This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2023.0322000

Deep Learning Using Preoperative Optical Coherence Tomography Images to Predict Visual Acuity Following Surgery for Idiopathic Full-Thickness Macular Holes

BURAK KUCUKGOZ[1,5], MUHAMMED MUTLU YAPICI[4], Declan C Murphy[2], Emma Spowart[3], David H. Steel[2], and Boguslaw Obara[1,2]

[1]School of Computing, Newcastle University, Newcastle upon Tyne, NE4 5TG UK
[2]Bioscience Institute, Newcastle University, Newcastle upon Tyne, NE2 4HH UK
[3]Sunderland Eye Infirmary, National Health Service, Sunderland, SR2 9HP UK
[4]Elmadag Vocational School, Ankara University, Ankara, 06780 TURKEY
[5]Institute of High-Performance Computing, A*STAR, Singapore, 138632, SINGAPORE

Corresponding author. Email: boguslaw.obara@newcastle.ac.uk

This work was supported in part by the Turkish Ministry of National Education, the Scientific and Technological Research Council of Turkey (TUBITAK) 2219-International Postdoctoral Research Fellowship Program (App. No:1059b192100378), and in part by Newcastle University

ABSTRACT

This study presents a fully automated image informatics framework. The framework is combined with a deep learning (DL) approach to automatically predict visual acuity outcomes for people undergoing surgery for idiopathic full-thickness macular holes using 3D spectral-domain optical coherence tomography (SD-OCT) images. To overcome the impact of high variation in real-world image quality on the robustness of DL models, comprehensive imaging data pre-processing, quality assurance, and anomaly detection procedures were utilised. We then implemented, trained, and tested nine state-of-the-art DL predictive models through our designed loss function with multiple 2D input channels on the imaging dataset. Finally, we quantitatively compared the models using four evaluation metrics. Overall, the predictive model achieved a MAE of 6.47 ETDRS letters score, demonstrating high predictability. This confirms that our fully automated approach with input from seven central SD-OCT images from each patient can robustly predict visual acuity measurements. Further research will focus on adapting 3D DL-based predictive models and the uncertainty of 2D and 3D DL-based predictive models.

INDEX TERMS Image analysis, machine learning, optical coherence tomography, visual acuity measurement.

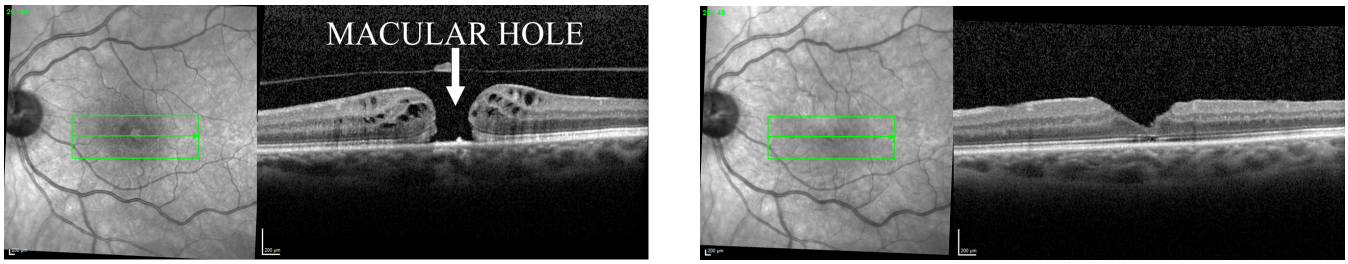
I. INTRODUCTION

IDIOPATHIC full-thickness macular holes (MHs) form secondary to age-related abnormalities of the vitreoretinal interface with a prevalence of up to 3 in 1000 people over the age of 55 [1]. They appear as a small dehiscence in the neurosensory retina at the centre of the fovea, a highly specialised part of the human retina responsible for fine acuity and colour vision [2]. Today's technology allows ophthalmologists to diagnose, classify and measure MHs using spectral-domain (SD) optical coherence tomography (OCT) scans. OCT is a non-invasive, high-resolution imaging technique that uses infrared light to provide 3D imaging of the retina [3] (see Fig. 1).

Macular holes can be effectively treated by closing the hole using vitrectomy surgery. Predicting the visual outcome after surgery is important to guide the decision to operate and manage patients' expectations. Several studies [4]–[6] have shown that postoperative visual acuity (VA) is highly correlated with preoperative VA, as well as a variety of measures of macular hole size that can be measured on SD-OCT. Various studies have attempted to precisely predict postoperative VA using manual 2D measurements of MHs and preoperative VA, although their predictive ability has been limited [7]. Three-dimensional automated image reconstruction has improved this ability [8], [9], but there are no current standards for shape, size, and resolution of

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676



(a) (b)

FIGURE 1: A preoperative 2D slice of a 3D SD-OCT image of a patient's eye with an idiopathic full-thickness macular hole and visual acuity (VA) of 42 ETDRS letters (a), and the postoperative 2D slice of a 3D SD-OCT image after successful surgery with closure of the hole, restoration of the foveal depression and a VA of 71 ETDRS letters (b).

OCT imaging data captured by different OCT devices for this task [10]. There are also many qualitative features and subtle alterations in retinal anatomy, for example, associated with chronicity, which may be predictive of acuity outcomes and that are difficult to measure [11], [12]. Additionally, image artefacts related to a patient's eye movement and media opacity pose a further challenge in developing image informatics methods [13]. Recently, some researchers have highlighted the low signal strength of OCT devices which results in issues such as image noises, blurriness and contrast reduction [14], [15]. Similarly, another study expanded the analysis to include scan centring and retinal region checks [16]. These challenges constitute our primary motivation.

To overcome those challenges, most machine learning (ML) and deep learning (DL) approaches have focused on the automated classification of macular diseases, such as age-related macular degeneration (AMD), diabetic macular oedema (DME), and MHs from OCT images data [17]–[22]. More recently, some DL approaches have improved the prediction of VA outcomes [23] using OCT data [24]–[26]. In particular, convolutional neural network (CNN) models have achieved high performance in OCT image analysis studies; however, there have only been a limited number of studies investigating VA measurements [23], [24], [27]. Considering the success of prominent CNN-based networks in medicine [28]–[30], they used a ResNet [31] in the [23], VGG [32] in the [33], and CBR-Tiny models [34] in the [27] as a backbone. These studies also presented that CNN-based networks excel in extracting spatial features from OCT images. Consequently, the implementation of CNN-based models for predicting VA measurements has gained significant importance for the next motivation. Subsequently, vision transformers (ViTs) [35] have recently demonstrated great potential in assisting clinicians with clinical diagnosis [36], particularly in OCT image analysis [37]. However, to the best of our knowledge, ViTs have not yet been thoroughly applied for predicting VA measurements. Since ViTs consider global context and dynamic attention, it revealed the need for comparing standard CNN-based state-of-art models (VGG, ResNet, Inception v3 and DenseNet, EfficientNetV2) and ViTs in this study, acting as another motivation.

This research presents a comprehensive image informatics framework for predicting both the preoperative and postoperative visual acuity measurements for patients with idiopathic full-thickness macular holes using an SD-OCT image dataset based on image preprocessing, image quality assessment, image anomaly detection, and deep learning models-based prediction.

The remainder of the article is organised as follows: Section II describes related works and our contributions, including regression models, classification models and predictive models of VA measurements based on image analysis approaches. In Section III, the benchmark OCT imaging dataset used in this study is introduced. Section IV summarises image preprocessing, quality measurement, anomaly detection, and the DL models-based visual acuity prediction methodology and experimental design. Section V details the study results, and Section VI concludes the study findings and provides recommendations for future research.

II. RELATED WORKS

Numerous image informatics approaches to assess macular diseases using OCT imaging data have been proposed in the published literature (see the summary in Table 1). As shown in Table 1, there has been only limited research investigating how to predict VA outcomes for specific retina diseases using OCT imaging data. In particular, the published literature aiming to predict postoperative vision for patients with idiopathic full-thickness macular holes is particularly sparse [24], [26]. We discuss the approaches that have been used to date.

A. IMAGE BASED CLASSIFICATION APPROACHES FOR DIFFERENT RETINAL DISEASES

Zhang et al. [18] proposed a binary classification of OCT image data based on kernel principal component analysis (PCA) model ensembles to predict patients with AMD-affected eyes from normal eyes. Also, a Bayesian network classifier was introduced by [39] and then tested on the same image dataset. Another study implemented the bag-of-words (BoW) model by keeping the most salient points corresponding to the top vertical gradient values calculated

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676

TABLE 1: The summary of the image informatics approaches focused on assessing visual acuity using OCT imaging data.

AUTHORS	TASK	DATASET	METHODS	DISEASE TYPE	OUTPUT	EVALUATION METHODS	RESULTS (Avg.)
Steel et al. [6]	Regression	3D OCT (1527 images)	Classical Statistics	MH	Postoperative VA	AUC	71.72%
Murphy et al. [17]	Regression	3D OCT (67 images)	Generalized Linear Model	MH	Postoperative VA	R-squared	0.45
Kawczynski et al. [23]	Regression Classification	3D OCT (1071 images)	ResNet-50	Others	BCVA 2 classes	RMSE AUC	9.01 0.91%
Obata et al. [24]	Regression Classification	3D OCT (259 images)	CNN	MH	BCVA 4 classes	R-squared Precision	0.46 75%,43%, 38%,50%
Rizzo et al. [26]	Classification	3D OCT-A (35 images)	Inception v3, VGG16, VGG19, SqueezeNet	MH	Postoperative VA	P value	0.005
Lachance et al. [27]	Classification	2D OCT (121 images)	CBR-Tiny	MH	Postoperative VA 2 classes	Accuracy	78.7±2.9%
Xu et al. [38]	Regression	3D OCT (56 images)	Segmentation Algorithm	MH	Geometric measurements and Postoperative VA	P value	0.0028
The proposed method	Regression	3D OCT (210 images)	ResNet-18 ResNet-50	MH	Postoperative VA	MAE R-squared RMSE	**6.00** **0.52** 9.23

in the OCT images [40]. However, this approach was limited by relying on key points and predicting only two classes: DME and normal eyes. Anantrasirichai et al. [19] proposed a support vector machine (SVM)-based approach to differentiate between normal eyes and eyes with glaucoma. Similarly, [41] used SVM to predict the presence of AMD and DME from normal eyes using a small image data set with fewer outliers. However, the obtained accuracy was extensively impacted by retinal layer discontinuities caused by the disease pathology and motion artefacts.

In addition, Lemaitre et al. [42] developed a local binary pattern (LBP) classifier to identify AMD and DME from normal eyes. Liu et al. [43] proposed an approach using image gradient information, LBP, and SVM with a radial basis function (RBF) kernel as a classifier. The approach first classified eyes as either normal or abnormal, and then was further sub-classified into either DME, AMD, MH or normal eyes.

Motozawa et al. [20] proposed two typical CNN-based models for classifying OCT data into AMD and normal, and they were also able to differentiate between wet and dry AMD. Likewise, Alqudah et al. [21] showed that by using image denoising and resizing and tuning a CNN model with an ADAM optimiser, higher accuracy and lower time cost could be achieved when classifying OCT images into five classes: choroidal neovascularisation (CNV, a feature of wet AMD), DME, dry AMD, drusen only (a feature of early AMD), and normal.

Another OCT image classification approach using a deep multi-scale CNN model was proposed by Rasti et al. [44]. The proposed model employed a prior decomposition and new cost function to discriminate and fast-learn representative image features. The authors used the modified versions of VGG, ResNet, and Inception models to detect normal, AMD, and DME features. Li et al. [45] suggested a novel DL model for predicting CNV, DME, drusen, and normal eyes, called OCTD_Net and based on modified DenseNet and ReLayNet models. Tsuji et al. [22] proposed a method using a capsule neural network (CapsNet) model to classify the same eye disorders by learning spatial information from the OCT images.

These image informatics approaches have several limitations, including (1) a limited ability to identify different pathologies affecting the macula, such as MHs, (2) they are typically time-consuming due to high computation costs, which means they are inappropriate for use in clinical practice where issues may need to be resolved in real-time, and (3) by only relying on a limited number of key points, other essential ocular characteristics may not be noticed during the classification tasks.

B. IMAGE BASED PREDICTION APPROACHES FOR VISUAL ACUITY AFTER MACULAR HOLE SURGERY

1) Classical Methods

Several authors have used regression to predict postoperative VA using routinely collected clinical data. For example, Steel et al. [6], using logistic regression and the univariate level using $\tilde{\chi}^2$ tests, achieved a model area under the receiver operator curve of 71.72% for predicting a visual acuity of 0.3 logMAR or better after surgery. Generalised linear modelling has been used to predict actual acuity using an automated multi-scale three-dimensional (3D) image analyser of OCT scans for MHs [17]. The study shows preoperative VA and MH height were important predictors of postoperative VA, achieving an R-squared value of 0.45. When preoperative vision was not included in the model and only OCT parameters were included, the most predictive model was 0.39. Interestingly, using only manual clinician-measured values, R-squared was only 0.20.

Other research teams have also investigated the three-dimensional parameters of MH using different methodologies, such as automatically calculating three dimensions based on

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676

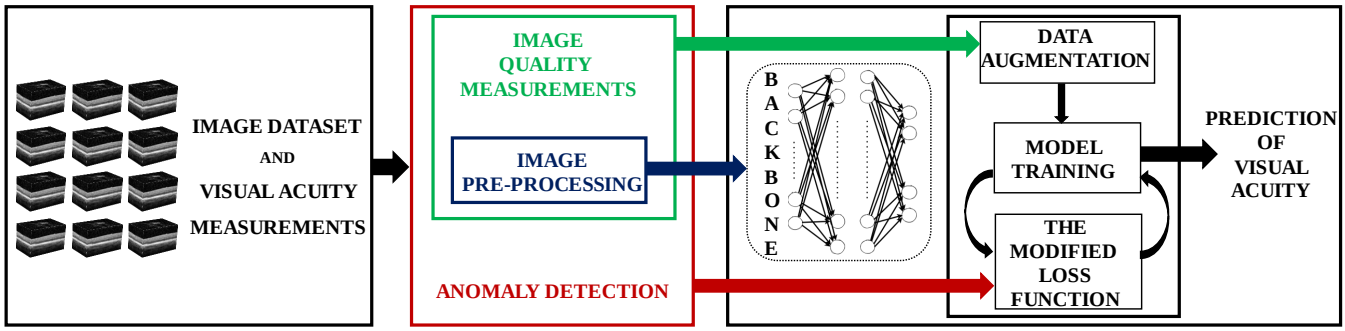


FIGURE 2: Workflow of the proposed image informatics framework. The first stage corresponds to the input OCT image dataset and VA measurements obtained by ophthalmologists, the second stage incorporates OCT data preparation (i.e. scaling, the centre of mass detection, and cropping), OCT image quality analysis (i.e. noise score, blurriness score, contrast score, motion score, and brightness-darkness score) and anomaly detection. With the obtained high-quality image dataset and labels, multiple state-of-the-art DL models are trained and optimized by our designed loss function to predict VA measurements in the final stages.

the sum of two-dimensional images [38]. The 3D macular hole size parameters, such as MH volume, base area, base diameter, and MH height, were significantly correlated to postoperative VA (P value from 0.0003–0.011). [38].

2) Deep Learning Based Methods

Some recent studies were not only able to classify eye disorders on OCT image datasets, but were also able to predict associated VA measurements and recommend potential treatment requirements. In particular, the study by [46] presented an end-to-end DenseNet-based model for recommending treatment options in patients with wet AMD, where the model's output range was low, intermediate and high treatment requirement scores. In a further study, Kawczynski et al. [23] proposed a ResNet-50 v2 model-based approach that predicted the best-corrected VA (BCVA) measurement for patients with wet AMD eyes following treatment. BCVA measurement was obtained from the regression model, and considering the regression model results, they classified higher than 69 letters and lower than 69 letters into two classes.

The research presented in [26] assessed the ability to predict VA in two groups of 35 people with surgically treated MH using unsupervised DL models, including Inception v3, VGG16, VGG19 and SqueezeNet. Similarly, Lachance et al. [27] proposed a hybrid model classifying VA as higher than 15 letters and lower than 15 letters. Another study proposed a model for predicting postoperative VA using a typical CNN-based model [24] and four classes: class A is higher than 85 letters, class B is between 75 and 80 letters, class C is between 60 and 75 letters, and class D is lower than 50 letters. The prediction of postoperative VA using DL was compared to 3 typical regression models using preoperative clinical data. The DL model had a superior precision value of 46% compared to approximately 40% for the regression models.

These DL-based image informatics approaches have limitations related to (1) image and label data preparation, (2) data volume, (3) data quality, and (4) low-level model robustness and generalisation when using a wide range of OCT machines at different hospitals. To address some of these limitations, we recently presented a more comprehensive image informatics framework utilising robust data preparation and anomaly detection approaches combined with state-of-art DL models on a closely allied OCT analysis problem of external limiting membrane detection [11].

Contributions: This paper adds to the literature by:

- Introducing a new 3D SD-OCT imaging benchmark dataset for 210 patients with idiopathic full-thickness macular holes (10,339 2D slices).
- Proposing a comprehensive image informatics framework to create a high-quality OCT image dataset used for a robust deep learning-based predictive model of visual acuity in patients following surgery with idiopathic full-thickness macular holes and presenting an automated solution for non-standardised OCT datasets (see Fig. 2). The method concludes the impact of the following surgery by predicting visual acuity.
- Quantitatively comparing nine 2D state-of-the-art DL-based predictive models of both preoperative and postoperative visual acuity using four evaluation metrics by optimizing the models with our designed loss function. To account for the 3D nature of the eye captured in 3D OCT imaging data, we used multiple image slices during the training phase.

III. MATERIALS

The proposed image informatics framework was designed, implemented and evaluated on two sets of SD-OCT imaging datasets, all captured using the Heidelberg Spectralis (Heidelberg, Germany) using the same imaging protocol at Sunderland Eye Infirmary, United Kingdom (UK) and Rigshospitalet, Copenhagen, Denmark (see full OCT image with fundus region Fig. 1 and without fundus region Fig. 3). Three different Spectralis cameras were used in the UK centre and one in Denmark. The images from the UK were collected as part of routine care between Jan 2017 and Jan 2021 under UK guidelines, and their use did not require ethical approval.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676

The images from Denmark were obtained from a previously published randomised controlled trial that obtained ethical approval (protocol Number: H-4-2013-091, Rigshospitalet, Copenhagen) and full informed consent was received from all participants.

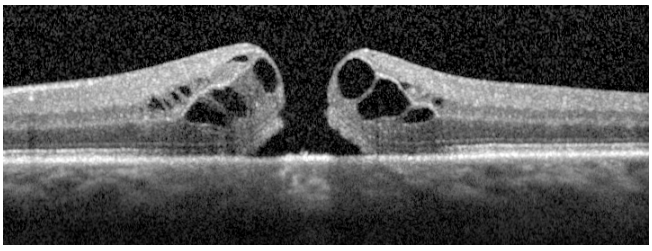


FIGURE 3: A random 2D slice from one of our 3D OCT images used during training and testing.

We included patients with a confirmed idiopathic full-thickness macular hole (on OCT) who had undergone vitrectomy and internal limiting membrane (ILM) peeling with gas tamponade surgery, successfully achieved primary hole closure (hole closure following a single surgery) evidenced by standardised OCT imaging two weeks after surgery, and had a best-corrected visual acuity recording at three months ($\pm$ two weeks) postoperatively. All patients were pseudophakic postoperatively, which prevents the confounding influence of cataracts on the visual acuity measurement. Patients who were phakic before surgery underwent combined phacovitrectomy if they were from the UK cohort, and phacoemulsification and intra-ocular lens insertion surgery one week before vitrectomy surgery if from the Denmark group.

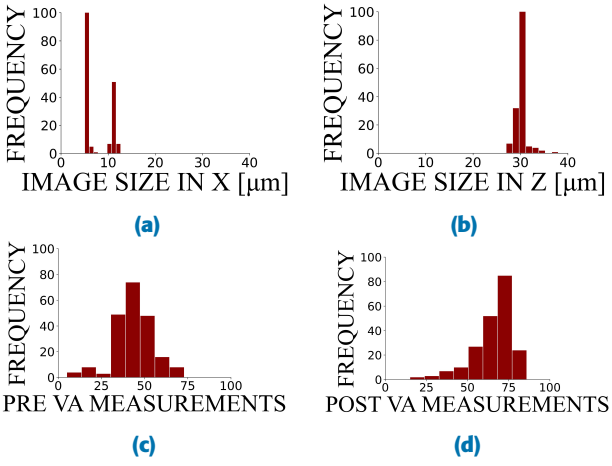


FIGURE 4: The distribution of OCT image sizes in *X* (a), *Z* (b), distribution of preoperative (c) and postoperative (d) visual acuity measurements. Image size in *Y* is $3.87\ \mu m$ for all images.

We excluded all secondary holes, non-full thickness holes, eyes with previous vitrectomy surgery and/or non-primary closure, and eyes with other co-existing causes for reduced vision, for example, AMD or amblyopia. This is because different medical treatments or operations might be needed. However, the techniques could be applied to other types of MH. For both image sets, the same standardised imaging protocol was used, namely a high-density central horizontal scanning protocol with 29–30 μm (microns) line spacing in the central 15 by 5 degrees. With 27–34 μm spacing between scans (Z-axis), there were typically 49 scans per dataset. The captured OCT images, however, had variable pixel widths, heights, and depths (X is from 178 to 497 px (pixels), Y is from 321 to 776 px, and Z is from 49 to 96 px) relating to different captured image resolutions. The pixel resolutions of the OCT images were, therefore, between 5.04 and $12.66\ \mu m$ per pixel in-width, but the same resolution of $3.87\ \mu m$ per pixel in-height (see Fig. 4a and 4b).

Image dataset and clinical data on 210 eyes from 210 patients meeting the inclusion and exclusion criteria were analysed: 67 from Denmark and 143 from the UK. The mean age was 70 years old (range 48–84), 172 (82%) were female, and 105 (50%) were right eyes. The mean minimum linear diameter of the holes was 383 μm, and the median duration of symptoms was six months. All scans used 16 automatic real-time settings, enabling multi-sampling and noise reduction over 16 images. While capturing the OCT images, the patients were asked to focus on a constant fixation object to minimise eye motion as a general procedure.

Visual acuity was measured in all patients using Early Treatment Diabetic Retinopathy Study (ETDRS) vision charts, with testing at four metres [47]. These charts have a group of five letters per row, with multiple rows with a reducing letter size of 0.1 logMAR per line. The VA measurement is calculated by how many letters can be correctly read on the chart. A score of 70 letters equates to 0.3 logMAR (or 20/40 Snellen visual acuity), whilst 35 letters equate to 1.0 logMAR (or a Snellen acuity of 20/200). The VAs were recorded by two experienced optometrists in a clinic unassociated with the study and were best-corrected VAs after refraction using a standardised protocol. It is ranged from 5 to 83, outlined in Fig. 4c and 4d. Our image pixel resolutions and VA measurements have imbalanced distributions.

IV. METHODS

In this section, we present a comprehensive description of the imaging data preprocessing steps, data quality assessment and anomaly detection methods to create a high-quality standardised 3D OCT image dataset for DL-based prediction of VA. This comprehensive approach significantly improved our proposed model's results.

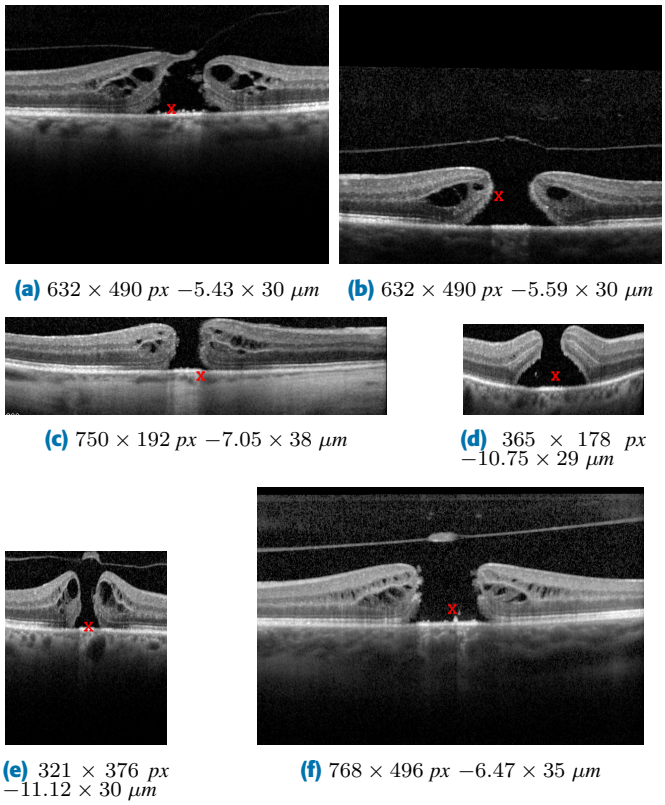
A. IMAGING DATA PREPARATION

1) Image Preprocessing

In 3D OCT images, due to ocular anatomy and acquisition distortions, the MHs may be scaled, shifted, and oriented randomly. Consequently, this causes high variability in the MH location and resolution, as shown in Fig. 5. To deal with those image acquisition issues, we used the following image preprocessing steps:

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676



(a) $632 \times 490\ px\ -5.43 \times 30\ \mu m$ **(b)** $632 \times 490\ px\ -5.59 \times 30\ \mu m$

(c) $750 \times 192\ px\ -7.05 \times 38\ \mu m$ **(d)** $365 \times 178\ px\ -10.75 \times 29\ \mu m$

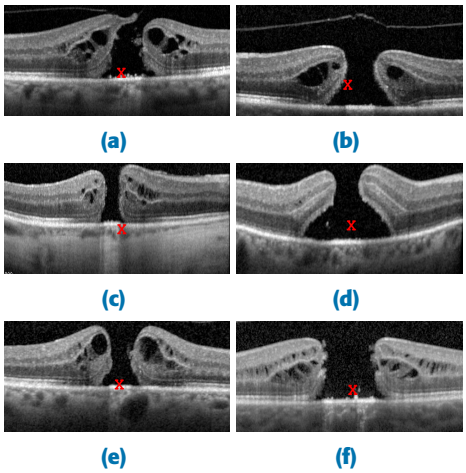
(e) $321 \times 376\ px\ -11.12 \times 30\ \mu m$ **(f)** $768 \times 496\ px\ -6.47 \times 35\ \mu m$

FIGURE 5: 2D mid-slices of six randomly selected 3D images, with their calculated centres of mass (red crosses), and with their corresponding image sizes in px and μm.

- Scaling (uniform resolution): All acquired OCT images (Fig. 5) were re-scaled across X, Y, and Z dimensions using the following sizes ($7.41 \times 3.87 \times 30.1\ \mu m$), and the resulting images are shown in Fig. 6. It ensured consistency in scale values for all OCT images.
- Intensity Weighted Centre of Mass: The MHs were located in a range of different positions in the 3D OCT image slices, as presented in Fig. 5. To centre the images around the positions of the MHs, we used pixel intensity weighted centres of mass calculated for each dimension as a focal point of the image, as shown in Fig. 6. It was used to ensure consistency in image acquisition position for all OCT images. The determining centres of mass were also considered when selecting the parameter for the data augmentation stage during the training of the DL model.
- Cropping: The scaled OCT images were then centred around intensity-weighted centres of mass (red cross) and cropped across the X, Y, and Z dimensions to the 3D size $452 \times 204 \times 49\ px$, as shown in Fig. 6). It ensured consistency in sizes for all OCT images.

2) Image Quality Assessment

The OCT images collected as part of routine clinical care inevitably differed in image quality related to patient movements, operator controls, and the OCT camera used.

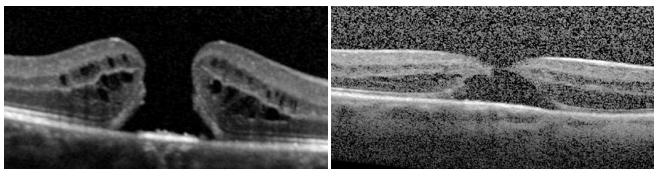


(a) **(b)** **(c)** **(d)** **(e)** **(f)**

FIGURE 6: Results of image preprocessing steps applied to images from Fig. 5. Final image size: ($452 \times 204 \times 49\ px\ -7.41 \times 3.87 \times 30.1\ \mu m$).

These resulted in several image imperfections, including speckle noise, contrast changes, and motion artefacts, which we measured using a variety of image quality measurement methods. As these methods mostly have no inherent upper limit, we scaled all features separately to a fixed range from 0 to 1. High scores mainly denote a heightened presence of the measured imperfection. To enable a high-quality dataset to be selected and to optimise the DL model, these methods served as a guide to detect and remove abnormal images. We measured image quality using the following evaluation metrics.

- Noise score: Noise can be a significant problem in OCT images. Many researchers have proposed wavelet transformations to assess the lower, average, and upper bound of noise in these images [48], [49]. We performed a wavelet-based estimator of the Gaussian noise standard deviation, which revealed significant noise variances, see Fig. 7.



(a) Low **(b) High**

FIGURE 7: Spectral-domain optical coherence tomography images demonstrating a small and large noise score.

- Blurriness score: Another important issue was the blur and sharpness of the OCT images. Recent studies have suggested the use of a Laplacian operator with Gaussian filters in measuring the blurriness and sharpness of images [50]–[52]. The Gaussian filter is defined as in Equation 1.

$$G(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x^2+y^2)}{2\sigma^2}}, \tag{1}$$

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676

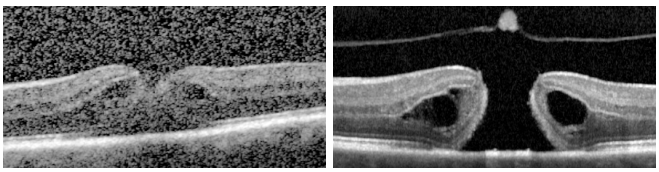
where x and y are coordinates of an image $I(x, y)$. σ is the Gaussian distribution standard deviation. The Gaussian scale-space representation L of an image $I(x, y)$ is defined as in the Equation 3.

$$L(x, y) = I(x, y) * G(x, y) \tag{2}$$

where $*$ is the convolution operator. Then, a Laplacian operator, which is expressed as the divergence of the gradient (∇I), and calculated for two dimensions as a sum of the second partial derivatives in the Cartesian coordinates:

$$\nabla^2 = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}, \tag{3}$$

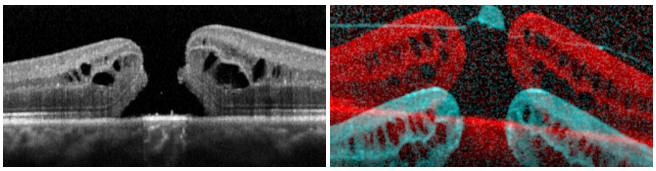
This is convolved with the image, resulting in rapid intensity changes. The variance in image intensity is then measured, where if low it is labelled as *blurry*, otherwise it is measured as *sharp* (see Fig. 8).



(a) Low **(b) High**

FIGURE 8: Spectral-domain optical coherence tomography images demonstrating blurriness and sharpness.

- Contrast score: The differences in the chromaticity and brightness of any pixel and any other pixels within the same scene represent image contrast [53]. We therefore measured the gradients of each image pixel, including their standard deviation, with marked differences in black and white luminance.
- Motion score: Due to the movement of the eye during OCT scanning, consideration has to be given to motion artefacts. Many researchers have proposed the Horn-Schunck optical flow motion estimation method [54]–[56]. This method implements first-order derivatives, allowing the velocity in the flow between sequential images to be measured with high accuracy and resolution. We measured the observed motion and perceived distortions in the smooth flow of information on the Z-axis of every 3D OCT image, as seen in Fig. 9. An RGB colour map was used for motion visualisation, with grey representing low motion and red or blue representing high motion between any two neighbouring 2D slices in a 3D image.
- Brightness-Darkness score: The brightness and darkness of images are associated with perceived luminance. Therefore, we calculated a luminance measurement [57]. Darkness is perceived if the luminance level is low, whereas brightness is perceived if the luminance is high.



(a) Low **(b) High**

FIGURE 9: An example of optical flow velocity vector magnitudes between two neighbouring 2D slices in a spectral-domain optical coherence tomography image with ((a) - grey colour) small and ((b) - red/blue colour) large motion.

Algorithm 1: 2D Image Anomaly Candidates and Anomaly Scores Calculation.

Require: $\{I\}$ //Set of 3D images
Ensure: $\{a\}, \{d\}$ //Sets of anomaly candidates and anomaly scores for all 2D image slices of 3D images
  $p \leftarrow |\{I\}|$ //Number of 3D images
  for $i \leftarrow 1$ **to** p **do**
    $s_{x,y,z} \leftarrow |I^i|$ //Image size for x, y, z
    for $z \leftarrow 1$ **to** s_z **do**
      $\{f_z^i\} \leftarrow quality(I_z^i)$ //2D image quality scores
    end for
  end for
  $\{f_z^i\} \leftarrow normalise(\{f_z^i\})$
  $\{a_z^i\}, \{d_z^i\} \leftarrow anomaly(\{f_z^i\})$

3) Anomaly Detection

To define a high-quality image dataset for training the DL-based VA predictive models, an anomaly detection method was used to eliminate low-quality images. This led to an improvement in our proposed model's results. According to [11], [58], although several anomaly detection methods have been developed, unsupervised anomaly detection methods are preferred. This is because they have the most flexible setup and do not require any labels or prior knowledge about the dataset [59]. Methods used for unsupervised anomaly detection include nearest-neighbour, clustering, statistical, subspace, and classifier-based methods [58].

This study used the nearest-neighbour method based on the local outlier factor (LOF) method. The LOF method computes the local density deviation of the entire dataset, showing how much a data point's local density differs from its neighbours. If the data has a significantly lower density compared to its neighbours, it has a high-density deviation, suggesting it may be abnormal. Here, we applied the elbow method to iteratively determine the optimal number of neighbours in the dataset, which was found to be 10. As discussed, we employed quality assessment measurements for each 2D image slice in every 3D image. Then, the quality scores were used as input to detect abnormal and normal images by the LOF method.

Algorithm 1 shows how the LOF-based approach determines normal and abnormal image candidates using anomaly prediction scores d and stored in a. Where f_z

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676

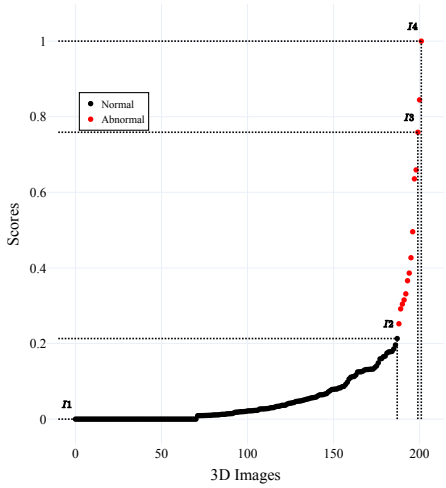


FIGURE 10: A graph depicting the 3D OCT image anomaly detection results: black and red points represent normal and abnormal images, respectively. I_1, I_2, I_3, and I_4 demonstrate randomly selected 3D images to be presented in Fig. 11.R1.4

corresponds to the image quality assessment measurement for each 2D image slice I_z in a 3D image $\{I^i\}$. $\{a\}$ corresponds to 2D image-based anomaly candidates and $\{d\}$ corresponds to 3D image-based anomaly scores calculated across the 3D image dataset $\{I\}$.

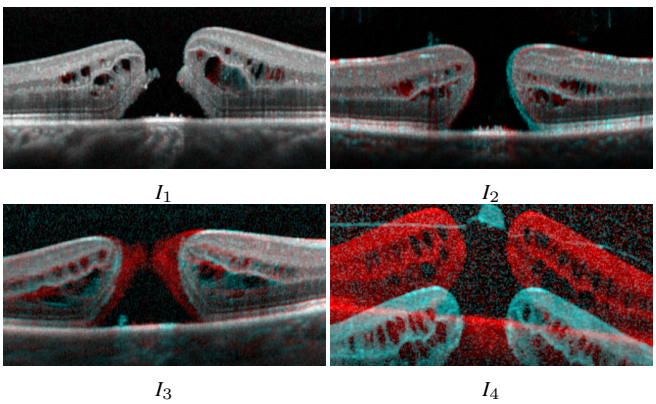


FIGURE 11: The best to worst 2D slices from 3D spectral domain optical coherence tomography images indicated as I_1, I_2, I_3, and I_4 in Fig. 10.

The distribution of normalised anomaly scores among 3D images was visualised in Fig. 10, where the red points represent the abnormal OCT image, and the black points represent normal OCT images. To confirm the accuracy of the proposed anomaly detection procedure, 2D slices of the selected 3D image samples in Fig. 10, from the best to the worst (I_1, I_2, I_3, and I_4), are displayed in Fig. 11 - where, red/blue colours represent a high anomaly score, and a grey colour represents a low anomaly score.

B. DEEP LEARNING MODELS:

Deep learning models, as an established but still continuously advancing technology, have significantly improved disease screening through medical imaging [27], [29], [60]. Numerous deep neural networks have been developed to classify, segment and predict various diseases since they have achieved comparable performance to human experts [28], [36]. To provide a comprehensive overview, nine well-known and state-of-the-art 2D DL-based predictive models were implemented, trained and tested in this study. The input to typical 2D convolutions is C * H * W, where C is the number of input channels, H is the height, and W is the width. The kernels move in two dimensions. In this study, we changed the first convolutional layers to feed the following DL-based predictive models with the input channels of single (1) and multiple OCT image slices (3, 5, 7, 9, 11, 13, 17, 19, and 21) since it led to better results:

1) Visual Geometry Group (VGG)

The VGG model [32], consisting of 5 groups of convolution layers and 1 group of a fully connected layer, provided the feature extraction process in the OCT images. During training, the input OCT image is propagated through convolution layers using 3×3 kernel-sized filters and a stride of 2. Although excessive depth can be computationally expensive and time-consuming, the most relevant salient features are extracted, including edges, corners, and interest points. In this study, this model was selected for its relatively small filter size compared with other models. VGG-based models with 11 and 19 layers were used.

2) Residual Neural Network (ResNet)

The ResNet is composed mainly of 5 groups of residual blocks and one fully connected layer [31]. These residual blocks were used to combine the information from the OCT images across multiple time points of the initiation phase. These residual blocks also link each other with skip connections. With this cross-layer connectivity, the convergence of deep networks is sped up. Thus, it prevents the problem of a diminishing gradient and provides a robust model against over-fitting in this study. The model used a 7×7 kernel-sized filter and a stride of 2 in the first convolution layer. ResNet-based models with 18, 34, and 50 layers were considered.

3) Inception v3

To suppress the high computational complexity problems encountered in VGG, ResNet and other mentioned models, the Inception v3 model with different kernel sizes, a max-pooling layer, and a stride, called Inception blocks, has been proposed by [61]. Due to this, it is remarkably useful for processing data in multiple resolutions and multilevel features, which makes this model suitable for OCT images. The model also overcomes high computation times by factorising the convolution (3×3) into asymmetric convolutions (3×1 and 1×3) [61].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676

4) Densely Connected Convolutional Networks (DenseNet)

DenseNet consists of sets of convolutional layers and direct connections from any convolutional layer to all subsequent layers [62]. Each layer receives a piece of collective information from all preceding layers with kernel sizes 1×1 and 3×3 in a dense and uses its feature maps as input. Then, 1×1 convolutional layer followed by 2×2 average pooling as the transition layers between sequential dense blocks. Thus, our OCT image dataset has not been exposed to a vanishing-gradient problem by showing a strengthened feature propagation during training.

5) EfficientNetV2

EfficientNetV2 has recently been further introduced with training faster and relatively smaller parameters, a new version of the well-known EfficientNet [63], [64]. Different from EfficientNet, EfficientNetV2 uses FusedMBConv in the earlier stages of the network, which replaces the depthwise convolutions (3×3) and expansion convolutions (1×1) in MBConv with single regular convolutions (3×3) [63], [64]. EfficientNetV2 also comes in different sizes. The large-sized (EfficientNetV2-L) model was selected as our OCT images' size is relatively large.

6) Vision Transformers (ViTs)

Vision Transformers (ViTs) have revolutionized computer vision using a transformer architecture [35]. They first split images into fixed-size patches. Each patch is linearly embedded into high-dimensional vectors. Then, to retain spatial information, positional embeddings are added to the patch embeddings. The embedded patch vectors, along with positional embeddings, are processed through a stack of Transformer encoder layers. Each layer includes self-attention mechanisms, allowing the model to capture relationships between different patches. The output of the Transformer encoder is typically a sequence of vectors. A classification head, often a linear layer, is added to obtain the final output for tasks. These steps allowed ViTs to discern global contextual information on OCT images. Therefore, we selected ViT-Base with an image patch size of 16 among several variations (ViT-B/16).

C. EXPERIMENTAL SETUP

This section introduces the DL models' training and evaluation details, including the k-fold cross-validation, the DL framework, data augmentation methods, parameter selections, and the evaluation criteria used.

TABLE 2: OCT imaging data used and splitting.

	Number of 3D Images	
	Initial	After data preparation
Train	168	152
Test	42	38

1) Training

Following image preprocessing, image quality assessment and anomaly detection procedures, twenty images were excluded from the initial OCT image dataset (see Table 2) (fifteen images - image quality assessment and anomaly detection, five images - image preprocessing). The final 3D OCT images used for the training were of the size of $452 \times 204 \times 49\,px$. The dataset was split uniformly into training and test sets using random five-fold cross-validation with a ratio of 80% and 20%. Specifically, each folded cross-validation consisted of 152 and 38 for training and testing, respectively.

Furthermore, DL models were trained using the same image size, with the use of single and multiple OCT image slices (1, 3, 5, 7, 9, 11, 13, 17, 19, and 21) at the first convolutional layer, centred around the mid-slice defined by the 3D intensity weighted centre of mass. To train the nine DL models used, we utilised Python 3.8.10, CUDA 11.4, cuDNN 8, PyTorch 1.9.0+cu102 running on a 64-bit Ubuntu operating system using a 3.4 GHz Intel Core-i9 with 32 GB of RAM and NVIDIA GTX 1080 Ti GPU with a frame buffer of 11 GB GDDR5X.

2) Data Augmentation

To enlarge the variety and amount of data artificially, we employed a range of image data augmentation techniques, helping us overcome over-fitting issues while maintaining the data properties that existed initially in the data. The data augmentation techniques used were:

- Rotation: To estimate the eye orientation during the scanning procedure and to define the data augmentation range for the rotation, we measured the orientation distribution across the image dataset and presented it in Fig. 12. The dominant orientation is around ± 0 degrees, with the rotation augmentation range between -22 and 15 degrees.
- Vertical and Horizontal Translation: Based on the calculated intensity weighed centres of mass for all OCT images (see Fig. 6), we defined a range of vertical and horizontal translation augmentations, $\pm 5px$ for the vertical translation (up-down) and $\pm 8px$ the horizontal translation (right-left) respectively.
- Horizontal Flip: Reversing all rows and columns of an OCT image's pixels allowed a mirror image OCT to be obtained, representing a fellow eye (right for left and left for right) and allowing the model to learn this unpredictable variability.
- Gaussian Blur: The calculated noise scores in Section IV-A2 were used to guide a Gaussian blur data augmentation technique.

3) Loss Function

The mean absolute error (MAE), the mean squared error (MSE), and the Huber loss (HL) functions were separately used. These functions are commonly used by the optimiser to minimise training errors. While MAE presents the average

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676

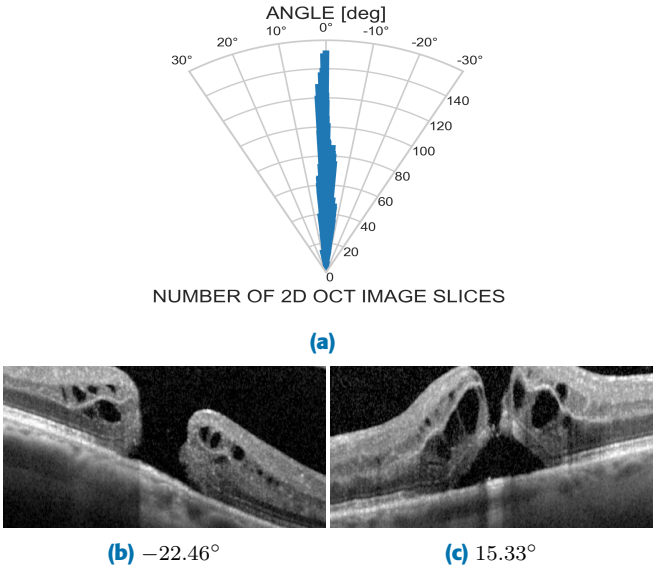


FIGURE 12: Eye orientation distribution in the OCT imaging dataset (a) and two sample images corresponding to the eye orientations, -22.46° (b) and 15.33° (c), respectively.

of the absolute differences between the actual and predicted visual acuity, scored as ETDRS letters, MSE loss is calculated as the average of the squared differences. We utilised HL as a combination of the MAE and the MSE, meaning that HL represents a quadratic behaviour for minor errors and a linear behaviour for significant errors. Moreover, HL and MAE are more robust to data with outliers. Minor error values calculated by these loss functions reflect a better model. In addition, to further improve the performance of DL models, we also modified the HL function, considering image quality assessment to guide the optimization. Therefore, while the optimizer considers the predictions at the first epochs, the model at the late epoch focuses on the model's robustness. The term of δ in the HL function was updated during model training based on the anomaly scores measured in Section IV-A3 and Algorithm 1 ($\{d\}$). A smaller δ made the loss function more sensitive to small errors, while a larger δ made it more robust to anomalies.

$$L_\delta = \begin{cases} \frac{1}{2}(y-\hat{y})^2 & \textit{if } |(y-\hat{y})| < \delta \\ \delta((y-\hat{y}) - \frac{1}{2}\delta) & \textit{otherwise} \end{cases} \tag{4}$$

where $\delta = \{d\}$ is the adjusted parameter of the HL function, which could provide more robustness to anomalies.

4) Parameter Selection

The stated loss functions were optimised via the ADAM algorithm with a fixed number of epochs ($n_{epoch} = 1000$), resulting in sufficient learning within fewer epochs in our experiments. The learning rate was set to 1×10^{-5} with weight decay ($w = 1\times10^{-5}$) and momentums ($\beta_1 = 0.5, \beta_2 = 0.9$), and an automatic learning schedule were added. The DL models were trained by dividing the dataset into 38 batches. Lastly, we saved the parameters of each DL model when the model's performance started to decrease since this reduction is a strong indication of over-fitting.

D. EVALUATION METRICS

To evaluate the performance of the models to predict both pre and postoperative visual acuity, we used the following metrics: R-squared, root-mean-square error (RMSE), mean absolute error (MAE), and the Pearson correlation coefficient. The R-squared value ranges from 0 to 1, with a higher value indicating a better fit between the predicted and actual values. RMSE and MAE values show better performance when they are closer to zero. Pearson correlation coefficients range between -1 and +1. A value of -1.0 shows a perfect negative correlation, whilst + 1.0 shows a perfect positive correlation. A zero correlation shows no relationship between the predicted and actual values. We used these four metrics to allow a comprehensive comparison with previous studies in the literature.

V. EXPERIMENTAL DESIGN AND TEST RESULTS

We present results for both preoperative and postoperative visual acuity. Preoperative visual acuity is known before surgery, but we were interested in assessing the model's performances for both measurements. To ensure the preoperative VA models were trained using the same data set as the postoperative VA models, we kept the same data augmentation methods and model parameters for both tasks. However, we did not keep the same five-fold cross-validation set-up, where each fold contained 152 and 38 OCT images as training and testing data sets, as the preoperative and postoperative VA measurements had different distributions. The modified HL function performed better than the other mentioned loss functions; therefore, we represented the results of optimizing the modified HL function in this section.

A. RESULTS BASED ON PREOPERATIVE VA

Table 3 shows the quantitative comparison between our trained DL models using a different number of OCT image slices, using MAE as the evaluation metric. All evaluation metric values are the means obtained using five-fold cross-validation. Statistically significant results are highlighted in bold. The obtained results clearly show that the majority of tested DL predictive models performed best with seven OCT image slices.

Table 4 illustrates the performance of nine DL models using the seven central OCT image slices, with all four evaluation metrics given as means obtained using five-fold cross-validation. The experimental results demonstrated that ResNet-18 was the most predictive in all scores, achieving 0.47 for R-squared, 7.34 for RMSE, 0.65 for the Pearson correlation coefficient, and 5.96 for MAE.

Table 5 shows the performance of the nine DL's best predictive models on the test dataset using seven input OCT image slices, with all four evaluation metrics: R-squared, RMSE, Pearson correlation coefficient, and MAE. The experimental results demonstrated that ResNet-18 was

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676

TABLE 3: The mean absolute error values, based on preoperative VA measurements, were obtained for nine state-of-the-art DL predictive models using a different number of OCT image slices through our designed loss function (the best results are highlighted in bold). All evaluation metric values are given as the means and standard deviations obtained using five-fold cross-validation.

MODELS	OCT IMAGE SLICES									
	1	**3**	**5**	**7**	**9**	**11**	**13**	**17**	**19**	**21**
VGG-11	**6.72±1.07**	6.88±1.11	7.09±0.98	7.21±0.89	7.48±0.82	7.21±0.83	7.67±0.69	7.56±0.87	7.42±0.62	7.46±0.77
VGG-19	6.56±1.12	**6.51±0.99**	6.74±0.97	6.88±0.82	7.07±0.55	6.94±0.91	7.20±0.66	7.41±0.89	7.05±0.78	7.11±0.77
ResNet-18	6.32±1.3	6.38±0.94	6.79±0.93	**5.96±0.72**	7.04±0.67	6.90±0.72	6.77±0.87	7.05±0.92	7.29±1.12	7.07±1.19
ResNet-34	6.27±1.20	6.60±1.23	6.78±1.17	**6.01±0.78**	6.96±0.79	7.03±0.68	7.26±0.65	6.72±0.89	6.94±0.91	6.75±0.97
ResNet-50	6.94±0.97	7.59±0.82	7.20±0.62	**6.75±0.89**	7.40±0.86	7.53±0.80	7.42±0.49	7.26±0.49	7.08±0.30	7.22±0.26
Inception v3	6.95±1.14	6.66±1.12	7.48±0.94	**6.42±0.83**	7.41±0.58	8.61±0.30	8.78±0.26	8.47±0.41	8.20±0.48	8.58±0.73
DenseNet-121	**6.08±0.91**	7.24±1.29	7.06±1.15	6.97±0.77	7.18±0.65	7.17±0.93	6.75±0.78	6.88±0.86	6.82±0.81	6.92±0.83
EfficientNetV2-L	6.58±1.01	6.83±0.97	6.82±1.01	**6.57±0.89**	6.90±1.15	6.98±1.23	6.95±1.18	7.00±0.93	6.97±0.95	7.06±1.11
ViT-B/16	7.05±0.70	6.88±0.62	6.93±0.62	6.90±0.63	6.98±0.60	6.82±0.64	**6.70±0.58**	6.71±0.51	6.87±0.51	6.95±0.52

TABLE 4: Quantitative comparison of nine state-of-the-art DL predictive models with seven OCT image slices, using four different evaluation metrics through our designed loss function, as the means and standard deviations obtained with five-fold cross-validation, based on preoperative VA measurements (the best results are highlighted in bold).

MODELS	R2	RMSE	Pearson	MAE
VGG-11	0.27	9.70	0.52	7.21±0.89
VGG-19	0.34	9.227	0.58	6.88±0.82
ResNet-18	**0.47**	**7.34**	**0.65**	**5.96±0.72**
ResNet-34	0.40	7.75	0.62	6.01±0.78
ResNet-50	0.25	9.04	0.39	6.75±0.89
Inception v3	0.38	8.27	0.61	6.42±0.83
DenseNet-121	0.30	9.37	0.55	6.97±0.77
EfficientNetV2-L	0.37	8.49	0.59	6.57±0.89
ViT-B/16	0.39	9.48	0.52	6.90±0.63

TABLE 5: Quantitative comparison of nine state-of-the-art DL's best predictive models on a uniform test dataset with seven OCT image slices, using four different evaluation metrics through our designed loss function, based on preoperative VA measurements (the best results are highlighted in bold).

MODELS	R2	RMSE	Pearson	MAE
VGG-11	0.47	7.64	0.68	6.06
VGG-19	0.48	7.55	0.69	5.93
ResNet-18	**0.59**	**6.67**	**0.73**	**5.25**
ResNet-34	0.35	6.76	0.59	5.35
ResNet-50	0.29	9.56	0.48	7.21
Inception v3	0.33	8.45	0.58	6.73
DenseNet-121	0.37	8.05	0.61	6.14
EfficientNetV2-L	0.41	7.45	0.61	6.13
ViT-B/16	0.35	8.95	0.52	6.91

again the most effective method, achieving 0.59 for R-squared, 6.67 for RMSE, 0.73 for the Pearson correlation coefficient, and 5.25 for MAE.

B. RESULTS BASED ON POSTOPERATIVE VA

Table 6 shows the quantitative comparison between our trained DL models using a different number of OCT image slices, with MAE as the evaluation metric (means obtained using five-fold cross-validation). Statistically significant results are highlighted in bold. The obtained results clearly show that the majority of tested DL predictive models performed best with the central seven OCT image slices.

Table 7 illustrates the performance of the nine DL models using seven OCT image slices, using all four outcome metrics: R-squared, RMSE, Pearson correlation coefficient, and MAE. The experimental results demonstrated that ResNet-50 was the most effective and superior in most evaluation scores, achieving 0.46 for R-squared, 9.01 for RMSE, 0.69 for the Pearson correlation coefficient, and 6.84 for MAE.

Table 8 shows the performance of the nine DL models on a uniform test dataset using seven input OCT image slices with all four evaluation metrics. The experimental results demonstrated that ResNet-50 was again the most effective and superior in most evaluation scores, achieving 0.52 for R-squared, 9.23 for RMSE, 0.71 for the Pearson correlation coefficient, and 6.47 for MAE.

In Fig. 13 and 14, we show descriptive results for the relationship between ground truth and predicted VA values and corresponding confidence intervals to show how close our best-trained model was able to predict VA values on the

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676

TABLE 6: The mean absolute error values, based on postoperative VA measurements, were obtained for nine state-of-the-art DL predictive models using a different number of OCT image slices through our designed loss function (the best results are highlighted in bold). All evaluation metric values are the means and standard deviations obtained using the five-fold cross-validation.

MODELS	OCT IMAGE SLICES									
	1	**3**	**5**	**7**	**9**	**11**	**13**	**17**	**19**	**21**
VGG-11	8.19±1.43	8.74±1.38	8.20±1.21	**7.94±1.11**	8.10±1.36	8.57±1.20	8.82±1.22	8.25±1.12	8.13±1.09	8.62±1.01
VGG-19	7.94±1.66	**7.76±1.12**	8.47±1.04	7.87±1.11	8.28±1.21	8.02±1.28	8.46±1.56	8.85±1.32	8.95±1.22	8.77±1.23
ResNet-18	7.29±1.76	7.31±1.25	7.67±1.01	**6.81±0.81**	7.52±1.24	8.09±1.47	7.45±1.50	7.45±0.95	7.57±1.12	7.35±0.80
ResNet-34	7.06±1.10	7.58±1.84	7.29±1.25	7.15±1.31	7.16±1.55	**6.96±1.59**	7.55±1.73	7.48±1.36	7.46±1.22	7.63±1.55
ResNet-50	7.37±1.61	7.70±1.56	7.37±1.51	**6.84±0.37**	7.26±0.54	7.58±0.90	7.59±1.23	7.64±1.30	7.41±1.14	7.75±1.62
Inception v3	**7.38±1.41**	7.57±1.28	8.01±1.43	7.57±1.51	7.60±1.12	7.78±1.01	8.01±1.51	7.66±1.60	8.07±1.35	7.91±1.23
DenseNet-121	7.87±1.19	7.75±1.27	7.68±1.01	**7.64±0.88**	8.02±1.22	7.79±1.37	7.60±1.41	8.01±1.29	7.67±1.41	7.90±1.39
EfficientNetV2-L	7.98±2.11	7.76±1.29	7.73±1.13	7.57±1.01	7.61±0.98	7.47±0.98	7.35±1.08	**6.97±1.16**	6.99±1.13	7.2±0.98
ViT-B/16	8.12±0.80	8.04±0.72	7.97±0.64	**7.89±0.61**	7.89±0.66	7.90±0.75	7.93±0.73	7.88±0.74	7.92±0.70	8.02±0.69

TABLE 7: Quantitative comparison of nine DL predictive models using seven OCT image slices, showing four different evaluation metrics through our designed loss function (with means obtained by five-fold cross-validation), as the means and standard deviations obtained with five-fold cross-validation, based on postoperative VA measurements (the best results are highlighted in bold).

MODELS	R2	RMSE	Pearson	MAE
VGG-11	0.27	10.77	0.52	7.94±1.11
VGG-19	0.25	11.52	0.45	7.87±1.11
ResNet-18	0.43	9.31	0.65	**6.81±0.81**
ResNet-34	0.36	9.95	0.60	7.15±1.31
ResNet-50	**0.46**	**9.01**	**0.69**	6.84±0.37
Inception v3	0.26	10.12	0.51	7.57±1.51
DenseNet-121	0.23	10.35	0.47	7.64±0.88
EfficientNetV2-L	0.36	10.05	0.53	7.57±1.01
ViT-B/16	0.38	10.79	0.60	7.89±0.61

TABLE 8: Quantitative comparison of nine state-of-the-art DL's best predictive models on a uniform test dataset with seven OCT image slices, using four different evaluation metrics through our designed loss function, based on postoperative VA measurements (the best results are highlighted in bold).

MODELS	R2	RMSE	Pearson	MAE
VGG-11	0.41	11.84	0.64	6.92
VGG-19	0.32	11.10	0.57	6.78
ResNet-18	0.49	9.99	0.70	**6.00**
ResNet-34	0.34	10.56	0.57	7.06
ResNet-50	**0.52**	**9.23**	**0.71**	6.47
Inception v3	0.41	10.17	0.64	6.20
DenseNet-121	0.17	10.74	0.40	7.55
EfficientNetV2-L	0.32	10.41	0.58	6.73
ViT-B/16	0.38	11.04	0.55	7.51

test dataset. Overall, the results confirm that our proposed fully automated image informatics framework can robustly predict both preoperative and postoperative visual acuity measurements for patients with idiopathic full-thickness macular holes using a high-quality SD-OCT image dataset.

VI. DISCUSSION AND CONCLUSION

We present a full image informatics approach to predict visual acuity outcomes in people undergoing surgery to treat MHs using preoperative SD-OCT images and deep learning-based predictive models.

To overcome the impact of high variations in real-world image quality on the robustness of the deep learning model, an extensive imaging data assessment and quality assurance procedure was implemented. Data preparation steps, including scaling, centre of mass detection, and cropping, were used to unify the imaging dataset's scale, size and centration. Further, data quality assessment measurements, including noise, blurriness, contrast, motion, and brightness-darkness scores, were calculated to identify and exclude abnormalities in the imaging dataset.

The resultant high-quality imaging dataset was then used to train nine state-of-the-art 2D deep learning-based predictive models for both pre and postoperative VA using multiple channels (2D+), followed by a quantitative performance comparison with our designed loss function.

All tested models were able to predict preoperative visual acuity with less than an 8.78 MAE letter score, with the best predictive model achieving a 5.96 MAE score with 0.47 for R-squared, 7.34 for RMSE, 0.65 for the Pearson correlation coefficient. Similarly, all tested models were able to predict

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676

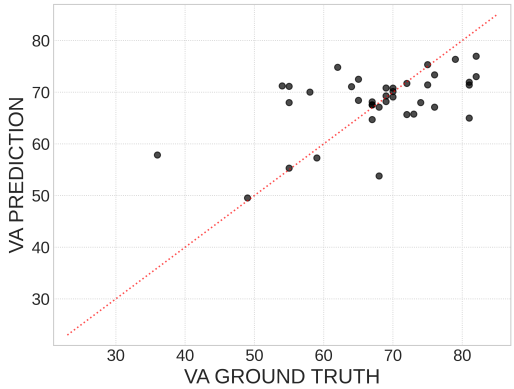


FIGURE 13: The scatter plot visualises the relationship between the ground truth and predicted postoperative VA measurements obtained by the ResNet-50 model on the test dataset (the highlighted result in Table 8). The red dotted line depicts the gold standard.

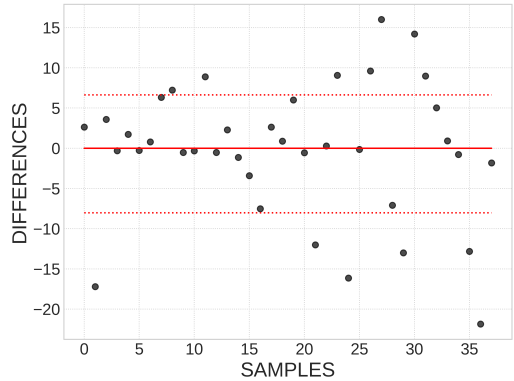


FIGURE 14: The 95% confidence interval between the ground truth and predicted postoperative VA values is shown with the red dotted lines (-8.02, 6.46) obtained by the ResNet-50 model on the test dataset (the highlighted result in Table 8). The solid red line depicts the gold standard.

postoperative visual acuity with less than an 8.95 letter MAE score, with the best predictive model achieving a 6.84 MAE score with 0.46 for R-squared, 9.01 for RMSE, 0.69 for the Pearson correlation coefficient.

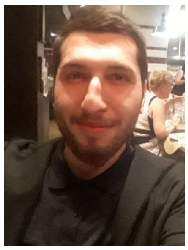
The CNN-based backbone networks mostly demonstrated high predictive performance, as evidenced by their competitive results in predicting preoperative and postoperative VA measurements. The reason might be that they leverage their inherent ability to analyse hierarchical features for complex structure analysis on OCT images [65]. Another might be that they likely recognize patterns regardless of the location of relevant structures in the input OCT image [66]. On the contrary, a certain standard has not been achieved regarding robustness when the number of input channels is altered.

In addition, the ViTs model did not perform as effectively as the CNN-based state-of-the-art models. This is likely because the ViT-based models improve training efficiency on large-scale datasets [35]. However, our dataset is relatively small. Furthermore, the ViTs come with expensive overhead, causing large parameter sizes [35], [67]. Hence, it is a computationally expensive and time-consuming process. On the contrary, the ViT models presented more robust performance. This is also because the ViT-based models may incorporate attention mechanisms to focus on the most relevant part of images and disregard irrelevant noise [36]. However, the introduction of redundant information and insufficient sparsity can impede the improvement of robustness in ViT-based models, leading to performance degradation [68].

As a result, ResNet architectures show slightly better results among nine state-of-the-art DL models. One of the reasons might be that it has residual blocks, leading to fewer vanishing problems. Additionally, ResNet-50 obtained the best results, which may be due to a deeper architecture and having a bottleneck. Overall, the prediction of preoperative visual acuity had better performance than postoperative VA in most of the metrics. As other studies have shown, however, preoperative VA is strongly correlated with postoperative VA.

Although other studies have reported similar results for the important and informative prediction of postoperative VA [17], [23], [24], our study provides more robust results, as we validated our results using an independent data set. Indeed, our best model achieved the highest metrics in all evaluation scores, achieving 0.52 for R-squared, 9.23 for RMSE, 0.71 for the Pearson correlation coefficient, and 6.47 for MAE, as shown in Table 1. These compare very favourably to previous traditional regression modelling methodologies. Furthermore, while we trained the predictive model on a limited dataset acquired by only one type of OCT imaging system, albeit in two hospitals and four different devices, the proposed 2D DL-based predictive approach contains a comprehensive image informatics framework with our designed loss function that can be applied across a breadth of many 3D medical image datasets.

To overcome those limitations, our future research work will focus on adapting full 3D deep learning-based predictive models, the uncertainty of 2D and 3D DL-based predictive models, and a substantive scale-up of the OCT imaging data size and types.



BURAK KUCUKGOZ received the B.Eng. degree in Electrical and Electronics Engineering from Gazi University, Ankara, Turkey, in 2014, the M.Sc. degree in Systems, Control and Signal Processing from University of Southampton, Southampton, U.K., in 2018. He is currently pursuing a PhD degree from Newcastle University, Newcastle upon Tyne, UK and the Agency for Science, Technology and Research (A*STAR) Institute, Singapore.

His research interests include image processing, medical imaging and image analysis, pattern recognition, computer vision, machine learning, and optimization algorithms for deep neural networks.

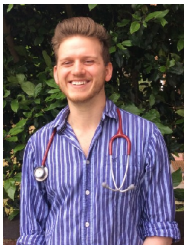
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676

MUHAMMED MUTLU YAPICI received the M.Sc. degree in Computer and Electronic Education from the Informatics Institute at Gazi University, Turkey, in 2012. He received a PhD in the Computer Engineering Department at Gazi University.

From 2014 to 2021, he worked as a lecturer in the Computer Technologies Department at Ankara University, Turkey. He was awarded the TUBITAK scholarship and worked as a Visiting Scholar in the School of Computing at Newcastle University, UK, for one year, from 2022 to 2023. Currently, he holds the position of Assistant Professor in the Computer Technologies Department at Ankara University.

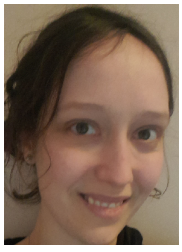
His primary research areas include Artificial Intelligence, Machine Learning, Deep Learning, Image Processing, Optimization Techniques, and their applications in medical research, signature verification, and object recognition.



DECLAN C MURPHY is an academic doctor in the United Kingdom with a specialist interest in Ophthalmology.

He has published over 60 peer-reviewed publications with over 450 citations. He was awarded a Bachelor of Medicine, Bachelor of Surgery (MBBS) with distinction from Norwich Medical School in 2019, as well as a Master in Research (MRes) in medical and molecular biosciences from Newcastle University, UK in 2018. His most recently held position is as an Academic Clinical Fellow in Ophthalmology in the North of England, affiliated with Newcastle University, and is a regular tutor at Sunderland Medical School.

Declan is passionate about developing actionable solutions to improve the diagnosis, investigation, and management of patients with a variety of illnesses relating to the eye and also improving healthcare service delivery, so such solutions can be provided to patients which achieve objective improvements to their care.



EMMA SPOWART is a speciality trainee in ophthalmology in the Northern deanery. She qualified in Medicine (MBBS) with a BSc in Medical Sciences with Neuroscience from University College London in 2016.

Following completion of foundation training in London, she moved to the North-East for ophthalmology training and has completed a PGCert in Medical Education (Newcastle, 2020).

She is currently one of the co-editors in the vitreoretinal section of the Royal College of Ophthalmology INSPIRE e-learning platform and is involved with the delivery of the undergraduate ophthalmology teaching programme at Newcastle University.

DAVID H STEEL is a consultant retinal specialist at Sunderland Eye Infirmary appointed in 1998 and Honorary Professor of Retinal Surgery at Newcastle University and Visiting Professor of Ophthalmology at the University of Liverpool.

He has published over 280 peer-reviewed publications with over 12,000 citations, and an H index of 53 in 2023. He was named the UK NIHR and Royal College of Ophthalmologists Researcher of the Year award in 2018. He is a past President of the British and Eire Association of Vitreoretinal Surgeons and an elected Executive Board member of the International renowned retinal society, the Club Jules Gonin.

As a clinical academic, he aims to help bridge the gap between the clinic and the laboratory. As such, his programme of research extends from the laboratory investigation of mechanisms of retinal disease in experimental models, through preclinical development of novel therapies, to trials of new medical and surgical interventions and systematic evidence reviews. He aims to reduce the prevalence of blindness, using translational research to improve the care of patients.

BOGUSLAW OBARA is a Professor of Image Informatics in the School of Computing and Dean of Business, Innovation and Skills at Newcastle University. Turing Fellow at The Alan Turing Institute. He obtained a first master's degree in computational physics from Jagiellonian University and a PhD in Image Processing from AGH University of Science and Technology.

Before joining Newcastle University as a professor; he held research assistant positions at the Polish Academy of Sciences and Computer Vision Laboratory, ETH, a Fulbright Fellowship at Vision Research Laboratory, postdoctoral positions at the Center for BioImage Informatics, University of California, and Oxford e-Research Centre, University of Oxford, and assistant, associate and professor positions at Department of Computer Science at Durham University. He was also an AstraZeneca Visiting Professor in Image Processing & Artificial Intelligence.

His research focuses on the design and implementation of complex image analysis and processing, pattern recognition, computer vision and machine learning solutions applied to a wide range of domains.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676

REFERENCES

[1] Seth R Flaxman, Rupert RA Bourne, Serge Resnikoff, Peter Ackland, Tasanee Braithwaite, Maria V Cicinelli, Aditi Das, Jost B Jonas, Jill Keeffe, John H Kempen, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *The Lancet Global Health*, 5(12):e1221–e1234, 2017.

[2] Tina Felfeli and Efrem D Mandelcorn. Macular hole hydrodissection: surgical technique for the treatment of persistent, chronic, and large macular holes. *Retina*, 39(4):743–752, 2019.

[3] Michael R. Hee, Carmen A. Puliafito, Carlton Wong, Jay S. Duker, Elias Reichel, Joel S. Schuman, Eric A. Swanson, and James G. Fujimoto. Optical coherence tomography of macular holes. *Ophthalmology*, 102(5):748–756, 1995.

[4] Hiroshi Kobayashi and Kaori Kobayashi. Correlation of quantitative three-dimensional measurements of macular hole size with visual acuity after vitrectomy. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 237(4):283–288, 1999.

[5] Sentaro Kusuhara and Akira Negi. Predicting visual outcome following surgery for idiopathic macular holes. *Ophthalmologica*, 231(3):125–132, 2014.

[6] DH Steel, PHJ Donachie, GW Aylward, DA Laidlaw, TH Williamson, and D Yorston. Factors affecting anatomical and visual outcome after macular hole surgery: findings from a large prospective uk cohort. *Eye*, 35(1):316–325, 2021.

[7] S Ullrich, C Haritoglou, C t Gass, M Schaumberger, MW Ulbig, and A Kampik. Macular hole size as a prognostic factor in macular hole surgery. *British Journal of Ophthalmology*, 86(4):390–393, 2002.

[8] Amar V Nasrulloh, Chris G Willcocks, Philip TG Jackson, Caspar Geenen, Maged S Habib, David HW Steel, and Boguslaw Obara. Multi-scale segmentation and surface fitting for measuring 3-d macular holes. *IEEE Transactions on Medical Imaging*, 37(2):580–589, 2017.

[9] Weina Jin, Xiaoxiao Li, Mostafa Fatehi, and Ghassan Hamarneh. Guidelines and evaluation of clinical explainable ai in medical image analysis. *Medical Image Analysis*, 84:102684, 2023.

[10] Ryan T Yanagihara, Cecilia S Lee, Daniel Shu Wei Ting, and Aaron Y Lee. Methodological challenges of deep learning in optical coherence tomography for retinal diseases: a review. *Translational Vision Science & Technology*, 9(2):11–11, 2020.

[11] Vivek Kumar Singh, Burak Kucukgoz, Declan C. Murphy, Xiaofan Xiong, David H. Steel, and Boguslaw Obara. Benchmarking automated detection of the retinal external limiting membrane in a 3D spectral domain optical coherence tomography image dataset of full thickness macular holes. *Computers in Biology and Medicine*, 140:105070, 2022.

[12] Burak Kucukgoz, M. Mutlu Yapici, David H Steel, and Boguslaw Obara. Evaluation of 2D and 3D deep learning approaches for predicting visual acuity following surgery for idiopathic full-thickness macular holes in spectral domain optical coherence tomography images. In *International Symposium on Image and Signal Processing and Analysis*, pages 1–6, 2023.

[13] Ahmadreza Baghaie, Zeyun Yu, and Roshan M D'Souza. State-of-the-art in retinal optical coherence tomography image analysis. *Quantitative Imaging in Medicine and Surgery*, 5(4):603, 2015.

[14] Jason Kugelman, David Alonso-Caneiro, Scott A Read, Stephen J Vincent, Fred K Chen, and Michael J Collins. Effect of altered oct image quality on deep learning boundary segmentation. *IEEE Access*, 8:43537–43553, 2020.

[15] Josef Kauer-Bonin, Sunil K Yadav, Ingeborg Beckers, Kay Gawlik, Seyedamirhosein Motamedi, Hanna G Zimmermann, Ella M Kadas, Frank Haußer, Friedemann Paul, and Alexander U Brandt. Modular deep neural networks for automatic quality control of retinal optical coherence tomography scans. *Computers in Biology and Medicine*, 141:104822, 2022.

[16] Yu Tian, Huanqiang Zeng, Junhui Hou, Jing Chen, and Kai-Kuang Ma. Light field image quality assessment via the light field coherence. *IEEE Transactions on Image Processing*, 29:7945–7956, 2020.

[17] Declan C Murphy, Amar Nasrulloh, Clare Lendrem, Clare Lendrem, Mark Alberti Morten la Cour, Boguslaw Obara, and David Steel. Predicting post-operative vision for macular hole with automated image analysis. *Ophthalmology Retina*, 2020.

[18] Yungang Zhang, Bailing Zhang, Frans Coenen, Jimin Xiao, and Wenjin Lu. One-class kernel subspace ensemble for medical image classification. *EURASIP Journal on Advances in Signal Processing*, 2014(1):1–13, 2014.

[19] Nantheera Anantrasirichai, Alin Achim, James E Morgan, Irina Erchova, and Lindsay Nicholson. Svm-based texture classification in optical coherence tomography. In *IEEE International Symposium on Biomedical Imaging*, pages 1332–1335, 2013.

[20] Naohiro Motozawa, Guangzhou An, Seiji Takagi, Shohei Kitahata, Michiko Mandai, Yasuhiko Hirami, Hideo Yokota, Masahiro Akiba, Akitaka Tsujikawa, Masayo Takahashi, et al. Optical coherence tomography-based deep-learning models for classifying normal and age-related macular degeneration and exudative and non-exudative age-related macular degeneration changes. *Ophthalmology and Therapy*, 8(4):527–539, 2019.

[21] Ali Mohammad Alqudah. Aoct-net: a convolutional network automated classification of multiclass retinal diseases using spectral-domain optical coherence tomography images. *Medical & Biological Engineering & Computing*, 58(1):41–53, 2020.

[22] Takumasa Tsuji, Yuta Hirose, Kohei Fujimori, Takuya Hirose, Asuka Oyama, Yusuke Saikawa, Tatsuya Mimura, Kenshiro Shiraishi, Takenori Kobayashi, Atsushi Mizota, et al. Classification of optical coherence tomography images using a capsule network. *BMC Ophthalmology*, 20(1):1–9, 2020.

[23] Michael G Kawczynski, Thomas Bengtsson, Jian Dai, J Jill Hopkins, Simon S Gao, and Jeffrey R Willis. Development of deep learning models to predict best-corrected visual acuity from optical coherence tomography. *Translational Vision Science & Technology*, 9(2):51–51, 2020.

[24] Shumpei Obata, Yusuke Ichiyama, Masashi Kakinoki, Osamu Sawada, Yoshitsugu Saishin, Taku Ito, Mari Tomioka, and Masahito Ohji. Prediction of postoperative visual acuity after vitrectomy for macular hole using deep learning–based artificial intelligence. *Graefe's Archive for Clinical and Experimental Ophthalmology*, pages 1–11, 2021.

[25] Yijun Hu, Yu Xiao, Wuxiu Quan, Bin Zhang, Yuqing Wu, Qiaowei Wu, Baoyi Liu, Xiaomin Zeng, Ying Fang, Yu Hu, et al. A multi-center study of prediction of macular hole status after vitrectomy and internal limiting membrane peeling by a deep learning model. *Annals of Translational Medicine*, 9(1), 2021.

[26] Stanislao Rizzo, Alfonso Savastano, Jacopo Lenkowicz, Maria Cristina Savastano, Luca Boldrini, Daniela Bacherini, Benedetto Falsini, and Vincenzo Valentini. Artificial intelligence and oct angiography in full thickness macular hole. new developments for personalized medicine. *Diagnostics*, 11(12):2319, 2021.

[27] Alexandre Lachance, Mathieu Godbout, Fares Antaki, Mélanie Hébert, Serge Bourgault, Mathieu Caissie, Éric Tourville, Audrey Durand, and Ali Dirani. Predicting visual improvement after macular hole surgery: a combined model using deep learning and clinical features. *Translational Vision Science & Technology*, 11(4):6–6, 2022.

[28] Pan Huang, Xiaoheng Tan, Xiaoli Zhou, Shuxian Liu, Francesco Mercaldo, and Antonella Santone. Fabnet: fusion attention block and transfer learning for laryngeal cancer tumor grading in p63 ihc histopathology images. *IEEE Journal of Biomedical and Health Informatics*, 26(4):1696–1707, 2021.

[29] Pan Huang, Xiaoli Zhou, Peng He, Peng Feng, Sukun Tian, Yuchun Sun, Francesco Mercaldo, Antonella Santone, Jing Qin, and Hualiang Xiao. Interpretable laryngeal tumor grading of histopathological images via depth domain adaptive network with integration gradient cam and priori experience-guided attention. *Computers in Biology and Medicine*, 154:106447, 2023.

[30] Catherine Park, Mohammad Dehghan Rouzi, Md Moin Uddin Atique, MG Finco, Ram Kinker Mishra, Griselda Barba-Villalobos, Emily Crossman, Chima Amushie, Jacqueline Nguyen, Chadi Calarge, et al. Machine learning-based aggression detection in children with adhd using sensor-based physical activity monitoring. *Sensors*, 23(10):4949, 2023.

[31] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[32] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[33] Migyeong Yang, Jinyoung Han, Ji In Park, Joon Seo Hwang, Jeong Mo Han, Jeewoo Yoon, Seong Choi, Gyudeok Hwang, and Daniel Duck-Jin Hwang. Prediction of visual acuity in pathologic myopia with myopic choroidal neovascularization treated with anti-vascular endothelial growth factor using a deep neural network based on optical coherence tomography images. *Biomedicines*, 11(8):2238, 2023.

[34] Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, and Samy Bengio. Transfusion: Understanding transfer learning for medical imaging. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3369676

R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[35] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

[36] Pan Huang, Peng He, Sukun Tian, Mingrui Ma, Peng Feng, Hualiang Xiao, Francesco Mercaldo, Antonella Santone, and Jing Qin. A vit-amc network with adaptive model fusion and multiobjective optimization for interpretable laryngeal tumor grading from histopathological images. *IEEE Transactions on Medical Imaging*, 42(1):15–28, 2022.

[37] Huajie Wen, Jian Zhao, Shaohua Xiang, Lin Lin, Chengjian Liu, Tao Wang, Lin An, Lixin Liang, and Bingding Huang. Towards more efficient ophthalmic disease classification and lesion location via convolution transformer. *Computer Methods and Programs in Biomedicine*, 220:106832, 2022.

[38] David Xu, Alex Yuan, Peter K Kaiser, Sunil K Srivastava, Rishi P Singh, Jonathan E Sears, Daniel F Martin, and Justis P Ehlers. A novel segmentation algorithm for volumetric analysis of macular hole boundaries identified with optical coherence tomography. *Investigative Ophthalmology & Visual Science*, 54(1):163–169, 2013.

[39] Abdulrahman Albarrak, Frans Coenen, Yalin Zheng, et al. Age-related macular degeneration identification in volumetric optical coherence tomography using decomposition and local feature extraction. In *International Conference on Medical Image, Understanding and Analysis*, pages 59–64, 2013.

[40] Freerk G Venhuizen, Bram van Ginneken, Bart Bloemen, Mark JJP van Grinsven, Rick Philipsen, Carel Hoyng, Thomas Theelen, and Clara I Sánchez. Automated age-related macular degeneration classification in oct using unsupervised feature learning. In *Medical Imaging: Computer-Aided Diagnosis*, volume 9414, page 94141I, 2015.

[41] Pratul P. Srinivasan, Leo A. Kim, Priyatham S. Mettu, Scott W. Cousins, Grant M. Comer, Joseph A. Izatt, and Sina Farsiu. Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomedical Optics Express*, 5(10):3568–3577, 2014.

[42] Guillaume Lemaitre, Mojdeh Rastgoo, Joan Massich, Shrinivasan Sankar, Fabrice Mériaudeau, and Désiré Sidibé. Classification of sd-oct volumes with lbp: application to dme detection. *Ophthalmic Medical Image Analysis International Workshop 2*, 2015.

[43] Yu-Ying Liu, Mei Chen, Hiroshi Ishikawa, Gadi Wollstein, Joel S Schuman, and James M Rehg. Automated macular pathology diagnosis in retinal oct images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding. *Medical Image Analysis*, 15(5):748–759, 2011.

[44] Reza Rasti, Hossein Rabbani, Alireza Mehridehnavi, and Fedra Hajizadeh. Macular oct classification using a multi-scale convolutional neural network ensemble. *IEEE Transactions on Medical Imaging*, 37(4):1024–1034, 2017.

[45] Xuechen Li, Linlin Shen, Meixiao Shen, Fan Tan, and Connor S Qiu. Deep learning based early stage diabetic retinopathy detection using optical coherence tomography. *Neurocomputing*, 369:134–144, 2019.

[46] David Romo-Bucheli, Ursula Schmidt Erfurth, and Hrvoje Bogunović. End-to-end deep learning model for predicting treatment requirements in neovascular amd from longitudinal retinal oct imaging. *IEEE Journal of Biomedical and Health Informatics*, 24(12):3456–3465, 2020.

[47] Christoph Kniestedt and Robert L Stamper. Visual acuity and its measurement. *Ophthalmology Clinics of North America*, 16(2):155–70, 2003.

[48] David L. Donoho and Iain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[49] Vikas Gupta, Rajesh Mahle, and Raviprakash S Shriwas. Image denoising using wavelet transform method. In *International Conference on Wireless and Optical Communications Networks*, pages 1–4, 2015.

[50] Raghav Bansal, Gaurav Raj, and Tanupriya Choudhury. Blur image detection using Laplacian operator and Open-CV. In *International Conference System Modeling & Advancement in Research Trends*, pages 63–67, 2016.

[51] L.M. Francis and N. Sreenath. Pre-processing techniques for detection of blurred images. In *International Conference on Computational Intelligence and Data Engineering*, pages 59–66, 2019.

[52] Hui Kong, Hatice Cinar Akakin, and Sanjay E. Sarma. A generalized laplacian of gaussian filter for blob detection and its applications. *IEEE Transactions on Cybernetics*, 43(6):1719–1733, 2013.

[53] Gabriele Simone, Marius Pedersen, and Jon Yngve Hardeberg. Measuring perceptual contrast in digital images. *Journal of Visual Communication and Image Representation*, 23(3):491–506, 2012.

[54] Berthold KP Horn and Brian G Schunck. Determining optical flow. In *Techniques and Applications of Image Understanding*, volume 281, pages 319–331, 1981.

[55] Enric Meinhardt-Llopis, Javier Sánchez Pérez, and Daniel Kondermann. Horn-schunck optical flow with a multi-scale strategy. *Image Processing on Line*, 3:151–172, 2013.

[56] Xiaoliang Gong and Stephan Bansmer. Horn–schunck optical flow applied to deformation measurement of a birdlike airfoil. *Chinese Journal of Aeronautics*, 28(5):1305–1315, 2015.

[57] Sergey Bezryadin, Pavel Bourov, and Dmitry Ilinih. Brightness calculation in digital image processing. In *International Symposium on Technologies for Digital Photo Fulfillment*, volume 2007, pages 10–15, 2007.

[58] Markus Goldstein and Seiichi Uchida. A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PloS One*, 11(4):e0152173, 2016.

[59] Filipe Falcão, Tommaso Zoppi, Caio Barbosa Viera Silva, Anderson Santos, Baldoino Fonseca, Andrea Ceccarelli, and Andrea Bondavalli. Quantitative comparison of unsupervised anomaly detection algorithms for intrusion detection. In *ACM/SIGAPP Symposium on Applied Computing*, pages 318–327, 2019.

[60] Amir Behzad Bagheri, Mohammad Dehghan Rouzi, Navid Alemi Koohbanani, Mohammad H Mahoor, MG Finco, Myeounggon Lee, Bijan Najafi, and Jayer Chung. Potential applications of artificial intelligence (ai) and machine learning (ml) on diagnosis, treatment, outcome prediction to address health care disparities of chronic limb-threatening ischemia (clti). In *Seminars in Vascular Surgery*. Elsevier, 2023.

[61] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.

[62] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2261–2269, 2017.

[63] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019.

[64] Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *International Conference on Machine Learning*, pages 10096–10106. PMLR, 2021.

[65] Mohammad Dehghan Rouzi, Behzad Moshiri, Mohammad Khoshnevisan, Mohammad Ali Akhaee, Farhang Jaryani, Samaneh Salehi Nasab, and Myeounggon Lee. Breast cancer detection with an ensemble of deep learning networks using a consensus-adaptive weighting method. *Journal of Imaging*, 9(11):247, 2023.

[66] Malliga Subramanian, M Sandeep Kumar, VE Sathishkumar, Jayagopal Prabhu, Alagar Karthick, S Sankar Ganesh, Mahseena Akter Meem, et al. Diagnosis of retinal diseases based on bayesian optimization deep learning network using optical coherence tomography images. *Computational Intelligence and Neuroscience*, 2022, 2022.

[67] Xiaoli Zhou, Chaowei Tang, Pan Huang, Sukun Tian, Francesco Mercaldo, and Antonella Santone. Asi-dbnet: an adaptive sparse interactive resnet-vision transformer dual-branch network for the grading of brain cancer histopathological images. *Interdisciplinary Sciences: Computational Life Sciences*, 15(1):15–31, 2023.

[68] Yutong Bai, Jieru Mei, Alan Yuille, and Cihang Xie. Are transformers more robust than CNNs? In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4