Deep Learning for Ultra-Widefield Imaging: A Scoping Review

Nishaant Bhambra
McGill University https://orcid.org/0000-0003-4966-6578
Fares Antaki
Université de Montréal
Farida El Malt
McGill University
AnQi Xu
Université de Montréal
Renaud Duval (renaud.duval@umontreal.ca)
Université de Montréal https://orcid.org/0000-0002-3845-3318

Research Article

Keywords: ultra-widefield imaging, deep learning, machine learning, artificial intelligence, quality assessment

Posted Date: October 27th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-992393/v1

Abstract

Purpose: This article is a scoping review of published and peer-reviewed articles using deep-learning (DL) applied to ultra-widefield (UWF) imaging. This study provides an overview of the published uses of DL and UWF imaging for the detection of ophthalmic and systemic diseases, generative image synthesis, quality assessment of images, and segmentation and localization of ophthalmic image features.

Methods: A literature search was performed up to August 31st, 20201 using PubMed, Embase, Cochrane Library, and Google Scholar. The inclusion criteria were as follows: (1) Deep Learning, (2) Ultra-Widefield Imaging. The exclusion criteria were as follows: (1) articles published in any language other than English, (2) articles not peer-reviewed (usually preprints) (3) no full-text availability (4) articles using machine learning algorithms other than deep learning. No study design was excluded from consideration.

Results: A total of 36 studies were included. 23 studies discussed ophthalmic disease detection and classification, 5 discussed segmentation and localization of UWF images, 3 discussed generative image synthesis, 3 discussed ophthalmic image quality assessment, and 2 discussed detecting systemic diseases via UWFI.

Conclusion: The application of DL to UWFI has demonstrated significant effectiveness in the diagnosis and detection of ophthalmic diseases including diabetic retinopathy, retinal detachment, and glaucoma. DL has been used with UWFI to also diagnose systemic diseases like Alzheimer's, and also applied in the generation of synthetic ophthalmic images. This scoping review highlights and discusses the current uses of DL with UWFI, and the future of DL applications in this field.

Key Messages

- **What is known:**
  - Deep learning has been applied to ophthalmic images for the diagnosis of ophthalmic diseases using fundus photography
  - The utility of deep learning applied to ultra-widefield imaging has not been summarized in a scoping review
- **New information:**
  - A scoping review of all published and peer-reviewed articles using deep learning with ultra-widefield imaging until August 2021 is provided here
  - This article discusses both the detection of ophthalmic diseases, as well as the use of deep learning technologies for the generation of new ophthalmic images, the detection of systemic diseases like Alzheimer's, and quality assessment of ophthalmic images.

Introduction

In 1926, the first fundus camera was introduced by Zeiss and Nordensen. At that time, the camera provided only a 20-degree field of view. Shortly thereafter, an improved camera provided practitioners with a 30-degree field of view of the fundus.[1]While a major advance at the time, these cameras provided ophthalmologists with a limited view of the retinal periphery. In 1981, the Diabetic Retinopathy Study provided an objective method to visualize up to 75-degrees of the retina by combining seven conventional 30-degree fundus images.[2, 3] This image type, known as 7 Standard Field (7SF) imaging, was the gold standard used in imaging for diagnosing diabetic retinopathy. This remained the gold standard until technical developments in widefield imaging (WFI) and ultra-widefield (UWF) imaging.

Widefield and Ultra-Widefield Imaging

WFI, referring to imaging providing over 50 degrees of field of view, was developed to see the periphery of the retina more widely. WFI utilizes a scanning laser ophthalmoscope (SLO), which separates the illuminating and imaging lasers used.[4] By separating the beams, WFI reduces artifacts produced from the interfaces in the ocular media.[4]

UWF imaging was introduced in 2000 by Optos®.[5] It specifically refers to imaging providing a greater field of view than WFI. UWF imaging can provide up to a 200-degree view of the retina, which allows for visualization of the optic disk and the peripheral retina in the same view.

Multiple WFI systems are available, with each differing in their technology and their field of view. Optos (Optos Inc, Dunfermline, UK) is a UWF imaging system that provides a capture of 200 degrees of the retina in a single image. The image provides coverage of approximately 82% of the retinal surface and does so without direct patient contact.[6]

Clinical Utility of UWF systems

As UWF imaging has provided a broader view of the retina, it has consistently been more effective at diagnosing retinal disease than previous imaging modalities. Ultra-widefield fluorescein angiography (UWF-FA), which combines UWF imaging with fluorescein angiography to visualize vessels, has been significantly more effective in diagnosing diabetic retinopathy (DR) than previous 7SF imaging.[7] UWF images (UWFIs) provide 3.2 times more retinal surface area than 7SF and allow for a more comprehensive assessment of peripheral lesions and nonperfusion in DR.[8, 9]

In retinal detachment, UWF imaging has provided improved assessment of peripheral retina breaks in comparison to indirect ophthalmoscopy.[10] In glaucoma, UWF imaging has been shown to have a high agreement with colour digital stereoscopy (CDS) in evaluating vertical cup-to-disc ratio and may be as effective in diagnosing glaucoma as CDS.[11] In patients with age-related macular degeneration (AMD), it was found that peripheral retina changes were

highly prevalent, indicating UWF imaging's greater value in diagnosing AMD than traditional fundoscopy.[12] From these findings, UWF imaging may be a window to the retina, as well as to the brain more broadly.

Machine Learning and the Different Types of Learning

Machine learning (ML) refers to the ability of machines to generate associations and patterns between variables, to learn in a sense similar to humans. By simulating the neural networks of human brains, ML networks generate probabilities and associations between variables to emulate human intelligence.[13] ML algorithms often can draw inferences between variables that are either imperceptible to humans or are too complex for human associations.[13] ML is divided into categories based on the approaches taken to assist computers in learning, on a spectrum between supervised learning and unsupervised learning.[14]

Supervised learning refers to ML from human-provided input and output pairs. For example, supervised learning for a task of classifying images would require a set of labelled images with their corresponding classification. By training a model on images and image classification, machines can learn to infer relationships between the two. The trained model should then be able to take unlabelled input data and determine its classification.[15] While this method is the most effective at training these associations, it also requires the most human involvement for training.

Between supervised and unsupervised learning lies semi-supervised learning, which refers to ML training from incompletely labelled training datasets. This approach provides the machine with an initial relationship between input and output data, without a fully labelled set.[16]

Unsupervised learning uses algorithms to learn patterns from data that lacks human labels and input. In a set of images, unsupervised learning hopes to learn the features that associate with different types of images.[17] For example, in training a model to classify images using an unsupervised approach, a successful unsupervised learning algorithm would determine features that correspond to a given cluster and categorize each into separate categories without input labels from humans.

Building a Deep Learning Model

Deep learning (DL) is a subset of machine learning that uses multiple layers of learning to identify features in data.[18] For example, in processing a fundus image, a lower layer may identify the edges of the vasculature, while higher layers may then utilize these edges of the vasculature in context to identify vessels as larger objects.

DL specifically associates variables along nodes in a computational neural network. By associating data along these nodes, the artificial neural network (ANN) assigns a positive weight to variables with positive correlations and negative weights to variables with negative correlations. These weights determine the contributory strength of an input variable to the outcome of the neural network.[19] This develops a network of probabilistic associations between input variables. This is analogous to biological neurons, where associations between neurons are strengthened or weakened with excitatory and inhibitory stimuli respectively.[20] By associating data along these ANNs, machines can learn, and train models based on input data. By associating features of the data across the nodes of the ANN, correlations between the data are strengthened or weakened. These connections between the nodes, known as *edges*, are analogous to the synapses in biological brains.[21]

In DL, nodes are associated into multiple layers. Each layer contains a set of nodes, and often perform different transformations on the input data. Each neural network contains an *input layer*, where the data enters untransformed, and an *output layer*, which produces the learned result. Between each is zero to multiple *hidden layers*, where further learning of data features occurs. Input data is processed forward from the input layer until it reaches the output layer.[22]

Convolutional Neural Networks (CNNs) build on ANNs by organizing data and nodes in three dimensions. Furthermore, CNNs separate feature extraction and classification into distinct layers. CNNs rely on a convolution layer, which performs a convolution operation on the data array or tensor. The convolution operation extracts high-level features from a data source, such as the edges of an image. By doing so, it reduces the spatial size of the data and flexibly adjusts to the features of the data that are deemed more important to higher-level processing.[23] For this reason, CNNs are especially useful in image processing, where their convolutional operation allows them to ignore noise and focus on higher-order image structures, like edges. Multiple CNN models exist, including LeNet, AlexNet, VGGNet, and InceptionResNetV2[24–27].

Training, Validation, and Testing

Datasets are split into training, validation, and testing sets such that each respective step has data that is similar to a model's intended input data. When datasets include more than one image per patient (right eye, left eye, and steered images), it is important to maintain patient-level splits ensuring that each image could be used either for training, validation, or testing but not all. This restriction eliminates the risk of train-test data contamination that can arise when DL models use non-clinically relevant patterns to drive predictions.

The training dataset is used to train the model to learn the weights and biases between dataset variables. This training dataset is the input data for the model to learn from. For this reason, the quality and quantity of this data will greatly impact the ability of the model to learn the features of interest.[28]

Validation datasets are used to tune the hyperparameters of the model. Hyperparameters are parameters that control the learning process of the model, while parameters are the node weights that are derived from training the model. The results of the validation set are used by the engineer to determine the optimal hyperparameters for the learning process. For this reason, this dataset is also known as the "development" dataset. The model does not learn from this dataset and does not develop weights or biases that would alter the model in an automated sense.[28]

Finally, the test dataset is used as the input data for evaluating the model. This dataset is used to determine the accuracy and effectiveness of the model. This test dataset is not used to adjust the model, nor does the model learn from it.[28]

The ratio of the datasetlocated to each subset depends on the goals of the model being trained and evaluated. In models that require many hyperparameters to be adjusted, a larger validation set is recommended. However, a validation set is optional if the user does not intend to tune these hyperparameters. Similarly, the amount of data to be allocated towards the training set depends on the complexity of the data and the amount of learning needed.[28]

Another approach to dividing the dataset is k-fold cross-validation. In this process, image data is divided into k groups, while k-1 groups are used as training data and 1 group is used as validation data. This repeats until each dataset becomes a validation dataset.

Image Preprocessing

Image preprocessing serves multiple purposes when training a DL model on an imaging task. The first purpose is to conserve computational resources by resizing images.[29] Often, large images (i.e. the standard 3900x3072 pixels of UWFIs) will be resized to significantly smaller, lower-resolution images (i.e. 227x227 pixels).[30]

The second purpose of image pre-processing is to increase the data available for training the model and to train the model on generalized cases. This is done via data augmentation, which increases the size of the data by performing transformations on it and producing new combinations of data to train on.[31] This provides the training set data that may match altered or changed input data. For example, one could augment an image dataset by adding noise to the images. This would help the model learn to classify noisy images with features of a given label correctly. As the goal is to train these models to be useful on real-world data, the training data should contain the same “errors” or adjustments that imperfect real-world input data does. This then allows for the model to become more robust in detecting the features of interest. Common image augmentation methods include adjusting brightness, gamma correction, histogram equalization, noise addition, and inversion.[32] Data augmentation can increase the size of the training set manyfold, often up to five or eighteen times the original dataset size.[32, 33]

Evaluating the Model

In ML classification tasks, predictions are classified into true positives, true negatives, false positives, and false negatives. From these values, the sensitivity and specificity values of the model are calculated. Sensitivity refers to the ability of the model to correctly predict positive observations, calculated as the number of true positives divided by the sum of the true positives and false negatives. Specificity refers to the ability to reject classifications for cases that do not fit the condition, calculated as the number of true negatives divided by the sum of the true negatives and number of false positives.[34]

Sensitivity and specificity depend on the thresholds used for detection. As a threshold for detection increases, sensitivity (predicting positive outcomes) decreases while specificity (rejecting positive outcomes) increases. By plotting the sensitivity or true positive rate as a function of the specificity or false positive rate, the receiver operating characteristic (ROC) curve is produced. The integral of the entire curve is referred to as the Area Under ROC curve (AUROC). The AUROC value serves as a measure of the model to correctly classify and predict based on the test data. A perfectly predictive model would score 1 while a perfectly inaccurate test would score 0.[35] The area under the precision-recall curve (AUPRC) is occasionally used and is the integral of the plot of positive predictive value (precision) as a function of the sensitivity (recall).[36]

Deep Learning In Ultrawide Field Imaging

Methodology

A literature search was performed up to August 31st, 2021 using the following online databases: PubMed, Embase, Cochrane Library, and Google Scholar. Article screening was done by the senior author (RD). The inclusion criteria were as follows: (1) Deep Learning (2) Ultra-Widefield Imaging. The exclusion criteria were as follows: (1) articles published in any language other than English, (2) articles not peer-reviewed (usually preprints) (3) no full-text availability (4) articles using machine learning algorithms other than deep learning. No study design was excluded from consideration. The detailed search methodology was as follows: (“deep learning” OR "artificial intelligence" OR "machine learning") AND ("Ultra-Widefield" OR "UWF" OR "UWFI" OR "Optos").

A total of 36 studies were included. A full listing of included studies, authors, and their respective digital object identifiers (DOIs) are listed in Table 1. A full listing of included studies and their respective architectures, datasets, and experimental results are listed in Table 2. A chart detailing the number of included publications by year is included in Figure 2. A map highlighting the number of publications by country is provided in Figure 3.

Disease Detection and Classification

Disease detection and classification have been the most thoroughly investigated uses for UWF imaging with DL. Specifically, DL has been used for disease detection and classification of diabetic retinopathy (DR), retinal detachment (RD), glaucoma, age-related macular degeneration (AMD), retinitis pigmentosa (RP), pachychoroid, retinal vein occlusion (RVO), idiopathic macular hole (IMH), retinal hemorrhage, and sickle cell retinopathy (SCR).

Diabetic Retinopathy

At the time of writing, six published peer-reviewed articles have explored DL with UWF imaging in DR patients.[32, 37–42] Five have specifically used UWFIs for the detection and classification of diabetic retinopathy.[32, 38–40, 42]

Wang *et al.* first used UWFIs to train a DL model for the detection of referrable diabetic retinopathy in 2018.[40] 754 UWFIs were graded by ophthalmologists, of which 643 were gradable and inputted into the algorithm. The study set a threshold of moderate non-proliferative diabetic retinopathy (NPDR) or higher (i.e. level 2 or higher on the International Clinical Diabetic Retinopathy scale) as sufficient to warrant a referral to an ophthalmologist. The study used the EYEART proprietary and closed-source algorithm to automatically detect and quantify DR lesions, such as hemorrhages, microaneurysms, lipid exudates, and wool spots. The EYEART algorithm, designed for standard flash colour images, was applied here to UWFIs.

The algorithm found 21.22% of the images contained referral-warranted DR while the graders determined 30.77% contained referral-warranted DR. When using both eyes to grade DR, the algorithm achieved a 91.7% sensitivity, 50.0% specificity, and 0.873 AUROC. When using individual eyes, the algorithm achieved a 90.3% sensitivity, 53.6% specificity, and 0.851 AUROC. While the authors were able to achieve high sensitivity, the low specificity indicates a high number of false positives using the EYEART algorithm. While the results were promising, a full understanding of their ML methods is unable to be determined as the algorithm is closed-source. As EYEART was designed on flash colour images, it is expected that algorithms designed and trained on UWFIs would be more effective.

Nagasawa *et al* published a study in 2019, which used DL for detecting treatment-naïve proliferative diabetic retinopathy (PDR) from UWFIs.[41] The authors used a deep convolutional network (DCNN) with PDR and non-PDR images.

The authors used the VGG-16 DCNN, which automatically learns the local features of an image and generates a classification model.[26] The authors used 40 deep learning models from 40 learning cycles and chose the model with the highest correct answer rate from test data as the DL model for the study. The DCNN selected achieved 94.7% sensitivity, 97.2% specificity, and 0.969 AUROC. Gradient-weighted class activation mapping (Grad-CAM) was utilized to visualize the image features used by the DCNN to classify images as containing referrable PDR.

The authors specifically used treatment-naïve PDR, which may have improved their results relative to Wang *et al.* Nonetheless, the authors were able to achieve high sensitivity and a high specificity, indicating that DCNN approaches trained on UWFIs may be superior to applying algorithms designed for colour fundus images (i.e. the EyeArt algorithm) to UWFIs for DR detection.

Using UWF-FA from PDR patients, Bawany *et al* utilized DL to correlate automated vessel density with visual acuity in 2020.[37] While not focusing on detection of DR generally, the goals of the study were to use a DL-quantified measure (retinal vessel density) and determine if it correlated with an outcome (visual acuity) known to be affected by PDR. Retinal blood vessels were first detected using a deep neural network (DNN) that was organized in the U-Net architecture. The study authors trained the dataset on UWF-FA images with corresponding ground-truth vessel maps trained using a human-in-the-loop procedure first demonstrated by Ding *et al.*[43] The output of the DNN was a vessel map where pixel intensity indicated the likelihood of a pixel being a vessel. The trained DNN achieved 0.930 AUPRC.

Vessel density was measured by calculating the percentage of vessel pixels in a circular area centered around the fovea. To study the correlation between vessel density and best corrected visual acuity (BCVA), UWF-FAs were analyzed using the trained model. The study found a statistically significant positive correlation between vessel density and BCVA of 0.4071 ($p = 0.0075$), but no statistically significant correlation between vessel density and central retinal thickness (CRT).

Tang *et al* published a study in 2021 which used DL to detect vision-threatening DR (VTDR) and referrable DR (RDR) from UWFIs.[39] Using UWFIs, they trained three CNNs to develop a pipeline for disease detection. The first CNN classified images as gradable or ungradable, the second for detecting VTDR, and the third for detecting RDR. The study used transfer learning and applied ResNet50 models pre-trained on ImageNet. Finally, the authors applied Class Activation Mapping (CAM) heatmaps for each result (true positive, true negative, false positive, and false negative) to assess DL performance.

The first CNN to determine gradeability achieved an 86.5% sensitivity, 82.1% specificity, and 0.923 AUROC on their primary dataset. The RDR detection CNN achieved 0.981 AUROC, 94.9% sensitivity, and 95.1% specificity. The VTDR CNN achieved 0.966 AUROC, 87.2% sensitivity, and 95.8% specificity. On four external datasets, the gradeability CNN achieved >0.82 AUROCs, >79.6% sensitivity, and >70.4% specificity while the RDR and VTDR CNNs achieved AUROCs and accuracies of >0.9 and >80% respectively.

Oh *et al* published a study investigating the early detection of DR using DL and UWFIs.[38] They compared the ability of CNNs to classify ETDRS 7SF vs. optic disk and macula-centered ETDRS F1-F2 images as containing diabetic retinopathy. They first trained a U-Net model with ResNet-18 for optic disk detection on the publicly available REFUGE dataset of colour FI.[44] The authors then used size and distance thresholds to determine macula locations. They then inputted UWFIs into this trained model to detect the optic disk and macula center. From these detected locations, they segmented the UWFIs into ETDRS 7SF images and F1-F2 images. 7SF ETDRS images contain 7 fields of 30 degrees each, while F1-F2 images contain only 30-degree overlapping circles centered on the optic disk and macula center.

The authors then trained a ResNet-34 model pre-trained on ImageNet and optimized their model using their dataset. In doing so, they achieved an 0.915 AUROC, 83.38% sensitivity, and 83.41% specificity on 7SF images. However, they achieved a 0.8867 AUROC, 80.60% sensitivity, and 80.61% specificity on F1-F2 images. 7SF images achieved results that were significantly greater for all three measures ($p < 0.001$) compared to those of F1-F2 images.

While the authors demonstrate that DL classification systems are more accurate using 7SF images, the achieved AUROC, sensitivity, and specificities have been greater in previously published studies using whole UWF. This indicates the greater utility of UWFIs over 7SF and F1-F2 images of the fundus.

Nagasawa *et al* published a second study on DR using DCNNs and UWFIs in April 2021.[42] They compared the accuracy of DL-based DR staging from UWFIs and OCTA images. UWFIs and OCT en face images of the superficial plexus, deep plexus, outer retina, choriocapillaris, and density map were extracted. OCTA

scans of a 6x6 mmregion were acquired for each patient. The OCTA and UWFIs were combined into a single image file, to form a third “imaging modality.” The severity of DR for each patient and their associated images were graded by three retinal specialists.

The study authors then trained a VGG-16 DCNN to first classify the images as containing DR, and the second to detect PDR. Each DCNN was tested on UWF, OCTA, and UWF-OCTA combined datasets. In detecting DR, the first DCNN achieved AUCs after training on the UWF, OCTA, and UWF-OCTA images of 0.790, 0.883, and 0.847 respectively. In detecting PDR, the second DCNN achieved AUCs after training on the UWF, OCTA, and UWF-OCTA images of 0.981, 0.928, and 0.964 respectively. This study demonstrates the ability of DL systems to detect DR and PDR but also demonstrates no additive benefit of combining imaging modalities (UWF and OCTA) to increase the accuracy of disease classification.

Retinal Detachment

Five studies have been published on retinal detachment detection from UWFIs.[33, 45–48] The first published, from Ohsugi *et al*, used UWFIs to detect rhegmatogenous retinal detachment (RRD).[45] The study used a CNN with 3 convolutional layers, of which each were followed by activation function (ReLU) layers and finished with two fully connected layers. The final output layer performed a binary classification using a softmax function. The trained model achieved an 0.988 AUROC, 97.6% sensitivity, and 96.5% specificity.

In 2019, Li *et al* developed a DL system for identifying specific characteristics of retinal detachment from UWFIs.[46] They developed a DL system for detecting notable peripheral retinal lesions (NPRLs), such as lattice degeneration and retinal breaks, which often lead to RRD. The authors then used and compared 4 CNNs: InceptionResNetV2, InveptionV3, ResNet50, and VGG-16. With each CNN, the authors explored three methods for improving the DL algorithm: i) no data augmentation, ii) data augmentation with brightness shifts, 45-degree rotation, and horizontal and vertical flipping, and iii) data augmentation with histogram brightness equalizations, 45-degree rotation, horizontal flipping, and vertical flipping. This led to 12 models trained and compared. The study found that the dataset trained on InceptionResNetV2 with the second data augmentation method achieved the greatest performance, with 98.7% sensitivity, and 99.2% specificity, and 99.1% total accuracy. This was significantly greater than comparisons with ophthalmologists in the study. The authors found that a general ophthalmologist with 5 years of experience had a 97.6% accuracy, 93.6% sensitivity ,and a 98.7% specificity, while one with 3 years experience had a 94.5% accuracy, 85.9% sensitivity and 96.8% specificity. These results are very promising for the continued use of DL in identifying NPRLs and the greater accuracy in comparison to trained ophthalmologists.

In 2020, Li *et al* then applied an InceptionResNetV2-based DL model to detect RD and discern macular status using UWFIs.[33] They first developed a DL system to detect RD. The model for detecting RD achieved a 96.1% sensitivity, 99.6% specificity, and 0.989 AUROC. A retina specialist with 3 years experience achieved 94.4% sensitivity and 99.1% specificity, while a specialist with 5 years experience achieved 95.4% sensitivity and 99.8% specificity. The RD images were then used as the dataset for the DL for macular status classification. This DL model achieved 93.8% sensitivity, 90.9% specificity, and 0.975 AUROC. The ophthalmologist with 3 years of training achieved sensitivities and specificities of 86.3% and 87.1%, while the more senior ophthalmologist achieved 91.3% and 92.4% respectively. The difference in discerning macular status between the DL and ophthalmologists is greater than their difference in RD detection. As macular status is an indication for emergency surgery, this difference is significant in demonstrating the utility and necessity of DL in ophthalmology.[33]

Zhang *et al*. developed a DL system for detecting lattice degeneration, retinal breaks, and RD in tessellated eyes. They then tested two image pre-processing techniques with the seResNext50 CNN. The first technique resized all images to 512x512, and when applied to the DL model, the model would output a positive number for each lesion per image. The second method used the cropping of patches of labelled lesions. The DL model applied to this dataset would then assign a positive score to each lesion and output the max score of all the image’s patches. Furthermore, they trained three distinct models for detecting lattice degeneration, retinal breaks, and RD respectively combining for a total of 6 tested models. In detecting lattice degeneration, the resizing method achieved 0.888 AUROC while the cropping method achieved 0.841. For retinal breaks, the resizing method had 0.843 AUROC while the cropping method achieved 0.953 AUROC. In RD, the resizing and cropping methods achieved AUROCs of 1.00 and 0.979 respectively. The use of the full image led to greater accuracy in all cases except for retinal breaks, where the cropping method was found to be superior.

In 2021, Antaki *et al*. published a study exploring the use of automated machine learning (AutoML) technologies for classifying RD, RP, and RVO from UWFIs.[48] They trained a DL model through the Google Cloud AutoML platform using RD and normal UWFIs. The binary classification of RD achieved an 89.77% sensitivity and 78.72% specificity when the confidence level of the system was set to 0.8. This model also achieved an AUPRC of 0.921.

Glaucoma

Two studies have investigated glaucoma with UWFIs.[49, 50] In 2018, Masumoto *et al*. applied a DL classifier to UWFIs to detect glaucoma in a patient dataset stratified by disease severity. The study authors first categorized glaucoma patients into early (-6 dB), moderate (-6 to -12 dB), and severe (-12dB or worse) based on visual field damage from Humphrey Field Analyzer measurements. In classifying any glaucoma, the DL model achieved a mean of 0.872 AUROC. For early, moderate, and severe glaucoma, the DL model achieved AUROCs of 0.830, 0.864, and 0.934 respectively. The DL model was similarly most sensitive and most specific in classifying severe glaucoma vs. healthy UWFIs. While the results are promising, the AUROC does not reach the 0.9 threshold, which was an acknowledged weakness by the study authors.

Li *et al*. used DL for automated glaucomatous optic neuropathy (GON) detection using UWFIs in 2020.[49] They trained a CNN based on the InceptionResNetV2 neural network. All UWFIs were classified as containing GON or not by glaucoma specialists, based on a vertical cup to disc ratio ≥ 0.7, rim width ≤ 0.1 of disc diameter, retinal nerve fiber layer defects, or disc splinter hemorrhages. The primary dataset achieved an AUROC, sensitivity, and specificity of 0.999, 97.5%, and 98.4% respectively. The range of AUROC, sensitivity and specificity achieved were 0.983-0.999, 97.5-98.2%, and 94.3-98.4% across the

primary and four external datasets. The methods demonstrated by Li *et al*. achieved significantly greater outcomes in detecting and classifying glaucoma on UWFIs than Masumoto *et al* likely due to the increased primary dataset size.

AMD

At the time of writing, two studies related to UWFIs and using DL to diagnose or detect AMD or its complications have been published.[51, 52] Matsuba *et al*. published a study in 2018 using DL to detect AMD on UWFIs.[51] In this study, they trained a deep CNN (DCNN) on UWFIs of healthy (no visible fundus disease) and patients with exudative AMD (wet AMD). The DCNN achieved a 0.976 average AUROC, with 100% average sensitivity, and 97.31% sensitivity in detecting wet AMD. Six ophthalmologists yielded a correct classification 81.9% of the time, with 71.4% and 92.5% sensitivity and specificity respectively. The study ophthalmologists averaged 11 minutes and 23.54 seconds for classification, while the DL model averaged 26.29 seconds.

The second study, published in 2021, comes from Li *et al*. who used DL for the automated detection of retinal exudates and drusen from UWFIs.[52] Images were labelled as containing retinal exudates and/or drusen (RED) or non-RED by four retina specialists. Two external datasets were then used for validation of the InceptionResNetV2 CNN model. On the primary dataset, 0.994 AUROC was achieved, with 94.2% sensitivity and 97.4% specificity. The external datasets achieved 0.972 and 0.988 AUROCs, with 94.9% and 95.1% sensitivities, and 96.5% and 97.3% specificities respectively.

Retinitis Pigmentosa

Masumoto *et al*. trained a deep CNN on UWFIs of retinitis pigmentosa (RP) in 2018.[53] Using UWF and UWF-FAF images, they trained a DCNN (VGG-16) to classify images based on whether they contained RP. UWFIs and UWF-FAFs from RP and healthy patients respectively were used in their dataset. The UWF DNN achieved 0.998 AUROC while that of the FAF achieved 1.00 AUROC. The UWF DCNN achieved 99.3% and 99.1% sensitivity and specificity scores respectively, while the UWF-FAF DCNN achieved 100% and 99.5% sensitivity and specificity scores. There were no statistically significant differences between the sensitivities and specificities of the UWF and the UWF-FAF DCNNs.

Antaki *et al*. published a study exploring the use of AutoML technologies for classifying RD, RP, and RVO from UWFIs.[48] They trained a DL model through the Google Cloud AutoML platform using RP and normal UWFIs.. The binary classification of RP achieved an 88.0% sensitivity and 100% specificity when the confidence level of the system was set to 0.5. This model also achieved 0.942 AUPRC. When repeated using the data from Masumoto *et al*[53] the system achieved an AUPRC of 1, with sensitivity, specificity, and PPV all increased to 100% with no misclassifications made by the AutoML model.

Pachychoroid

A single peer-reviewed study on using UWFIs in detecting pachychoroid disease has been published. Kim *et al*. used an AutoML platform to classify UWFIs based on their presence of pachychoroid disease.[54] Specifically, the authors trained the Google AutoML Vision on UWF indocyanine green angiography (UWF-ICGA) of healthy and pachychoroid patients. Pachychoroid and non-pachychoroid UWF-ICGA images were uploaded. They trained two models, the first of which used all images in their original orientation and the second of which horizontally flipped left eye images such that all images were of the same laterality. The first model achieved precision, accuracy, sensitivity, and specificity values of 0.8182, 0.8367, 0.8182, and 0.8519 respectively, while the second model achieved 0.8636, 0.8776, 0.8636, 0.8889 respectively. However, the mean precision, accuracy, sensitivity, and specificity scores of three retina specialists were 0.9048, 0.9388, 0.9500, and 0.9643. These results indicate that training the AutoML model with images of the same laterality led to better results, but that the current training did not reach the levels of precision or recall of retina specialists.

RVO

Three peer-reviewed studies exist on using DL on UWFIs in RVO, two of which were published by Nagasato *et al*, and the third from Antaki *et al*.[48, 55, 56] The first, published in 2018, uses UWFIs to classify and detect central retinal vein occlusion (CRVO). The study used UWFIs from CRVO and non-CRVO healthy subjects. A VGG-16 based DNN was trained on the dataset, along with fine-tuning using parameters borrowed from ImageNet. After comparing 40 DL models obtained in 40 learning cycles, they used the DL model with the highest rate of correct answers for evaluation. The model achieved 0.989 AUROC, 98.4% sensitivity, and 97.9% specificity. They similarly used a support vector machine (SVM) algorithm to detect CRVO from UWFIs. The SVM achieved 0.895 AUROC, 84.0% sensitivity, and 87.5% specificity. The DL model achieved significantly greater results in all measures compared to the SVM ($p < 0.001$).

In 2019, the same group completed a similar study using a DL model on UWFIs of branch retinal vein occlusion (BRVO) patients. In this study, they used the same model (VGG-16), and DNN parameters on a BRVO dataset. Specifically, they trained the DNN on BRVO and non-BRVO healthy UWFIs. They similarly tested an SVM model. In this study, the DNN achieved 0.976 AUROC, 94.0% sensitivity, and 97.0% specificity. The SVM model achieved 0.857 AUROC, 80.5% sensitivity, and 84.3% specificity. The authors demonstrated the ability of a DNN to accurately detect BRVO and the superiority of a DNN over SVMs in detecting BRVO.

In 2021, Antaki *et al*. published a study exploring the use of the Google Cloud AutoML platform for classifying RD, RP, and RVO from UWFIs.[48] The binary classification of RVO achieved a 84.9% sensitivity and 100% specificity when the confidence level of the system was set to 0.5. This model also achieved 0.967 AUPRC. While the sensitivity was lower than that of Nagasato *et al*, their model achieved comparable specificities.[55, 56]

Myopia

In 2020, Shi *et al*. published a study where they studied the ability of a DL system to detect myopia using UWFIs.[57] For this task, they used a custom CNN, known as the Myopia Detection network (MDNet). This network combined dense connection and Residual Squeeze-and-Excitation attention for detecting myopia. The CNN combined attention dense blocks, transition blocks, convolutional layers, max-pooling layers, and a dense layer to make full use of shallow features and improve information flow.

They trained the CNN on left and right UWFIs. The study defined severe myopia as having a spherical equivalent (SE) less than -6 diopters (D), mild myopia as between -6D and -3D, and mild myopia as SE between -3D and 0D. Images were then cropped for a region of interest around the optic disk of 400x400 pixels, centered on the optic disk and including the macula.

In evaluating the model, the study authors used mean absolute error (MAE) as the main evaluation index, as well as root-mean-square error (RMSE) and mean-absolute-percent error (MAPE). The CNN achieved optimal results at an MAE of 1.1150 D and RMSE and MAPE of 1.4520 D and 24.99% respectively. These results show that myopia is effectively detected within reasonable error using DL and UWFIs.

Idiopathic Macular Hole

A single peer-reviewed study on detecting idiopathic macular hole (IMH) using UWFIs and DL has been published. In 2018, Nagasawa *et al* trained a deep CNN on normal and IMH images.[32] The CNN achieved an 0.9993 AUROC, 100% accuracy, 100% sensitivity, and 99.5% specificity. The CNN was able to classify images at an average speed of 32.80±7.36 seconds for a series of 50 test images. They similarly tested the ability of human ophthalmologists to detect IMH from the same UWF test images. The ophthalmologists were able to achieve an 80.6±5.9% accuracy, 69.5±15.7% sensitivity, and 95.2±4.3% specificity, and required an average time of 838±199.16 seconds to classify the same 50 images. From this study, it is clear that IMH is more accurately and rapidly diagnosed using CNNs than trained ophthalmologists.

Retinal Hemorrhage

Li *et al.* published a study using a DL system to screen retinal hemorrhage (RH) from a dataset of RH and non-RH UWFIs.[58] The study used InceptionResNetV2, with weights pre-trained for ImageNet classification for CNN initialization. On the primary dataset, the CNN achieved an 0.999 AUROC, 98.9% sensitivity, 99.4% specificity, and 99.3% accuracy. Two external datasets were used for further testing, which achieved 0.998 and 0.997 AUROCs, 96.7% and 97.6% sensitivities, 98.7% and 98.0% specificities, and 98.4% and 98.0% accuracies respectively. On an external dataset, an ophthalmologist with five years of training achieved a 95.9% sensitivity and a 99.5% specificity, while an ophthalmologist with three years of training achieved 92.6% and 98.9% respectively. Here, the ophthalmologists scored sensitivities lower than the trained CNN, but specificities were close to the specificity of the CNN.

Sickle Cell Retinopathy

A single study, published in 2020, explores using DL with UWFIs to diagnose sickle cell retinopathy. Specifically, the study from Cai *et al.* explored the detection of sea fan neovascularization (SFN) from UWFIs of patients with sickle cell hemoglobinopathy.[59] The study notes that the detection of potentially asymptomatic SFN provides the opportunity for prophylactic scatter laser photocoagulation, which can help to reduce the rates of proliferative sickle cell retinopathy (PSR) vision loss. An InceptionV4 CNN, pre-trained on the ImageNet dataset, was trained on the image set for 100 iterations. After training, the CNN achieved an 0.988 AUROC, 97.0% accuracy, 97.4% sensitivity, and 97.4% specificity. Only a single image received a false-negative classification from the CNN, due to a severe lid artifact obscuring the retinal vasculature.

Quality Assessment

Three studies have been published on using DL methods for quality assessment of UWFIs.[60–62] The first, published in 2020 by Calderon-Auza *et al* focuses on using CNNs as a teleophthalmology support system, to determine the quality of images provided. Specifically, the system they proposed uses four steps to determine UWFIs quality. First, the system detects the optic disc (OD), performs quality analysis on the OD, determines obstruction (i.e. eyelash shadows) detection of the region of interest (ROI), and then segments the vessels of the image. For OD detection, faster R-CNN (FR-CNN) for feature extraction along with the AlexNet CNN architecture was used. On their dataset, this CNN configuration achieved an accuracy, sensitivity, and specificity of 0.9254, 0.9643, and 0.4424 respectively. Images determined as containing an OD were then used as the dataset for the OD quality analysis step. For this step, VGG-16 was used and achieved a 0.8612 accuracy, 0.9113 sensitivity, and 0.8064 specificity in detecting classifying ODs by quality. For obstruction analysis in the ROI, centered on the optic disk and macula, a SegNet-trained CNN achieved a 1.0 accuracy due to the low number of artifacts in the training and test sets. Finally, for vessel segmentation in the ROI, a SegNet architecture with VGG-16 proposed by the authors achieved an 0.9784 accuracy, 0.7169 sensitivity, and 0.9816 specificity on their dataset.

While the study from Calderon-Auza *et al* is a proof of concept of the uses of DL for detecting low-quality UWFIs, it also provides examples of these techniques in practice. For this reason, this study provides readers with a clear implementation of the uses of DL in UWFIs for a multi-step process in tele-ophthalmology.

Li *et al.* proposed and designed a classification system using “U-Net style” CNN and UWF-FAF in 2020.[61] UWF-FAF were graded as ungradable, poor, good, or best by ophthalmologists. The CNN achieved 90.5% sensitivity and 87.0% specificity for distinguishing between gradable and ungradable images and a sensitivity and specificity of 78.9% and 94.1% for distinguishing between optimal quality (good, best) and limited quality (poor, ungradable) images. The authors calculated the overall accuracy of the classifier as 89.0% for gradable vs. ungradable classification and 89.3% for recognizing optimal quality versus limited quality. The model also achieved an 0.920 AUROC.

In 2020, Li *et al* proposed a DL-based image filtering system (DLIFS) to filter out poor-quality UWFIs in an automated fashion, such that only images of sufficient quality would be used in subsequent AI diagnostic systems.[62] Images were identified as poor-quality or good-quality images by 4 retina specialists. Image quality was categorized as poor if more than 1/3 of the fundus was obscured, macular vessels could not be identified or >50% of the macular area was obscured, or if the vessels within a 1-disc diameter of the OD margin could not be identified. From this dataset, they trained InceptionResNetV2 with weights pre-trained for ImageNet. The CNN would classify the image quality for each inputted UWFI. The trained DLIFS achieved an 0.996 AUROC, 96.9% sensitivity, and 96.6% specificity. Two external datasets were used for testing, with which the DLIFS achieved 0.994 and 0.997 AUROCs, 95.6% and 96.6% sensitivities, and 97.9% and 98.8% specificities respectively.

Segmentation and Localization

Five peer-reviewed studies have been published on segmentation and localization using UWFIs, all of which focus on vessel segmentation.[43, 63–66] The first, from Ding *et al* presented a method to detect retinal vessels in UWF fluorescent angiography (UWF-FA).[43] In this study, the authors developed a method to produce vessel segmentation maps without previously labelled ground-truth datasets. They primarily relied on cross-modality transfer and human-in-the-loop (HITL) learning. The HITL approach, a form of semi-supervised learning, allowed the DL system to predict the vessels and respond to human feedback regarding whether it had segmented the vessels correctly and accurately. Over multiple iterations, this led to complete segmentation of the vessels. The authors were able to reduce manual annotation effort by first using morphological analysis to segment the vessels in a preliminary fashion. This was followed by a cross-modality approach that transferred vessel maps from UWF colour images to UWF-FA using robust chamfer alignment in an Expectation-Maximization framework. These were combined using the HITL iterative DL process for detection of retinal vessels.

The first step in the pipeline, relying on cross-modality transfer, trained a DNN on a dataset of ground truth colour UWFIs with UWF-FA from the same patient eye taken at the same time. Specifically, the DNN was trained on existing labelled UWFIs to extract the vessel maps from unlabelled UWFIs. These detected vessel maps were then geometrically aligned and transferred to the UWF-FA. These new vessel maps, aligned to UWF-FA, served as the approximate ground truth for training a DNN for vessel detection in UWF-FA images. From this point, the DNN for detecting vessel segmentation was continually run starting from the approximate ground truth from the UWF colour image, until the DNN did not produce maps with new changes or more vessels segmented.

The process of producing vessel maps was approached as one that would be best suited for a generative adversarial network (GAN), in producing an output image of vessel segmentation from an input image of UWF-FA.

The authors then evaluated their method of reducing the burden of annotation by calculating the number of pixels added and removed at each iteration. After 7 iterations, approximately 19,300 (2.0%) new pixels were added, and 14,100 (1.4%) of pixels were removed. In validating their approach on an external dataset, they achieved a maximal 0.987 AUROC, with significant improvements over traditional morphological techniques for vessel segmentation.

The same team of Ding *et al* then published a method to segment vessels from colour UWFIs via iterative multi-modal registration and learning.[63] In this project, they similarly utilized concurrently captured UWF-FA to segment the vessels from UWFIs. The first step requires multi-modal registration of the vessels segmented first from UWF-FA using a pre-trained DNN to the UWFIs, using parametric chamfer alignment. The second step utilized a learning method to mitigate the noisy labels due to the differences in the UWF-FA and UWFIs modalities. The detected UWFIs vessel maps are then used for the registration in the following iteration, allowing for iterative improvement until the segmented vessel maps are accurate. After this training, the DNN can detect vessels from UWFIs without concurrently captured UWF-FA. After training their DNN, they evaluated the model on an external dataset of UWFIs, achieving an AUROC PR of 0.886.

Nunez do Rio *et al*. published a study in 2020 that explored the use of DL-based segmentation for quantification of retinal capillary non-perfusion using UWF-FA.[64] Capillary non-perfusion (CNP) is a metric that is useful in determining retinal ischemia. For this reason, they sought to use UWF-FA, which is a high-resolution image with a clearly defined retinal vasculature, to quantify this. For this process, they trained a U-Net-style CNN on 75 UWF-FA that were manually graded for CNP to segment and extract the vasculature of these images. 20 images were also segmented by an expert grader manually. To standardize the CNP measurement, a circular grid of rings of increasing radius was centered on the foveal avascular zone (FAZ). The segmentation model achieved an 0.82 AUROC. Between the manually graded images and the automatically segmented images, an inter-grader dice similarity coefficient (DSC) of 65.51 was achieved. In comparing the assessment of CNP between the CNN model and the grader, a Kappa score of 0.55 was achieved. The authors conclude that this automatic segmentation method allows for a DL-based segmentation of CNP and a quantifiable measurement of CNP from UWF-FA.

Wang, Z *et al*. published a study in 2020 that utilized a multi-task Siamese network for separating retinal arteries from retinal veins using deep convolution.[65] They did so on an FI dataset (DRIVE), a UWFI dataset (WIDE), and an OCT dataset (INSPIRE). Using these datasets, they first segmented the vessels using a CNN-based approach, followed by skeletonization of the vessels. Next, they built a graph representing the vascular network by finding branching and end points on the skeleton map. Next, errors such as twinborn nodes produced by overlapping vessels were removed by morphological analysis through the skeleton. This produced a refined vascular graph. They then used Convolution Along Vessel (CAV) to extract visual features by convolving the image along the vessel segments and the geometric features of the vessels, by tracking the direction of blood flow in the vessels. Following this, the Siamese network was trained to learn to classify vessel types by visual features of vessel segments, and by estimating the similarity of every two connected segments by comparing their visual and geometric features. This was done to separate the vasculature types into individual trees of arteries and veins. On the WIDE dataset of UWFI, they were able to achieve an accuracy value of 94.5%.

In 2021, Sevgi *et al*. published a study that explored the ability to extract the cumulative retinal vessel areas (RVA) from UWF-FA images using CNN-based DL segmentation. For this study, they extracted the RVA from the available UWF-FA image frames. Images that contained the maximum RVA were considered the optimum early phase, while a frame that was taken ≥ 4 minutes after that closely mirrored the RVA from the early image was considered the late phase frame. Image analysts then evaluated the selected pairs. 1578 UWF-FA sequences from 66 sessions were used to create cubic splines and a total of 13,980 UWF-FA sequences from 462 sessions were used for evaluation. 85.2% of the sessions had appropriate images for both phases successfully identified. 90.7% of early and 94.6% of late frames were successfully identified.

Generative Image Synthesis using GANs

At the time of writing, four studies involving generative adversarial networks (GANs) and UWFIs have been published.[67–70] GANs, designed in 2014 by Goodfellow *et al* are ML frameworks that utilize two competing neural networks to generate new data.[71] The first neural network, named the ”generator”, generates random data. The second neural network, the “discriminator”, is trained on data that is to be modelled and produced. As the generator produces

data, the discriminator will reject synthesized data from the generator that does not sufficiently represent the training data source. Through iterative processes, the generator becomes more effective at generating data that effectively represents the target data, until synthetic data that appears close to the ground-truth dataset is produced.[72] These approaches have been effective at generating data such as human faces that appear realistic.[73]

Ju *et al.* published another study in 2020, where they utilized GANs to produce labelled datasets of UWFIs from labelled fundus image datasets. They noted that due to the differences in fundus images (FIs) and UWFIs, labelled datasets of FIs could not be used for UWFIs. For this reason, they used a GAN to generate synthetic UWFIs for training. Using a consistency regularization method, they ensured that the pathologies present in labelled FIs were similarly present in corresponding generated UWFIs. The first step in this process required using target UWFIs to train a target-task model, which helps to regulate the quality of generated data. Following this, pseudo-labels were generated for the generated UWFIs. Finally, they used the original UWFIs samples and the generated samples to train the target-task model together. To test that the generated UWFIs were properly pseudo-labelled and carried the disease pathology of interest, they then classified the images that contained DR using a Res-Net50 based residual neural network. They then similarly validated their synthetic UWFIs by testing them with vessel segmentation and lesion detection tasks. The study authors effectively succeeded at producing high-quality UWFIs that mirrored FIs in pathology, and mirrored "natural" UWFIs in image quality and complexity.

In 2020, Xie *et al.* published a study where they proposed a GAN which used an attention encoder (AE) and generation flow network to build a UWFIs classifier for retinal pathologies found in patients under the age of eighteen (i.e. Coats, FEVR, morning glory syndrome, retinitis pigmentosa, and diabetic retinopathy).[69] The goal of this project was to harness the adversarial learning that occurs between the generator and the discriminator to build robustness into their classification model. Their proposed method achieved higher classification accuracy (84.75% and 97.25%) compared to classifiers based on a standard CNN architecture such as ResNet-50 (77.35% and 87.95%).

In 2020, Yoo *et al* used GANs in a way opposite to Ju *et al.*[68, 70] In their study, they utilized a GAN architecture to produce synthetic FIs from UWFIs. Specifically, they used the CycleGAN architecture to translate the UWFIs to FIs while maintaining the structure, pathology, and lesions specific to the original FIs without generating new or fake features into the FI. The authors began by using a dataset of UWFIs and FIs, which were reviewed by ophthalmologists for image quality. The GAN was trained on the dataset of UWFIs and FIs, and then tested on the test dataset of UWFIs to generate synthetic FIs. Image registration was applied to crop the region of interest on the input UWFIs, focused on the optic disk and fovea, for conversion into an FI. After training the CycleGAN model for 40 epochs, the model was able to successfully transfer the image from UWFI to FIs with high fidelity to the original UWFI structure and pathologies. For example, UWFIs with diabetic retinopathy microaneurysms and blot hemorrhages, glaucomatous optic nerves, retinal detachment, CRVO, Drusen, and retinal atrophy all had their specific lesions transferred to FIs successfully. Finally, they calculated structural similarity (SSIM) indices between the generated FIs and the ground truth FIs, and achieved an average SSIM level of 0.802, indicating strong similarities between the image produced and the ground truth image.

Systemic Diseases

UWF imaging is also being used in conjunction with DL for the prediction of non-ocular and neurological factors. As UWF imaging is a rich image format, exploratory studies have been conducted to determine if retinal changes can be associated with features like an individual's age, vascular changes, and neurological status.

Age and Brachial-Ankle Pulse-Wave Velocity

Nagasato *et al.* published a study in 2020 demonstrating an ability to predict both patient age and their brachial-ankle pulse-wave velocity (baPWV) using UWFIs and DL.[74] For each patient included in the study, they also recorded patient baPWV. They then processed these images to contain the entire image (the total image), a cropped region of the optic disk and macula (the central region), and the total image with the central region covered in black pixels (the peripheral region). Each of these processed images were used as separate datasets for the model and compared by the study authors in their performance for DL prediction of age and baPWV. They then used patient baPWV, UWFI, and age as input data for a VGG-16 based CNN. The results showed that the total, central, and peripheral images were all able to predict the age and baPWV of a patient with statistical significance. Specifically, the statistical significance of the correlation between predicted and actual age and baPWV were both $p<0.001$ for all three datasets. Conclusively, the authors show that UWFIs can be used to make clear and specific predictions of a patient's age and baPWV specifically, which is itself a marker of vascular health.

Alzheimer's Disease

In 2020, Wisely *et al.* used multiple imaging modalities to train a DL model to identify symptomatic Alzheimer's disease (AD).[75] In this study, the authors used UWFIs, UWF-FAFs, colour maps of ganglion cell-inner plexiform layer (GC-IPL) thickness, and superficial capillary plexus (SCP) *en face* optical coherence tomography angiography (OCTA) for their training. They included these imaging modalities from eyes from cognitively healthy subjects and patients with symptomatic AD. The DL model designed took the three imaging modalities as input, as well as OCT and OCTA numerical and patient data. The model used a shared-weight image feature extractor to extract modality-agnostic features that were then used in a modality-specific function in a fully connected layer. After training the model, they then tested the model on each imaging modality individually, as well as combinations of the data. They found UWFIs to lead to an 0.450 AUROC when inputted alone, and UWF-FAF to achieve an 0.618 AUROC when inputted alone. On their own, OCTA achieved an 0.582 AUROC, and GC-IPL achieved an 0.809 AUROC. All images when inputted together achieved an 0.829 AUROC, while all images along with quantitative data achieved an 0.830 AUROC. All images and all data achieved an 0.836 AUROC, while GC-IPL, quantitative data, and patient data together achieved the highest AUROC of 0.841. These findings indicate that GC-IPL has the strongest individual predictive value of symptomatic AD and that the inclusion of more imaging modalities (i.e. OCTA, UWF-FAF, and UWF) do not improve the predictive value in this case. As well, the predictive value of UWF imaging alone for symptomatic AD is low.

Discussion

Across a broad range of domains and diseases, DL has been demonstrated to be useful when used in conjunction with UWF imaging. In the detection and classification of disease, DL models have been accurate, sensitive, and specific across a variety of ophthalmic disorders. In this review, we summarized the use of DL in detecting diabetic retinopathy, retinal detachment, glaucoma, AMD, retinitis pigmentosa, pachychoroid, retinal vein occlusion, idiopathic macular hole, retinal hemorrhage, and sickle cell retinopathy from UWFIs. While disease detection has been the most published application of DL in UWFIs, its use in quality assessment of UWFIs and segmenting and localizing the structures of the retina should not be overlooked.

Similarly, the high-resolution provided by each UWFI allows for novel generative uses with GANs. Finally, authors have demonstrated the novel utility of UWF's high-resolution imaging of the eye for diagnosing systemic and neurological disorders like Alzheimer's, estimating a patient's age, and vascular health via brachial-ankle pulse wave velocity.

Benefits and Risks of Deep Learning with UWF Imaging

As shown by the studies discussed above, the diagnostic potential of DL when used with UWF imaging is accurate and often exceeds the accuracy of trained ophthalmologists in many cases. The detection of ophthalmic diseases is a clear use for DL in UWF imaging. With automated DL systems for detection and diagnosis, the likelihood of detecting vision-threatening pathology early is greater. As visual acuity often does not return as many ophthalmic disorders progress, the early and high sensitivity detection of these disorders is beneficial to patient care. In multiple studies discussed above, DL systems achieved sensitivities greater than trained ophthalmologists but in some cases, specificities that were lower than their human counterparts.[33] For this reason, while DL models have achieved impressive accuracies, a reduced specificity compared to humans may lead to greater false-positive diagnoses. From this, the risk of unnecessary medical or surgical intervention increases if DL models are followed without questioning their results. For cases like this, where DL models are more sensitive but less specific than ophthalmologists, clinical use of DL models is still beneficial. In these cases, the combined increased sensitivity of DL models paired with the greater or equal specificity of ophthalmologists may lead to a synergistic effect in accuracy. In a clinical environment, this pairing can help to reduce both false negatives via the DL model's superior sensitivity and false positives via human specificity equal or greater than the DL models'.

As DL models continue to learn associations between disease classifications and UWFI features, the risk of poor explainability is possible. As DL models typically do not explain the associations they form, the possibility of associations being formed between unexpected image features and disease classification is possible. For example, the biases present in the training datasets may become codified in the DL models' associations. This reproduction of human biases has been seen in AI implementations in healthcare in the past.[76] However, this can be mitigated by using large datasets with multiple graders and reviewers, to minimize individual human bias. Furthermore, to ensure that ophthalmologists can understand the image features leading to disease classification, Grad-CAM can be applied to show heatmaps of the image regions and lesions leading to stronger associations with a disease type.[77, 78]

The Future of Ultra-Widefield Imaging and Deep Learning

In this review, several novel methods using DL and UWFIs have emerged. In particular, the use of GANs is useful to translate UWFIs to other imaging modalities, such as FIs with high fidelity.[70] Furthermore, the ability of DL to translate existing FIs to UWFIs provides the opportunity to translate decades of FIs available to ophthalmologists into a novel imaging modality, which can further strengthen DL training in the future.

UWF imaging and DL also have the potential to lead to an improved understanding of the pathophysiology of ophthalmic and non-ophthalmic disorders. While not yet demonstrated by the papers included here, using DL as an exploratory method can be proposed. For example, with a large enough dataset carefully classified, GradCAM interpretability models could demonstrate lesions on UWFIs that may predispose or increase the likelihood of a specific ophthalmic disorder. When combined with other data types and other imaging modalities, as exemplified by Wisely *et al* in their study for the detection of AD, [75] the use of DL as an exploratory option for further understanding ophthalmic pathophysiology becomes possible.

As demonstrated by the quality assessment studies discussed in this review, DL may also be useful for expanded use of tele-ophthalmology. If general practitioners or technicians can access UWF imaging for patients in remote areas, quality assessment methods can be used to determine if the UWFI is of sufficient quality for an ophthalmologist to review remotely or at a later date. For this reason, DL's utility is not simply in diagnosing and detecting diseases, but also in improving access to ophthalmic services for individuals in remote regions or with minimal access to ophthalmology expertise.

Future Directions

For future studies, we hope to explore the differences of each study systematically and compare the accuracies of the proposed DL models quantitatively using meta-analysis methods. However, due to the heterogeneity of data and aims with UWF imaging and DL, such analysis was not feasible at this time.

Declarations

Statements and Declaration: The authors have no relevant financial or non-financial interests to disclose.

Acknowledgments: None.

Meeting Presentation: None.

Word Count: 9894

References

1. Agarwal A (2007) Fundus Fluorescein and Indocyanine Green Angiography: A Textbook and Atlas. SLACK Incorporated

2. (1981) Diabetic retinopathy study. Report Number 6. Design, methods, and baseline results. Report Number 7. A modification of the Airlie House classification of diabetic retinopathy. Prepared by the Diabetic Retinopathy. Invest Ophthalmol Vis Sci 21:1–226
3. Kumar V, Surve A, Kumawat D et al (2021) Ultra-wide field retinal imaging: A wider clinical perspective. Indian J Ophthalmol 69:824–835. https://doi.org/10.4103/ijo.IJO_1403_20
4. Kaines A, Oliver S, Reddy S, Schwartz SD (2009) Ultrawide Angle Angiography for the Detection and Management of Diabetic Retinopathy. Int Ophthalmol Clin 49:53–59. https://doi.org/10.1097/IIO.0b013e31819fd471
5. Witmer MT, Kiss S (2012) The Clinical Utility of Ultra-Wide-Field Imaging. Review of Ophthalmology
6. Optos.com - Optos products. https://www.optos.com/products/. Accessed 21 Aug 2021
7. Wessel MM, Aaker GD, Parlitsis G et al (2012) ULTRA–WIDE-FIELD ANGIOGRAPHY IMPROVES THE DETECTION AND CLASSIFICATION OF DIABETIC RETINOPATHY. RETINA 32:785–791. https://doi.org/10.1097/IAE.0b013e3182278b64
8. Antaki F, Coussa RG, Mikhail M et al (2020) The prognostic value of peripheral retinal nonperfusion in diabetic retinopathy using ultra-widefield fluorescein angiography. Graefe's Archive for Clinical Experimental Ophthalmology 258:2681–2690
9. Liu TA, Arevalo JF (2019) Wide-field imaging in proliferative diabetic retinopathy. International journal of retina vitreous 5:1–4
10. Fogliato G, Borrelli E, Iuliano L et al (2019) Comparison Between Ultra-Widefield Pseudocolor Imaging and Indirect Ophthalmoscopy in the Detection of Peripheral Retinal Lesions. Ophthalmic Surg Lasers Imaging Retina 50:544–549. https://doi.org/10.3928/23258160-20190905-02
11. Quinn NB, Azuara-Blanco A, Graham K et al (2018) Can ultra-wide field retinal imaging replace colour digital stereoscopy for glaucoma detection? Ophthalmic Epidemiol 25:63–69. https://doi.org/10.1080/09286586.2017.1351998
12. Forshaw TRJ, Minör ÅS, Subhi Y, Sørensen TL (2019) Peripheral Retinal Lesions in Eyes with Age-Related Macular Degeneration Using Ultra-Widefield Imaging: A Systematic Review with Meta-analyses. Ophthalmology Retina 3:734–743. https://doi.org/10.1016/j.oret.2019.04.014
13. El Naqa I, Murphy MJ (2015) What Is Machine Learning? In: El Naqa I, Li R, Murphy MJ (eds) Machine Learning in Radiation Oncology: Theory and Applications. Springer International Publishing, Cham, pp 3–11
14. Sathya R, Abraham A (2013) Comparison of supervised and unsupervised learning algorithms for pattern classification. International Journal of Advanced Research in Artificial Intelligence 2:34–38
15. Norvig P, Russell S (2021) Artificial Intelligence: A Modern Approach, Global Edition, 4th edition. Pearson, Harlow
16. Zhou X, Belkin M (2014) Chapter 22 - Semi-Supervised Learning. In: Diniz PSR, Suykens JAK, Chellappa R, Theodoridis S (eds) Academic Press Library in Signal Processing. Elsevier, pp 1239–1269
17. Hinton G (1999) Unsupervised Learning: Foundations of Neural Computation, 1st edition. Bradford Books, Cambridge, Mass
18. Deng L (2014) Deep Learning: Methods and Applications. FNT in Signal Processing 7:197–387. https://doi.org/10.1561/2000000039
19. Dongare AD, Kharde RR, Kachare AD (2012) Introduction to artificial neural network. International Journal of Engineering Innovative Technology (IJEIT) 2:189–194
20. Zhou V (2019) Machine Learning for Beginners: An Introduction to Neural Networks. In: Medium. https://towardsdatascience.com/machine-learning-for-beginners-an-introduction-to-neural-networks-d49f22d238f9. Accessed 21 Aug 2021
21. Fornito A, Zalesky A, Bullmore ET (2016) Chapter 2 - Nodes and Edges. In: Fundamentals of Brain Network Analysis. Academic Press, San Diego, pp 37–88
22. Ciresan DC, Meier U, Masci J et al Flexible, High Performance Convolutional Neural Networks for Image Classification. 6
23. Saha S (2018) A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way. In: Medium. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53. Accessed 21 Aug 2021
24. LeCun Y, Boser B, Denker JS et al (1989) Backpropagation Applied to Handwritten Zip Code Recognition. Neural Comput 1:541–551. https://doi.org/10.1162/neco.1989.1.4.541
25. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet Classification with Deep Convolutional Neural Networks. In: Advances in Neural Information Processing Systems. Curran Associates, Inc
26. Alippi C, Disabato S, Roveri M (2018) Moving Convolutional Neural Networks to Embedded Systems: The AlexNet and VGG-16 Case. In: 2018 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN). pp 212–223
27. Pretrained Inception-ResNet-v2 convolutional neural network - MATLAB inceptionresnetv2. https://www.mathworks.com/help/deeplearning/ref/inceptionresnetv2.html;jsessionid=ae4b2abd60579ecab16f783b4f26. Accessed 21 Aug 2021
28. Shah T (2020) About Train, Validation and Test Sets in Machine Learning. In: Medium. https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7. Accessed 21 Aug 2021
29. Talebi H, Milanfar P (2021) Learning to Resize Images for Computer Vision Tasks. arXiv:210309950 [cs]
30. CNN-Based Quality Assessment for Retinal Image Captured by Wide Field of View Non-Mydriatic Fundus Camera | IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8769037?casa_token=3eNYAqzlF2oAAAAA:68ENuGgsl_7lZnBiRUeSJ1sg_TBSU_cPqfPPhUvroGFUNzl6j3D1OrvqEoiJmrAf2v3aeMY2. Accessed 21 Aug 2021
31. Perez L, Wang J (2017) The Effectiveness of Data Augmentation in Image Classification using Deep Learning. arXiv:171204621 [cs]
32. Nagasawa T, Tabuchi H, Masumoto H et al (2018) Accuracy of deep learning, a machine learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting idiopathic macular holes. PeerJ 6:e5696. https://doi.org/10.7717/peerj.5696

33. Li Z, Guo C, Nie D et al (2020) Deep learning for detecting retinal detachment and discerning macular status using ultra-widefield fundus images. Communications Biology 3:. https://doi.org/10.1038/s42003-019-0730-x
34. Parikh R, Mathai A, Parikh S et al (2008) Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol 56:45–50
35. Mandrekar JN (2010) Receiver Operating Characteristic Curve in Diagnostic Test Assessment. Journal of Thoracic Oncology 5:1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d
36. Boyd K, Eng KH, Page CD (2013) Area under the Precision-Recall Curve: Point Estimates and Confidence Intervals. In: Blockeel H, Kersting K, Nijssen S, Železný F (eds) Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, pp 451–466
37. Bawany MH, Ding L, Ramchandran RS et al (2020) Automated vessel density detection in fluorescein angiography images correlates with vision in proliferative diabetic retinopathy. PLOS ONE 15:e0238958. https://doi.org/10.1371/journal.pone.0238958
38. Oh K, Kang HM, Leem D et al (2021) Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images. Sci Rep 11:1897. https://doi.org/10.1038/s41598-021-81539-3
39. Tang F, Luenam P, Ran AR et al (2021) Detection of Diabetic Retinopathy from Ultra-Wide Field Scanning Laser Ophthalmoscope Images: A Multi-Center Deep-Learning Analysis. Ophthalmology Retina S246865302100035X. https://doi.org/10.1016/j.oret.2021.01.013
40. Wang K, Jayadev C, Nittala MG et al (2018) Automated detection of diabetic retinopathy lesions on ultrawidefield pseudocolour images. Acta Ophthalmol 96:e168–e173. https://doi.org/10.1111/aos.13528
41. Nagasawa T, Tabuchi H, Masumoto H et al (2019) Accuracy of ultrawide-field fundus ophthalmoscopy-assisted deep learning for detecting treatment-naïve proliferative diabetic retinopathy. Int Ophthalmol 39:2153–2159. https://doi.org/10.1007/s10792-019-01074-z
42. Nagasawa T, Tabuchi H, Masumoto H et al (2021) Accuracy of Diabetic Retinopathy Staging with a Deep Convolutional Neural Network Using Ultra-Wide-Field Fundus Ophthalmoscopy and Optical Coherence Tomography Angiography. Journal of ophthalmology 2021
43. Ding L, Bawany MH, Kuriyan AE et al (2020) A Novel Deep Learning Pipeline for Retinal Vessel Detection In Fluorescein Angiography. IEEE Trans Image Process 29:6561–6573. https://doi.org/10.1109/TIP.2020.2991530
44. Diaz-Pinto A, Xu Y, Li X et al (2020) REFUGE challenge: a unified framework for evaluating automated methods for glaucoma assessment from fundus photographs
45. Ohsugi H, Tabuchi H, Enno H, Ishitobi N (2017) Accuracy of deep learning, a machine-learning technology, using ultra–wide-field fundus ophthalmoscopy for detecting rhegmatogenous retinal detachment. Sci Rep 7:. https://doi.org/10.1038/s41598-017-09891-x
46. Li Z, Guo C, Nie D et al (2019) A deep learning system for identifying lattice degeneration and retinal breaks using ultra-widefield fundus images. Annals of Translational Medicine 7:618–618. https://doi.org/10.21037/atm.2019.11.28
47. Zhang C, He F, Li B et al (2021) Development of a deep-learning system for detection of lattice degeneration, retinal breaks, and retinal detachment in tessellated eyes using ultra-wide-field fundus images: a pilot study. Graefes Arch Clin Exp Ophthalmol. https://doi.org/10.1007/s00417-021-05105-3
48. Antaki F, Coussa RG, Kahwati G et al (2021) Accuracy of automated machine learning in classifying retinal pathologies from ultra-widefield pseudocolour fundus images. British Journal of Ophthalmology
49. Li Z, Guo C, Lin D et al (2020) Deep learning for automated glaucomatous optic neuropathy detection from ultra-widefield fundus images. British Journal of Ophthalmology bjophthalmol-2020-317327. https://doi.org/10.1136/bjophthalmol-2020-317327
50. Masumoto H, Tabuchi H, Nakakura S et al (2018) Deep-learning Classifier With an Ultrawide-field Scanning Laser Ophthalmoscope Detects Glaucoma Visual Field Severity. J Glaucoma 27:647–652. https://doi.org/10.1097/IJG.0000000000000988
51. Matsuba S, Tabuchi H, Ohsugi H et al (2019) Accuracy of ultra-wide-field fundus ophthalmoscopy-assisted deep learning, a machine-learning technology, for detecting age-related macular degeneration. Int Ophthalmol 39:1269–1275. https://doi.org/10.1007/s10792-018-0940-0
52. Li Z, Guo C, Nie D et al (2021) Automated detection of retinal exudates and drusen in ultra-widefield fundus images based on deep learning. Eye 1–6. https://doi.org/10.1038/s41433-021-01715-7
53. Masumoto H, Tabuchi H, Nakakura S et al (2019) Accuracy of a deep convolutional neural network in detection of retinitis pigmentosa on ultrawide-field images. PeerJ 7:e6900. https://doi.org/10.7717/peerj.6900
54. Kim IK, Lee K, Park JH et al (2020) Classification of pachychoroid disease on ultrawide-field indocyanine green angiography using auto-machine learning platform. Br J Ophthalmol bjophthalmol-2020-316108. https://doi.org/10.1136/bjophthalmol-2020-316108
55. Nagasato D, Tabuchi H, Ohsugi H et al (2018) Deep Neural Network-Based Method for Detecting Central Retinal Vein Occlusion Using Ultrawide-Field Fundus Ophthalmoscopy. Journal of Ophthalmology 2018:1–6. https://doi.org/10.1155/2018/1875431
56. Nagasato D, Tabuchi H, Ohsugi H et al (2019) Deep-learning classifier with ultrawide-field fundus ophthalmoscopy for detecting branch retinal vein occlusion. Int J Ophthalmol 12:94–99. https://doi.org/10.18240/ijo.2019.01.15
57. Shi Z, Wang T, Huang Z et al (2021) A method for the automatic detection of myopia in Optos fundus images based on deep learning. International Journal for Numerical Methods in Biomedical Engineering e3460
58. Li Z, Guo C, Nie D et al (2020) Development and Evaluation of a Deep Learning System for Screening Retinal Hemorrhage Based on Ultra-Widefield Fundus Images. Translational Vision Science Technology 9:3. https://doi.org/10.1167/tvst.9.2.3
59. Dai L, Wu L, Li H et al (2021) A deep learning system for detecting diabetic retinopathy across the disease spectrum. Nat Commun 12:3242. https://doi.org/10.1038/s41467-021-23458-5
60. Calderon-Auza G, Carrillo-Gomez C, Nakano M et al (2020) A Teleophthalmology Support System Based on the Visibility of Retinal Elements Using the CNNs. Sensors 20:2838. https://doi.org/10.3390/s20102838

61. Li HH, Abraham JR, Sevgi DD et al (2020) Automated Quality Assessment and Image Selection of Ultra-Widefield Fluorescein Angiography Images through Deep Learning. Trans Vis Sci Tech 9:52–52. https://doi.org/10.1167/tvst.9.2.52
62. Li Z, Guo C, Nie D et al (2020) Deep learning from “passive feeding” to “selective eating” of real-world data. npj Digital Medicine 3:. https://doi.org/10.1038/s41746-020-00350-y
63. Ding L, Kuriyan AE, Ramchandran RS et al (2020) Weakly-Supervised Vessel Detection in Ultra-Widefield Fundus Photography Via Iterative Multi-Modal Registration and Learning. IEEE Trans Med Imaging 1–1. https://doi.org/10.1109/TMI.2020.3027665
64. Nunez do Rio JM, Sen P, Rasheed R et al (2020) Deep Learning-Based Segmentation and Quantification of Retinal Capillary Non-Perfusion on Ultra-Wide-Field Retinal Fluorescein Angiography. Journal of Clinical Medicine 9:2537. https://doi.org/10.3390/jcm9082537
65. Wang Z, Jiang X, Liu J et al (2020) Multi-Task Siamese Network for Retinal Artery/Vein Separation via Deep Convolution Along Vessel. IEEE Trans Med Imaging 39:2904–2919. https://doi.org/10.1109/TMI.2020.2980117
66. Sevgi DD, Srivastava SK, Wykoff C et al (2021) Deep learning-enabled ultra-widefield retinal vessel segmentation with an automated quality-optimized angiographic phase selection tool. Eye 1–6. https://doi.org/10.1038/s41433-021-01661-4
67. Ju L, Wang X, Zhou Q et al (2020) Bridge the Domain Gap Between Ultra-wide-field and. Traditional Fundus Images via Adversarial Domain Adaptation
68. Ju L, Wang X, Zhao X et al (2020) Leveraging Regular Fundus Images for Training. UWF Fundus Diagnosis Models via Adversarial Learning and Pseudo-Labeling
69. Xie H, Lei H, Zeng X et al (2020) AMD-GAN: Attention encoder and multi-branch structure based generative adversarial networks for fundus disease detection from scanning laser ophthalmoscopy images. Neural Netw 132:477–490. https://doi.org/10.1016/j.neunet.2020.09.005
70. Yoo TK, Ryu IH, Kim JK et al (2020) Deep learning can generate traditional retinal fundus photographs using ultra-widefield images via generative adversarial networks. Comput Methods Programs Biomed 197:105761. https://doi.org/10.1016/j.cmpb.2020.105761
71. Goodfellow I, Pouget-Abadie J, Mirza M et al (2014) Generative Adversarial Nets. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.
72. (2016) Generative Models. In: OpenAI. https://openai.com/blog/generative-models/. Accessed 18 Aug 2021
73. (2021) NVlabs/stylegan. NVIDIA Research Projects
74. Nagasato D, Tabuchi H, Masumoto H et al (2020) Prediction of age and brachial-ankle pulse-wave velocity using ultra-wide-field pseudo-color images by deep learning. Sci Rep 10:. https://doi.org/10.1038/s41598-020-76513-4
75. Wisely CE, Wang D, Henao R et al (2020) Convolutional neural network to identify symptomatic Alzheimer’s disease using multimodal retinal imaging. British Journal of Ophthalmology bjophthalmol-2020-317659. https://doi.org/10.1136/bjophthalmol-2020-317659
76. Parikh RB, Teeple S, Navathe AS (2019) Addressing Bias in Artificial Intelligence in Health Care. JAMA 322:2377–2378. https://doi.org/10.1001/jama.2019.18058
77. Selvaraju RR, Cogswell M, Das A et al (2020) Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. Int J Comput Vis 128:336–359. https://doi.org/10.1007/s11263-019-01228-7
78. Hanif AM, Beqiri S, Keane PA, Campbell JP (2021) Applications of interpretability in deep learning models for ophthalmology. Curr Opin Ophthalmol 32:452–458. https://doi.org/10.1097/ICU.0000000000000780

Tables

Table 1: Listing of all included studies, sorted alphabetically by publication author

Author	Title	Country of Origin	Category	Subcategory	Publication Title	DOI	Date of Publicat
Antaki *et al.*	Accuracy of automated machine learning in classifying retinal pathologies from ultra-widefield pseudocolour fundus images	Canada	Disease Detection and Classification	Retinal Detachment, Retinitis Pigmentosa, Retinal Vein Occlusion	British Journal of Ophthalmology	10.1371/journal.pone.0238958	August 2021
Bawany, M *et al.*	Automated vessel density detection in fluorescein angiography images correlates with vision in proliferative diabetic retinopathy	USA	Disease Detection and Classification	Diabetic Retinopathy	PLOS ONE	10.1371/journal.pone.0238958	Septem 11, 202
Cai, S *et al.*	Deep Learning Detection of Sea Fan Neovascularization From Ultra-Widefield Color Fundus Photographs of Patients With Sickle Cell Hemoglobinopathy	USA	Disease Detection and Classification	Sickle Cell Retinopathy	JAMA Ophthalmology	10.1001/jamaophthalmol.2020.5900	Decemb 30, 202
Calderon-Auza, G *et al.*	A Teleophthalmology Support System Based on the Visibility of Retinal Elements Using the CNNs	Mexico	Quality Assessment		Sensors	10.3390/s20102838	January 2020
Ding, L *et al.*	A Novel Deep Learning Pipeline for Retinal Vessel Detection In Fluorescein Angiography	USA	Segmentation and Localization	Vessel Segmentation	IEEE Transactions on Image Processing	10.1109/TIP.2020.2991530	January 2020
Ding, L *et al.*	Weakly-Supervised Vessel Detection in Ultra-Widefield Fundus Photography Via Iterative Multi-Modal Registration and Learning	USA	Segmentation and Localization	Vessel Segmentation	IEEE Transactions on Medical Imaging	10.1109/TMI.2020.3027665	Septem 29, 202
Ju, L *et al.*	Leveraging Regular Fundus Images for Training UWF Fundus Diagnosis Models via Adversarial Learning and Pseudo-Labeling	Australia	Generative Image Synthesis using GANs		IEEE Transactions on Medical Imaging	10.1109/TMI.2021.3056395	Novemb 27, 202
Kim, I *et al.*	Classification of pachychoroid disease on ultrawide-field indocyanine green angiography using auto-machine learning platform	South Korea	Disease Detection and Classification	Pachychoroid	British Journal of Ophthalmology	10.1136/bjophthalmol-2020-316108	July 3, 2020
Li Z *et al.*	Automated detection of retinal exudates and drusen in ultra-widefield fundus images based on deep learning	USA	Disease Detection and Classification	AMD	Eye	10.1038/s41433-021-01715-7	August 2021
Li, H *et al.*	Automated Quality Assessment and	China	Quality Assessment		Translational Vision Science &	10.1167/tvst.9.2.52	January 28, 202

	Image Selection of Ultra-Widefield Fluorescein Angiography Images through Deep Learning				Technology		
Li, Z *et al.*	Deep learning for automated glaucomatous optic neuropathy detection from ultra-widefield fundus images	China	Disease Detection and Classification	Glaucoma	British Journal of Ophthalmology	10.1136/bjophthalmol-2020-317327	Septemi 16, 202
Li, Z *et al.*	A deep learning system for identifying lattice degeneration and retinal breaks using ultra-widefield fundus images	China	Disease Detection and Classification	Retinal Detachment	Annals of Translational Medicine	10.21037/atm.2019.11.28	Novemb 1, 2019
Li, Z *et al.*	Deep learning for detecting retinal detachment and discerning macular status using ultra-widefield fundus images	China	Disease Detection and Classification	Retinal Detachment	Communications Biology	10.1038/s42003-019-0730-x	Decemb 1, 2020
Li, Z *et al.*	Development and Evaluation of a Deep Learning System for Screening Retinal Hemorrhage Based on Ultra-Widefield Fundus Images	China	Disease Detection and Classification	Retinal Hemmorhage	Translational Vision Science & Technology	10.1167/tvst.9.2.3	January 29, 202
Li, Z *et al.*	Deep learning from "passive feeding" to "selective eating" of real-world data	China	Quality Assessment		npj Digital Medicine	10.1038/s41746-020-00350-y	Decemb 1, 2020
Masumoto, H *et al.*	Deep-learning Classifier With an Ultrawide-field Scanning Laser Ophthalmoscope Detects Glaucoma Visual Field Severity:	Japan	Disease Detection and Classification	Glaucoma	Journal of Glaucoma	10.1097/IJG.0000000000000988	July 1, 2018
Masumoto, H *et al.*	Accuracy of a deep convolutional neural network in detection of retinitis pigmentosa on ultrawide-field images	Japan	Disease Detection and Classification	Retinitis Pigmentosa	PeerJ	10.7717/peerj.6900	May 7, 2019
Matsuba, S *et al.*	Accuracy of ultra-wide-field fundus ophthalmoscopy-assisted deep learning, a machine-learning technology, for detecting age-related macular degeneration	Japan	Disease Detection and Classification	AMD	International Ophthalmology	10.1007/s10792-018-0940-0	June 1, 2019
Nagasato, D *et al.*	Prediction of age and brachial-ankle pulse-wave velocity using ultra-wide-field pseudo-color images by deep learning	Japan	Systemic Diseases	Age and Brachial-Ankle Pulse-Wave Velocity	Scientific Reports	10.1038/s41598-020-76513-4	Decemb 1, 2020
Nagasato, D *et al.*	Deep-learning classifier with ultrawide-field	Japan	Disease Detection	BRVO	International Journal of Ophthalmology	10.18240/ijo.2019.01.15	January 18, 201

	fundus ophthalmoscopy for detecting branch retinal vein occlusion		and Classification				
Nagasato, D *et al.*	Deep Neural Network-Based Method for Detecting Central Retinal Vein Occlusion Using Ultrawide-Field Fundus Ophthalmoscopy	Japan	Disease Detection and Classification	CRVO	Journal of Ophthalmology	10.1155/2018/1875431	Novemb 1, 2018
Nagasawa *et al.*	Retinopathy Staging with a Deep Convolutional Neural Network Using Ultra-Wide-Field Fundus Ophthalmoscopy and Optical Coherence Tomography Angiography	Japan	Disease Detection and Classification	Diabetic Retinopathy	Journal of Ophthalmology	10.1155/2021/6651175	April 5, 2021
Nagasawa, T *et al.*	Accuracy of ultrawide-field fundus ophthalmoscopy-assisted deep learning for detecting treatment-naÃ¯ve proliferative diabetic retinopathy	Japan	Disease Detection and Classification	Diabetic Retinopathy	International Ophthalmology	10.1007/s10792-019-01074-z	October 2019
Nagasawa, T *et al.*	Accuracy of deep learning, a machine learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting idiopathic macular holes	Japan	Disease Detection and Classification	Idiopathic Macular Hole	PeerJ	10.7717/peerj.5696	October 22, 201
Nunez do Rio, J *et al.*	Deep Learning-Based Segmentation and Quantification of Retinal Capillary Non-Perfusion on Ultra-Wide-Field Retinal Fluorescein Angiography	UK	Segmentation and Localization	Vessel Segmentation	Journal of Clinical Medicine	10.3390/jcm9082537	August 2020
Oh, K *et al.*	Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images	South Korea	Disease Detection and Classification	Diabetic Retinopathy	Scientific Reports	10.1038/s41598-021-81539-3	January 21, 202
Ohsugi, H *et al.*	Accuracy of deep learning, a machine-learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting rhegmatogenous retinal detachment	Japan	Disease Detection and Classification	Retinal Detachment	Scientific Reports	10.1038/s41598-017-09891-x	Decemb 1, 2017
Sevgi, D *et al.*	Deep learning-enabled ultra-widefield retinal vessel segmentation with an automated	USA	Segmentation and Localization	Vessel Segmentation	Eye	10.1038/s41433-021-01661-4	January 2021

	quality-optimized angiographic phase selection tool						
Shi *et al.*	A method for the automatic detection of myopia in Optos fundus images based on deep learning	China	Disease Detection and Classification	Myopia	International Journal for Numerical Methods in Biomedical Engineering	10.1002/cnm.3460	March 2 2021
Tang, F *et al.*	Detection of Diabetic Retinopathy from Ultra-Wide Field Scanning Laser Ophthalmoscope Images: A Multi-Center Deep-Learning Analysis	China	Disease Detection and Classification	Diabetic Retinopathy	Ophthalmology Retina	10.1016/j.oret.2021.01.013	Februar 1, 2021
Wang, K *et al.*	Automated detection of diabetic retinopathy lesions on ultrawidefield pseudocolour images	USA	Disease Detection and Classification	Diabetic Retinopathy	Acta Ophthalmologica	10.1111/aos.13528	March 1 2018
Wang, Z *et al.*	Multi-Task Siamese Network for Retinal Artery/Vein Separation via Deep Convolution Along Vessel	China	Segmentation and Localization	Vessel Segmentation	IEEE Transactions on Medical Imaging	10.1109/TMI.2020.2980117	Septemi 1, 2020
Wisely, C *et al.*	Convolutional neural network to identify symptomatic Alzheimer's disease using multimodal retinal imaging	USA	Systemic Diseases	Alzheimer's Disease	British Journal of Ophthalmology	10.1136/bjophthalmol-2020-317659	Novemb 26, 202
Xie, H *et al.*	AMD-GAN: Attention encoder and multi-branch structure based generative adversarial networks for fundus disease detection from scanning laser ophthalmoscopy images	China	Generative Image Synthesis using GANs		Neural Networks	10.1016/j.neunet.2020.09.005	Decemb 1, 2020
Yoo, T *et al.*	Deep learning can generate traditional retinal fundus photographs using ultra-widefield images via generative adversarial networks	South Korea	Generative Image Synthesis using GANs		Computer Methods and Programs in Biomedicine	10.1016/j.cmpb.2020.105761	Decemb 1, 2020
Zhang, C *et al.*	Development of a deep-learning system for detection of lattice degeneration, retinal breaks, and retinal detachment in tessellated eyes using ultra-wide-field fundus images: a pilot study	China	Disease Detection and Classification	Retinal Detachment	Graefe's Archive for Clinical and Experimental Ophthalmology	10.1007/s00417-021-05105-3	Februar 4, 2021

Table 2: A full listing of included studies and their respective deep learning architectures, dataset sizes and imaging types, AUROC and/or AUPRC, sensitivity, specificity, precision, and accuracy when reported.

Author	Title	Architecture(s)	Total Dataset Size	Train/Validate/Test	AUROC or AUPRC	Sensitivity
Antaki et al	Accuracy of automated machine learning in classifying retinal pathologies from ultra-widefield pseudocolour fundus images	Google Cloud AutoML	2137 UWFI	80/10/10	**AUPRC:** RVO: 0.9670 RP: 0.9420 RD: 0.9210	RVO: 0.8491 RP: 0.8800 RD: 0.8977
Bawany, M et al.	Automated vessel density detection in fluorescein angiography images correlates with vision in proliferative diabetic retinopathy	U-NET	42 UWF-FA	19/0/81	**AUPRC:** 0.9300	
Cai, S et al.	Deep Learning Detection of Sea Fan Neovascularization From Ultra-Widefield Color Fundus Photographs of Patients With Sickle Cell Hemoglobinopathy	InceptionV4 CNN	1182 UWFI	70/10/20	**AUROC:** 0.988	0.9740
Calderon-Auza, G et al.	A Teleophthalmology Support System Based on the Visibility of Retinal Elements Using the CNNs	OD Detection: FR-CNN AlexNet OD Quality Analysis: VGG-16 Obstruction Analysis: SegNet Vessel Segmentation: SegNet with VGG-16	1288 UWFI	OD Detection and Quality Analysis: 70/0/30 Obstruction Analysis: 80/0/20 Vessel Segmentation: 90/0/10		OD Detection: 0.9643 OD Quality Analysis: 0.9113 Obstruction Analysis: N/A Vessel Segmentation: 0.7169
Ding, L et al.	A Novel Deep Learning Pipeline for Retinal Vessel Detection In Fluorescein Angiography	GAN with UNet followed by human-in-the-loop approach	8 UWFFA	k-fold cross validation (k=8)	**AUROC:** 0.987 **AUPRC:** 0.93	
Ding, L et al.	Weakly-Supervised Vessel Detection in Ultra-Widefield Fundus Photography Via Iterative Multi-Modal Registration and Learning	Custom DNN	PRIME-FP20 dataset: 15 concurrently captured UWFI and UWF-FA DRIVE dataset: 40 FP STARE dataset: 40 FP	PRIME-FP20 dataset: k-fold cross validation (k=5) DRIVE dataset: trained DRIVE, tested STARE STARE dataset: trained on STARE, tested on DRIVE	**AUPRC:** PRIME-FP20: 0.8450 DRIVE: 0.8860 STARE: 0.8840	
Ju, L et al.	Leveraging Regular Fundus Images for Training UWF Fundus Diagnosis Models via Adversarial Learning and Pseudo-Labeling	Transfer of FP to UWFI: modified Cycle-GAN Evaluation: Classification: ResNet-50 Detection: YOLOv3 Segmentation: U-Net with ResNet50	Classification: 552 UWFI; 2500 FP Detection: 242 UWFI; 1169 FP Segmentation: 175 UWFI; 1000 FP	60/15/25		Detection: 0.9101
Kim, I et al.	Classification of pachychoroid disease on ultrawide-field indocyanine green angiography using auto-machine learning platform	Google AutoML Vision	783 UWF-ICGA	80/10/10		Model 1: 0.8182 Model 2: 0.8636
Li, Z et al.	Automated detection of retinal exudates and drusen in ultra-	InceptionResNetV2	19891 UWFI	70/15/15	**AUROC:** CMAAI: 0.994 ZOC: 0.972 XOH: 0.988	CMAAI: 0.9420 ZOC: 0.9490 XOH: 0.9510

	widefield fundus images based on deep learning						
Li, H et al.	Automated Quality Assessment and Image Selection of Ultra-Widefield Fluorescein Angiography Images through Deep Learning	"U-Net style" CNN	3935 UWF-FAF	90/0/10	**AUROC:** Test Set 1 Gradeability: 0.9200	Test Set 1 Gradeability: 0.9050 Quality: 0.7890 Test Set 2 Gradeability: 0.9860 Quality: 1.000	
Li, Z et al.	Deep learning for automated glaucomatous optic neuropathy detection from ultra-widefield fundus images	InceptionResNetV2	22972 UWFI	70/15/15	**AUROC:** CMAAI: 0.9990 ZOC: 0.9830 XOH: 0.9900 TOPP: 0.9900	CMAAI: 0.9750 ZOC: 0.9790 XOH: 0.9820 TOPP: 0.9790	
Li, Z et al.	A deep learning system for identifying lattice degeneration and retinal breaks using ultra-widefield fundus images	InceptionResNetV2, InceptionV3, ResNet50, VGG16	5005 UWFI	70/15/15	**AUROC:** InceptionV3: 0.9970 ResNet50: 0.9890 InceptionResNetV2: 0.9990 VGG16: 0.9980	InceptionV3: 0.9870 ResNet50: 0.9680 InceptionResNetV2: 0.9870 VGG16: 0.9740	In I
Li, Z et al.	Deep learning for detecting retinal detachment and discerning macular status using ultra-widefield fundus images	InceptionResNetV2	11087 UWFI	70/15/15	**AUROC:** RD detection: 0.9890 Macular status classification: 0.9750	RD detection: 0.9610 Macular status classification: 0.9380	cl
Li, Z et al.	Development and Evaluation of a Deep Learning System for Screening Retinal Hemorrhage Based on Ultra-Widefield Fundus Images	InceptionResNetV2	16827 UWFI	70/15/15	**AUROC:** CMAAI: 0.9990 ZOC: 0.9980 XOH: 0.9970	CMAAI: 0.9890 ZOC: 0.9670 XOH: 0.9760	
Li, Z et al.	Deep learning from "passive feeding" to "selective eating" of real-world data	InceptionResNetV2	40562 UWFI	70/15/15	**AUROC:** CMAAI: 0.9960 ZOC: 0.9940 XOH: 0.9970	CMAAI: 0.9690 ZOC: 0.9560 XOH: 0.9660	
Masumoto, H et al.	Deep-learning Classifier With an Ultrawide-field Scanning Laser Ophthalmoscope Detects Glaucoma Visual Field Severity:	Custom DCNN	1399 UWFI	80/0/20	**AUROC:** Normal vs All: 0.872 Normal vs Early: 0.830 Normal vs Moderate: 0.864 Normal vs Severe: 0.934	Normal vs All: 0.813 Normal vs Early:0.838 Normal vs Moderate: 0.775 Normal vs Severe: 0.909	N
Masumoto, H et al.	Accuracy of a deep convolutional neural network in detection of retinitis pigmentosa on ultrawide-field images	VGG-16	373 UWFI 373 UWF-FAF	k-fold cross validation (k=5)	**AUROC:** UWFI: 0.9980 UWF-FAF: 1.000	UWFI: 0.9930 UWF-FAF:1.000	
Matsuba, S et al.	Accuracy of ultra-wide-field fundus ophthalmoscopy-assisted deep learning, a machine-learning technology, for detecting age-related macular degeneration	Custom DCNN	364 UWFI	70/0/30	**AUROC:** 0.9976	1.000	
Nagasato, D et al.	Prediction of age and brachial-ankle pulse-wave velocity using	VGG-16	170 UWFI (central, peripheral, total)	k-fold cross validation (k=5)			

	ultra-wide-field pseudo-color images by deep learning						
Nagasato, D et al.	Deep-learning classifier with ultrawide-field fundus ophthalmoscopy for detecting branch retinal vein occlusion	DL Model: VGG-16 SVM: Soft Margin	466 UWFI	k-fold cross validation (k=9)	**AUROC:** DL: 0.9760 SVM: 0.8570	DL:0.9400 SVM: 0.8050	
Nagasato, D et al.	Deep Neural Network-Based Method for Detecting Central Retinal Vein Occlusion Using Ultrawide-Field Fundus Ophthalmoscopy	DL Model: VGG-16 SVM: Soft Margin	363 UWFI	k-fold cross validation (k=9)	**AUROC:** DL: 0.9890 SVM: 0.8950	DL: 0.9840 SVM: 0.8400	
Nagasawa et al.	Accuracy of Diabetic Retinopathy Staging with a Deep Convolutional Neural Network Using Ultra-Wide-Field Fundus Ophthalmoscopy and Optical Coherence Tomography Angiography	VGG-16	491 UWFI and OCTA	k-fold cross validation (k=5)	**AUROC:** DR Detection: Optos: 0.7900 OCTA: 0.8830 Optos-OCTA: 0.8470 PDR Detection: Optos: 0.9810 OCTA: 0.9280 Optos-OCTA: 0.9640	DR Detection: Optos: 0.8090 OCTA: 0.8390 Optos-OCTA: 0.7860 PDR Detection: Optos: 0.9020 OCTA: 0.7450 Optos-OCTA: 0.8040	
Nagasawa, T et al.	Accuracy of deep learning, a machine learning technology, using ultra-wide-field fundus ophthalmoscopy for detecting idiopathic macular holes	Custom DCNN	910 UWFI	70/0/30	**AUROC:** 0.9993	1.000	
Nagasawa, T et al.	Accuracy of ultrawide-field fundus ophthalmoscopy-assisted deep learning for detecting treatment-naive proliferative diabetic retinopathy	VGG-16	378 UWFI	k-fold cross validation (k=9)	**AUROC:** 0.9690	0.9470	
Nunez do Rio, J et al.	Deep Learning-Based Segmentation and Quantification of Retinal Capillary Non-Perfusion on Ultra-Wide-Field Retinal Fluorescein Angiography	"U-Net style" CNN	75 UWF-FAF	k-fold cross validation (k=5)	**AUROC:** 0.82 **AUPRC**: 0.73	0.6609	
Oh, K et al.	Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images	ResNet-34	13271 UWFI	k-fold cross validation (k=10)	**AUROC:** ETDRS 7SF: 0.9150 ETDRS F1–F2: 0.8867	ETDRS 7SF: 0.8338 ETDRS F1–F2: 0.8060	E
Ohsugi, H et al.	Accuracy of deep learning, a machine-learning technology, using ultrawide-field fundus ophthalmoscopy	Custom CNN	831 UWF	80/0/20	**AUROC:** 0.988	0.9760	

	for detecting rhegmatogenous retinal detachment						
Sevgi, D et al.	Deep learning-enabled ultra-widefield retinal vessel segmentation with an automated quality-optimized angiographic phase selection tool	Custom CNN	RVA Extraction: 7787 UWF-FA patches Phase Selection: 13,980 UWF-FA sequences Cubic Spline Creation: 1578 UWF-FA sequences	90/0/10			
Shi et al.	A method for the automatic detection of myopia in Optos fundus images based on deep learning	MDNet	7141 UWFI	95/2.5/2.5			
Tang, F et al.	Detection of Diabetic Retinopathy from Ultra-Wide Field Scanning Laser Ophthalmoscope Images: A Multi-Center Deep-Learning Analysis	ResNet50	9392 UWFI	80/20/0	**AUROC:** Gradeability: 0.9230 RDR Detection: 0.9810 VTDR Detection: 0.9660 **AUPRC:** RDR Detection: 0.9990 VTDR Detection: 0.9370	Gradeability: 0.8650 RDR Detection: 0.9490 VTDR Detection: 0.8720	G
Wang, K et al.	Automated detection of diabetic retinopathy lesions on ultrawidefield pseudocolour images	EYEART	754 UWFI	50/0/50	**AUROC:** both eyes: 0.873 Unilateral Eye: 0.851	Bilateral Eyes: 0.917 Unilateral Eye: 0.903	
Wang, Z et al.	Multi-Task Siamese Network for Retinal Artery/Vein Separation via Deep Convolution Along Vessel	Custom CNN: Convolution Along Vessel (CAV) utilizing ResNet-18	WIDE: 30 UWFI DRIVE: 40 FP INSPIRE: 40 FP	k-fold cross validation (k=2)		WIDE: 0.9600 DRIVE: 0.9690 INSPIRE: 0.9730	
Wisely, C et al.	Convolutional neural network to identify symptomatic Alzheimer's disease using multimodal retinal imaging	ResNet18	284 UWF 284 UWF-FAF 284 GC-IPL 284 OCTA	AD: 35/10/55 Control: 21/11/68	**AUROC:** UWFI: 0.4500 UWFFA: 0.6180 GC-IPL: 0.8090 OCTA: 0.5820 All: 0.8290 All + quantitative data: 0.8300 All + all data: 0.8360 GC-IPL + all data: 0.8410		
Xie, H et al.	AMD-GAN: Attention encoder and multi-branch structure based generative adversarial networks for fundus disease detection from scanning laser ophthalmoscopy images	Custom GAN ResNet-34	SLO-1-2228 UWFI SLO-2: 2362 UWFI	80/0/20	**AUROC:** SLO-1: 0.8890 SLO-2: 0.9940	SLO-1: 0.8215 SLO-2: 0.9736	
Yoo, T et al.	Deep learning can generate traditional retinal fundus photographs using ultra-widefield images via generative	CycleGAN	451 UWFI 745 FI	90/0/10			

	adversarial networks						
Zhang, C et al.	Development of a deep-learning system for detection of lattice degeneration, retinal breaks, and retinal detachment in tessellated eyes using ultra-wide-field fundus images: a pilot study	seResNext50	911 UWFI	60/20/20	**AUROC:** Original Resizing Method: LD: 0.8880 RB: 0.8430 RD: 1.0000 Cropping Method: LD: 0.8410 RB: 0.9530 RD: 0.9790	Original Resizing Method: LD: 0.7710 RB: 0.6250 RD:0.8750 Cropping Method: LD: 0.8290 RB: 0.8750 RD: 0.8120	

Abbreviations: AUROC: area under receiver operating curve, AUPRC: area under precision recall curve, CAV: convolution along vessel, CMAAI: Chinese Medical Alliance for Artificial Intelligence, CNN: convolutional neural network, DNN: deep neural network, FI: fundus image, LD: lattice degeneration, OAG: open-angle glaucoma, OD: Optic Disc, RB: retinal breaks, RD: retinal detachment, RD: Retinal detachment, RP: Retinitis Pigmentosa, RVO: Retinal Vein Occlusion, SLO: Scanning Laser Ophthalmoscopy, SVM: support vector machine, TOPP: Tsukazaki Optos Public Project, UWF-FA: ultra-widefield fluorescein angiography, UWF-FAF: ultrawide field fundus autofluorescence, XOH: Xudong Ophthalmic Hospital, ZOC: Zhongshan Ophthalmic Center., UWFI: ultrawide field image(s), UWF-ICGA: UWF indocyanine green angiography

Figures

Figure 1

Comparison of Optos Ultra-Widefield Imaging (200 degrees field of view) to Colour Fundus Photography (45 degrees field of view). A: Optos ultra-widefield optomap colour image of the left fundus of a patient with diabetic retinopathy with an overlaid colour fundus photograph from the same patient eye over the optic disc and macula region.

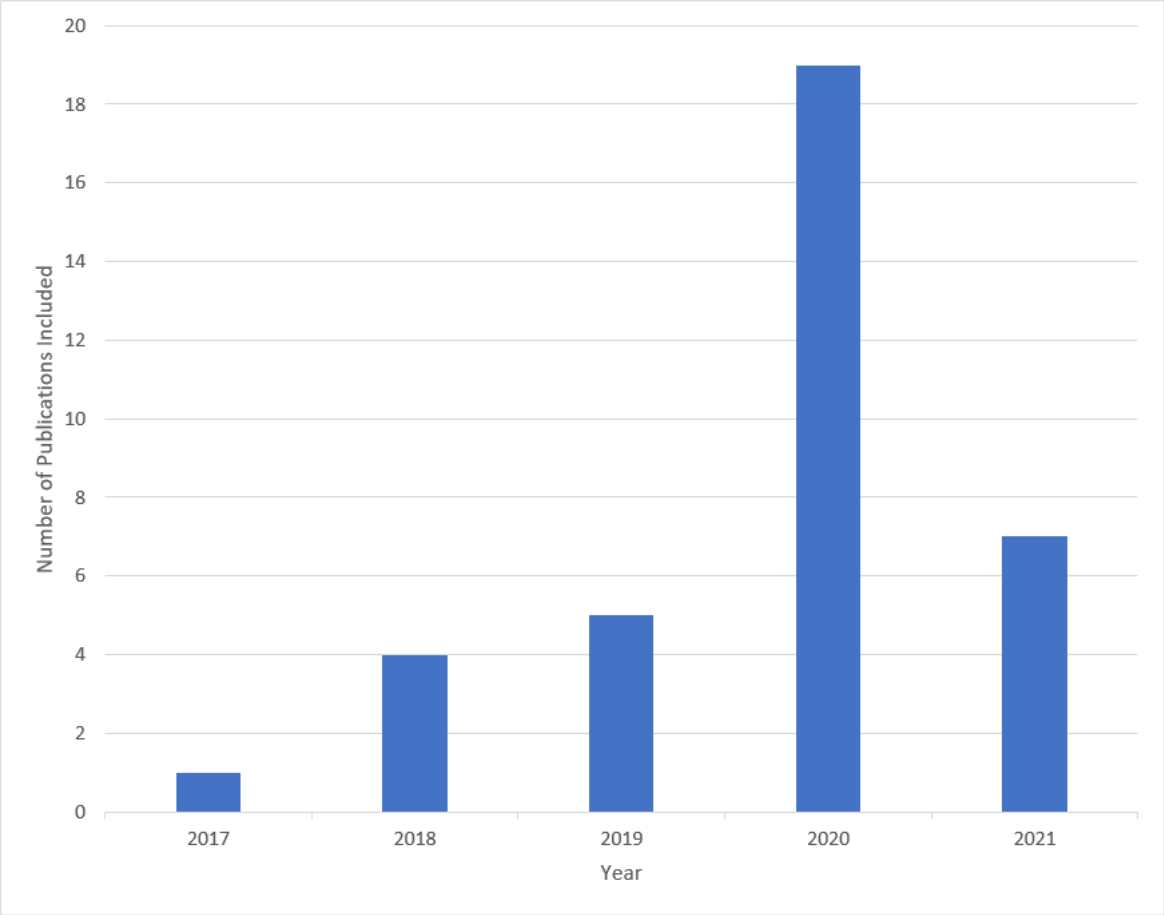

Figure 2

Number of included studies by year of publication

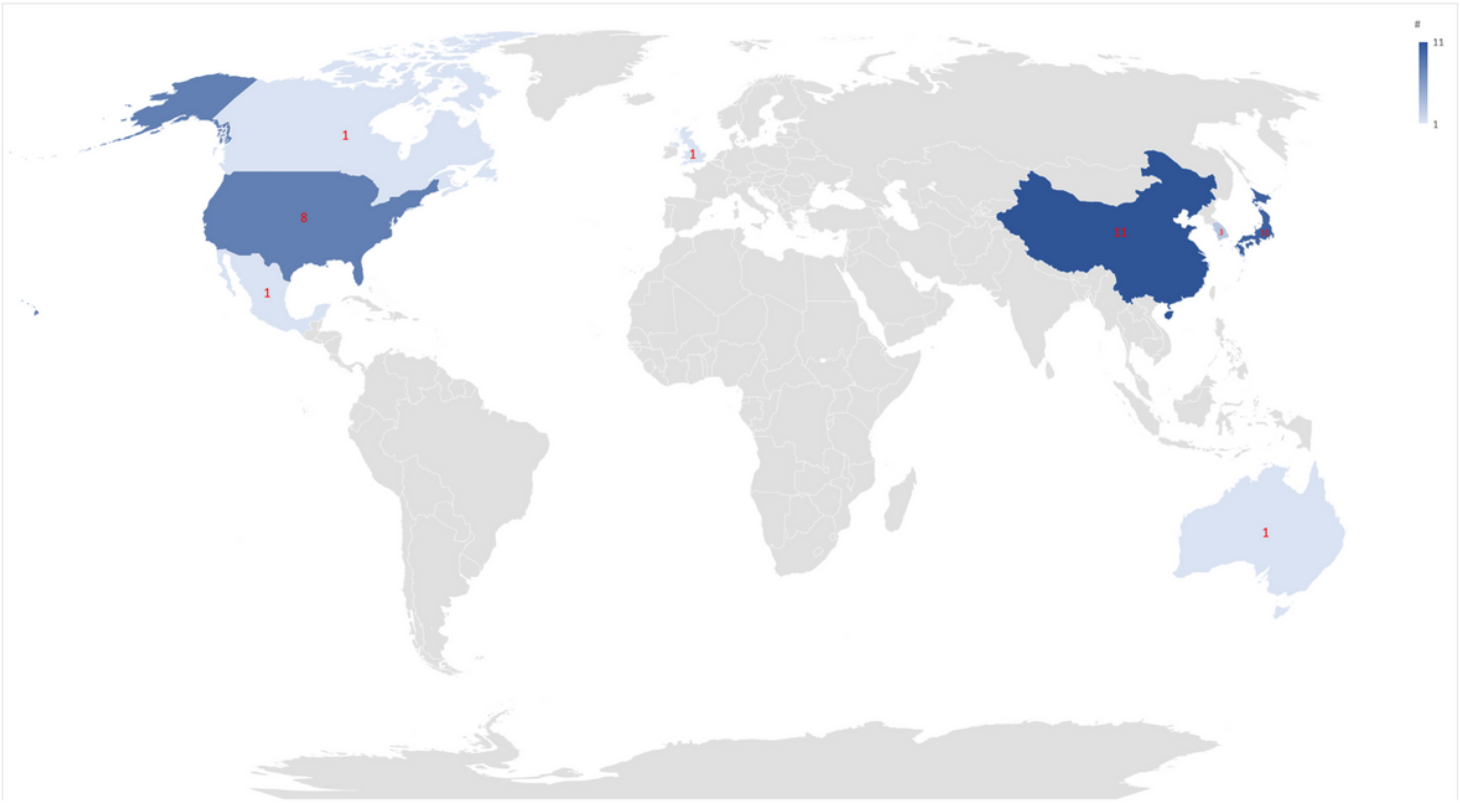

Figure 3

The countries of origin of all included studies. Each country with an included study is highlighted in blue, with the number of studies from each written in red.