally offer better generalization. The latter strategy is also effective at addressing implementation and computational challenges. These findings reinforce the importance of comparing model performance and efficiency when choosing the most effective approach. The use of multimodal data (i.e., combining the data of OCT with other imaging modalities) is another approach that can be implemented to achieve improved predictions and enhance clinical decision-making. The reviewed studies showed that the availability of rich, well-annotated datasets for some eye diseases, such as DR and DME, led to improved disease understanding and facilitated the development of deep learning models. Conversely, less common diseases, such as retinal vein occlusion (RVO) and inherited retinal diseases, remain understudied, and thus calls for multidimensional and multimodal data that can provide complementary information about a disease to overcome the data scarcity problems [23]. It must also be mentioned that although [64] introduced a dual-modality framework that combines OCTA and OCT images at the patient level, representing a clinically significant advancement, its relatively small dataset limits its clinical applicability. In addition, the imbalanced class distribution, the absence of cross-center validation, and the lack of clarity regarding the final selection of images for training and testing raise concerns about potential selection bias and the overall generalizability of the multimodal approach. Meanwhile, Sekar *et al.* [55] departed from conventional methods and multimodal fusion approaches by introducing a lesion proxy module that learns lesion-aware features from pixel-level annotations for DR severity prediction. This approach eliminates the need for manual lesion segmentation, hence reducing annotation costs, while the identified lesion-related regions may provide clinically meaningful information to support diagnosis.

Despite technological advances, numerous challenges persist in the existing deep learning-based studies of retinal disease diagnosis. Among these challenges is class imbalance. Although dataset downsampling strategy was designed to improve training efficiency through the creation of a balanced dataset, this method has, however, reduced data variation, negatively impacting the model's generalizability [25]. In addition, the "black boxes" property of deep learning models remains a significant challenge to model interpretability. Deep learning models used a series of nonlinear transformations to automatically extract important features from input data, creating a mechanism that is uninterpretable to end-users, thus greatly limiting their applicability in clinical settings [27]. Integrating AI systems into the practical clinical workflows is also regarded as a considerable challenge due to differences in imaging protocols and the need for real-time processing [22].

This article identifies several shortcomings in earlier publications. First, the analysis of past studies used varying subsets of publicly available OCT and OCTA datasets, consisting of diverse patient demographics, data acquisition, imaging performance, and disease severities. Therefore, the findings of this study do not represent the overall diagnostic performance. Furthermore, the outcomes of these earlier studies are rarely validated in real-world clinical settings, impeding the understanding of their clinical significance. Besides, the lack of multicenter studies is a key concern for the reproducibility of the reported results. Lastly, most existing systems exhibit a complex architecture that demands high computational power, restricting their use to resource-rich settings. Several possibilities can, therefore, be explored in future research. One of the possible attempts is to combine multiple sources of publicly available OCT or OCTA datasets, which should be analyzed together to enhance the clinical usability of the proposed systems. Standardized evaluation procedures should also be defined to enhance the reliability of AI-based models [23]. Optimal computational efficiency that can be achieved through the design of lightweight architecture and model compression techniques can be further explored to facilitate their implementation in

low-resource environments. Besides, despite the growing attention given to hyperparameter optimization, detailed investigations into parameter sensitivity remain limited. Many earlier studies provide little discussion regarding model convergence behavior, rendering it difficult to assess the stability and robustness of the training process across different experimental settings. It was also found that most studies that employed hybrid models, such as [53], did not conduct ablation studies, which are essential for systematically evaluating the contribution of each individual component within the model architecture.

This review has its own limitations. This literature review is limited to articles from Springer Nature, IEEE Xplore, and ScienceDirect databases, published between 2020 and June 2026. They were selected manually based on their relevance to the topic; therefore, the process may have excluded some important studies published in other databases or topics that fall outside the scopes considered in this review. Although a second reviewer verified the selections to minimize errors, the screening and synthesis were primarily performed by a single reviewer, which may have introduced subjectivity in study selection and potential bias in interpretation. These factors should be considered when interpreting the findings presented in this article. In addition, there remain many other retinal diseases that have not been included in the analysis of this study, such as RVO, hypertensive retinopathy, inherited retinal dystrophies, stroke-related changes, and neurodegenerative diseases (e.g., Alzheimer's), due to space limitations and the scope of this review. Future reviews could employ more rigorous methodologies, such as adopting systematic review design, expanding the search to include additional databases and longer publication periods, and applying formal quality assessment to ensure greater transparency and reproducibility.

9. CONCLUSION

Following the growing demand for early detection of retinal disease using deep learning technology, this paper reviewed and compiled recent studies related to the classification of retinal conditions, specifically DR, AMD, CNV, DME, drusen, and normal conditions, using OCT/OCTA images. Earlier reports showed that the OCT and OCTA imaging methods, combined with deep learning, produce superior prediction accuracy, sensitivity, and specificity, with values ranging between 71% to 99%. This review also examines the tradeoffs of using pretrained models and task-specific models. While the former offer better generalizability and promote rapid deployment, the latter yield better classification performance with improved model interpretability. Although existing studies have demonstrated promising results using deep learning techniques for retinal disease detection, several limitations remain. Firstly, the variability in the datasets used and the diseases being studied hinder straightforward performance comparisons. Secondly, there are also challenges associated with varying hardware requirements and model complexity, which make the practical implementation of different models proposed for the task uncertain. Overcoming these challenges will facilitate the development of edge AI devices that are useful in resource-constrained settings. Other potential efforts also include introducing a multimodal imaging approach for improved diagnoses using deep learning. This is in addition to the rigorous clinical validation that can contribute to the reliability and usability of these systems in real-world clinical settings. This study concluded that early disease detection can be achieved by further advancing the existing technology and integrating an efficient workflow, i.e., through optimizing the model design and implementing innovative processing methods. This system could also optimize treatment planning and support improved patient outcomes in retinal healthcare.

ACKNOWLEDGEMENT AND FUNDING

This research was supported by the Ministry of Higher Education (MOHE) through the Fundamental Research Grant Scheme (FRGS/1/2024/TK07/UTHM/02/5). The authors also wish to express their gratitude to the Universiti Teknikal Malaysia Melaka (UTeM) for supporting the research and publication.

DECLARATION OF CONFLICTING INTERESTS

The authors declare no potential conflicts of interest with respect to the research and publication of this article.

REFERENCES

[1] T. W. Hsu, Y. M. Bai, S. J. Tsai, T. J. Chen, C. S. Liang and M. H. Chen, Risk of retinal disease in patients with bipolar disorder: a nationwide cohort study, *Psychiatry and Clinical Neurosciences*, 76, 2022, 106–113.

[2] Q. Tang, X. Li, J. Wang, Y. Zhang, X. Wang, Y. Liu, J. Zhou and Y. Chen, Study on the Interaction between the characteristics of retinal microangiopathy and risk factors for cerebral small vessel disease, *Contrast Media and Molecular Imaging*, 2022, 1-12.

[3] World Health Organization, Blindness and vision impairment. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment, 2023 (accessed 05.01.2024.

[4] T. Y. Wong and T. E. Tan, The diabetic retinopathy 'pandemic' and evolving global strategies: The 2023 Friedenwald lecture, *Investigative Ophthalmology & Visual Science*, 64, 2023.

[5] C. J. T. Ong, M. Y. Z. Wong, K. X. Cheong, J. Zhao, K. Y. C. Teo and T. E. Tan, Optical coherence tomography angiography in retinal vascular disorders, *Diagnostics*, 13, 2023, 1–20.

[6] Y. Hirano, N. Suzuki, T. Tomiyasu, R. Kurobe, Y. Yasuda, Y. Esaki, T. Yasukawa, M. Yoshida and Y. Ogura, Multimodal imaging of microvascular abnormalities in retinal vein occlusion, *Journal of Clinical Medicine*, 10, 2021, 405.

[7] C. Iovino, C. M. Iodice, D. Pisani, L. Damiano, V. Di Iorio, F. Testa and F. Simonelli, Clinical applications of optical coherence tomography angiography in inherited retinal diseases: An up-to-date review of the literature, *Journal of*

Clinical Medicine, 12, 2023, 3170.
[8] L. von der Emde, M. Saßmannshausen, O. Morelle, G. Rennen, F. G. Holz, M. W. M. Wintergerst and T. Ach, Reliability of retinal layer annotation with a novel, high-resolution optical coherence tomography device: a comparative study, *Bioengineering*, 10, 2023, 438.
[9] R. K. Meleppat, C. R. Fortenbach, Y. Jian, E. S. Martinez, K. Wagner, B. S. Modjtahedi, M. J. Motta, D. L. Ramamurthy, I. R. Schwab and R. J. Zawadzki, In vivo imaging of retinal and choroidal morphology and vascular plexuses of vertebrates using swept-source optical coherence tomography, *Translational Vision Science & Technology*, 11, 2022, 11.
[10] Y. Nakamichi, Gradient mapping of multi-timescale optical coherence tomography angiography signals for enhancing signal-to-noise ratio of flow detection, *Journal of Biomechanical Science and Engineering*, 18, 2023, 1–11.
[11] T. Rabinovitch, V. Yehezkeli, D. Goldenberg, A. Loewenstein and E. Moisseiev, Evaluation of accuracy and agreement of optical coherence tomography angiography interpretation of common retinal findings and diagnoses, *Ophthalmologica*, 244, 2021, 141–149.
[12] B. Chawla and R. Singh, Recent advances and challenges in the management of retinoblastoma, *Indian Journal of Ophthalmology*, 65, 2017, 133–139.
[13] M. S. Ramakrishnan, J. L. Kovach, C. C. Wykoff, A. M. Berrocal and Y. S. Modi, American society of retina specialists clinical practice guidelines on multimodal imaging for retinal disease, *Journal of VitreoRetinal Diseases*, 8, 2024, 234–246.
[14] A. Invernizzi, M. Pellegrini, E. Cornish, K. Y. C. Teo, M. Cereda and J. Chhablani, Imaging the choroid: From indocyanine green angiography to optical coherence tomography angiography, *Asia-Pacific Journal of Ophthalmology*, 9, 2020, 335–348.
[15] K. Agarwal, A. Vinekar, P. Chandra, T. R. Padhi, S. Nayak, S. Jayanna, B. Panchal, S. Jalali and T. Das, Imaging the pediatric retina: An overview, *Indian Journal of Ophthalmology*, 69, 2021, 812–823.
[16] J. J. Pieczynski, P. Kuklo and A. Grzybowski, The role of telemedicine, in-home testing and artificial intelligence to alleviate an increasingly burdened healthcare system: Diabetic retinopathy, *Ophthalmology and Therapy*, 10, 2021, 445–464.
[17] N. Govindaswamy, D. Ratra, D. Dalan, S. Doralli, A. A. Tirumalai, R. Nagarajan, T. Mochi, N. Shetty and A. Sinha Roy, Vascular changes precede tomographic changes in diabetic eyes without retinopathy and improve artificial intelligence diagnostics, *Journal of Biophotonics*, 13, 2020, e202000107.
[18] X. Yao, M. N. Alam, D. Le and D. Toslak, Quantitative optical coherence tomography angiography: A review, *Experimental Biology and Medicine*, 245, 2020, 301–312.
[19] S. Sivasubramaniam and S. P. Balamurugan, Early detection and prediction of heart disease using wearable devices and deep learning algorithms, *Multimedia Tools and Applications*, 84, 2025, 6187–6201.
[20] T. Liu, Y. Hu, Z. Liu, Z. Jiang, X. Ling, X. Zhu and W. Li, Deep learning-based DCE-MRI automatic segmentation in predicting lesion nature in BI-RADS category 4, *Journal of Imaging Informatics in Medicine*, 38, 2025, 2053–2062.
[21] J. Kaur, S. Bhatti, K. Tan, O. R. Popoola, M. A. Imran, R. Ghannam, Q. H. Abbasi and H. T. Abbas, Contextual beamforming: exploiting location and AI for enhanced wireless telecommunication performance, *APL Machine Learning*, 2, 2024, 016113.
[22] T. Miftahushudur, H. M. Sahin, B. Grieve and H. Yin, A survey of methods for addressing imbalance data problems in agriculture applications, *Remote Sensing*, 17, 2025, 454.
[23] M. H. Akpinar, A. Sengur, O. Faust, L. Tong, F. Molinari and U. R. Acharya, Artificial intelligence in retinal screening using oct images: A review of the last decade (2013–2023), *Computer Methods and Programs in Biomedicine*, 254, 2024, 108253.
[24] A. Hayati, M. R. Abdol Homayuni, R. Sadeghi, H. Asadigandomani, M. Dashtkoohi, S. Eslami and M. Soleimani, Advancing diabetic retinopathy screening: A systematic review of artificial intelligence and optical coherence tomography angiography innovations, *Diagnostics*, 15, 2025, 737.
[25] B. Hassan, H. Raja, T. Hassan, M. U. Akram, H. Raja, A. A. Abd-Alrazaq, S. Yousefi and N. Werghi, A comprehensive review of artificial intelligence models for screening major retinal diseases, *Artificial Intelligence Review*, 57, 2024, 1.
[26] C. Lam, Y. L. Wong, Z. Tang, X. Hu, T. X. Nguyen, D. Yang, S. Zhang, J. Ding, S. K. H. Szeto, A. R. Ran and C. Y. Cheung, Performance of artificial intelligence in detecting diabetic macular edema from fundus photography and optical coherence tomography images: A systematic review and meta-analysis, *Diabetes Care*, 47, 2024, 304–319.
[27] M. Ennab and H. Mcheick, Enhancing interpretability and accuracy of AI models in healthcare: a comprehensive review on challenges and future directions, *Frontiers in Robotics and AI*, 11, 2024, 1444763.
[28] M. El Habib Daho, Y. Li, R. Zeghlache, Y. C. Atse, H. Le Boité, S. Bonnin, D. Cosette, P. Deman, L. Borderie, C. Lepicard and R. Tadayoni, Improved automatic diabetic retinopathy severity classification using deep multimodal fusion of UWF-CFP and OCTA images, *International Workshop on Ophthalmic Medical Image Analysis*, Cham, Switzerland, 2023, 11–20.
[29] N. Ban, A. Shinojima, K. Negishi and T. Kurihara, Drusen in AMD from the perspective of cholesterol metabolism and hypoxic response, *Journal of Clinical Medicine*, 13, 2024, 2608.
[30] M. Kropp, O. Golubnitschaja, A. Mazurakova, L. Koklesova, N. Sargheini, T. K. S. Vo, E. de Clerck, J. Polivka Jr., P. Potuznik, J. Polivka, I. Stetkarova, P. Kubatka and G. Thumann, Diabetic retinopathy as the leading cause of blindness and early predictor of cascading complications—Risks and mitigation, *EPMA Journal*, 14, 2023, 21–42.
[31] G. Nkrumah, M. Paez-Escamilla, S. R. Singh, M. A. Rasheed, D. Maltsev, A. Guduru and J. Chhablani, Biomarkers for central serous chorioretinopathy, *Therapeutic Advances in Ophthalmology*, 12, 2020, 2515841420950846.
[32] C. Desai, O. K. Radhakrishnan, N. J. Cardoza, K. Mohankumar and M. Mohan, A study conducted in western

Maharashtra to evaluate the risk factors for diabetic macular edema in patients with type 2 diabetes mellitus, *Indian Journal of Clinical and Experimental Ophthalmology*, 2021.

[33] G. Nagamani and T. Sudhakar, An improved dynamic-layered classification of retinal diseases, *IAES International Journal of Artificial Intelligence*, 13, 2024, 417–429.

[34] M. Adhi and M. Reinoso, Spontaneous regression and quiescence of choroidal neovascularization secondary to traumatic choroidal rupture depicted on OCT angiography, *European Journal of Ophthalmology*, 2021, 11206721211059030.

[35] Z. G. Lu, A. May, B. Dinh, V. Lin, F. Su, C. Tran, H. Adivikolanu, R. Ehlen, B. Che, Z. H. Wang, D. H. Shaw, S. Borooah and P. X. Shaw, The interplay of oxidative stress and ARMS2-HTRA1 genetic risk in neovascular AMD, *Vessel Plus*, 5, 2021, 4.

[36] Z. Tang, Y. Ju, X. Dai, N. Ni, Y. Liu, D. Zhang, H. Gao, H. Sun, J. Zhang and P. Gu, HO-1-mediated ferroptosis as a target for protection against retinal pigment epithelium degeneration, *Redox Biology*, 43, 2021, 101971.

[37] C. Farinha, M. L. Cachulo, R. Coimbra, D. Alves, S. Nunes, I. Pires, J. P. Marques, J. Costa, A. Martins, I. Sobral, P. Barreto, I. Laíns, J. Figueira, L. Ribeiro, J. Cunha-Vaz and R. Silva, Age-related macular degeneration staging by color fundus photography vs. multimodal imaging—epidemiological implications (The Coimbra Eye Study—Report 6), *Journal of Clinical Medicine*, 9, 2020, 1329.

[38] M. Kiruthika and G. Malathi, A comprehensive review on early detection of drusen patterns in age-related macular degeneration using deep learning models, *Photodiagnosis and Photodynamic Therapy*, 51, 2025, 104454.

[39] M. Belmouhand, S. P. Rothenbuehler, J. Bjerager, S. Dabbah, J. B. Hjelmborg, I. C. Munch, C. Dalgård and M. Larsen, Heritability and risk factors of incident small and large drusen in the Copenhagen twin cohort eye study: A 20-year follow-up, *Ophthalmologica*, 245, 2022, 421–430.

[40] S. Yang, Z. Gao, H. Qiu, C. Zuo, L. Mi, H. Xiao and X. Liu, Low-reflectivity drusen with overlying RPE damage revealed by spectral-domain OCT: Hint for the development of age-related macular degeneration, *Frontiers in Medicine*, 8, 2021, 706502.

[41] A. Grzybowski, D. P. Rao, P. Brona, K. Negiloni, T. Krzywicki and F. M. Savoy, Diagnostic accuracy of automated diabetic retinopathy image assessment softwares: IDx-DR and medios artificial intelligence, *Ophthalmic Research*, 66, 2023, 1286–1292.

[42] M. Karabeg, G. Petrovski, S. N. Hertzberg, M. G. Erke, D. S. Fosmark, G. Russell, M. C. Moe, V. Volke, V. Raudonis, R. Verkauskiene, J. Sokolovska, I. K. Haugen and B. E. Petrovski, A pilot cost-analysis study comparing AI-based EyeArt® and ophthalmologist assessment of diabetic retinopathy in minority women in Oslo, Norway, *International Journal of Retina and Vitreous*, 10, 2024, 40.

[43] M. Li and C. Wan, The use of deep learning technology for the detection of optic neuropathy, *Quantitative Imaging in Medicine and Surgery*, 12, 2022, 2129–2143.

[44] M. Hafner, S. G. Priglinger, B. von Livonius and M. J. Gerhardt, Quantitative comparison of a novel swept-source optical coherence tomography angiography device with three established systems, *Scientific Reports*, 15, 2025, 20129.

[45] J. I. Lim, C. D. Regillo, S. R. Sadda, E. Ipp, M. Bhaskaranand, C. Ramachandra, K. Solanki, H. Dubiner, G. Levy-Clarke, R. Pesavento and M. D. Sherman, Artificial intelligence detection of diabetic retinopathy: Subgroup comparison of the EyeArt system with ophthalmologists' dilated examinations, *Ophthalmology Science*, 3(1), 2023, 100228.

[46] J. Lin and M. D. Smucker, How do users find things with PubMed? Towards automatic utility evaluation with user simulations, *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Singapore, 2008, 19–26.

[47] M. Marocchi, L. Abbott, Y. Rong, S. Nordholm and G. Dwivedi, Abnormal heart sound classification and model interpretability: a transfer learning approach with deep learning, *Journal of Vascular Diseases*, 2, 2023, 438–459.

[48] F. Y. Shih and H. Patel, Deep learning classification on optical coherence tomography retina images, *International Journal of Pattern Recognition and Artificial Intelligence*, 34, 2020, 2052002.

[49] D. Le, M. Alam, C. K. Yao, J. I. Lim, Y. T. Hsieh, R. V. P. Chan, D. Toslak and X. Yao, Transfer learning for automated OCTA detection of diabetic retinopathy, *Translational Vision Science & Technology*, 9, 2020, 35.

[50] N. B. Khalaf, H. K. Aljobouri and M. S. Najim, Identification and classification of retinal diseases by using deep learning models, *2023 International Conference on Smart Applications, Communications and Networking (SmartNets)*, Istanbul, Türkiye, 2023, 1–5.

[51] Y. Wang, M. Lucas, J. Furst, A. A. Fawzi and D. Raicu, Explainable deep learning for biomarker classification of OCT images, *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, Cincinnati, OH, USA, 2020, 204–210.

[52] I. Khalil, A. Mehmood, H. Kim and J. Kim, OCTNet: A modified multi-scale attention feature fusion network with InceptionV3 for retinal OCT image classification, *Mathematics*, 12, 2024, 3003.

[53] K. Gencer, G. Gencer, T. H. Ceran, A. E. Bilir and M. Doğan, Photodiagnosis with deep learning: A GAN and autoencoder-based approach for diabetic retinopathy detection, *Photodiagnosis and Photodynamic Therapy*, 53, 2025, 104552.

[54] A. O. Asia, Z. Cai, A. Alasri and A. ELrashidi, Multimodel deep transfer learning and hybrid BiLSTM-AM-DCN with Walrus optimization algorithm for multiclass retinal disease classification in OCT images, *Egyptian Informatics Journal*, 34, 2026, 100988.

[55] P. Sekar, K. S. Suba Raja and R. Krishnaraj, DRCNN-Lesion Proxy: A hybrid CNN with lesion-inspired feature simulation for diabetic retinopathy severity classification, *Scientific Reports*, 15, 2025, 37954.

[56] U. S. Khan and S. U. R. Khan, Boost diagnostic performance in retinal disease classification utilizing deep ensemble classifiers based on OCT, *Multimedia Tools and Applications*, 84, 2025, 21227–21247.

[57] Q. Zhang, Z. Liu, J. Li and G. Liu, Identifying diabetic macular edema and other retinal diseases by optical coherence tomography image and multiscale deep learning, *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, 2020, 4787–4800.
[58] D. Kermany, K. Zhang and M. Goldbaum, Large dataset of labeled optical coherence tomography (OCT) and chest X-ray images, Version 3, *Mendeley Data*, 2018.
[59] A. M. Ismail, F. E. Abd El-Samie, O. A. Omer and A. S. Mubarak, Ensemble transfer learning networks for disease classification from retinal optical coherence tomography images, *Journal of Optics*, 2024, 1–16.
[60] A. Adel, M. M. Soliman, N. E. M. Khalifa and K. Mostafa, Automatic classification of retinal eye diseases from optical coherence tomography using transfer learning, *2020 16th International Computer Engineering Conference (ICENCO)*, Cairo, Egypt, 2020, 37–42.
[61] P. T. Mooney, Retinal OCT images (Kermany2018), *Kaggle Dataset*, 2018. https://www.kaggle.com/datasets/paultimothymooney/kermany2018 (accessed 19.12.2025.
[63] H. Yang, L. Chen, J. Cao and J. Wang, HRS-Net: A hybrid multiscale network model based on convolution and transformers for multiclass retinal disease classification, *IEEE Access*, 12, 2024, 144219–144229.
[65] L. Hamid, A. Elnokrashy, E. H. Abdelhay and M. M. Abdelsalam, A deep learning LSTM-based approach for AMD classification using OCT images, *Neural Computing and Applications*, 36, 2024, 19531–19547.
[64] Ö. F. Aydın, F. B. Tek and Y. Turkan, Retinal disease classification from bimodal OCT and OCTA using a CNN–ViT hybrid architecture, *2025 10th International Conference on Computer Science and Engineering (UBMK)*, Instanbul, Türkiye, 2025, 260–264.
[66] J. Kim and L. Tran, Ensemble learning based on convolutional neural networks for the classification of retinal diseases from optical coherence tomography images, *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, Rochester, MN, USA, 2020, 532–537.
[67] Y. Mori and N. Modi, Enhancing OCT image classification for retinal disease diagnosis: a novel approach using squeeze-and-excitation ResNet, *2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)*, Gwalior, India, 2024, 940–945.
[62] A. Laouarem, C. Kara-Mohamed, E. B. Bourennane and A. Hamdi-Cherif, HTC-Retina: A hybrid retinal diseases classification model using transformer-convolutional neural network from optical coherence tomography images, *Computers in Biology and Medicine*, 178, 2024, 108726.
[68] K. G. Lee, S. J. Song, S. Lee, H. G. Yu, D. I. Kim and K. M. Lee, A deep learning-based framework for retinal fundus image enhancement, *PLoS ONE*, 18, 2023, e0282416.
[69] K. S. Swarnalatha, U. A. Nayak, N. A. Benny, H. B. Bharath, D. Shetty and S. D. Kumar, Detection of diabetic retinopathy using convolution neural network, *Emerging Research in Computing, Information, Communication and Applications: Proceedings of ERCICA 2022*, Singapore: Springer Nature Singapore, 2022, 427–439.
[70] P. Zang, L. Gao, T. T. Hormel, J. Wang, Q. You, T. S. Hwang and Y. Jia, DcardNet: Diabetic retinopathy classification at multiple levels based on structural and angiographic optical coherence tomography, *IEEE Transactions on Biomedical Engineering*, 68, 2020, 1859–1870.
[71] A. Roy, R. Abdullah, F. Ahmed, S. Mashfi, S. H. Khan and D. Z. Karim, RetNet: retinal disease detection using convolutional neural network, *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Chittagong, Bangladesh, 2023, 1–6.
[72] G. Altan, DeepOCT: An explainable deep learning architecture to analyze macular edema on OCT images, *Engineering Science and Technology, an International Journal*, 34, 2022, 101091.
[73] A. M. Alqudah, AOCT-NET: a convolutional network automated classification of multiclass retinal diseases using spectral-domain optical coherence tomography images, *Medical & Biological Engineering & Computing*, 58, 2020, 41–53.
[74] M. Opoku, B. A. Weyori, A. F. Adekoya and K. Adu, CLAHE-CapsNet: efficient retina optical coherence tomography classification using capsule networks with contrast limited adaptive histogram equalization, *PLoS ONE*, 18, 2023, e0288663.
[75] X. Huang, Z. Ai, H. Wang, C. She, J. Feng, Q. Wei, B. Hao, Y. Tao, Y. Lu and F. Zeng, GABNet: Global attention block for retinal OCT disease classification, *Frontiers in Neuroscience*, 17, 2023, 1143422.
[76] X. Wang, F. Tang, H. Chen, C. Y. Cheung and P. A. Heng, Deep semisupervised multiple instance learning with self-correction for DME classification from OCT images, *Medical Image Analysis*, 83, 2023, 102673.
[77] X. Wang, F. Tang, H. Chen, L. Luo, Z. Tang, A. R. Ran, C. Y. Cheung and P. A. Heng, UD-MIL: Uncertainty-driven deep multiple instance learning for OCT image classification, *IEEE Journal of Biomedical and Health Informatics*, 24, 2020, 3431–3442.