Assessing the Performance of Artificial Intelligence Systems for the Screening of Diabetic Retinopathy: A Systematic Review and Meta-Analysis

Ryan Sadjadi (ORCID: 0000-0001-9664-0704)

The Primary Care Unit, Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridge, UK

Corresponding author: Ryan Sadjadi

Primary: Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge, Institute of Public Health, Cambridge, UK, CB2 0SR

Email: rs2104@medschl.cam.ac.uk

Phone: +1 (408)-826-7840

Abstract

Diabetic retinopathy is the most common microvascular complication of diabetes mellitus and one of the leading causes of blindness globally. Due to the progressive nature of the disease, earlier detection and timely treatment can lead to substantial reductions in the incidence of irreversible vision-loss. Artificial intelligence (AI) screening systems have offered clinically acceptable and quicker results in detecting diabetic retinopathy from retinal fundus and optical coherence tomography (OCT) images. Thus, this systematic review and meta-analysis of relevant investigations was performed to document the performance of AI screening systems that were applied to fundus and OCT images of patients from diverse geographic locations including North America, Europe, Africa, Asia, and Australia. A systematic literature search on Medline, Global Health, and PubMed was performed and studies published between October 2015 and January 2020 were included. The search strategy was based on the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) reporting guidelines, and AI-based investigations were mandatory for studies inclusion. The abstracts, titles, and full-texts of potentially eligible studies were screened against inclusion and exclusion criteria. Twenty-one studies were included in this systematic review; 18 met inclusion criteria for the meta-analysis. The pooled sensitivity of the evaluated AI screening systems in detecting diabetic retinopathy was 0.93 (95% CI: 0.92-0.94) and the specificity was 0.88 (95% CI: 0.86-0.89). The included studies detailed training and external validation datasets, criteria for diabetic retinopathy case ascertainment, imaging modalities, DR-grading scales, and compared AI results to those of human graders (e.g., ophthalmologists, retinal specialists, trained nurses, and other healthcare providers) as a reference standard. The findings of this study showed that the majority AI screening systems demonstrated clinically acceptable levels of sensitivity and specificity for detecting referable diabetic retinopathy from retinal fundus and OCT photographs. Further improvement depends on the continual development of novel algorithms with large and gradable sets of images for training and validation. If cost-effectiveness ratios can be optimized, AI can become a financially sustainable and clinically effective intervention that can be incorporated into the healthcare systems of low-to-middle income countries (LMICs) and geographically remote locations. Combining screening technologies with treatment interventions such as anti-VEGF therapy, acellular capillary laser treatment, and vitreoretinal surgery can lead to substantial reductions in the incidence of irreversible vision-loss due to proliferative diabetic retinopathy.

Keywords

Table of Contents

List of abbreviations

ADRSS	Automated Diabetic Retinopathy Screening System
AI	Artificial Intelligence
AMD	Age-related Macular Degeneration
AUC	Area Under the Curve
BMI	Body Mass Index
CNN	Convolutional Neural Network
DL	Deep Learning
DLA	Deep Learning Algorithm
DLS	Deep Learning System
DME	Diabetic Macular Edema
DR	Diabetic Retinopathy
ETDRS	Early Treatment Diabetic Retinopathy Study
FN	False Negative
FOP	Fundus on Phone
GPU	Graphics Processing Unit
ICDRS(S)	International Clinical Diabetic Retinopathy Severity (Scale)
ICER	Incremental Cost-Effectiveness Ratio
ICO	International Council of Ophthalmology
IDP	Iowa Detection Program
IDRID	Indian Diabetic Retinopathy Image Dataset
IRMA	Intraretinal Microvascular Abnormalities
LMIC	Low-to-Middle-Income Country
ML	Machine Learning
MeSH	Medical Subject Heading
NHMRC	National Health and Medical Research Council
NHS	National Health Service
NIH	National Institutes of Health
OCT	Optical Coherence Tomography
RPE	Retinal Pigmented Epithelium
ReLU	Rectified Linear Units
ResNet	Residual Neural Network
SROC	Summary Receiver Operating Characteristic (curve)
SiDRP	Singapore Diabetic Retinopathy Screening Program
TP	True Positive
US(A)	United States (of America)
VEGF	Vascular Endothelial Growth Factor
VGG	Visual Geometry Group
mtmDR	More Than Mild Diabetic Retinopathy

1. Introduction

Diabetes mellitus is a global epidemic that affects approximately 422 million people globally and has been increasing rapidly in recent decades (22). The traditional approach to caring for diabetes mellitus in diverse health settings, including primary, secondary and tertiary care facilities, has been ineffective in addressing diabetes-induced complications, resulting in limited access to screening resources, increasing disease incidence rates, and unfavorable outcomes in low, middle, and high-income countries (22). Currently, innovative and novel methods of care must be developed in an effort to address the systemic effects of diabetes mellitus on the health of patients. Diabetic retinopathy, which affects one-third of diabetes mellitus patients, is the most prevalent diabetes-induced complication and causes preventable vision-loss if left untreated (23). Specifically, diabetic retinopathy is a microvascular complication of diabetes mellitus that leads to the development of lesions that progressively damage to the retina over time (23). Detecting diabetic retinopathy during its early stages is essential to prevent progressive vision loss. With an approximate incidence range of 2.4 to 13.1%, diabetic retinopathy is the leading cause of vision loss in low-to-middle-income countries (LMICs), with adults aged 18-64 constituting the most at-risk groups for the condition (24). Public health interventions aimed at managing and identifying diabetic retinopathy in its early stages and increasing participation and access to screening and treatment services are crucial.

Although substantial data regarding diabetic retinopathy pathology exists and a comprehensive guide has been developed for ophthalmologists and internists by the International Council of Ophthalmology (ICO) detailing evidence-based principles for diagnosis, definition, screening and referral criteria, follow-up, and management options, a lack of screening programs in many countries is contributing to increasing rates of preventable vision loss (25). National diabetic retinopathy screening programs are not commonly incorporated in many countries because the implementation and maintenance of such programs requires substantial resources, and many patients are unaware that they present the condition (23, 25). Furthermore, attending recommended follow-ups for patients with diabetes is a challenge for those under financial pressure, or for those lacking any mode of transportation (23, 25). An additional contributor to the global ubiquity of diabetic retinopathy is the lack of resource availability for care and the cost of essential long-term treatment. For example, diabatic macular edema, which occurs when there is abnormal leakage of fluid in the macula from damaged blood vessels in the retina and commonly caused by diabetic retinopathy, requires long-term and expensive treatments including vascular endothelial growth factor (VEGF) injections (26). Variable responses to treatment make caring for diabetic retinopathy-induced complications even more difficult and less reliable (24,26).

1.1 Framework for diabetic retinopathy management and clinical gaps

Reducing the burden of diabetic retinopathy necessitates a balance of individual and collective preventive measures including intensive medical treatment for diabetes patients experiencing progressive vision loss, screening for undiagnosed diabetic retinopathy, to changes in transport or economic policies affecting the majority of the population (27). The management of diabetic retinopathy can be improved by implementing particular interventions at different settings ranging from individual to population care (23, 27).

Preventing diabetes mellitus is one of the most upstream methods of preventing diabetic retinopathy (28). Strategies for doing so include but are not limited to tight glycemic control through dietary modifications, earlier detection by screening, physical activity promotion for high-risk groups, worksite behavioral interventions, changes to internal built environments and transport infrastructure, or fiscal policy to support access to healthy food (27, 28).

Additional strategies include improving public awareness, developing evidence-based clinical guidelines and screening programs, and optimizing the utilization of anti-VEGF for progressive diabetic retinopathy (29).

The lack of sustainable diabetic retinopathy screening programs has formed a serious clinical gap for managing the condition (23, 29). In order to close this gap, sustainable screening programs must be developed globally so that diabetes mellitus patients can be cared for in accessible primary care settings (30). Additionally, the advancement of risk prediction methods is necessary to ascertain which patients will likely develop vision-threatening diabetic retinopathy (28, 30). At present, not even ophthalmologists can predict which group of diabetic retinopathy patients are at increased risk of vision loss. Lastly, improvements in accessibility and treatments for those experiencing vision-loss from diabetic retinopathy is essential (30). Such treatments include acellular capillary laser treatments, anti-VEGF injections, and vitreoretinal surgeries (26).

Despite the existence of solutions to address clinical gaps, significant barriers stand in the way of their implementation (31). Establishing financially sustainable screening programs is a key challenge, particularly for LMICs that lack the resources to develop and maintain such efforts (31). Additionally, the recruitment and training of retinal specialists remains a significant challenge as their demand far exceeds the number of professionals available to screen retinal images of diabetes mellitus patients (32).

With the incidence of diabetes mellitus and vision-threatening diabetic retinopathy rapidly increasing globally, it is imperative that novel screening methods that are accurate, cost-effective, and sustainable are developed.

1.2 Artificial intelligence application in Ophthalmology

In the field of ophthalmology, the application of artificial intelligence (AI) using machine learning (ML) and deep learning (DL) has been extensively investigated. Specific ocular conditions that have been assessed with the use of AI technologies include glaucoma, age-related macular degeneration, and non-proliferative and proliferative retinopathies (33). Several applications of AI in optical coherence tomography (OCT) and retinal fundus photography have demonstrated high performance levels comparable to those of manual retinal graders and ophthalmologists (27, 33). DL technologies have become useful in identifying macular edema based on OCT images, which can be particularly useful when screening for late-stage diabetic retinopathy (34).

1.3 Artificial intelligence application for diabetic retinopathy

One of the most promising methods for the large-scale management of diabetic retinopathy is the use of AI screening technologies. The success of AI screening systems largely depends on the presence of accessible nationwide screening programs in which diabetes mellitus patients are reminded to attend routine appointments (33, 35). The accuracy of such technological developments, as assessed by prior investigations, has surpassed clinically acceptable thresholds of sensitivity (number of true positive assessments over the number of all positive assessments reported) and specificity (number of true negative assessments over the number of all negative assessments reported). With recent advancements in ML and DL, countries across the globe are revisiting the incorporation of AI systems for diabetic retinopathy screening (30, 35).

Diabetic retinopathy is primed for AI. Screening for the condition depends solely on the use of a single image, whether it be color retinal fundus photographs or OCT images (27). Regardless of whether a diabetes mellitus patient is symptomatic or asymptomatic for diabetic retinopathy, the aforementioned imaging techniques will display the presence of hallmark morphological lesions so long as the condition is developing (36). Trained AI

systems are thus sufficient to screen for retinal lesions and can serve as an effective and efficient solution for the scarcity of diabetic retinopathy management (25, 34). An additional upside to automated diabetic retinopathy screening is that it does not replace the role of eye care professionals (35, 36). By establishing wide-spread and accessible screening programs, the rate of disease detection will likely increase in parallel with the subsequent demand for tertiary care from ophthalmologists (36). Thus, the accessibility of eye care will increase and can encourage future investigations that seek to optimize and incorporate automated screening technologies into clinical settings. Furthermore, AI can be easily incorporated into diabetic retinopathy screening programs and thus acts as an enhancer rather than a disruptor to traditional screening methods (37).

1.4 The current progress of applying AI for diabetic retinopathy screening

Studies conducted by research groups across the globe have demonstrated clinically acceptable performances of AI screening systems in detecting diabetic retinopathy (38). This research extends to multifaceted AI as well, with some groups developing a single system that screens for multiple eye diseases at once including diabetic retinopathy, age-related macular degeneration (AMD), and glaucoma (31, 25, 38). Large external validation image sets consisting of retinal fundus or OCT photographs from diabetes mellitus patients in different countries have been used to demonstrate the performance of various AI systems (27). Countries and continents which have provided validation sets include the United States, Europe, Africa, Australia, India, China, Korea, and Thailand (26, 27, 32).

In addition to demonstrating clinically acceptable performance levels, AI for diabetic retinopathy screening has met the standards of the Food and Drug Administration in the United States, the Health Sciences Authority in Singapore, and Conformité Européene-marking in the European Economic Area (39, 40). Securing regulatory approval serves as an important contributor to the progress of making AI application in eye care settings a commonplace practice (39).

Lastly, economic studies evaluating the financial feasibility of AI implementation have shown promising cost-effectiveness results. Based on the calculation of incremental cost-effectiveness ratios (ICERs), AI solutions demonstrate cost-saving benefits when compared to traditional, manual methods of retinal grading (41). These results support the claim that automated screening services for diabetic retinopathy prevention are not only clinically proven to detect signs of disease with high accuracy, but also are economically sustainable and would benefit primary care settings that choose to adopt them (38, 41).

1.5 Diabetic retinopathy progression and pathology

Generally, diabetic retinopathy progresses according to particular parameters. Glucose and glycated hemoglobin (HbA1c) levels, blood pressure, lipid level, and smoking have near linear relationships with retinopathy progression (42). Pregnancy may also cause rapid deterioration of the retina in those developing diabetic retinopathy (43). In humans, it takes several years for diabetic retinopathy to reach a stage where it could threaten a person's sight (42, 43). The retina itself is a light-sensitive layer of cells at the back of the eye, which converts incident light into electrical signals that are sent to the brain for image generation (44). In order for the retina to function properly, it needs a constant supply of blood which it receives though a network of capillaries (42,44). Over time, uncontrolled and consistent high blood glucose levels can damage retinal vasculature in three notable stages (45). The first stage is known as background retinopathy during which tiny bulges, classified as aneurysms, develop in the blood vessels (46). These bulges may cause bleeding, however at this stage of diabetic retinopathy development, they usually do not affect a person's vision (43, 46). The

second stage, pre-proliferative retinopathy, is characterized by more considerable bleeding due to greater damage to retinal vasculature and potential hemorrhaging (47). At this stage, vision will likely be impacted. The third stage, called proliferative retinopathy, demonstrates scar tissue and neovascularization in the form of minimally or nonfunctional acellular capillaries (42, 47). These new vessels that develop on the retina are structurally weak, cause further bleeding, and eventually progressive vision loss (47). In addition to the aforementioned lesions, microglial infiltration, lipemia retinalis, intraretinal microvascular abnormalities (IRMAs), and ischemia serve as other hallmark indicators of diabetic retinopathy progression (48). Anyone with type 1 or type 2 diabetes mellitus is at risk of developing diabetic retinopathy, however, early detection with effective screening systems and subsequent treatment can prevent progressive vision loss. Figure 1 presents isolated healthy retinal vasculature. Figure 2 compares the vasculature of a non-diabetic versus a diabetic retina. The arrows indicate incident acellular capillaries that have resulted from neovascularization.

Figure 1. — Trypsin digest micrograph of isolated non-diabetic retinal vascular
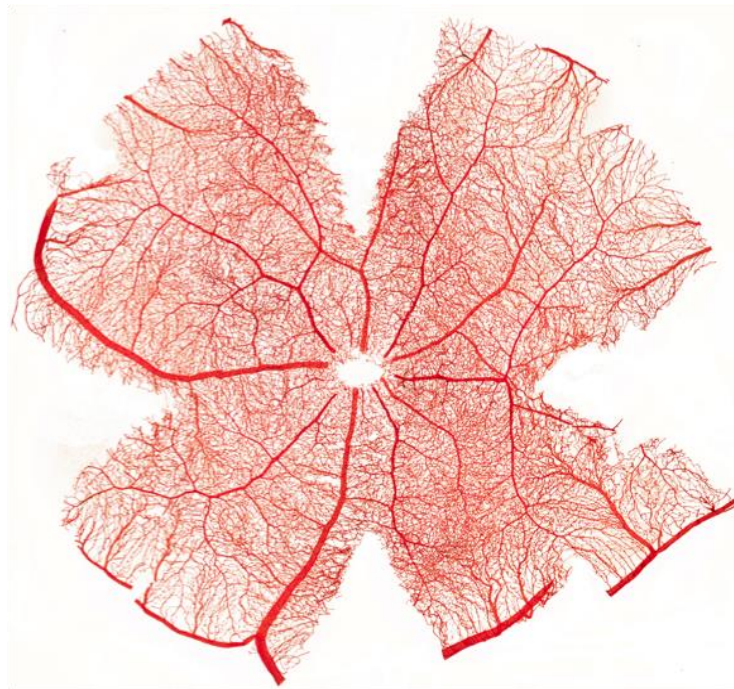
Figure 2. — Comparison of healthy versus diabetic retinal vasculature
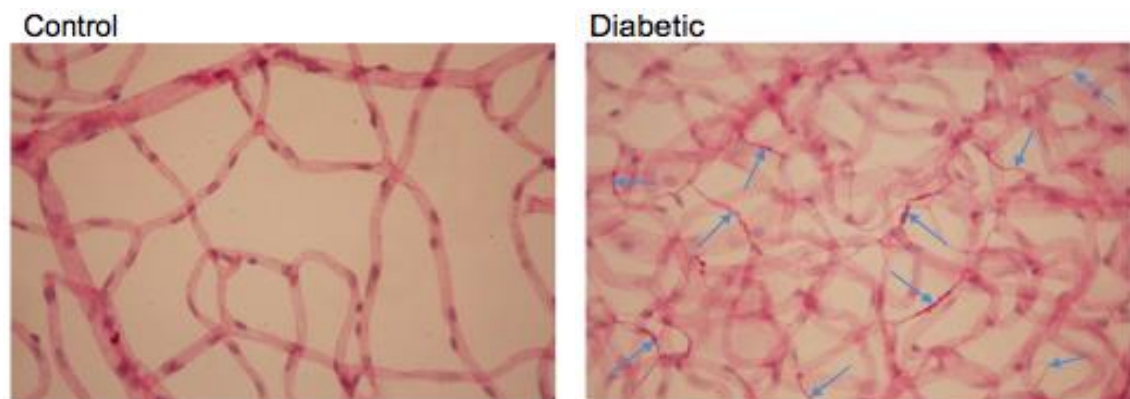
Retinal architecture at 200x magnification. The left image (presented for comparative purposes) presents vasculature in a non-diabetic retina. The right image displays neovascularization during the proliferative retinopathy stage. Arrows indicate areas of neovascularization.

1.6 Different types of imaging techniques used for artificial intelligence-based screening of diabetic retinopathy

Studies investigating the performance of AI systems in screening for diabetic retinopathy use two primary methods of retinal imaging: retinal fundus photography and OCT imaging.

OCT allows for high resolution imaging in the axial direction of the retina, resulting in cross-section visualization of vasculature, retinal cell layers, and limiting membranes (49). Additionally, OCT has the capability of capturing retinal reflectance, in which light is delivered through the pupil and images are formed from the light reflected back from the retina (49). The detection of reflectance allows studies to investigate biomarkers, such as inflammatory cytokines or neurotoxins released by microglial cells, that may affect visual function at the cellular level (49, 50). This is particularly useful when screening for small changes in the retinal cell layers. Thus, OCT imaging can capture the location, nature of retinal changes, thickness of the retina, and integrity of the surrounding structures (49, 50).

Retinal fundus photographs document the current ophthalmoscopic appearance of a patient's retina without detailed visualization of the retinal cell layers (51). It is useful for detecting significant or large changes in retinal cell layers, however, is limited in its ability to detect small changes (unlike OCT imaging). A fundus camera is a specialized low power microscope with an attached camera that sends light rays through the pupil upon image capture (52). If the illumination system of the fundus camera and the produced image are focused and aligned, the illuminating light rays will reflect off the retina and back into the objective lens of the camera. A retinal fundus image is subsequently generated (51, 52).

Retinal fundus photography was used more widely prior to the optimization of OCT imaging (52). Nowadays, OCT is commonly used due to its ability to detect subtle changes in the retina (52). In the context of screening for diabetic retinopathy, neovascularization is an important biomarker for detection in OCT and fundus photographs, whereas detecting changes in the retinal pigmented epithelium (RPE), as well as other cell layers, is more suited for OCT imaging.

1.7 Datasets and research communities used for the development and training of artificial intelligence screening systems

Four notable datasets and data scientist research communities were used across the various studies included in this review to externally validate diabetic retinopathy AI screening systems: Messidor-2, EyePACS, Kaggle, and E-Ophtha.

The Messidor-2 dataset is a collection of 874 diabetic retinopathy examinations (1,748 fundus images) each consisting of two macula-centered fundus images (one per eye). It does not include annotations that define a diabetic retinopathy ground truth, which allows researchers to unbiasedly externally validate their respective AI systems of interest (53). The *Messidor-Original* dataset consists of 529 examinations (1,058 retinal fundus images) that come in pairs or as single images (53, 54). In order to generate new Messidor datasets, diabetic patients were recruited Brest University Hospital in France between October 16, 2009 and September 6, 2010 (53). The hospital's Ophthalmology Department imaged eye fundi, without inducing dilation, using a Topcon TRC NW6 non-mydriatic fundus camera at a 45-degree field of view (53). Only macula-centered images were incorporated in the dataset in order to remain consistent with *Messidor-Original* (54).

The EyePACS database consists of over five million retinal fundus images from diverse populations presenting different degrees of diabetic retinopathy severity (55). Such a large variety helps AI algorithms recognize diverse retinas that exist in real-world settings globally. Major automated screening development studies have and are currently using EyePACS datasets to train and externally validate algorithms.

Kaggle is an online community of data scientists and machine learning researchers that provide a large set of high-resolution retinal fundus images taken under a variety of imaging conditions (56, 57). Two images, one of each eye, from each subject are included in the dataset. Some images are displayed as one would see the retina anatomically (56). For example, when viewing an image of the right eye, one would see the optic nerve on the right and the macula on the left side of the image (56, 57). Other images are demonstrated as seen through a condensing lens on a microscope: inverted as one sees in typical live eye exams. The Kaggle dataset is known for containing some images that contain artifacts and are out of focus, underexposed, or overexposed. (57) AI systems must function and provide accurate outputs in the presence of such noise and variation to be deemed clinically acceptable (43).

E-ophtha is a database of color retinal fundus images used specifically for diabetic retinopathy research. E-ophtha contains two datasets consisting of 463 fundus images that demonstrate either exudates, microaneurysms, or hemorrhages (58). The exudate database contains 47 images with exudates and 35 images with no lesions. The microaneurysm set contains 148 images with microaneurysms or small hemorrhages and 233 images with no lesions (59). Thus, this dataset is particularly useful for training algorithms to recognize exudates, microaneurysms, and hemorrhages in fundus images (58, 59).

1.8 Convolution neural networks computation and training methodology

Convolution neural networks (CNNs) form the base of deep learning (DL), a subset of machine learning (ML) where the algorithms are inspired by the structure of the human brain (60, 61). CNNs take in data, train themselves to recognize the patterns in the data, then predict an output. They are made up of layers of neurons (61). The first layer is known as the input layer which receives the input, the output layer predicts the final output, and the in between layers perform the majority of the computations required by the neural network (62).

In the context of diabetic retinopathy, CNNs are trained to recognize retinal lesions by training their algorithms with fundus and OCT images (61). Figure 3 demonstrates a general example of how a trained CNN computes probabilistic lesion outputs and makes a correct prediction (60). A section of retinal vasculature is presented to the CNN at 200 times

magnification (63). The arrows indicate diabetes-induced acellular capillary growth (62, 63). Each pixel of the image is fed as input to each neuron of the first layer. Neurons of one layer are connected to neurons of the next layer through channels, each of which is assigned a numerical value known as weights (64). The inputs are multiplied to the corresponding weights and their sum is sent as input to the neurons in the hidden layer (63, 64, 65). Each of these neurons is associated with a numerical value called the bias, which is then added to the input sum (64, 65). This value is then passed through a threshold function called the activation function. The result of the activation function determines if the particular neuron will get activated or not (66). An activated neuron transmits data to the neurons of the net layer over the channels. In this manner, the data is propagated through the network (63, 66). This is called forward propagation. In the output layer, the neuron with the highest value fires and determines the output. The values represent a probability. In this example, the neuron correctly associated with acellular capillary recognition has the highest probability, hence that is the most likely output predicted by the neural network.

If the neuron had associated with an incorrect output, such as aneurysm or hemorrhage, it would be an indication that further training is necessary. During the training process, along with the input, the CNN also has the output fed to it (67). The predicted output, whether correct or incorrect, is compared against the actual output to realize the error in prediction. The magnitude of the error indicates how wrong the CNN is and a positive or negative value suggests that the predicted value is either higher or lower than expected, respectively (67). This information is then transferred backward through the neural network, known as back propagation (68). Now based on this information, the weights are adjusted. This cycle of forward propagation and backward propagation is repeatedly performed with multiple inputs (68, 69). The process is continued until the weights are assigned such that the neural network can predict retinal lesions correctly in most cases (69). This brings the training process to an end.

Figure 3. — Convolutional neural network analysis of diabetic retinopathy-induced acellular capillary growth
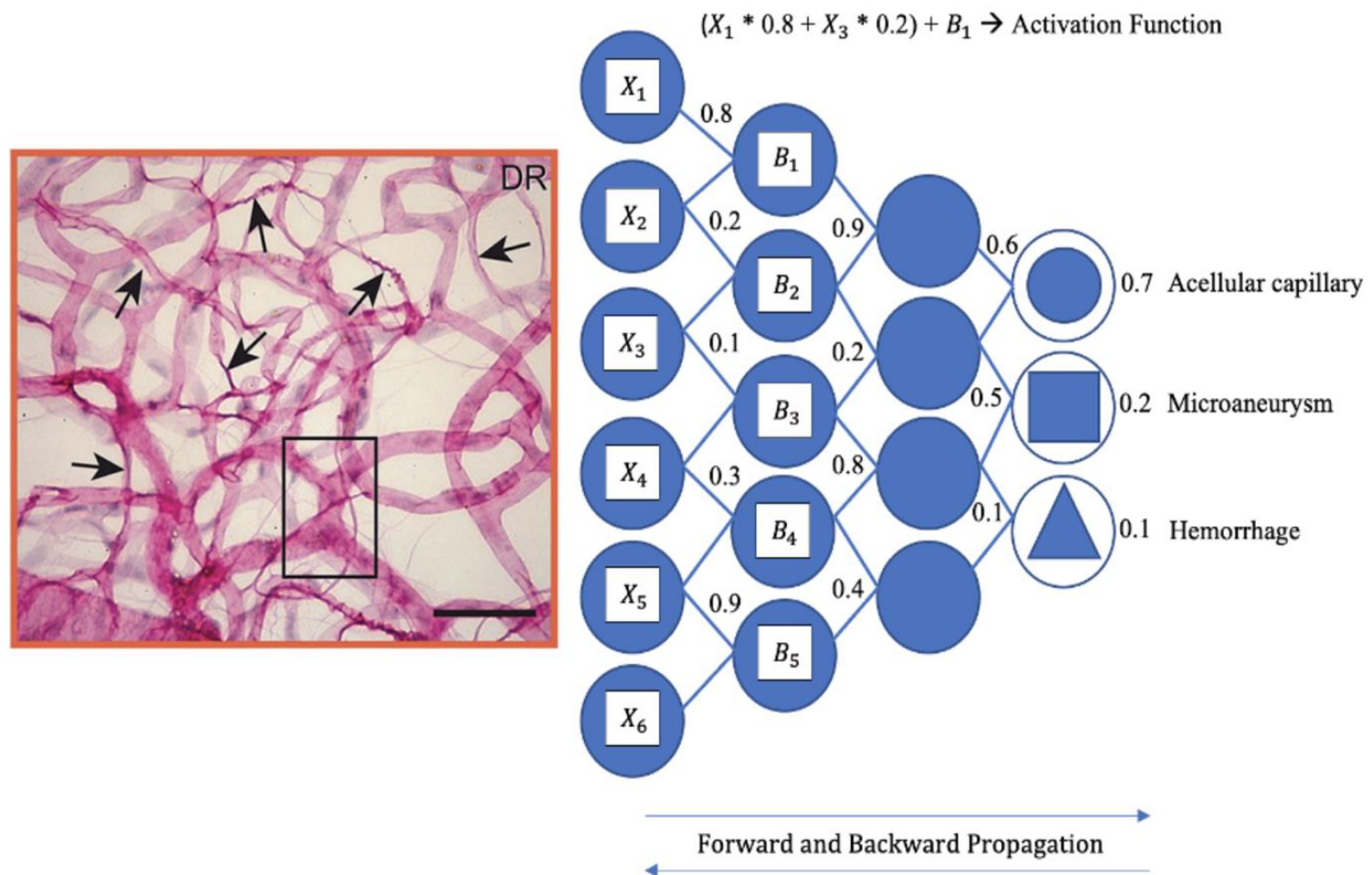
The left-hand image displays the formation of acellular capillaries in the vascular architecture of a diabetic retina.

1.9 Convolutional neural networks assessed by the included studies

Five different CNN architectures were used by the studies included in this review: AlexNet, Inception, Iowa Detection Program (IDP), Visual Geometry Group (VGG), and EyeArt.

The AlexNet architectures consists of eight distinct layers: five convolutional layers and three fully connected layers (70). AlexNet has three features that make it unique compared to other existing CNNs: overlapping pooling, rectified linear units (ReLU) nonlinearity, and multiple graphics processing units (GPUs). Normally, CNNs pool outputs of adjacent groups of neurons with no overlapping (70). However, when overlapping was introduced in AlexNet, researchers observed a reduction in error by approximately 0.5% and found that it is more difficult for architectures with overlapping pooling to provide inaccurate output predictions (71). AlexNet uses ReLU instead of the hyperbolic tangent (tanh) function, which was traditionally used by CNNs (70, 71). Incorporating ReLU is particularly advantageous to quickening AI training time; ReLU-based systems are able to reach a 25% error on CIFAR-10, EyePACS, and Messidor datasets six times faster than systems using the tanh function (72). In addition to faster training times, AlexNet also has the capacity to analyze larger models. AlexNet allows for multi-GPU training by putting half of its neurons on one GPU and the other half on another GPU (72).

Inception is a DL architecture consisting of CNNs that are 27 neuronal layers deep (73). Inception V3 and Inception V4, which are constituents of the Inception family and referred to in this review by the eligible studies, possess important ML features including label smoothing, factorized 7 x 7 convolutions, and the use of auxiliary classifiers to propagate input information to lower down in the network (73). Label smoothing is a regularization technique for classification problems to prevent the Inception model from predicting outcomes too confidently during training and generalizing poorly (74). Factorized 7 x 7 convolutions includes changes that factorize the first 7 x 7 convolutional layer into a sequence of 3 x 3 convolutional layers. The term convolution itself refers to the mathematical combination of two functions to produce a third function (merging two sets of information). In the case of Inception V3 and Inception V4, the convolution performed on the input data, a fundus or OCT image, helps produce a feature map from which the CNNs can distinguish lesions that they have or are being trained to recognize (73, 74). Auxiliary classifiers are a component of the Inception architecture that improves the propagation of computations made by the large and deep Inception neural networks when receiving an input (75). In the context of diabetic retinopathy, including auxiliary classifiers in the AI screening system improves the efficiency of translating an input into a probabilistic outcome of the identity of a retinal lesion (73, 74, 75).

IDP is an algorithm based on expert designed image analysis that uses wavelet transformations (76). A wavelet is a mathematical function that is useful in image processing (76, 77). Wavelet compression works by analyzing an image and converting it into a set of mathematical expressions that can be decoded by a neural network to identify features of an image (76, 77). This is particularly useful when a CNN is fed an image containing large quantities or easily mistakable pieces of information (77). In the context of diabetic retinopathy, if a fundus or OCT image is taken of an eye with many and diverse lesions, for example from the proliferative stage, IDP has the ability to distinguish between morphological structures with considerable accuracy with its wavelet feature (77).

VGG is a classical CNN based on an analysis of how to increase the depth of such networks (78). VGG is characterized by its simplicity as it uses small 3 x 3 filters, pooling layers, and a fully connected layer (78). Applying 3 x 3 convolutions on images with a 3 x 3 filter allows for the analysis of three-dimensional images (79). Additionally, they are used for blurring, sharpening, edge detection, and the embossing of images (78, 79). The pooling

feature of VGG allows the architecture to reduce the size of images while preserving their important characteristics (78). Fully connected layers are simply the connection between one layer of neurons to another, as is a defining feature of CNNs in general (78,79).

The EyeArt system is a cloud-based AI eye screening technology used to detect different stages of diabetic retinopathy through automated analysis of patients' color fundus images (80). It is commonly used amongst endocrinologists, general practitioners, and diabetologists in primary care settings to rapidly and accurately screen for signs diabetic retinopathy within minutes (29). EyeArt uses morphological image analysis with DL techniques to create an automated diabetic retinopathy screening system (ADRSS) engineered for large-scale deployment in the cloud (80). EyeArt is known for its speed and accuracy as it is able to screen 100,000 patients in less than 45 hours whereas human graders can screen retinal images of only 8 to 12 patients per hour (80, 81).

1.10 Population of Interest

People who have been diagnosed with type 1 or type 2 diabetes mellitus comprise the population of interest. The populations that have contributed to the performance results of AI screening systems in this review are from diverse countries and continents including the United States, the United Kingdom, Africa, India, China, Thailand, and Australia. Participants were recruited from a variety of healthcare settings including primary care practices, screening units and programs in urban centers, endocrinology outpatient services, and tertiary care diabetes and general hospitals. Patients with type 1 or type 2 diabetes mellitus were identified either through diabetes or pharmacy registers and were invited for screening studies. Participants consented for the studies and had retinal photographs taken of their eyes.

1.11 Case definition and other important terms

Due to varying criteria in different countries regarding whether a person has diabetic retinopathy, there is no global standard or checklist of symptoms that have been defined. However, based on the studies included in this review, there are a set of general indicators that are commonly used to screen for diabetic retinopathy: types 1 and 2 diabetes mellitus, diabetic macular edema, drusen, exudative retinal detachment, microvascular abnormalities, and retinal vessel occlusion (48, 82). Diabetic retinopathy results from microvascular lesions in the retinas of patients suffering from type 1 or type 2 diabetes mellitus (83). Type 1 diabetes is an autoimmune reaction that attacks one's beta cells in the pancreas, leading to an inability to produce enough insulin and subsequently leads to consistently high blood glucose levels; it can have both genetic and environmental origins (84). Type 2 diabetes occurs when one's body becomes resistant to insulin and is associated with genetics and lifestyle choices (28). Diabetic macular edema, a contributor to the progression of diabetic retinopathy, occurs when leaky vessels cause fluid to build up in the macula at the center of the retina; it is commonly screened for during diabetic retinopathy examinations (85). Drusen is a defining feature of retinal degeneration and appear as small yellow or white spots on the retina that can be detected by ophthalmologists and trained AI screening systems with retinal photography (86). Exudative retinal detachment develops when fluid collects in the subretinal space. This often follows the development of diabetic macular edema in diabetic retinopathy patients as fluid builds up on the retina (83, 84, 85, 86). Microvascular abnormalities associated with diabetic retinopathy include microaneurysms and hemorrhaging of retinal capillaries and neovascularization, the formation of new and structurally weak vessels (acellular capillaries). Lastly retinal vein occlusion, the blockage of blood vessels in the fundus of one's eye, is a potential indicator of diabetic retinopathy that is screened for during examinations (87). Occlusion could relate to the development of hyperlipidemia and

hypertension in diabetes patients, which lead to subsequent microvascular complications (82, 88).

1.12 Meaningful measures of AI screening system performance

Sensitivity and specificity are the measures of AI performance that are assessed by the eligible studies. In this review, sensitivity values are reported as the percentage of screened participants with diabetic retinopathy who are correctly identified as positive by the AI screening system of interest (89). Specificity values are reported as the percent of screened participants without diabetic retinopathy who are correctly identified as negative by the system of interest (89). The "Royal Devon and Exeter National Health Service (NHS) Standards" are that a diabetic retinopathy screening program must achieve a sensitivity and specificity of ≥80% to be deemed clinically acceptable (16, 89). Area under a receiver operating characteristic (ROC) curve (AUC) was also reported by some of the included studies. In the context of using AI screening for diabetic retinopathy, the AUC is a measure of a particular screening system (90). Specifically, the AUC can be interpreted as the average value of sensitivity for all possible values of specificity (90). Alternatively, it can be understood as the probability that a randomly selected participant with diabetic retinopathy has a screening result indicating a greater likelihood of presenting the condition than that of a randomly chosen subject with diabetic retinopathy (90, 91). ROC curves demonstrate the sensitivity plotted as a function of the specificity. Each point on the ROC curve signifies a true positive-true negative pair (90, 91).

1.13 Databases and search strategy

Systematic search methods were performed using Medline, Global Health, and PubMed and with MeSH terms as appropriate. Prior to finalizing a search methodology, pilot examination of studies was carried out in order to identify key MeSH terms used in relevant literature. Search filters were not used when selecting studies to avoid the exclusion of potentially admissible studies. Terms utilized in literature searches are as follows:

Table 1. — Databases and respective search terms used for literature search

---

Medline
(Assessment OR Evaluation OR Testing) AND (Artificial Intelligence OR Convolutional Neural Network OR Deep Learning OR Machine Learning) AND (Screening) AND (fundus OR retina) AND (Imaging OR Optical Coherence Tomography) AND (Diabetic Retinopathy OR Diabetes-Induced Vision Loss OR Diabetes) AND (Progressive OR Advanced OR Worsening) AND (Prevention OR Protection)

Global Health
(Assessment OR Evaluation OR Testing) AND (Artificial Intelligence OR Convolutional Neural Network OR Deep Learning OR Machine Learning) AND (Screening) AND (fundus OR retina) AND (Imaging OR Optical Coherence Tomography) AND (Diabetic Retinopathy OR Diabetes-Induced Vision Loss OR Diabetes) AND (Progressive OR Advanced OR Worsening) AND (Prevention OR Protection)

Pubmed
((Assessment[MeSH] OR Evaluation[MeSH] OR Testing[MeSH]) AND (Artificial Intelligence OR Convolutional Neural Network OR Deep Learning OR Machine Learning) AND (Screening[MeSH]) AND (fundus[MeSH] OR retina[MeSH]) AND (Imaging[MeSH] OR Optical Coherence Tomography[MeSH]) AND (Diabetic Retinopathy[MeSH] OR Diabetes-Induced Vision Loss[MeSH] OR Diabetes[MeSH]) AND (Progressive OR Advanced OR Worsening) AND (Prevention[MeSH] OR Protection[MeSH]))

1.14 Selection of studies

341 studies were initially gathered using the aforementioned search terms in their designated databases. The studies were imported to Mendeley and duplicate literature was discarded, leaving 224 records for assessment. The remaining records were screened by title-abstract review according to the inclusion and exclusion outlined in Table 2. Of the 224 records, 57 titles and abstracts were chosen for full-text evaluation. Following the completion of full-text evaluation and further consideration of inclusion-exclusion criteria, 21 studies were deemed eligible for inclusion in this study. Figure 4 illustrates the methodology used for the identification, screening, and eligibility-determination of included publications.

Table 2. — Inclusion and exclusion criteria used to screen preliminary publications collected from literature search

Inclusion Criteria

Study uses AI to detect DR
Study provides grading comparisons between AI and manual graders
Study provides proof of DR development with retinal photographs
Study specifies retinal imaging technique(s) used for data collection
Study provides sensitivity and specificity values for included AI systems
Study uses a sample size of greater than 100 participants for real-world and external validation
Study uses real-world validation data set(s) to assess neural network performance
Study participants did not have a history of laser treatments or surgeries of the retina or injection into either eye
Study participants were not participating in another investigational eye study or actively receiving investigation product for DR or DME
Study provides a pathway of regulatory approval for AI screening system
Study specifies DR-induced lesions ascertained by AI screening system

Exclusion Criteria

Study is irrelevant to diabetic retinopathy
Study does not use neural networks for retinal analysis
Study does not provide grading comparisons between AI and manual graders
Study does not specify retinal imaging technique(s) used for data collection
Study does not provide sensitivity or specificity values for included AI systems
Study uses a sample size of less than 100 participants for real-world and external validation
Study does not use real-world validation data set(s) to assess neural network performance
Study participants have a history of laser treatments or surgeries of the retina or injection into either eye
Study participants were participating in another investigational eye study or actively receiving investigation product for DR or DME
Study does not provide a pathway of regulatory approval for AI screening system
Study does not specify DR-induced lesions ascertained by AI screening system

Figure 4. — Screening method used for selection of studies

Identification

Records identified through database search of Medline (n = 93)

Records identified through database search of Global Health (n = 107)

Records identified through database search of PubMed (n = 141)

Records after duplicates removed (n = 224)

Screening

Records screened (n = 224)

Exclusion of records based on abstract and title review

Irrelevant to Diabetic Retinopathy (n = 42)
Neural networks are not used for retinal analysis (n = 29)
Screening accuracy comparison between AI versus primary graders not provided (n = 18)
Imaging type not specified (n = 27)
Sensitivity and/or specificity are not provided (n = 17)
Sample size is less than 800 participants (n = 38)
Real-world validation data set is not used (n = 33)

Total excluded (n = 204)

Eligibility

Full-text articles assessed for eligibility (n = 57)

Exclusion of records based on full-text review

History of laser treatment of the retina or injections into either eye, or any history of retinal surgery (n = 8)
Participants are currently participating in another investigational eye study or actively receiving investigational product for DR or DME (n = 9)
Pathway for regulatory approval (ex. FDA) not provided (n = 12)
Lesions identified with AI system not specified (n = 7)

Total excluded (n = 36)

Included

Studies included in review (n = 21)

2. Results

2.1 Study quality

Quality assessment of the included studies was performed using the National Institutes of Health (NIH) quality assessment tool. The guidelines of the tool were used to provide a number score out of 14 and overall rating for each of the included studies. The guidelines used for scoring consist of 14 "yes" or "no" questions regarding the clarity, validity, design, methods, and cohort populations of the included studies. After assessing all appropriate study components, if the number of "yes" answers is equal to or greater than seven, a "Good" overall rating is assigned to the reviewed study. Scores from four to six or less than three are designated as "Medium" and "Poor" rated studies, respectively. Of the 20 included studies, 13 were designated as "Good" studies, seven as "Medium", and none as "Poor" quality. "Yes" and "no" determinations were made to the best of the reviewer's ability with consideration to all aspects of every study in order to decrease the likelihood of subjective errors.

Table 3. — Quality assessment of included studies using the National Institutes of Health (NIH) tool

3. Author	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Final Quality Score	Rating
Abramoff 2018 (1)	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	NA	NA	N	10	Good
Bellemo 2019 (2)	Y	Y	Y	Y	Y	N	Y	N	Y	N	Y	NA	NA	N	8	Good
De Fauw 2018 (3)	Y	Y	Y	N	Y	Y	N	Y	Y	N	Y	NA	NA	N	8	Good
Gargeya 2017 (4)	Y	Y	Y	Y	Y	N	N	Y	Y	N	Y	NA	NA	N	8	Good
Gulshan 2016 (5)	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	NA	NA	N	10	Good
Hansen 2015 (6)	Y	Y	Y	Y	Y	Y	N	N	Y	N	Y	NA	NA	N	8	Good
Kanagasingam 2018 (7)	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y	NA	NA	N	9	Good
Keel 2018 (8)	Y	Y	Y	Y	Y	Y	N	N	Y	N	Y	NA	NA	N	8	Good
Kermany 2018 (9)	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	NA	NA	N	10	Good
Li 2018 (10)	Y	Y	Y	Y	Y	N	Y	N	Y	N	Y	NA	NA	N	8	Good
Li 2019 (11)	Y	Y	Y	Y	Y	N	N	N	Y	N	Y	NA	NA	N	7	Good
Lim 2019 (12)	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	NA	NA	N	10	Good
Rajalakshmi 2018 (13)	Y	Y	Y	Y	Y	N	Y	N	Y	N	Y	NA	NA	N	8	Good

3. Author	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Final Quality Score	Rating
Raju 2017 (14)	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	NA	NA	N	9	Good
Raumviboonsuk 2019 (15)	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y	NA	NA	N	9	Good
Sayres 2019 (16)	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	NA	NA	N	9	Good
Son 2020 (17)	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y	NA	NA	N	9	Good
Ting 2017 (18)	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	NA	NA	N	9	Good
Torre 2019 (19)	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	NA	NA	N	9	Good
Tufail 2016 (20)	Y	Y	Y	Y	Y	Y	N	N	Y	N	Y	NA	NA	N	8	Good
Zeng 2019 (21)	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y	NA	NA	N	9	Good

Y: Yes, N: No, NA: Not applicable. (Q1. Was the research question or objective in this paper clearly stated? Q2. Was the study population clearly specified and defined? Q3. Was the participation rate of eligible persons at least 50%? Q4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants? Q5. Was a sample size justification, power description, or variance and effect estimates provided? Q6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured? Q7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed? Q8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)? Q9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants? Q10. Was the exposure(s) assessed more than once over time? Q11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants? Q12. Were the outcome assessors blinded to the exposure status of participants? Q13. Was loss to follow-up (response rate) after baseline 20% or less? Q14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)? Rating—(Good, Medium or Poor), Good = 7–14 yes; Medium = 4–6 yes; Poor = 0-3.

2.2 Study characteristics

Relevant study characteristics for all 21 eligible studies were collected and are detailed in Table 4. The year of publication of the included studies ranges from 2015 to 2020. Studies were conducted in diverse continents including the North America, Europe, Africa, Asia, and Australia. Four of the included studies used OCT as the imaging modality (1, 3, 9, 11) while the other 17 implemented fundus photography (2, 4-8, 10, 12-21). Three studies reported national screening programs implemented by their respective countries including Australia's National Health and Medical Research Council (NHMRC; 8), Thailand's Ministry of Public Health Program (15), and the United Kingdom's National Health Service (NHS) Diabetic Eye Screening Program (20). Diverse CNNs were assessed by the eligible studies: AlexNet (1), VGGNet (2, 11, 12, 18), Inception V3 (5, 7-10), IDP (6), EyeArt (13, 20), and Inception V4 (15). Datasets for the development and external validation of diabetic retinopathy AI screening systems were implemented by the included studies. The following sets were used for development: Messidor-2 (1, 6, 15), the Singapore Diabetic Retinopathy Screening Program (SiDRP; 2), LabelMe (8), EyePACS (14, 16, 19, 21), and independent datasets using real life screening data (20). The following datasets or sources were used for external validation: primary care practices (1), mobile screening units(2), E-ophtha (4, 17), Messidor-2 (4, 5), EyePACS (4, 12, 16, 19, 21), national eye studies (6), outpatient services (8), tertiary care diabetes hospital (13), Kaggle (14), general hospitals (15), the Indian Diabetic Retinopathy Image Dataset (IDRID; 17), and regional screening programs (20). Sensitivity, specificity, and AUC values reported by the included studies are also detailed in Table 4. If multiple datasets were used for the external validation of an AI screening system, separate sensitivity and specificity values are detailed for each set. Two different scales were implemented by various studies to grade the severity of diabetic retinopathy upon screening: Early Treatment Diabetic Retinopathy Study (ETDRS) and the International Clinical Diabetic Retinopathy and Diabetic Macular Edema Severity (ICDRS) scale.

Table 4. — Characteristics of included studies

Study	Country or Continent	Year of Publication	Imaging	DR National Screening	Deep Learning AI Techniques	Data Set for Development	Data Set for Real-World/External Validation	Referable DR diagnostic performance	Grading scale
Abramoff 2018 (1)	USA	2018	OCT	No	AlexNet	Messidor-2	10 primary care practices from across the USA	87.2% sensitivity; 90.7% specificity	ETDRS
Bellemo 2019 (2)	Africa	2019	Fundus	No	VGGNet	SiDRP 2010-2013	5 mobile screening units in urban centers in Zambia	0.973 AUC	ICDRS
De Fauw 2018 (3)	UK	2018	OCT	NR	3D U-Net architecture for segmentation network; Customized CNNs for classification network	877 segmented scans for the segmentation network; 14,884 scans with diagnoses and referral decisions for the classification network	997 subjects	0.990 AUC	NR
Gargeya 2017 (4)	USA	2017	Fundus	NR	Customized CNNs	75,137 images	E-ophtha: 405 images Messidor-2: 1,748 images	E-Ophtha (no DR vs. mild DR) sensitivity/ specificity: 90%/94% Messidor-2 (no DR vs. any stage of DR) sensitivity/ specificity: 93%/87%	NR

Study	Country or Continent	Year of Publication	Imaging	DR National Screening	Deep Learning AI Techniques	Data Set for Development	Data Set for Real-World/External Validation	Referable DR diagnostic performance	Grading scale
Gulshan 2016 (5)	USA	2016	Fundus	NR	Inception-V3 architecture	128,175 images	Messidor-2: 1,748 images EyePACS-1: 9,963 images	Messidor-2 sensitivity/specificity: 96.1%/93.9%

EyePACS-1 sensitivity/specificity: 97.5%/93.4% | NR |
Hansen 2015 (6)	UK	2015	Fundus	NR	Iowa Detection Program (IDP)	Messidor-2	Nakuru Eye Study in Kenya: 3,460 patients	91% sensitivity 69.9% specificity 0.878 AUC	NR
Kanagasingam 2018 (7)	Australia	2018	Fundus	NR	Inception V3 architecture	30,000 images	193 participants	100% sensitivity 92% specificity	NR
Keel 2018 (8)	Australia	2018	Fundus	NHMRC program	Inception V3	LabelMe	196 patients from 2 urban endocrinology outpatient services	92.3% sensitivity 93.7% specificity	NR
Kermany 2018 (9)	USA and China	2018	OCT	NR	Inception V3	108,312 images	1,000 images	97.8% sensitivity 97.4% specificity 0.999 AUC	NR
Li 2018 (10)	Australia	2018	Fundus photograph	NR	Inception V3	71,043 images	35,201 images	92.5% sensitivity 98.5% specificity 0.955 AUC	NR
Li 2019 (11)	USA and China	2019	OCT	NR	VGG-16 network	109,312 images	1,000 images	98.8% sensitivity 98.8% specificity 0.999 AUC	NR
Lim 2019 (12)	Canada	2019	Fundus	NR	VGGNet	40,542 images	EyePACS-1: 101,710 images	91.3% sensitivity 91.1% specificity 0.965 AUC	NR
Rajalakshmi 2018 (13)	India	2018	Fundus	NR	EyeArt™ software	NR	1 tertiary care diabetes hospital in Chennai, Remidio Fundus on Phone	95.8% sensitivity 80.2% specificity	ICDRS

Study	Country or Continent	Year of Publication	Imaging	DR National Screening	Deep Learning AI Techniques	Data Set for Development	Data Set for Real-World/External Validation	Referable DR diagnostic performance	Grading scale
Raju 2017 (14)	India	2017	Fundus	NR	CNN	EyePACS, 35,126 participants	Kaggle data (n = 53,126)	80.3% sensitivity 92.3% specificity 93.3% AUC	ICDRSS
Raumviboonsuk 2019 (15)	Thailand	2019	Fundus	Ministry of Public Health program	Inception V4 (Google AI)	Messidor-2: 1,748 images	7,517 patients from hospitals of health centers in 13 regions in Thailand	96.9% sensitivity 95.3% specificity 0.988 AUC	ICDRS
Sayres 2019 (16)	USA	2019	Fundus	NR	CNN	EyePACS, 140,000 images	EyePACS-2, n = 1,958 participants	91.6% sensitivity 94.7% specificity 0.884 AUC	ICDRSS
Son 2020 (17)	Korea	2020	Fundus	NR	CNN	95,350 images	IDRID: 143 images E-ophtha: 434 images	IDRid sensitivity/specificity/ AUC: 92.6%/94.0%/0.98 E-Ophtha sensitivity/specificity/ AUC: 93.6%/97.1%/0.965	NR
Ting 2017 (18)	Zambia	2019	Fundus	NR	Ensemble of VGGNet with ResNet	76,370 images	4,504 images	92.3% sensitivity 89.0% specificity 0.973 AUC	NR
Torre 2019 (19)	Spain	2019	Fundus	NR	CNN	EyePACS data set, 75,650 images	EyePACS dataset: 3,000 images as testing and 10,000 images as validation	91.1% sensitivity 90.8% specificity	ICDRSS

Study	Country or Continent	Year of Publication	Imaging	DR National Screening	Deep Learning AI Techniques	Data Set for Development	Data Set for Real-World/External Validation	Referable DR diagnostic performance	Grading scale
Tufail 2016 (20)	UK	2016	Fundus	Yes, NHS Diabetic Eye Screening Program	EyeArt	Independent evaluations using real life screening data from a regional screening service	Region of London, real world screening data; EyeArt study consisting of 30,000 patients across 3 different regional screening programs	93.8% sensitivity	NHS Diabetic Eye Screening
Zeng 2019 (21)	China	2019	Fundus	NR	CNN	EyePACS, 28,104 participants	EyePACS: 7,024 participants	82.2% sensitivity 70.7% specificity 0.951 AUC	NR

NR: Not Reported, ETDRS: Early Treatment Diabetic Retinopathy Study, ICDRS(S): Internatinoal Clinical Diabetic Retinopathy Severity (Scale), CNN: Convolutional Neural Network, NHMRC: National Health and Medical Research Council, OCT: Optical Coherence Tomography, VGG: Visual Geometry Group, ResNet: Residual Neural Network.

2.3 Grading scales used to assign diabetic retinopathy severity upon screening

ETDRS and ICDRS are the two grading scales reported by various studies in this review. ETDRS levels are classified as follows: level 10 (no evidence of diabetic retinopathy), level 15 (probable diabetic retinopathy, no microaneurysms), level 20 (microaneurysms only), level 35 (mild no proliferative diabetic retinopathy), levels 43 though 47 (moderate non-proliferative diabetic retinopathy), level 53 (severe non-proliferative diabetic retinopathy), levels 60 through 71 (proliferative diabetic retinopathy), level 80 (proliferative diabetic retinopathy with vitreous hemorrhage), and levels 90 through 95 (ungradable images). ICDRS grading scores are as follows: 0 (no apparent retinopathy), 1 (mild non-proliferative diabetic retinopathy), 2 (moderate non-proliferative diabetic retinopathy), 3 (severe non-proliferative diabetic retinopathy), and 4 (proliferative diabetic retinopathy). 16 studies reported manual grading of fundus or OCT images in order to act as references or to evaluate AI screening accuracy against that of human analysis (2-6, 8, 10-13, 15-20). Manual graders included ophthalmologists, retinal specialists, and trained nurses (who received 3 to 6 months of training before receiving retinal grading accreditation).

2.4 Case Ascertainment

The studies included in this review utilize diverse criteria for the ascertainment of diabetic retinopathy cases. All 21 studies use a prior diagnosis of diabetes mellitus (either type 1 or type 2) as a potential contributor to the development of diabetic retinopathy (1-21). The following are other indicators that were considered for case ascertainment: macular edema (19 studies; 1-6, 8-16, 18-21), neovascularization (8 studies; 3, 6, 9, 11, 13, 15, 16, 18), microvascular abnormalities including but not limited to microaneurysms and hemorrhaging (8 studies; 6, 10, 13, 14, 16-18, 21), drusen (6 studies; 1, 7, 9, 11, 17, 18), exudates (4 studies; 4, 7, 17, 21), increased retinal thickness (3 studies; 1, 6, 9), retinal detachment (2 studies; 4, 10), increased HbA1c levels (2 studies; 1,2), retinal pigmented epithelium (RPE) atrophy (1 study; 1), enlarged optic cup to disc ratio (1 study; 1), vitreous inflammation (1 study; 4), choroidal infiltrate (1 study; 4), retinal vessel occlusion (1 study; 10), and choroidal infiltrate (1 study; 4).

Table 5. — Case definition and case ascertainment used in included studies

Reference	Case definition: symptomatic indicators included as diagnostic criteria if specified	Examination, Laboratory and/or imaging technique? If yes: type of examination and/or imaging	Grading methods of case ascertainment; analysis protocols; additional data sources
Abramoff 2018 (1)	Diabetic Retionpathy (DR); diagnosis of diabetes, enlarged cup to disc ratio, any drusen, any retinal pigment epithelium atrophy, increased retinal thickness, macular edema, HbA1C level analysis	Yes, Optical Coherence Tomography (OCT), retinal fundus	More than mild diabetic retinopathy (mtmDR) participants were graded according to FPRC reading protocol; DR severity levels were ascertained according to Early Treatment Diabetic Retinopathy Study Severity Scale (ETDRS) and to Diabetic Macular Edema (DME)
Bellemo 2019 (2)	DR; diagnosis of diabetes, HbA1C level analysis, macular edema	Yes, Two-field (macular centered and retinal centered), color, non-stereo, 45° retinal fundus photographs were taken for each eye using the Digital Retinopathy System, CentreVue	The retinal images were assessed by nurses and imaging technicians of non-medical background who were trained and supervised by ophthalmologists over a series of visits Images were graded at the time of photography The dataset provided patient demographics and risk factors (ex. age, sex, diabetes duration, diabetes type, BMI, systolic blood pressure, and diastolic blood pressure

Reference	Case definition: symptomatic indicators included as diagnostic criteria if specified	Examination, Laboratory and/or imaging technique? If yes: type of examination and/or imaging	Grading methods of case ascertainment; analysis protocols; additional data sources
De Fauw 2018 (3)	DR; diagnosis of diabetes, neovascularization, macular edema	Yes, OCT (Topcon 3D OCT)	Data were selected from a retrospective cohort of all patients who attended Moorfields Eye Hospital NHS Foundation Trust, which consists of 32 clinic sites serving urban, mixed socioeconomic and ethnicity population centered around London, United Kingdom A validation subset of 993 scans (993 patients) was graded separately by three junior graders and ophthalmologists with disagreement in clinical labels arbitrated by a senior retinal specialist with over 10 years of experience and image reading center certification for OCT segmentation
Gargeya 2017 (4)	DR; vitritis, optic disc edema, macular "leopard-print" choroidal infiltrate, exudative retinal detachment, diagnosis of diabetes	Yes, Retinal fundus	Deep learning network learned data-driven features from training data set, characterizing DR based on an expert-labelled ground truth; a panel of retinal specialists determined the ground truth for our data set before experimentation Training data set was used to differentiate healthy fundi from those with DR Information learned in automated screening method was visualized readily through an automatically generated abnormality heatmap, highlighting subregions within each input fundus image for further clinical review

Reference	Case definition: symptomatic indicators included as diagnostic criteria if specified	Examination, Laboratory and/or imaging technique? If yes: type of examination and/or imaging	Grading methods of case ascertainment; analysis protocols; additional data sources
Gulshan 2016 (5)	DR; macular edema, diagnosis of diabetes	Yes, Retinal fundus	DR severity was graded according to the International Clinical Diabetic Retinopathy Severity (ICDRS) scale Each retinal image was graded 3 to 7 times for DR, DME, and image gradability by a panel of 54 US licensed ophthalmologists and ophthalmology senior residents between May and December 2015 The AI algorithm was validated in January and February 2016 using 2 separate data sets, both graded by at least 7 US board-certified ophthalmologists with high intragrader consistency
Hansen 2015 (6)	DR; macular edema, microvascular abnormalities (microaneurysms and/or hemorrhages), neovascularization, retinal thickening, diagnosis of diabetes	Yes, Retinal fundus	First, human grading was performed for the presence or absence of DR, and for those with DR this was sub-divided in to referable or non-referable DR The human graders were masked to the patient's status to having or not having diabetes The automated Iowa Diabetic Retinopathy (IDP) detection software was deployed to identify those with DR and also to categorize the severity of DR
Kanagasingam 2018 (7)	DR; drusen, exudates (around the optic disc, the papillomacular area, and the macula), diagnosis of diabetes	Yes, Retinal fundus	The study manually classified retinal images from the data sets into 2 classes: adequate image quality for DR grading and inadequate image quality for DR grading. Then only adequate quality images were used to train the convolutional neural network (CNN) model Potential bias: there were some images whose quality was ambiguous between adequate and inadequate, which was expected to influence some outcomes

Reference	Case definition: symptomatic indicators included as diagnostic criteria if specified	Examination, Laboratory and/or imaging technique? If yes: type of examination and/or imaging	Grading methods of case ascertainment; analysis protocols; additional data sources
Keel 2018 (8)	DR; macular edema, diagnosis of diabetes	Yes, Single-field, non-mydriatic retinal fundus photographs	Adults with diabetes were recruited from two urban endocrinology outpatient clinics and fundus photographs were graded for referrable DR (≥ pre-proliferative DR) Each participant underwent: (1) automated screening model; (2) manual model where retinal images were transferred to a retinal grading center and manual grading outcomes were distributed to the patients within 2 weeks of assessment Each image was randomly assigned to a single ophthalmologist for initial grading and, following this, sequentially assigned to individual graders until three consistent grading outcomes were achieved The mean assessment time for automated screening was 6.9 minutes
Kermany 2018 (9)	DR; neovascularization, subretinal fluid, retinal thickening, drusen, macular edema, diagnosis of diabetes	Yes, OCT images	An independent test set of 1,000 images from 633 patients was used to compare the AI network's referral decisions with the decisions made by human experts An occlusion test was performed on 491 images to identify the areas contributing most to the neural network's assignment of the predicted diagnosis; this testing successfully identified the region of interest in 94.7% of images that contributed the highest importance to the deep-learning algorithm

Reference	Case definition: symptomatic indicators included as diagnostic criteria if specified	Examination, Laboratory and/or imaging technique? If yes: type of examination and/or imaging	Grading methods of case ascertainment; analysis protocols; additional data sources
Li 2018 (10)	DR; macular edema, retinal hemorrhaging, intraretinal microvascular abnormalities, retinal vessel occlusion, retinal detachment, diagnosis of diabetes	Yes, Retinal fundus, two-field 45° images were taken	Each image in the local data set was graded between three and eight times, with a mean agreement of 87.3% among 21 ophthalmologists Graders were blind to the previous grading outcomes and a given image could be assigned to a grader only once; the consensus grading outcome was assigned as the final, conclusive grade of each image The external data set contained retinal images from 14,520 eyes from three population-based studies
Li 2019 (11)	DR; macular edema, neovascularization, drusen, diagnosis of diabetes	Yes, OCT	A total of 207,130 retinal OCT images between 2013 and 2-17 were selected from retrospective cohorts of 5,319 adult patients Only retinal images with a clear consensus annotation between ophthalmologists were taken into the sample and imported into the database To avoid any errors in grading, the resulting evaluation set was further checked by another senior retina expert
Lim 2019 (12)	DR; macular edema, diagnosis of diabetes	Yes, Undialated-2field retinal fundus photography (macula centered and disk centered images); dilated 4-wide field stereoscopic retinal fundus photography	The EyeArt AI eye screening system was evaluated against the clinical reference standard based on adjudicated grading of the 4-wide field photographs by expert graders at the Wisconsin Fundus Photograph Reading Center using the Early Treatment Diabetic Retinopathy Study (ETDRS) Severity Scale

Reference	Case definition: symptomatic indicators included as diagnostic criteria if specified	Examination, Laboratory and/or imaging technique? If yes: type of examination and/or imaging	Grading methods of case ascertainment; analysis protocols; additional data sources
Rajalakshmi 2018 (13)	DR; macular edema, neovascularization, vitreous/preretinal hemorrhage, microaneurysms, diagnosis of diabetes	Yes, Retinal fundus photography taken using a smartphone-based device	301 patients with type 2 diabetes underwent retinal photography with Remidio 'Fundus on phone' (FOP), a smartphone-based device, at a tertiary care diabetes center in India Grading of DR was performed by the ophthalmologists using ICDRS The sensitivity and specificity of automated grading were assessed and validated against the ophthalmologists' grading
Raju 2017 (14)	DR; macula edema, intraretinal microvascular abnormalities, diagnosis of diabetes	Yes, Retinal fundus	Using the CNN based approach for automatic screening of diabetic retinopathy, the prediction network was trained based on the publicly available Kaggle dataset
Raumviboonsuk 2019 (15)	DR; macular edema, neovascularization, diagnosis of diabetes	Yes, Retinal fundus	Grades adjudicated by a panel of international retinal specialists served as the reference standard The study was conducted to assess the screening performance of the Deep Learning (DL) algorithm compared to real-world graders for classifying multiple clinically relevant severity levels of DR in the national screening program for DR in Thailand
Sayres 2019 (16)	DR; macular edema, neovascularization, microaneurysm, hemorrhage, diagnosis of diabetes	Yes, Retinal fundus	Ten ophthalmologists (5 general ophthalmologists, 4 retina specialists, 1 retina fellow) read images for DR severity based on the ICDRS scale in each of 3 conditions: (1) unassisted; (2) grades only; (3) grades plus heatmap

Reference	Case definition: symptomatic indicators included as diagnostic criteria if specified	Examination, Laboratory and/or imaging technique? If yes: type of examination and/or imaging	Grading methods of case ascertainment; analysis protocols; additional data sources
Son 2020 (17)	DR; hemorrhage, hard exudate, cotton-wool patch, drusen, macular hole, myelinated nerve fiber, chorioretinal atrophy or scar, membrane abnormalities, any vascular abnormality, retinal nerve fiber layer defect, glaucomatous disc change, nonglaucomatous disc change, diagnosis of diabetes	Yes, Retinal fundus	57 ophthalmologists (including 16 certified retina specialists, 9 certified glaucoma specialists, and 3 certified cornea specialists) were recruited as readers of macular-centered retinal fundus images 3 independent readings from 3 different readers were collected for each image; after completion of annotation, all datasets from the 3 reading systems were aggregated regardless of the subspecialities of the readers
Ting 2017 (18)	DR; macular edema, neovascularization, drusen, thinning or notching of the neuroretinal rim, hemorrhages, diagnosis of diabetes	Yes, Retinal fundus	The Singapore National Diabetic Retinopathy Screening Program (SIDRP) uses digital retinal photography, a tele-ophthalmology platform, and assessment of diabetic retinopathy by a team of trained professional graders For each patient, 2 retinal photographs (optic disc and fovea) were taken of each eye; All trained graders received 3 to 6 months of training before certification and underwent annual reaccreditation
Torre 2019 (19)	DR; macular edema, diagnosis of diabetes	Yes, Retinal fundus	For every patient, right and left eye images are reported; all images are classified by ophthalmologists according to a standard severity scale The dataset is split into two disjoint sets containing eye images of different patients, one for training and the other for testing
Tufail 2016 (20)	DR; macular edema, diagnosis of diabetes	Yes, Retinal fundus	Retinal images were manually graded following a standard national protocol; discrepancies between manual grades and AI system results were sent to a reading center for arbitration
Zeng 2019 (21)	DR; macular edema, hard exudates, hemorrhages, microaneurysms, diagnosis of diabetes	Yes, Retinal fundus	The proposed AI model accepts binocular fundus images as inputs and learns their correlation to help to make a prediction

DR: Diabetic Retionpathy, OCT: Optical Coherence Tomography, mtmDR: more than mild Diabetic Retinopathy, ETDRS: Early Treatment Diabetic Retinopathy Study, DME: Diabetic Macular Edema, BMI: Body Mass Index, HbA1c: Glycated Hemoglobin, ICDR: International Clinical Diabetic Retinopathy, DL: Deep Learning.

2.5 Quality analysis of the included studies: Limitations and Strengths

The primary objective of the included studies was to assess the abilities of particular AI systems of interest to screen for diabetic retinopathy in people with type 1 and type 2 diabetes mellitus. Each study demonstrated a notable limitation in some form relating to either its external validity or use of imaging methods to capture the fundus of OCT photographs.

There are three key limitations regarding the external validity of the results reported by the included studies to the burden of diabetic retinopathy in their respective countries. The first limitation is not specifying the setting, environment, or type of community (for example, urban, suburban, or rural) from which study participants were recruited (1, 2, 8, 12). Environmental factors play a key role in the pathogenesis of diabetes mellitus. Such factors include air pollution, soil, water, stress, lack of physical activity, unhealthy diet, vitamin D deficiency, and exposure to particular pathogens. Although a genetic basis also exists, the time of onset of diabetes mellitus, and subsequently diabetic retinopathy, depends largely on the aforementioned environmental factors. Risk factors for diabetic retinopathy are complex, and past studies have demonstrated that the neighborhood environment in which patients live influences retinal microvascular complications associated with diabetes. The assessment of environmental factors that contribute to diabetic retinopathy is important in AI screening studies that do not specify the severity of the disease present in study participants. The severity or stage of the disease may impact the AI system's reported sensitivity and specificity values, and so considering confounding variables such as setting, environment, and types of communities is important to produce accurate and generalizable results. The second limitation to the external validity of various studies is the recruitment of participants solely from a single location or limited geographic region (6, 7, 13). Reporting performance metrics based on a limited population produces results that may not be representative or generalizable to a larger area of interest, for example the country in which the study was conducted. A study that reports sensitivity and specificity results of an AI screening system based on participants from a single city may be representative of the population of that geographic region. However, further studies assessing such performance metrics must be conducted on a larger scale in order to develop national AI screening programs that are based on nationwide data and evaluation. The third limitation to external validity of particular studies included in this review is not providing the geographic or demographic information of recruited participants (16, 19, 20, 21). Even if a particular AI screening system presents exceptionally high sensitivity and specificity results, because the geographic scope and demographic breakdown of its participants are unknown, one cannot estimate how the system will perform when presented with retinal scans of diabetes mellitus patients from a different region and demographic composition.

The majority of the included studies considered diverse symptomatic indicators of diabetic retinopathy, used proper imaging methods, and properly reported the quantity and source of fundus or OCT images (4, 5, 7-21). Reliable datasets designed for laboratory and clinical research that were used by the included studies for training and external validation are as follows: E-ophtha (4, 17), Messidor-2 (4, 5), EyePACS (5, 12, 16, 19), DiaRetDB1 (7), Kaggle (7, 14), and IDRID (17). However, four studies (1, 2, 3, 6) only provided the source of retinal images and failed to detail the sample size of the dataset used for the development, training, and external validation of their respective AI screening systems. For instance, *Abramoff et al.* reported that it obtained OCT images from 10 primary care practices from across the United States, however, did not report the quantity of images analyzed by its AlexNet AI system. The quantity of images evaluated by AI screening systems directly effects how well it learns to recognize particular lesions. Although the generalizability of results may seem

promising from *Abramoff et al.*, as it considers clinical settings from across the US, if the quantity of images analyzed was low, AlexNet may not have received sufficient training to recognize particular disease indicators, which could lessen the validity of its reported sensitivity and specificity results.

Despite presenting limitations, each study demonstrated notable strengths. Nine studies utilized large quantities of fundus or OCT images for the training and external validation of their respective AI systems (4, 5, 8, 10, 11, 14, 15, 18, 19). Quantities ranged from 25,000 to over 600,000 images. Using larger sample sizes allows for a more precise estimate of sensitivity and specificity results, can be more representative of the sample's population, and can be used to better generalize results. In addition to using a sample size of greater than 200,000 retinal images, *Li et al.* (2019), assessed the performance of its AI system of interest (VGG-16) over a four-year period, to assess the consistency of sensitivity and specificity results. Three studies recruited population-based cohorts, which allowed for the estimation of AI screening performance (sensitivity and specificity) values in the reference populations (1, 2, 15). For example, *Raumviboonsuk et al.* implemented over 25,000 retinal images from a community-based nationwide diabetic retinopathy screening program in Thailand and reported severity distributions based on the stage of diabetic retinopathy predicted by its Inception V4 AI system. Additionally, one study applied its AI system to screen for a disease other than diabetic retinopathy for external validation purposes (9). Such an approach to external validation was unique to this study amongst all included studies in this review. *Kermany et al.* demonstrated the general applicability of its Inception V3 system to screen for pediatric pneumonia using chest X-Rays to help externally validate its reported results of 97.8% sensitivity and 97.4% specificity. Furthermore, this demonstrates the versatility of the Inception V3 system. There are additional strengths amongst the included studies that are noteworthy. *De Fauw et al.* used less prohibitive training data requirements, which allowed researchers to develop their U-Net AI screening systems using retinal images from across multiple real-world settings. *Hansen et al.* and *Sayres et al.* assessed their respective AI screening systems across different stages and severity levels of diabetic retinopathy, which was unique to these two studies. Providing stage and severity breakdown is an important feature for AI systems that is being assessed for future clinical implementation, as diabetic retinopathy is a progressive rather than a binary disease. *Kanagasingam et al.* assesses its Inception V3 AI screening system against an established gold standard grading protocol that the study developed in collaboration with ophthalmologists who specialize in retinal diseases. The alignment of AI screening results with a grading protocol from a reliable source supports the demonstrated performance of the system of interest. *Son et al.* developed its CNN to recognize diverse lesions, which increases the potential of screening systems to differentiate between stages and severity levels of progressing diabetic retinopathy. *Lim et al.* collected ungradable EyeArt AI results, dilated the unreadable images, and repeated screening assessments. Doing so improves the gradeability rate of EyeArt and allows for the greater use of available retinal images. Lastly, *Tufail et al.* reported the incremental cost-effectiveness ratio (ICER) for its EyeArt AI screening system, which is particularly useful for informing public health interventions that plan to implement of accurate, top-performing, and economical AI systems for large-scale DR screening programs.

Table 6. — Quality of included studies

Reference	Main Objective	Limitations regarding external validity	Limitations in symptomatic indicators / imaging methods and reporting / further issues	Strengths
Abramoff 2018 (1)	Trial of an AI system to detect diabetic retinopathy in people with diabetes	Urban, suburban, and rural classifications are not specified for the 10 primary care practices included in the study; community type could have a potential impact on the frequency of certain lesions and/or stages of DR in the study population, therefore the study's external validity is limited to its areas of interest	The number of retinal images gathered from the 10 primary care practices across the US is not specified; this creates uncertainty in the validity of sensitivity and specificity values reported by the study as the number of retinal images used to train the AI algorithm directly effects how well it can recognize lesions during the external validation phase	The study includes 10 primary care practices from across the US rather than within a single region
Bellemo 2019 (2)	To evaluate the accuracy of an AI model using deep learning in a population-based diabetic retinopathy screening program in Zambia, a lower-middle-income country.	Urban, suburban, and rural classifications are not specified for the 5 mobile screening clinics included in the study; community type could have a potential impact on the frequency of certain lesions and/or stages of DR in the study population, therefore the study's external validity is limited to its areas of interest	The number of retinal images gathered from the 5 mobile screening clinics across Zambia is not specified; this creates uncertainty in the validity of sensitivity and specificity values reported by the study as the number of retinal images used to train the AI algorithm directly effects how well it can recognize lesions during the external validation phase	The study includes 5 mobile screening units from across Zambia rather than within a single region
De Fauw 2018 (3)	To apply a novel deep learning architecture to a clinically heterogeneous set of three-dimensional optical coherence tomography scans from patients referred to a major eye hospital	N/A: population-based cohorts were used to externally validate the study's DLA	The number of retinal images in the external validation data set used to assess the deep learning architecture is not provided; this creates uncertainty in the validity of sensitivity and specificity values reported by the study as the number of retinal images used to train the AI algorithm directly effects how well it can recognize lesions during the external validation phase	The study removes previous barriers to wider clinical use without prohibitive training data requirements for pathologies across multiple real-world settings

Reference	Main Objective	Limitations regarding external validity	Limitations in symptomatic indicators / imaging methods and reporting / further issues	Strengths
Gargeya 2017 (4)	To develop robust diagnostic technology to automate DR screening	N/A: 405 E-ophtha and 1,748 Messidor-2 retinal fundus images collectively serve as a sufficiently large data set for external validation	N/A: E-ophtha and Messidor-2 are trusted databases of color retinal fundus images designed for laboratory and clinical research	A large data set of 75,137 publicly available retinal fundus images from diabetic patients were used to train and test an AI model to differentiate healthy fundi from those with DR
Gulshan 2016 (5)	To apply deep learning to create an algorithm for automated detection of diabetic retinopathy and diabetic macular edema in retinal fundus photographs	N/A: 9,963 EyePACS-1 and 1,748 Messidor-2 retinal fundus images collectively serve as a sufficiently large data set for external validation	N/A: EyePACS-1 and Messidor-2 are trusted databases of color retinal fundus images designed for laboratory and clinical research	A large data set of 128,175 retinal fundus images from diabetic patients were used to train and test an AI model to differentiate healthy fundi from those with DR; each image was graded 3 to 7 times for DR by 54 US licensed ophthalmologists and ophthalmology senior residents over an 8-month period
Hansen 2015 (6)	To compare the Iowa Detection Program (IDP) ability to detect diabetic eye diseases to human grading carried out at Moorfields Reading Center on the population of Nakuru Study from Kenya	All 3,460 subjects whose retinal images were used are long-term residents of Nakuru, which limits the generalizability of sensitivity and specificity results to the study's geographic area of interest	The number of retinal images gathered from the Nakuru Eye Study from Kenya is not specified; this creates uncertainty in the validity of sensitivity and specificity values reported by the study as the number of retinal images used to train the AI algorithm directly effects how well it can recognize lesions during the external validation phase	Grading conducted by retinal specialists and the Iowa Detection Program (the study's AI system) broke down screening results by severity of DR: no-apparent retinopathy, mild, moderate, severe, and proliferative; doing so allows sensitivity and specificity values to be assessed for different stages of DR, which leads to a more holistic evaluation of the AI system

Reference	Main Objective	Limitations regarding external validity	Limitations in symptomatic indicators / imaging methods and reporting / further issues	Strengths
Kanagasingam 2018 (7)	To describe the performance of an AI system for diabetic retinopathy deployed in a primary care practice	This study enrolled diabetes patients from a single primary care practice in Western Australia; providing results from a single location with a limited number of participants (193) limits the external validity of the results to the area of interest	N/A: DiaRetDB1, Kaggle (EyePACS), and the study's own Australian Tele-eye care DR databases of color retinal fundus images are designed for laboratory and clinical research Proper symptomatic indicators, imaging methods, and reporting techniques were considered and used in this study	The binary reading (disease or no disease) by ophthalmologists was used as the gold standard and compared with the grading obtained from the AI system, which contributes to the validity of sensitivity and specificity findings reported by this study
Keel 2018 (8)	To evaluate the feasibility and patient acceptability of a novel AI-based DR screening model within endocrinology outpatient settings	This study enrolled diabetes patients from two urban endocrinology departments in in Melbourne, Australia; because differing environments, community types, and locations of residence can potentially contribute to the frequency of particular lesions or stages of DR, providing results from only two locations with a limited number of participants (196) limits the external validity of the results	N/A: proper symptomatic indicators, imaging methods (retinal fundus), and reporting techniques were considered and used in this study	The deep learning algorithm was developed using a large number of retinal photographs (66,790) acquired from an online database (http://www.labelme.org, Guangzhou, China) designed for laboratory and clinical research
Kermany 2018 (9)	To establish a diagnostic tool based on a deep-learning framework for the screening of patients with common treatable blinding retinal diseases	The model was tested with 1,000 retinal images belonging to 633 patients from an unspecified location; uncertainty with regards to the environments, community types, and locations of residence of the study participants limits the study's ability to generalize its findings to retinal images belonging to patients outside of the study	N/A: proper symptomatic indicators, imaging methods (OCT), and reporting techniques were considered and used in this study	Study provides a more transparent and interpretable diagnosis of DR by highlighting the regions recognized by the neural network Study further demonstrates the general applicability of its AI system to screen for pediatric pneumonia using chest X-ray images

Reference	Main Objective	Limitations regarding external validity	Limitations in symptomatic indicators / imaging methods and reporting / further issues	Strengths
Li 2018 (10)	To describe the development and validation of an AI-based, deep learning algorithm (DLA) for the detection of referable DR	N/A: population-based cohorts were used to externally validate the study's DLA	N/A: proper symptomatic indicators, imaging methods (retinal fundus), and reporting techniques were considered and used in this study	For external validation, the study tested its DLA using 35,201 images of 14,520 eyes from population-based cohorts of Malays, Caucasian Australians, and Indigenous Australians
Li 2019 (11)	To explore the use of deep transfer learning method based on the visual geometry group 16 (VGG-16) network for classifying age-related macular degeneration (AMD) diabetic macular edema (DME) in OCT images accurately and automatically	N/A: A large set of images derived from population-based cohorts were used in each of the study locations (the Shiley Eye Institute of the University of California, San Diego, the California Retinal Research Foundation, Medical Center Ophthalmology Associates, the Shanghai First People's Hospital, and the Beijing Tongren Eye Center	N/A: diverse symptomatic indictors including neovascularization, diabetic macular edema, and drusen were considered; proper imaging methods (OCT), and reporting techniques were applied in this study	For external validation, the study tested its DLA using 207,130 retina OCT images between 2013 and 2017 from retrospective cohorts of 5,319 adult patients from healthcare institutions in the United States and China; the use of these population-based cohorts strengthens the external validity of sensitivity and specificity findings for patients residing in California, Shanghai, and Beijing
Lim 2019 (12)	Evaluate AI system to screen people with diabetes at point-of-care for DR including DME	Urban, suburban, and rural classifications are not apparent or considered in this multi-center study; community type could have a potential impact on the frequency of certain lesions and/or stages of DR in the study population, therefore the study's external validity is limited to its areas of interest	N/A: EyePACS-1 is a trusted database of color retinal fundus images designed for laboratory and clinical research; proper imaging methods (retinal fundus), and reporting techniques were applied in this study	Dilated images were used for subjects with ungradable EyeArt AI results on undilated images, which improves the gradeability rate of the EyeArt AI system and allows for the greater use of available retinal images
Rajalakshmi 2018 (13)	To assess the role of AI-based automated software for detection of DR and sight-threating DR by retinal fundus photography taken using a smartphone-based device and validate it against ophthalmologist's grading	This study enrolled diabetes patients from a single tertiary care diabetes center in India; providing results from a single location limits the external validity of the results to the area of interest	N/A: proper symptomatic indicators, imaging methods (retinal fundus), and reporting techniques were considered and used in this study	This is the first time that an automated AI DR screening software has been successfully tested for its accuracy for smartphone-based retinal imaging

Reference	Main Objective	Limitations regarding external validity	Limitations in symptomatic indicators / imaging methods and reporting / further issues	Strengths
Raju 2017 (14)	To use deep learning applications to classify the stages of DR and detect the laterality of the eye using retinal funduscopic images	N/A: population-based cohorts were used to externally validate the study's DLA	N/A: Kaggle is a trusted database of color retinal fundus images designed for laboratory and clinical research; proper imaging methods (retinal fundus), and reporting techniques were applied in this study	A large data set of ~53,000 retinal fundus images from diabetic patients was used as the validation dataset along with 8,810 images for the training data set, which increases the power of the AI screening test
Raumviboo 2019 (15)	To validate a deep learning algorithm on a large-scale clinical population, and compare the algorithm performance with that of human graders	N/A: population-based cohorts were used to externally validate the study's DLA	N/A: proper symptomatic indicators, imaging methods (retinal fundus), and reporting techniques were considered and used in this study	A total of 25,326 gradable retinal images of patients with diabetes from the community-based, nationwide screening program of DR in Thailand were analyzed for DR severity and referable DME
Sayres 2019 (16)	To understand the impact of deep learning DR algorithms on physician readers in computer-assisted settings	The study utilized 1,796 retinal fundus images from 1,612 unique patients, however, the source, geographic distribution, and other relevant information regarding the diversity of the participant population is not specified; this creates uncertainty and limits the external validity of the study's results	N/A: EyePACS-2 is a trusted database of color retinal fundus images designed for laboratory and clinical research; proper imaging methods (retinal fundus), and reporting techniques were applied in this study	This study assessed the sensitivity and specificity of its deep learning DR algorithm across a 5-point scale on DR severity: no apparent retinopathy, mild nonproliferative DR, moderate nonproliferative DR, severe nonproliferative DR, and proliferative DR; such information is valuable when determining treatment plans for patients with diverse stages of DR

Reference	Main Objective	Limitations regarding external validity	Limitations in symptomatic indicators / imaging methods and reporting / further issues	Strengths
Son 2020 (17)	To develop and evaluate deep learning models that screen multiple abnormal findings in retinal fundus images	N/A: population-based cohorts were used to externally validate the study's DLA; macula-centered retinal fundus images from the Seoul National University Bundang Hospital Retina Image Archive were obtained at the health screening center and ophthalmology outpatient clinic at Seoul national University Bundang Hospital	N/A: the Indian Diabetic Retinopathy Image Dataset (IDRID) and E-Ophtha dataset are trusted databases of color retinal fundus images designed for laboratory and clinical research; proper imaging methods (retinal fundus), and reporting techniques were applied in this study	12 major findings were assessed (hemorrhage, hard exudate, cotton-wool patch, drusen, membrane, macular hole, myelinated nerve fiber, chorioretinal atrophy or scar, any vascular abnormality, retinal nerve fiber layer defect glaucomatous disc change, and nonglaucomatous disc change) with their region information using DLA; this method of assessment demonstrate the versatility of the DLA
Ting 2017 (18)	To evaluate the performance of a deep learning system (DLS) in detecting referable DR, vision-threatening DR, possible glaucoma, and age-related macular degeneration (AMD) in community and clinic-based multiethnic populations with diabetes	N/A: Validation of the DLS was completed using a primary validation data set in the Singapore National Diabetic Retinopathy Screening Program and 10 multiethnic cohorts with diabetes; using national and ethnically diverse datasets strengthens the external validity of the study's results for the population of Singapore	N/A: The Singapore National Diabetic Retinopathy Screening Program; is a trusted database of color retinal fundus images designed for laboratory and clinical research; proper imaging methods (retinal fundus), and reporting techniques were applied in this study	A large data set of 494,661 retinal fundus images from diabetic patients was used as the validation dataset along with 112,648 images for the training data set, which increases the power of the AI screening test
Torre 2019 (19)	To present a DR deep learning interpretable classifier, determine its performance in classifying retinal images into different levels of severity and explain the classification results by assigning a score for each point in the hidden and input spaces	Source of external validation set is not provided; creates uncertainty as to the strength of the study's external validity	N/A: EyePACS is a trusted database of color retinal fundus images designed for laboratory and clinical research; proper imaging methods (retinal fundus), and reporting techniques were applied in this study	A large data set of 75,650 retinal fundus images from diabetic patients was used as the validation dataset along with 10,000 images for the training data set, which increases the power of the AI screening test

Reference	Main Objective	Limitations regarding external validity	Limitations in symptomatic indicators / imaging methods and reporting / further issues	Strengths
Tufail 2016 (20)	To determine whether automated DR image assessment systems (ARIAS) can be safely introduced into DR screening pathways to replace human graders	Source of external validation set is not provided; creates uncertainty as to the strength of the study's external validity	N/A: proper imaging methods (retinal fundus), and reporting techniques were applied in this study	The study provides incremental cost-effectiveness ratios (ICERs) for AI screening systems assessed; informs future public health efforts in implementing effective and economical AI systems for the large-scale screening of DR
Zeng 2019 (21)	A computer-aided diagnosis method based on DLA is proposed to automatically diagnose the referable DR by classifying color retinal fundus photographs into two grades	Source of external validation set is not provided; creates uncertainty as to the strength of the study's external validity	N/A: proper imaging methods (retinal fundus), and reporting techniques were applied in this study	Different from previous studies assessing AI screening systems, the proposed system in this paper accepts binocular retinal fundus images as inputs and learns their correlation to help to make a prediction

DR: diabetic retinopathy, E-ophtha: a database of color fundus images designed for scientific research, Messidor-2: a collection of diabetic retinopathy examinations, each consisting of two macula-centered eye fundus images (one per eye), EyePACS-1: a database of retinal images used by clinicians to validate diabetic retinopathy screening systems, Kaggle: a large set of high-resolution retina images taken under a variety of imaging conditions, OCT: Optical Coherence Tomography, DLA: deep learning algorithm, DLS: deep learning system, N/A: not applicable.

2.6 Meta-Analysis

2.6.1   Sensitivity of AI screening systems reported by the eligible studies

Sensitivity is reported in this review as the percentage of screened participants with diabetic retinopathy who are correctly identified as positive by the AI screening system of interest. Amongst the 21 eligible studies included in this review, 19 reported sensitivity estimates for their respective AI architectures that were applied to retinal images from diverse populations (see Figure 6 on the next page; 1, 4-21). The median sensitivity amongst the studies is 92.5% true-positive rate, with a total range of 80.3% to 100%. The quartile 1 (Q1) to quartile 3 (Q3) quartile range is 91.4% to 96.1% (see Figure 5, below). The mean of the reported results is 92.4% sensitivity. Sensitivity estimates depended largely on the quantity of retinal images used to train, develop, and externally validate AI screening systems, which was determined by the studies' respective authors. In addition to specificity variation being influenced by image quantity, the imaging modality (fundus photography or OCT), geographic area of recruited participants, and the number of participants may have affected the results. Furthermore, Table 7 compares sensitivity values to the sample size of training datasets applied to AI screening systems. It is noteworthy that the subgroup with the larger sample size of $\geq$ 75,000 retinal images showed higher sensitivity (94.0%; 95% CI: 91.3% to 96.7%) than the smaller sample size of < 75,000 images (90.5%; 95% CI: 87.1% to 93.9%).

Figure 5. — Box-plot of overall sensitivity reported by eligible studies
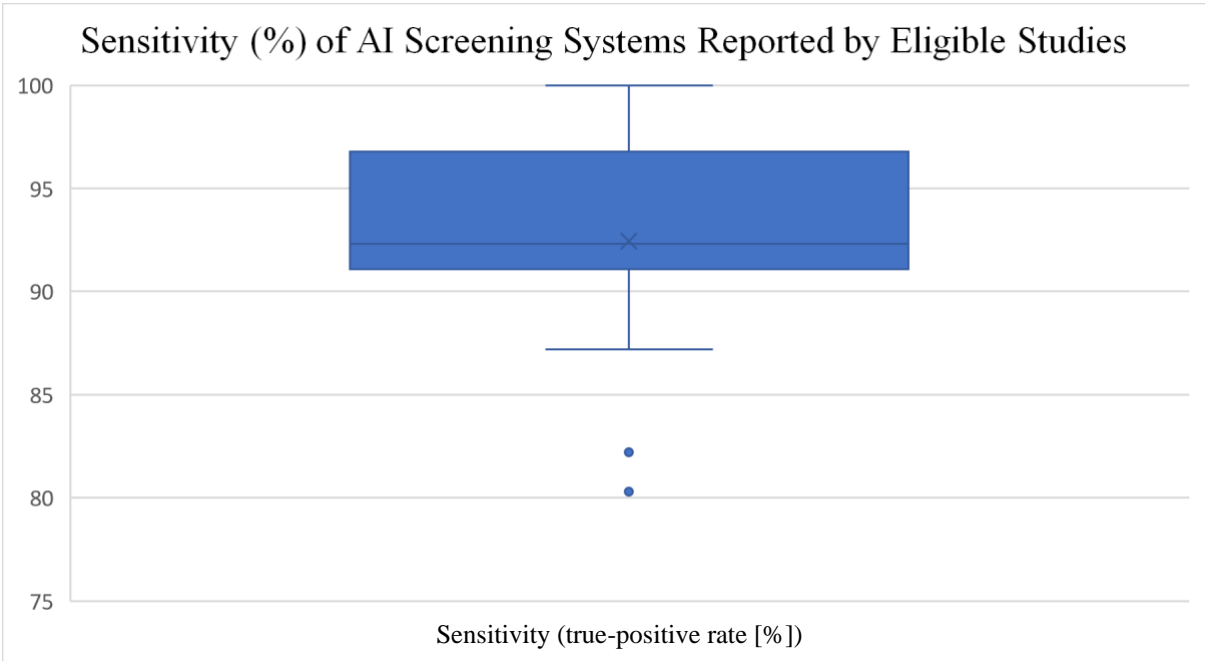

Sensitivity (true-positive rate [%])

Table 7. — Subgroup analysis of training set sample size and sensitivity of automatic detection of diabetic retinopathy using AI screening systems

| Sample size of training set	Number of eligible studies	Sensitivity, % (95 CI)
< 75,000	6	90.5 (87.1 to 93.9)
$\geq$ 75,000	8	94.0 (91.3 to 96.7)

Figure 6. — Sensitivity (%) of AI screening systems and external validation data set (if specified) reported by the included studies
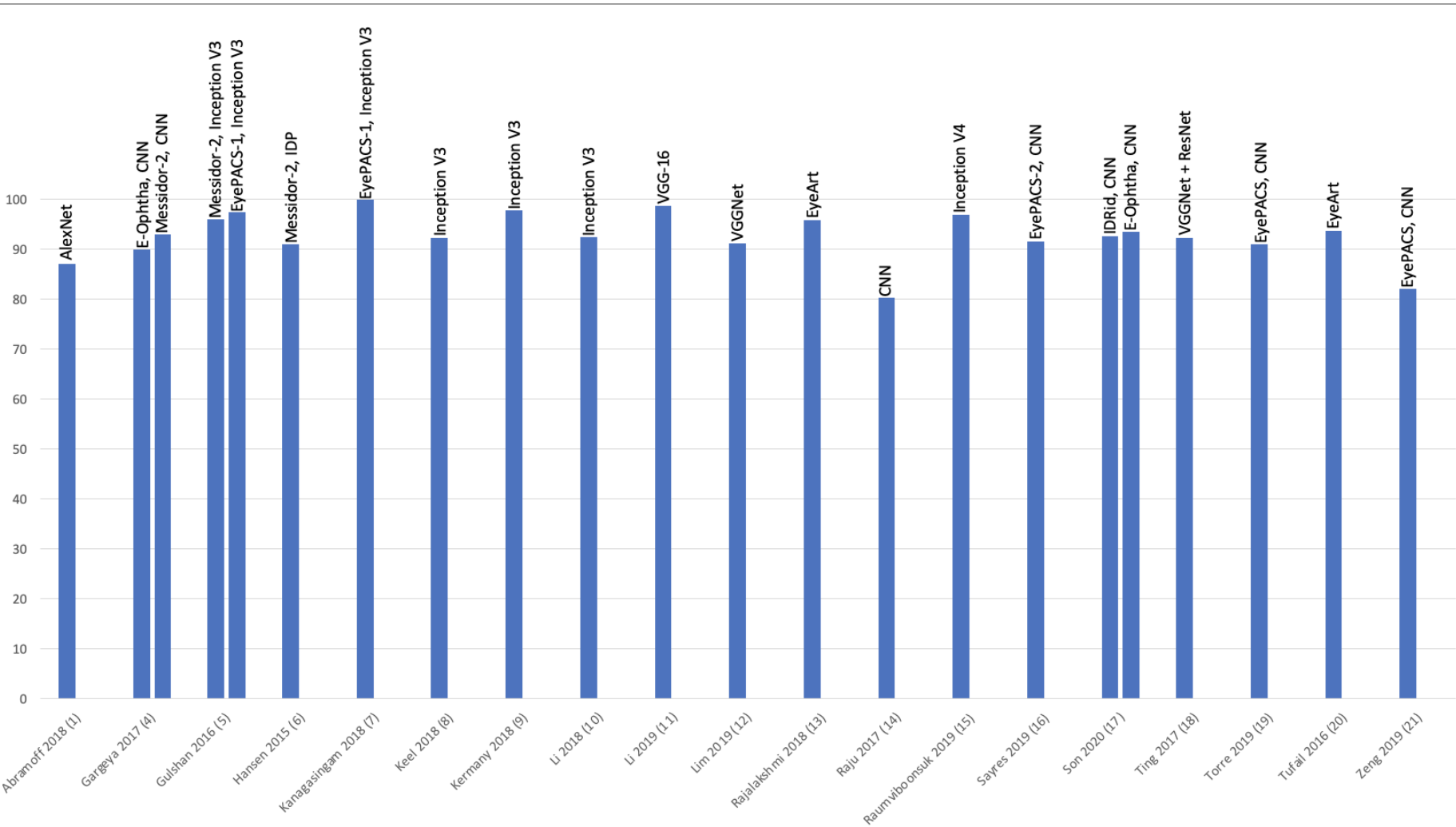
2.6.2   Sensitivity of AI screening systems according to type of architecture implemented

Different AI architectures may produce different sensitivity results due to many potential reasons including varying algorithms, reference standards, diversity of recognizable retinal lesions, various stages and severities of diabetic retinopathy presented to the architectures, type of imaging modality, quality and quantity of retinal images, geographic area in which assessments are conducted, demographic breakdown of participants, and different training, development, and external validation datasets. Figure 7 presents the mean sensitivity of each AI system that was assessed by the included studies. From lowest to highest sensitivity, the order of the reported results is as follows: 87.2% (AlexNet), 89.3% (unspecified CNNs), 91.0% (IDP), 94.1% (VGG), 94.8% (EyeArt), 96.0% (Inception V3), 96.9% (Inception V4). The pooled sensitivity amongst all studies is 92.8%. Due to the aforementioned differences between studies, precise comparisons of sensitivity are limited.

Figure 7. — Mean Sensitivity (%) of Each AI Screening System Reported by the Included Studies
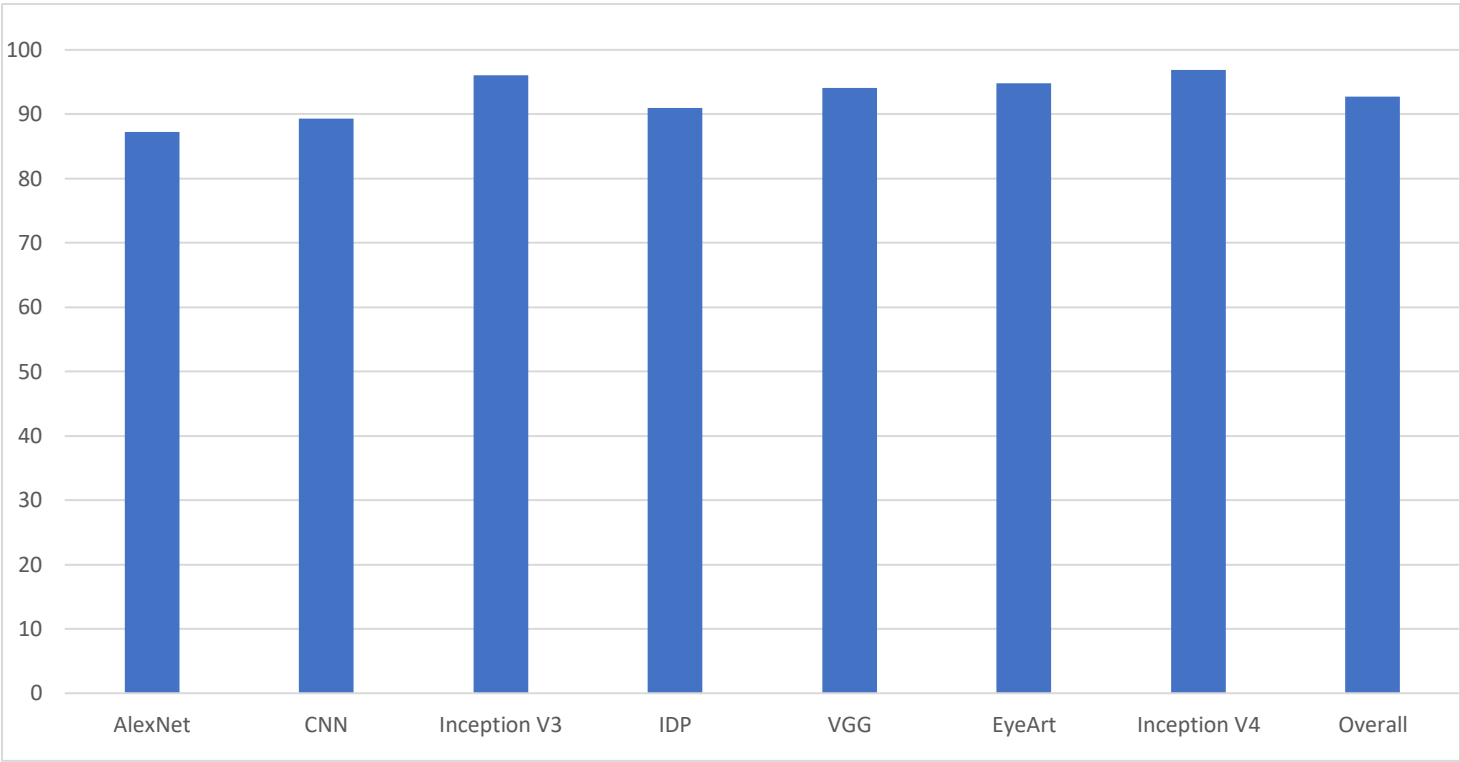
2.6.3   Specificity of AI screening systems reported by the eligible studies

Specificity is reported in this review as the percentage of screened participants without diabetic retinopathy who are correctly identified as negative by the AI screening system of interest. Amongst the 21 eligible studies included in this review, 18 reported specificity estimates for their respective AI architectures applied to retinal images from diverse populations (see Figure 9 on the next page; 1, 4-21). The median specificity amongst the studies is 92.2% false-positive rate, with a total range of 69.9% to 98.8%. The Q1 to Q3 quartile range is 90.6% to 95.2% (see Figure 8, below). The mean of the reported results is 90.3% specificity. Specificity estimates depended largely on the quantity of retinal images used to train, develop, and externally validate AI screening systems, which was determined by the studies respective authors. In addition to specificity variation being influenced by image quantity, the imaging modality (fundus photography or OCT), geographic area of recruited participants, and the number of participants may have affected the results. Furthermore, Table 8 demonstrates compares specificity values to the sample size of training data sets applied to AI screening systems. It is noteworthy that the subgroup with the larger sample size of $\geq$ 75,000 retinal images showed higher specificity (93.7%; 95% CI: 90.6% to 96.8%) than the smaller sample size of < 75,000 images (90.0%; 95% CI: 85.4% to 94.6%).

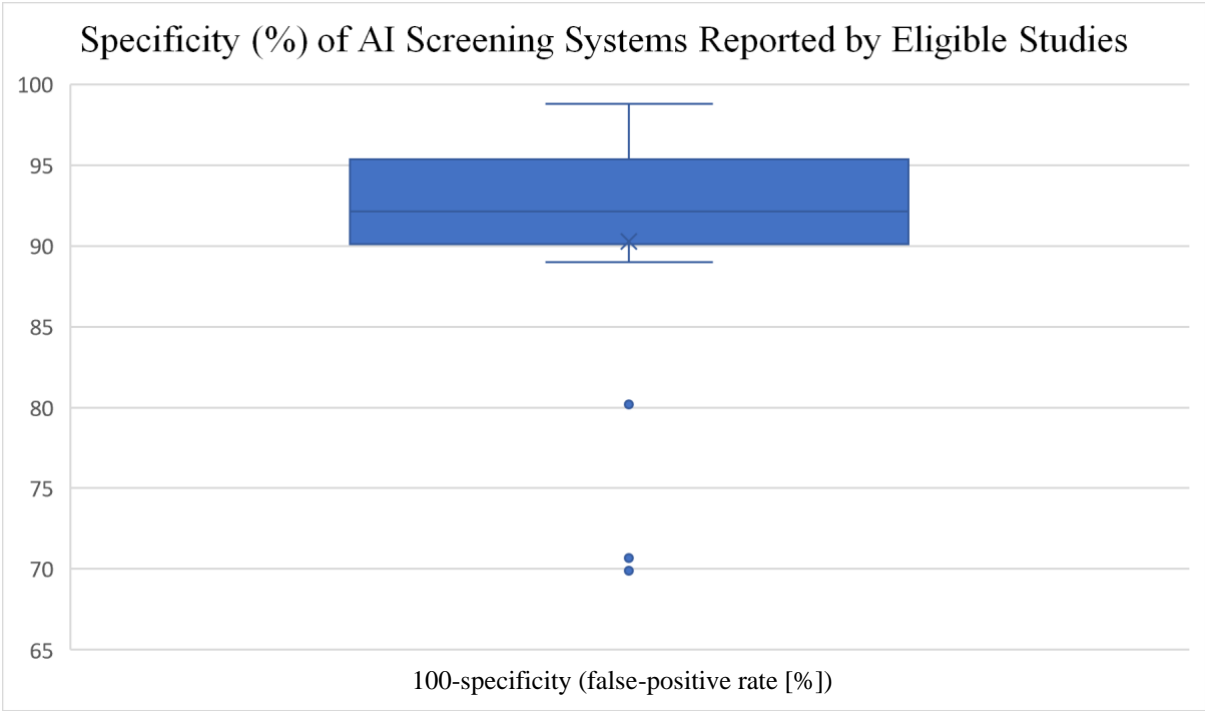**Figure 8. — Box-plot of overall specificity reported by eligible studies**

Table 8. — Subgroup analysis of training set sample size and specificity of automatic detection of diabetic retinopathy using AI screening systems

| Sample size of training set	Number of eligible studies	Specificity, % (95 CI)
< 75,000	6	90.0 (85.4 to 94.6)
$\geq$ 75,000	8	93.7 (90.6 to 96.8)

Figure 9. — Specificity (%) of AI Screening Systems Reported by the Included Studies
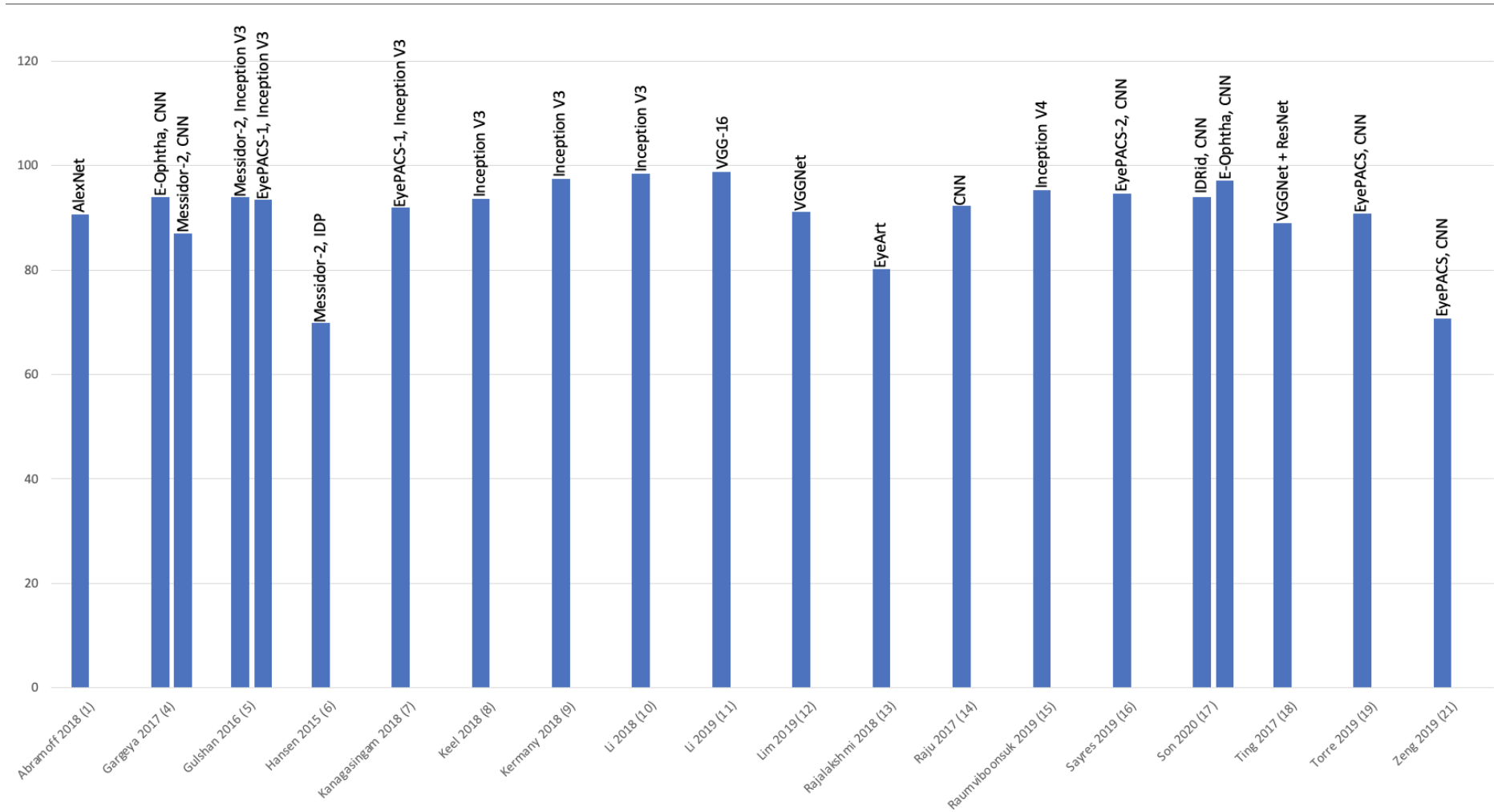
2.6.4   Specificity of AI screening systems reported by the eligible studies

Different AI architectures may produce different specificity results due to many potential reasons including varying algorithms, reference standards, diversity of recognizable retinal lesions, various stages and severities of diabetic retinopathy presented to the architectures, type of imaging modality, quality and quantity of retinal images, geographic area in which assessments are conducted, demographic breakdown of participants, and different training, development, and external validation datasets. Figure 10 presents the mean specificity of each AI system that was assessed by the included studies. From lowest to highest specificity, the order of the reported results is as follows: 69.9% (IDP), 80.2% (EyeArt), 90.1% (unspecified CNNs), 90.7% (AlexNet), 93.0% (VGG), 94.8% (Inception V3), and 95.3% (Inception V4). The pooled specificity amongst all studies is 87.7%. Due to the aforementioned differences between studies, precise comparisons of specificity are limited.
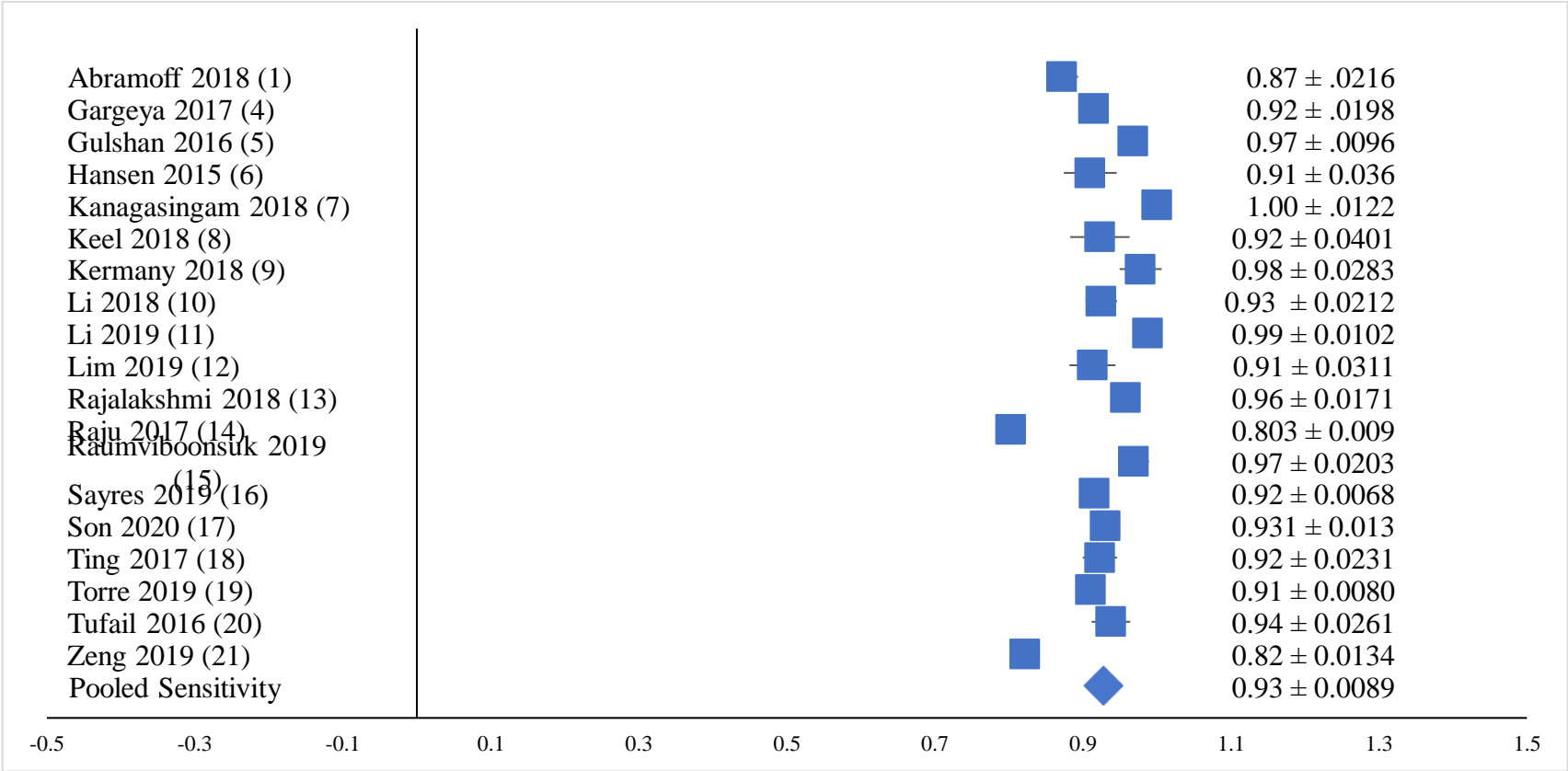
Figure 10. — Mean Specificity (%) of Each AI Screening System Reported by the Included Studies
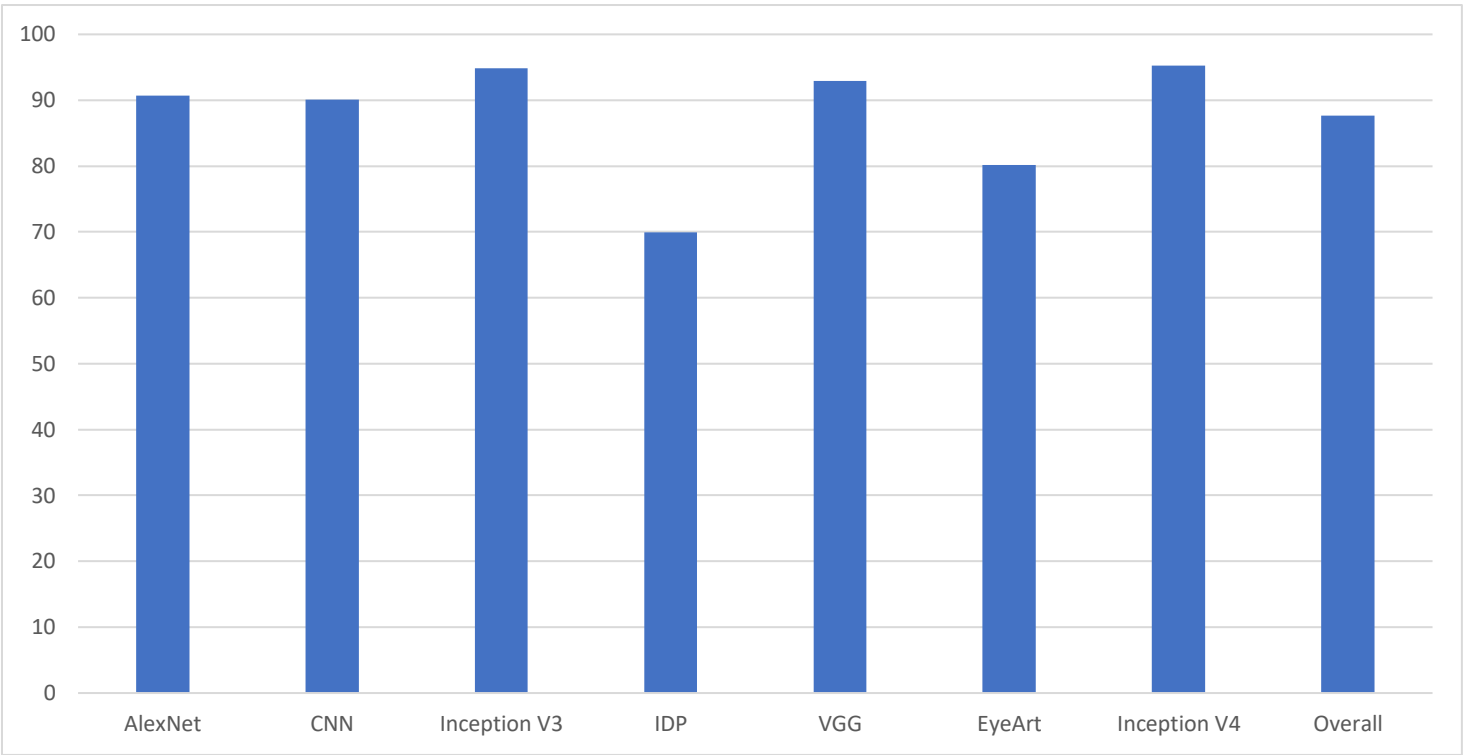
### 2.6.5   Forest plot analyses of sensitivity and specificity

Forest plot analyses were conducted to demonstrate reported sensitivity and specificity values from individual studies. Figures 11 and 12 display notable heterogeneity between studies assessing the performance of diverse AI systems in screening for diabetic retinopathy.

Figure 11. — Forest Plot for Reported Sensitivity Values

Study	Value
Abramoff 2018 (1)	$0.87 \pm .0216$
Gargeya 2017 (4)	$0.92 \pm .0198$
Gulshan 2016 (5)	$0.97 \pm .0096$
Hansen 2015 (6)	0.91 ± 0.036
Kanagasingam 2018 (7)	$1.00 \pm .0122$
Keel 2018 (8)	0.92 ± 0.0401
Kermany 2018 (9)	0.98 ± 0.0283
Li 2018 (10)	0.93 ± 0.0212
Li 2019 (11)	0.99 ± 0.0102
Lim 2019 (12)	0.91 ± 0.0311
Rajalakshmi 2018 (13)	0.96 ± 0.0171
Raju 2017 (14)	0.803 ± 0.009
Raumviboonsuk 2019 (15)	0.97 ± 0.0203
Sayres 2019 (16)	0.92 ± 0.0068
Son 2020 (17)	0.931 ± 0.013
Ting 2017 (18)	0.92 ± 0.0231
Torre 2019 (19)	0.91 ± 0.0080
Tufail 2016 (20)	0.94 ± 0.0261
Zeng 2019 (21)	0.82 ± 0.0134
Pooled Sensitivity	0.93 ± 0.0089

The squares and horizontal lines correspond to the study-specific sensitivity and 95% confidence intervals (CIs), respectively. The diamond represents the pooled sensitivity and 95% CI. The overall pooled sensitivity is 92.8% (95% CI: 91.9%-93.7%).

Figure 12. — Forest Plot for Reported Specificity Values
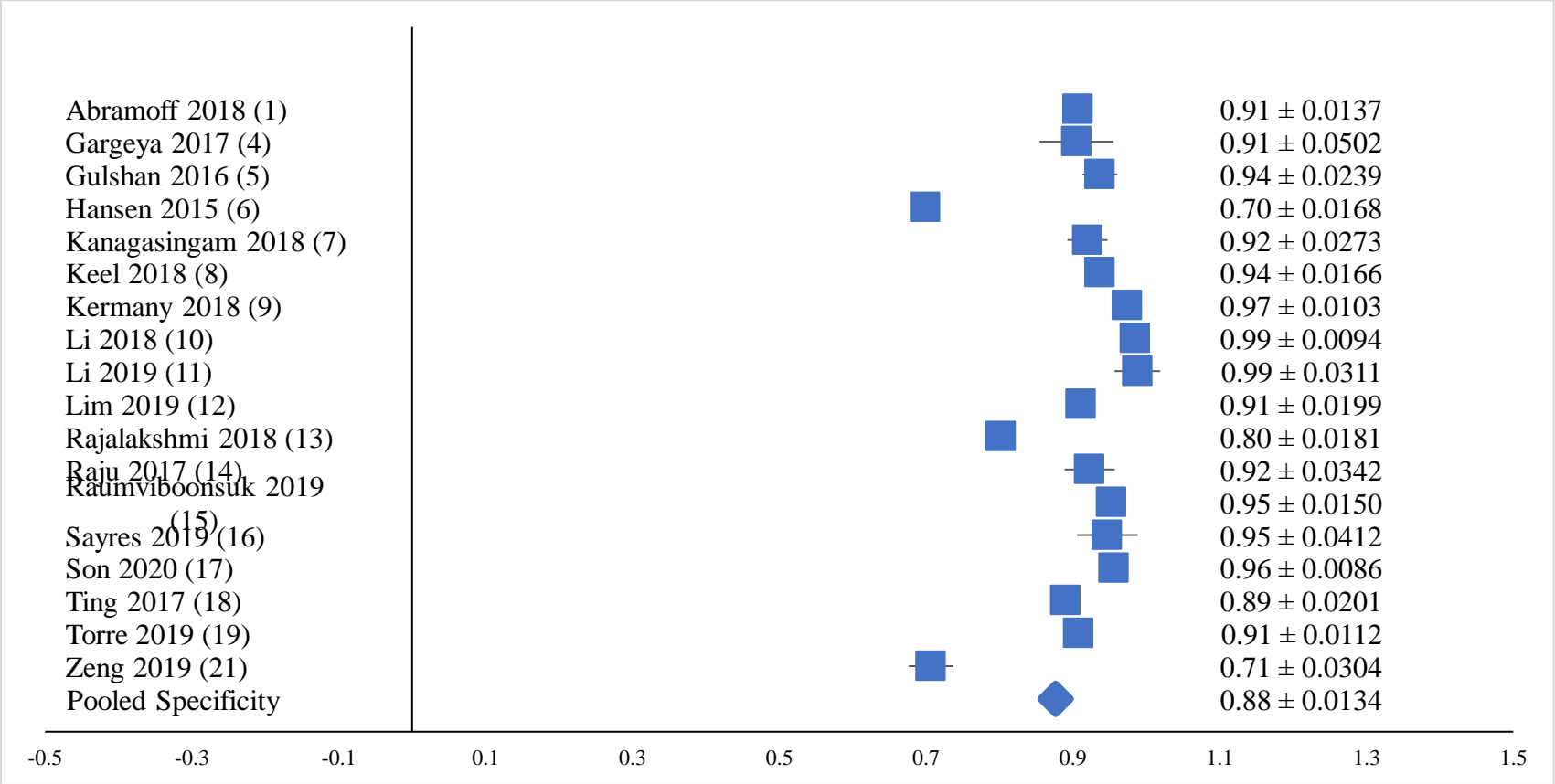
The squares and horizontal lines correspond to the study-specific specificity and 95% confidence intervals (CIs), respectively. The diamond represents the pooled specificity and 95% CI. The overall pooled specificity is 87.7% (95% CI: 86.4% to 89.0%).

2.6.6   Summary receiver operating characteristic (SROC) curve analysis

Figure 13 displays an SROC curve of the included studies. The dashed line indicates the 95% prediction region. This SROC curve shows the relationship between reported sensitivity and specificity values for each study. High sensitivity corresponds to a high negative predictive value and is the ideal factor of a "rule-out" test for diabetic retinopathy, while a high specificity corresponds to a high positive predictive value and is the ideal factor for a rule-in test.

Figure 13. — Summary receiver operating characteristics (SROC) curves of eligible studies
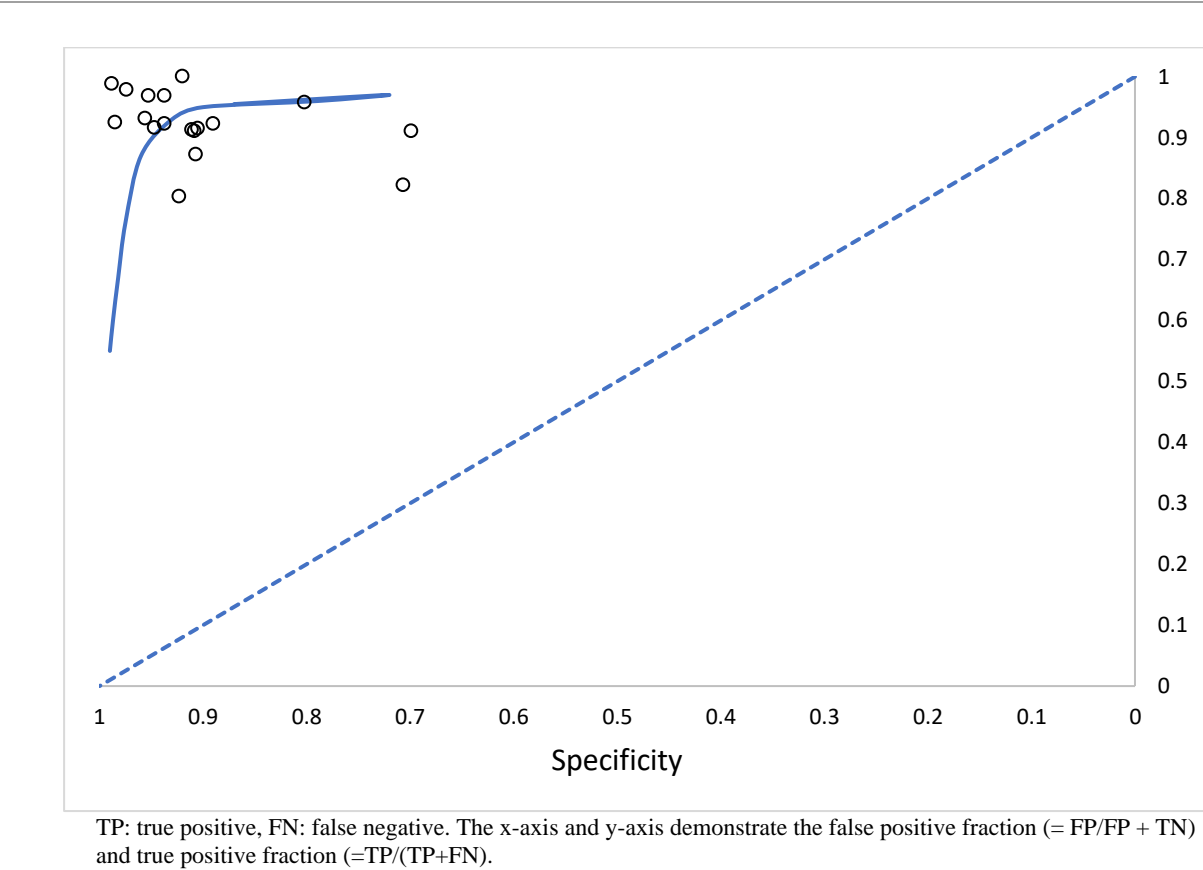
TP: true positive, FN: false negative. The x-axis and y-axis demonstrate the false positive fraction (= FP/FP + TN) and true positive fraction (=TP/(TP+FN)).

3. Discussion

To date, this is the largest systematic review and meta-analysis to assess the utility of neural networks for diabetic retinopathy screening. This study shows that the neural architecture method can correctly detect 92.8% (95% CI: 91.9% to 93.7%) of diabetes mellitus patients with referable diabetic retinopathy and exclude 87.7% (95% CI: 86.4% to 89.0%) of those without referable diabetic retinopathy. These results are superior to pooled sensitivity and specificity results reported in previous meta-analyses on AI screening system performance and surpasses the 80% acceptability threshold needed for AI screening systems to be applied in clinical settings. The sensitivity and specificity of six CNN models were compared in subgroup analyses. No significant differences in sensitivity were found among the included studies and all reported values were clinically acceptable. The mean specificity of IDP (71%) was reported to be of lower specificity than the 80% acceptance threshold, while all other AI screening systems demonstrated mean values of 80% or greater. The majority of studies recruiting diabetes mellitus patients from diverse backgrounds were conducted to assess the performance of their respective AI screening systems for clinical use, but there is a lack of studies that summarized those results quantitatively. This review provides quantitative evidence of the accuracy of such systems. Additionally, the results of this study showed that CNNs have great potential in clinical application as high screening accuracy was demonstrated amongst diverse neural architectures. Since it is quite expensive and time-consuming to develop and train algorithms with a large quantity of high-resolution and labeled retinal images, it would be cost-effective for future investigations to ascertain a gold standard for size of development and training sets and image resolution. The findings from these investigations could be particularly useful for using neural networks to detect rare diseases, of which there are only a limited number of cases. It is noteworthy that this meta-analysis did demonstrate that the subgroups with larger sample sizes showed higher performances in terms of sensitivity and specificity. However, further research is still crucial before concluding that the sample size of the training dataset influences sensitivity and specificity results. The screening accuracy of CNNs may not be affected by the criteria in experts' standards for screening diabetic retinopathy, which is reasonable because ICDRS was developed using ETDRS. However, because ICDRS is easier and more commonly used in clinical settings, it may be preferable to consider ICDRS criteria when developing, testing, and validating automated screening systems, though ETDRS is still treated as the gold standard.

With regard to imaging modalities, the majority of studies that have and currently investigate AI screening performance use fundus photography as the modality of choice and a smaller number utilize OCT imaging. Fundus photography produces a two-dimensional image of the three-dimensional structure of the retina, while OCT captures the cross-sectional axial of the retina through light coherence. The main shortcoming of using fundus photography for screening is that it only produces two-dimensional images, while the structure of the retina is three-dimensional. For this reason, it is advantageous to use OCT for diabetic retinopathy visualization as OCT can produce clear three-dimensional images of thick samples by rejecting background signals while collecting light directly reflected from retinal surfaces. Thus, researchers should focus on training algorithms to interpret OCT images in order to better screen for diabetic retinopathy. Alternatively, training neural networks to recognize and analyze both OCT and fundus images might lead to higher screening performance.

A comprehensive literature search was conducted in global health and biomedical databases. High quality studies that met a specific set of eligibility criteria were included in this study. Performance metrics of sensitivity and specificity were meta-analyzed to assess

the performance of diverse AI screening systems. The heterogeneity of results was also graphically demonstrated through forest plot and SROC curve analyses.

3.1 Limitations for review

This review possesses several limitations that must be considered. First, in the subgroup analyses assessing the sensitivity and specificity of AI screening systems, only one study contributed results to the AlexNet, IDP, and Inception V4 subgroups. Reporting performance levels based on a limited number of studies weakens the credibility of this study's meta-analytic findings. Second, several studies used the same datasets for training and validation of their respective AI screening systems. Assessing the impact of using overlapping data sources on the outcome of the results is challenging to evaluate because the contents of each dataset was not discussed in the included studies. Third, CNNs lack standardized cut-off points or thresholds with which to designate the severity of diabetic retinopathy. Due to such an absence, this review could not make strong comparisons between their abilities to screen for disease severity. Fourth, there was a strong risk of selection bias amongst the included studies with regard to participant recruitment. It is unclear whether participant data was included in multiple datasets, so the overall performance metrics may be underestimated or overestimated due to potential changes in said participants' severities of diabetic retinopathy between data sources.

4. Conclusions

This review and meta-analysis demonstrates clinically acceptable performances from the majority AI systems used in diabetic retinopathy screening. Although the majority of neural networks showed clinically acceptable performance levels, further improvement depends on the continual development of novel algorithms with large and gradable sets of images for training and validation. With the rapidly growing global burden of diabetic retinopathy, AI screening systems can increase the ability for disease prevention by allowing for early detection. If cost-effectiveness ratios can be optimized, AI can become a financially sustainable and clinically effective intervention that can be incorporated into the healthcare systems of LMICs and geographically remote locations. AI screening can increase the efficiency of eye care and diabetes services and optimize the care of patients within healthcare systems that provide large-scale services on a population level. Combining screening technologies with treatment interventions such as anti-VEGF therapy, acellular capillary laser treatment, and vitreoretinal surgery can lead to substantial reductions in the incidence of irreversible vision-loss due to proliferative diabetic retinopathy. With further advancement, AI will inform and improve primary, secondary, and tertiary care settings' approaches to diabetic retinopathy management.

5. References

1. Abràmoff, Michael D. "Pivotal Trial of an Autonomous AI-Based Diagnostic System for Detection of Diabetic Retinopathy in Primary Care Offices." *Digital Medicine*, Nature, 28 Aug. 2018.
2. Bellemo, Valentina. "JavaScript Needs to Be Enabled for This Application to Run Correctly." *Elsevier Enhanced Reader*, The Lancet, May 2019.
3. Fauw, Jeffrey De. "Clinically Applicable Deep Learning for Diagnosis and Referral in Retinal Disease." *Nature*, Nature Medicine, 13 Aug. 2018.
4. Gargeya, Rishab. "Automated Identification of Diabetic Retinopathy Using Deep Learning." *ScienceDirect*, American Academy of Ophthalmology, July 2017.
5. Gulshan, Varun. "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs." *EndNote Click*, The Journal of the American Medical Association, 29 Nov. 2016.
6. Hansen, M. B. "Results of AUTOMATED Retinal Image Analysis for Detection of Diabetic RETINOPATHY from THE Nakuru STUDY, KENYA." *EndNote Click*, PLOS One, 1 Oct. 2015.
7. Kanagasingam, Yogesan. "Evaluation of Artificial Intelligence–Based Grading of Diabetic Retinopathy in Primary Care." *JAMA Network*, The Journal of the American Medical Association, 28 Sept. 2018.
8. Keel, Stuart. "Feasibility and Patient Acceptability of a Novel Artificial Intelligence-Based Screening Model for Diabetic Retinopathy at Endocrinology Outpatient Services: a Pilot Study." *Nature Research*, Scientific Reports, 12 Mar. 2018.
9. Kermany, Daniel S. "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning." *ScienceDirect*, Cell, 22 Feb. 2018.
10. Li, Zhixi. "An Automated Grading System for Detection of Vision-Threatening Referable Diabetic Retinopathy on the Basis of Color Fundus Photographs." *Diabetes Journal*, Diabetes Care, Dec. 2018.
11. Li, Feng. "Fully Automated Detection of Retinal Disorders by Image-Based Deep Learning." *National Library of Medicine*, Clinical and Experimental Ophthalmology, 4 Jan. 2019.
12. Lim, Jennifer. "Artificial Intelligence Screening for Diabetic Retinopathy: Analysis from a Pivotal Multi-Center Prospective Clinical Trial." *Eyenuk*, EyeArt, 2019.
13. Rajalakshmi, Ramachandran. "Automated Diabetic Retinopathy Detection in Smartphone-Based Fundus Photography Using Artificial Intelligence." *Nature*, Eye, 9 Mar. 2018.
14. Raju, Manoj. "Development of a Deep Learning Algorithm for Automatic Diagnosis of Diabetic Retinopathy." *Scopus*, IOS Press, 21 Aug. 2017.
15. Ruamviboonsuk, Paisan. "Deep Learning versus Human Graders for Classifying Diabetic Retinopathy Severity in a Nationwide Screening Program." *Nature*, Npj Digitial Medicine, 10 Apr. 2019.
16. Sayres, Rory. "Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy." *ScienceDirect*, American Academy of Ophthalmology, Apr. 2019.
17. Son, Jaemin. "Development and Validation of Deep Learning Models for Screening Multiple Abnormal Findings in Retinal Fundus Images." *ScienceDirect*, American Academy of Ophthalmology, 1 Jan. 2020.
18. Ting, Daniel Shu Wei. "Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From

Multiethnic Populations With Diabetes." *JAMA Network*, The Journal of the American Medical Association, 12 Dec. 2017.

19. Torre, Jordi. "A Deep Learning Interpretable Classifier for Diabetic Retinopathy Disease Grading." *ScienceDirect*, Neurocomputing, 5 July 2020.

20. Tufail, Adnan. "Automated Diabetic Retinopathy Image Assessment Software: Diagnostic Accuracy and Cost-Effectiveness Compared with Human Graders." *ScienceDirect*, American Academy of Ophthalmology, Mar. 2017.

21. Zeng, Xianglong. "Automated Diabetic Retinopathy Detection Based on Binocular Siamese-Like Convolutional Neural Network." *IEEE Xplore*, IEEE Access, 5 Mar. 2019.

22. "Diabetes Key Facts." *World Health Organization*, World Health Organization, 13 Apr. 2021.

23. Lee, Ryan. "Epidemiology of Diabetic Retinopathy, Diabetic Macular Edema and Related Vision Loss." *BMC*, Eye and Vision, 2015.

24. Bain, Stephen C. "Worsening of Diabetic Retinopathy with Rapid Improvement in Systemic Glucose Control: A Review." *Wiley*, Diabetes, Obesity & Metabolism, 22 June 2018.

25. "ICO Guidelines for Diabetic Eye Care." *International Council of Ophthalmology*, International Council of Ophthalmology, 2017.

26. Virgili, Gianni. "Anti-Vascular Endothelial Growth Factor (Anti-VEGF) Drugs for Diabetic Macular Oedema." *Cochrane*, Cochrane, 16 Oct. 2018.

27. Wong, Tien Yin. "Strategies to Tackle the Global Burden of Diabetic Retinopathy: From Epidemiology to Artificial Intelligence." *Karger*, Ophthalmologica, 13 Aug. 2019.

28. "Prevention: Diabetic Retinopathy." *NHS Choices*, NHS, 30 Oct. 2018.

29. Khandekar, Rajiv. "Screening and Public Health Strategies for Diabetic Retinopathy in the Eastern Mediterranean Region." *NCBI*, Middle East African Journal of Ophthalmology, June 2012.

30. Bora, Ashish. "Predicting the Risk of Developing Diabetic Retinopathy Using Deep Learning." *The Lancet*, The Lancet Digital Health, 26 Nov. 2020.

31. Bascaran, Covadonga. "Effectiveness of Task-Shifting for the Detection of Diabetic Retinopathy in Low- and Middle-Income Countries: a Rapid Review Protocol." *Springer Nature*, BMC, 4 Jan. 2021.

32. Oh, Kangrok. "Early Detection of Diabetic Retinopathy Based on Deep Learning and Ultra-Wide-Field Fundus Images." *Nature*, Scientific Reports, 21 Jan. 2021.

33. Zapata, Miguel Angel. "Artificial Intelligence to Identify Retinal Fundus Images, Quality Validation, Laterality Evaluation, Macular Degeneration, and Suspected GlaucomaMi." *Dovepress*, Clinical Ophthalmology, 13 Feb. 2020.

34. Heydon, Peter. "Prospective Evaluation of an Artificial Intelligence-Enabled Algorithm for Automated Diabetic Retinopathy Screening of 30 000 Patients." *BMJ*, British Journal of Ophthalmology, 30 June 2020.

35. Grzybowski, Andrzej. "Artificial Intelligence for Diabetic Retinopathy Screening: a Review." *NCBI*, Eye, 5 Sept. 2019.

36. Xie, Qiwei. "An Innovative Method for Screening and Evaluating the Degree of Diabetic Retinopathy and Drug Treatment Based on Artificial Intelligence Algorithms." *ScienceDirect*, Pharmacological Research, 2 June 2020.

37. Wang, Yue-Lin. "Progress of Artificial Intelligence in Diabetic Retinopathy Screening." *Wiley*, Diabetes/Metabolism Research and Reviews, 3 Oct. 2020.

38. Lim, Gilbert. "Different Fundus Imaging Modalities and Technical Factors in AI Screening for Diabetic Retinopathy: a Review." *Springer Link*, Eye and Vision, 14 Apr. 2020.

39. Padhy, Srikanta Kumar. "Artificial Intelligence in Diabetic Retinopathy: A Natural Step to the Future." *NCBI*, Indian Journal of Ophthalmology, 13 Mar. 2019.

40. Arbelaez, Maria Clara. "Clinical Outcomes of Laser in Situ Keratomileusis with an Aberration-Neutral Profile Centered on the Corneal Vertex Comparing Vector Planning with Manifest Refraction Planning for the Treatment of Myopic Astigmatism." *ScienceDirect*, Journal of Cataract & Refractive Surgery, Dec. 2017.

41. Nguyen, Hai. "Cost-Effectiveness of a NationalTelemedicine Diabetic RetinopathyScreening Program in Singapore." *ScienceDirect*, American Academy of Ophthalmology, Dec. 2016.

42. Lechner, Judith. "The Pathology Associated with Diabetic Retinopathy." *ScienceDirect*, Vision Research, 29 Apr. 2017.

43. Axer-Siegel, Ruth. "Diabetic Retinopathy during Pregnancy." *ScienceDirect*, Ophthalmology, 7 Oct. 2017.

44. Jensen, Ralph J. "Ganglion Cells and (Dye-Coupled) Amacrine Cells in the Turtle Retina That Have Possible Synaptic Connection." *ScienceDirect*, Brain Research, 11 Mar. 2003.

45. Barber, A J. "Neural Apoptosis in the Retina during Experimental and Human Diabetes. Early Onset and Effect of Insulin." *The Journal of Clinical Investigation*, The Journal of Clinical Investigation, 15 Aug. 1998.

46. Watkins, Peter J. "Retinopathy." *BMJ*, Clinical Review, 26 Apr. 2003.

47. Patel, Vinod. "Retinal Blood Flow in Diabetic Retinopathy." *BMJ*, BMJ, 19 Sept. 1992.

48. Toh, Huishi. "Vascular Changes in Diabetic Retinopathy—a Longitudinal Study in the Nile Rat." *Nature*, Laboratory Investigation , 17 May 2019.

49. Ishibazawa, Akihiro. "Optical Coherence Tomography Angiography in Diabetic Retinopathy: A Prospective Pilot Study." *ScienceDirect*, American Journal of Ophthalmology, July 2015.

50. Hwang, Thomas S. "Optical Coherence Tomography Angiography Features of Diabetic Retinopathy." *NCBI*, Retina, Nov. 2015.

51. Williams, George A. "Single-Field Fundus Photography for Diabetic Retinopathy Screening: A Report by the American Academy of Ophthalmology." *ScienceDirect*, Ophthalmology, May 2004.

52. Riva, Charles E. "Fundus Camera Based Retinal LDV." *OSA Publishing*, Applied Optics, 1 Jan. 1981.

53. Abràmoff, Michael D. "Messidor-2." *ADCIS*, ADCIS, Mar. 2013.

54. Solanki, Kaushal. "EyeArt: Automated, High-Throughput, Image Analysis for Diabetic Retinopathy Screening." *ARVO*, Investigative Ophthalmology & Visual Science, June 2015.

55. "EyePACS Data Analysis." *EyePACS*, EyePACS.

56. Graham, Ben. "Kaggle Diabetic Retinopathy Detection Competition Report." *SparseConvNet*, SparseConvNet, 6 Aug. 2015.

57. Quellec, Gwenolé. "Deep Image Mining for Diabetic Retinopathy Screening." *ScienceDirect*, Medical Image Analysis, July 2017.

58. Sambyal, Nitigya. "Modified U-Net Architecture for Semantic Segmentation of Diabetic Retinopathy Images." *ScienceDirect*, Biocybernetics and Biomedical Engineering, 8 June 2020.

59. Amin, Javeria. "A Method for the Detection and Classification of Diabetic Retinopathy Using Structural Predictors of Bright Lesions." *ScienceDirect*, Journal of Computational Science, 15 Jan. 2017.

60. Albawi, Saad. "Understanding of a Convolutional Neural Network." *IEEE Xplore*, IEEE, 8 Mar. 2018.

61. Shaban, Mohamed. "A Convolutional Neural Network for the Screening and Staging of Diabetic Retinopathy." *PLOS One*, PLOS One, 22 June 2020.

62. Pratt, Harry. "Convolutional Neural Networks for Diabetic Retinopathy." *ScienceDirect*, Procedia Computer Science, 8 July 2016.

63. Tan, Jen Hong. "Segmentation of Optic Disc, Fovea and Retinal Vasculature Using a Single Convolutional Neural Network." *ScienceDirect*, Journal of Computational Science, 27 Feb. 2017.

64. Radu, Valentin. "Performance Aware Convolutional Neural Network Channel Pruning for Embedded GPUs." *IEEE Xplore*, IEEE, 19 Mar. 2020.

65. Li, Hongyang. "Multi-Bias Non-Linear Activation in Deep Neural Networks." *PMLR*, Proceedings of Machine Learning Research, 2016.

66. Lin, Guifang. "Research on Convolutional Neural Network Based on Improved Relu Piecewise Activation Function." *ScienceDirect*, Procedia Computer S, 11 May 2018.

67. Xu, Bing. "Empirical Evaluation of Rectified Activations in Convolutional Network." *ArXiv*, Machine Learning, 27 Nov. 2015.

68. Li, Hongsheng. "Highly Efficient Forward and Backward Propagation of Convolutional Neural Networks for Pixelwise Classification." *ArXiv*, Computer Vision and Pattern Recognition, 16 Dec. 2014.

69. Hemanth, Jude D. "A Modified Deep Convolutional Neural Network for Abnormal Brain Image Classification." *IEEE Xplore*, IEEE, 10 Dec. 2018.

70. Sadhana, S. "An Intelligent Technique for Detection of Diabetic Retinopathy Using Improved Alexnet Model Based Convoluitonal Neural Network." *IOS Press*, Journal of Intelligent & Fuzzy Systems, 12 Apr. 2021.

71. Khalifa, Nour Eldeen M. "Deep Transfer Learning Models for Medical Diabetic Retinopathy Detection." *NCBI*, Acta Informatica Medica, 27 Dec. 2019.

72. Yamashita, Rikiya. "Convolutional Neural Networks: an Overview and Application in Radiology." *Springer Open*, Insights into Imaging, 22 June 2018.

73. Mohammadian, Saboora. "Comparative Study of Fine-Tuning of Pre-Trained Convolutional Neural Networks for Diabetic Retinopathy Screening." *IEEE Xplore*, IEEE, 9 Aug. 2018.

74. Galdran, Adrian. "Residual Networks for Pulmonary Nodule Segmentation and Texture Characterization." *Springer Link*, Lecture Notes in Computer Science, 17 June 2020.

75. Saxena, Gaurav. "Improved and Robust Deep Learning Agent for Preliminary Detection of Diabetic Retinopathy Using Public Datasets." *ScienceDirect*, Intelligence-Based Medicine, Dec. 2020.

76. Abràmoff, Michael D. "Improved Automated Detection of Diabetic Retinopathy on a Publicly Available Dataset Through Integration of Deep Learning." *ARVO*, Investigative Ophthalmology & Visual Science, Oct. 2016.

77. Abràmoff, Michael D. "Automated Analysis of Retinal Images for Detection of Referable Diabetic Retinopathy." *JAMA Network*, JAMA Ophthalmology, Mar. 2013.

78. Mateen, Muhammad. "Fundus Image Classification Using VGG Architecture with PCA and SVD." *MDPI*, Symmetry, 20 Dec. 2018.

79. Sun, Yunlei. "The Neural Network of One-Dimensional Convolution-An Example of the Diagnosis of Diabetic Retinopathy." *IEEE Xplore*, IEEE, 15 May 2019.

80. Bhaskaranand, Malavika. "The Value of Automated Diabetic Retinopathy Screening with the EyeArt System: A Study of More Than 100,000 Consecutive Encounters from People with Diabetes." *Liebert Publishers*, Diabetes, Technology, & Therapeutics, 21 Oct. 2019.

81. Olvera-Barrios, Abraham. "Diagnostic Accuracy of Diabetic Retinopathy Grading by an Artificial Intelligence-Enabled Algorithm Compared with a Human Standard for Wide-Field True-Colour Confocal Scanning and Standard Digital Retinal Images." *BMJ Journals*, British Journal of Ophthalmology, 21 Jan. 2021.

82. Colagiuri, Stephen. "Glycemic Thresholds for Diabetes-Specific Retinopathy Implications for Diagnostic Criteria for Diabetes." *American Diabetes Association: Diabetes Care*, Pathophysiology/Complications, Jan. 2011.

83. Curtis, T M. "Microvascular Lesions of Diabetic Retinopathy: Clues towards Understanding Pathogenesis?" *Nature*, Eye, 15 May 2009.

84. Soria, B. "From Stem Cells to Beta Cells: New Strategies in Cell Therapy of Diabetes Mellitus." *SpringerLink*, Diabetologia, Apr. 2001.

85. Arevalo, Fernando. "Classification of Diabetic Retinopathy and Diabetic Macular Edema." *NCBI*, World Journal of Diabetes, 15 Dec. 2013.

86. Niemeijer, Meindert. "Automated Detection and Differentiation of Drusen, Exudates, and Cotton-Wool Spots in Digital Color Fundus Photographs for Diabetic Retinopathy Diagnosis." *ARVO*, Investigative Ophthalmology & Visual Science, May 2007.

87. Silva, Paolo S. "Hemorrhage and/or Microaneurysm Severity and Count in Ultrawide Field Images and Early Treatment Diabetic Retinopathy Study Photography." *ScienceDirect*, American Academy of Ophthalmology, 20 Mar. 2017.

88. O'Mahoney, Paul R. A. "Retinal Vein Occlusion and Traditional Risk Factors for Atherosclerosis." *JAMA Network*, Epidemiology, May 2008.

89. Wong, Tien Yin. "Artificial Intelligence With Deep Learning Technology Looks Into Diabetic Retinopathy Screening." *JAMA Network*, American Medical Association, 13 Dec. 2016.

90. Ramachandran, Nishanthan. "Diabetic Retinopathy Screening Using Deep Neural Network." *Wiley*, Clinical & Experimental Ophthalmology , 7 Sept. 2017.

91. Seoud, Lama. "Red Lesion Detection Using Dynamic Shape Features for Diabetic Retinopathy Screening." *IEEE Xplore*, IEEE, Apr. 2016.