Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 9, No.3, September : 2024

USING VOLUMETRIC IMAGES, ATTENTION-GUIDED 3D-CNN FOR GLAUCOMA IDENTIFICATION AND STRUCTURAL-FUNCTIONAL ASSOCIATION

Anshul jain, Research Scholar, Department of CSE, LNCT University Bhopal M.P, India, aanshul0912@gmail.com
Dr. Vikas Sakalle Professor, Department of CSE, LNCT University Bhopal M.P, India, vikassakalle@gmail.com

ABSTRACT—
The direct analysis of 3D Optical CoherenceTomography (OCT) volumes enables deep learning models(DL) to learn spatial structural information and discovernew bio-markers that are relevant to glaucoma. Downsampling3D input volumes is the state-of-art solution toaccommodate for the limited number of training volumesas well as the available computing resources. However,this limits the network's ability to learn from small retinalstructures in OCT volumes. In this paper, our goal isto improve the performance by providing guidance to DLmodel during training in order to learn from finer ocularstructures in 3D OCT volumes. Therefore, we propose anend-to-end attention guided 3D DL model for glaucomadetection and estimating visual function from retinal structures.The model consists of three pathways with the samenetwork architecture but different inputs. One input is theoriginal 3D-OCT cube and the other two are computedduring training guided by the 3D gradient class activationheatmaps. Each pathway outputs the class-label and thewhole model is trained concurrently to minimize the sumof losses from three pathways. The final output is obtainedby fusing the predictions of the three pathways. Also, toexplore the robustness and generalizability of the proposedmodel, we apply the model on a classification task forglaucoma detection as well as a regression task to estimatevisual field index (VFI) (a value between 0 and 100). A 5-fold cross-validation with a total of 3782 and 10,370 OCTscans is used to train and evaluate the classification andregression models, respectively. The glaucoma detectionmodel achieved an area under the curve (AUC) of 93.8%compared with 86.8% for a baseline model without theattention-guided component. The model also outperformedsix different featurebased machine learning approachesthat use scanner computed measurements for training. Further,we also assessed the contribution of different retinallayers that are relevant to glaucoma. The VFI estimationmodel achieved a Pearson correlation and median absoluteerror of 0.75 and 3.6%, respectively, for a test set of size3100 cubes.

INDEX TERMS—
3D convolutional neural networks, opticalcoherence tomography, gradient-weighted class activationmaps, glaucoma detection, visual field estimation, attentionguided deep learning

I. INTRODUCTION

GLaucoma isthe leading cause of irreversible blindnessworldwide. The number of worldwide glaucomapatients, aged 40-80 years, is estimated to beapproximately 80 million in 2020 with about 20 millionincrease since 2010 [1]. Glaucoma is associated with opticnerve damage, functional vision loss and death of retinalganglion cells [2]. Structural and functional methods areutilized jointly to determine the severity of glaucomaand monitor its progression [3]. One of the functionaltests utilized is called visual field test (VFT), and it isused to evaluate vision loss due to glaucoma and otheroptic nerve diseases [4]. VFT, however, is costly, timeconsumingand shows poor repeatability as it is greatly affected by cataracts, visual acuity, glaucoma medications,severity of glaucoma, learning effect, distractionand other factors [5], [6].On the other hand, structural measurements are objectiveand based on the imaging of the optic nerve head(ONH), macula and surrounding regions. It enables thequantification of retinal structures relevant to glaucomasuch as the retinal nerve fiber layer (RNFL) and ganglioncell-inner plexiform layer (GCIPL) complex [3]. Manyresearchers have investigated the relationships betweenvisual field test results and structural measures thatare produced by optical

coherence tomography (OCT)scanners [7]–[9]. For instance in [7], [8], RNFL thicknesswas found to be linearly related to visual field loss atadvanced disease stage. However, finding such relationshipis very challenging as sometimes the optic nervechanges before the visual field loss [10]–[13], and othertimes visual field loss occurs prior to structural damageat the optic nerve [14].Deep learning (DL) approaches have been previouslyused with fundus 2D-colour images for ocular diseasedetection and diagnosis [15], [16]. This includes segmentationof retinal vessels [17]–[20], optic disc and optic cupsegmentation [18], [21], [22], classification of glaucoma
[23]–[28], and image registration [29]. A more recent 3Dimaging modality is the spectral-domain OCT technologythat provides clinicians with high-resolution imagesand quantified measurements of the retinal structures. Inclinics, OCT scans are the standard for eye care and areemployed for diagnosing and monitoring various retinaldiseases, evaluating progression, and assessing responseto therapy [30]. This technology enables the use of 3D DLtechniques to learn new structural parameters useful forthe diagnosis and management of glaucoma, quantify itsrelevant ocular structures (such as the individual retinallayers, optic nerve head, choroid, and lamina cribrosa)[31], and investigate whether functional measurementssuch as visual field index (VFI) or mean deviation (MD)can be inferred from structure (i.e. OCT volumes).Further, the literature shows that most of DL initiativesin OCT glaucoma detection have primarily dependedon scanner measurements of different retinal structuressuch as the thickness of RNFL and the ganglion cellcomplex (GCC), limiting the generalizability of DL modelsto measurements from different commercial scannerssince they are calculated differently. This also limitsthe ability of DL models to discover new structuralbiomarkers which are not quantified by the scanners. Forexample, in [32], glaucoma was diagnosed by trainingDL model using thicknesses maps for both RNFL andGCIPL with an AUC of 93.7%. Also in [33], AlexNetpretrained model [34] was used for feature extractionusing probability and thickness maps of RNFL andGCIPL layers, followed by random forest classifier [35]to discriminate between healthy and glaucomatous eyes.The best performance was achieved using RNFL probabilitymap with an accuracy of 93.1%. In another studyby An et al. [36], VGG19-based transfer learning modelwas performed to detect glaucoma using both thicknessand deviation maps for each of RNFL and GCC layers.Then, a random forest classifier combining features fromdifferent inputs achieved an AUC of 96%. Further, Wanget al. [37] proposed S-D net that has two parts, Snetfor segmentation of 6 retinal layers and D-net forthe diagnosis of glaucoma according to RNFL thicknessvector of length 1024 calculated from the segmentationresults. The method achieved a dice coefficient of 0.959for S-net and an accuracy of 85.4% for D-net.The only end-to-end DL model that relied on the 3Dscans as an input was presented by Maetschke et al.[38]. A 3D-CNN model composed of 5 convolutionallayers with ReLU activation, batch-normalization usinginput volumes that were downsampled by a factor ofnearly 80 (64_64_128 (b-scans_a-scans_depth) vs originalsize of 200_200_1024). The highest achieved AUCwas 94% which outperformed classical machine learning(ML) techniques. In [39], we extended the 3D-CNNmodel proposed by Maetschke et al. [38] to investigatewhether utilizing larger input volumes would improvethe network performance or not. We used an inputvolumes of size 128_128_256 to train a network with8 convolutional layers. We obtained an AUC of 97% using the same dataset used in [38]. Further, in [40],
Maetschke et al. extended his work to assess structuralfunctionalcorrelation using 3D-CNN model. Specifically,VFI and MD functional measures were estimated directlyfrom 3D raw OCT scans. The highest achieved PearsonCorrelation (r) was 0.88 compared with 0.74 for the bestperforming classical ML algorithms.Another important aspect is the clinical interpretabilityand transparency [41] of the developed DL models.In this regard, class activation maps (CAMs) [42] andgradient-weighted class activation maps (grad-CAMs)[43] have been recently proposed to reveal insights intothe decisions of deep learning models. Both of thesetechniques identify areas of the images that the networkrelied on heavily to generate the classification. However,CAM requires a specific network architecture, namelythe use of a global average pooling layer prior to theoutput layer. Grad-CAM is a generalized form of CAMand can be used with any CNN-based architecture withoutany additional

requirements.Further, the visualization of DL models for glaucomadetection has been studied in three papers [36], [38],[39]. An et al. [36] identified pathologic regions in 2Dthickness maps using grad-CAM, which have shown tobe in agreement with the important decision makingregions used by physicians. Also, Maetschke et al. [38]implemented 3D-CAM to identify the important regionsfor detecting glaucoma in 3D OCT volumes. The mapswere however, in a coarse resolution that matched thedownsampled input image. This method also employedspecific architecture changes to accommodate the requirementsof CAM generation. It is also noteworthythat neither of these approaches analyzed the CAMs inany systematic fashion, and merely used the heat mapsto validate findings in a small number of images thatwere qualitatively assessed. Lastly, in our previous work[39], we used 3D grad-CAM to visualize the importantdecision regions in a higher resolution than was availablebefore. One of the conducted experiments was to quantitativelyvalidate grad-CAM results for 3D OCT volumesby occluding important decision regions identified in theheat maps and assessing the impact of this on the performanceof the model. Occluding the most importantdecision regions in grad-CAM heatmaps dropped theperformance by nearly 40% while occluding the leastimportant areas only resutled in a 4% drop in the performance.

The paper also included a quantitative comparisonbetween CAM and grad-CAM heatmaps with thelater significantly outperforming CAM heatmaps. Thishas motivated us to use grad-CAM heatmaps to provideguidance to DL model during training and improve theperformance by learning the finer ocular structures in 3DOCT volumes associated with disease as well as visualfunction.In this paper, we propose an end-to-end attentionguided DL framework for glaucoma detection and estimatingVFI. The model is trained directly on 3D volumesusing three inputs, one is the original 3D-OCT
cube and the other two are computed during trainingguided by 3D grad-CAM heatmaps [43]. The modelconsists of three pathways that have the same networkarchitecture. First pathway uses the original volumes asan input after downsampling to size 256_64_64. Thengrad-CAM heatmaps are generated to identify retinalstructures in the original volumes, which the networkrelies on for detecting glaucoma. Occlusion of the lessimportant retinal structures in original cubes is used asan input for the second pathway. The input for the thirdpathway is obtained by cropping the region with themost important structures. The contribution of this workcan be summarized as follows:_ The proposed approach continues to avoid thedependency on segmented structural thicknessesthrough direct analysis of raw OCT scans, and alsoimproves on previously approached techniques byfocusing on the important decision areas, identifiedby grad-CAM heatmaps, to learn more about fineocular structures._ The performance of the model is evaluated for twodifferent tasks: i) A classification task for glaucomadetection and ii) A regression task for VFI functionalparameter estimation._ The proposed DL framework provides analysis of3D attention maps in a higher resolution than wasavailable before. This facilitates the understandingand interpretation of the network's decision forglaucoma detection and diagnosis.

_ For the first time, we provide a quantitative clinicalassessment for the contribution of different ocularstructures that the network relied on when detectingglaucoma._ Intensive experiments are conducted to demonstratethe effectiveness of the proposed approach and itwas compared with another 3D-CNN and classicalML approaches trained on scanner computed measurements.

The rest of the paper is organized as follows. SectionII explains the proposed network architecture andDL framework. In Sections III and IV, we describethe dataset and experimental setup used for trainingand testing each of the glaucoma detection and VFIestimation models, respectively. Section V discusses theexperimental results and the performed clinical assessmenttechniques. Finally, we conclude and outline futureresearch directions in Section VI.II. ATTENTION-GUIDED NETWORK ARCHITECTURE

The framework of the proposed attention-guided DLmodel (AG-OCT) is presented in Figure 1. The modelconsists of three pathways called global, focused andlocal OCT structure pathways. They have same networkarchitecture but different inputs with resolutionof 256_64_64 (depth_b-scans_a-scans). Also, the firsttwo pathways share same trainable weights, while thethird one has its own learned

weights. This is becausethe first two pathways have the same field of view, wherethe inputs are centered on the ONH and cover an areaof 6_6_2 mm3. While the third pathway has a smallerfield of view, i.e. different coverage area, as it focuseson a small region of the original area. The networkarchitecture contains eight 3D-convolutional layers, eachis followed by ReLU activation [44], batch-normalization[45] and max-pooling in order. The 3D convolutional, layers have incremental number of filters of 16-16-32-32-32-32-64-128 with kernel size of 3, and stride of 1 for alllayers. Also, 3D max-pooling layers have size of 2 andstride of 2. This is followed by global average poolinglayer and a fully-connected output layer in order). Detailsabout each pathway and the model loss are provided inthe following sub-sections.

A. Pathway#1: Global OCT StructureThis pathway learns the global OCT retinal structuresthat are relevant to the target task, i.e. glaucoma or VFIestimation. It receives its input by downsampling theoriginal 3D-OCT cubes to size 256 _64_64. We implement3D grad-CAM to generate heatmaps that highlightthe important decision areas in input volumes, followingthe explanation provided in [43]. In this context, grad-CAM heatmap is compute for conv#2 feature map that is128_32_32)(Lay. 2 in Figure 1). The generated heatmapis then used to derive the input of the other two pathwaysduring training. We do not use CAM as it restrictsthe network architecture design. Further, CAM wouldgenerate heatmap visualization only for conv#8 featuremap, which in our case has a size of 4_2_2. Hence,when resizing and overlaying on the original cube ofsize 1024_200_200 (depth_b-scans_a-scans) will notprovide any meaningful results.B. Pathway#2: Focused OCT StructureThe aim of the second pathway is to learn the correctoutput (e.g. glaucoma or healthy) using occluded cubes.The least important regions in the original cube arehidden, guiding the network to learn the location of theimportant decision areas. The rational is that if grad-CAM yields the correct decision areas, then hiding theleast important decision areas should not have a greatimpact on the network performance results, since theseareas are not important and most likely refer to noiseand/or redundant information that are contained in theOCT volumes.To do this, the input volumes are occluded by zeroingthe rows and columns with the lowest heatmap weights.Specifically, we extract a set of indices with the lowestweights per each dimension using average pooling forspatial dimension reduction. For example, a heat mapwith size 1024_200_200 is reduced to a vector of size1024_1_1 by averaging the values of each 200_200 mapto get a rank of weights for the first dimension, i.e.depth. The indices of the lowest x values (i.e. weights) inthe resultant vector represent the least important region

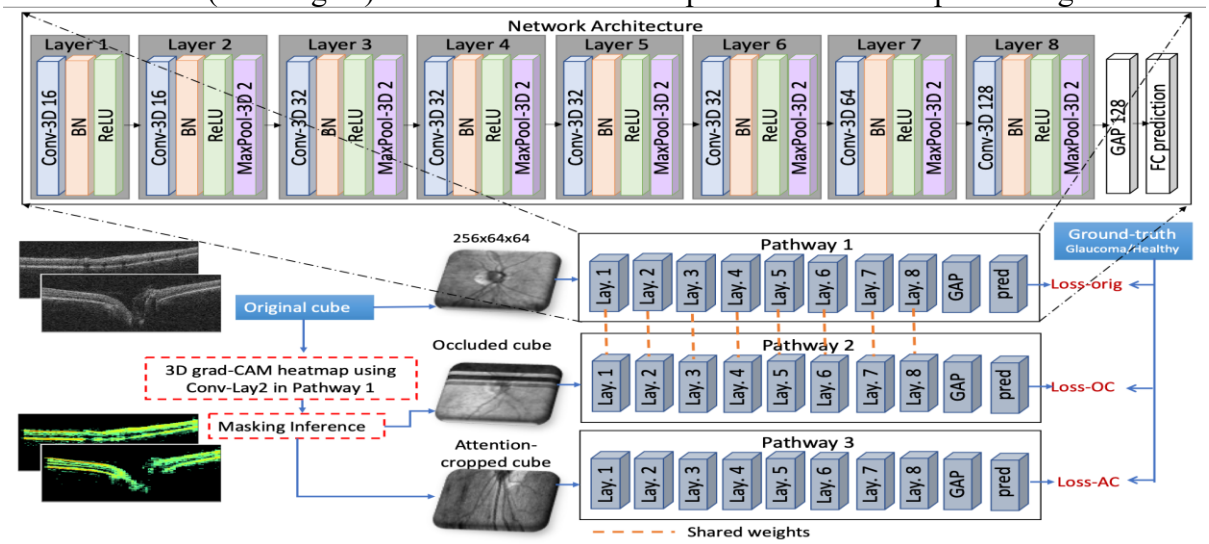


Fig. 1. Framework of the proposed attention-guided DL model using 3D OCT volumes (AG-OCT)

for this dimension. We apply this process on the bscansand depth dimensions with x values of 64 and256 respectively (both values are chosen to match thedesired input shape of the network), while we considerthe 200 a-scan columns are all important. This means thata fixed region of size 256_64_200 is occluded for eachvolume in its original resolution. The occluded-cube isdownsampled

to size 256_64_64 and is used as inputfor the second pathway of the network.C. Pathway#3: Local OCT StructureThe third pathway enables the network to learn moreabout the local structures in the OCT volumes by retainingdetail and image resolution in the important areas(i.e. a close up zoom into the important ocular structures).In this context, we use the generated grad-CAMheatmap to find the most important 3D sub-region in theinput volume, which we call the attention-cropped cube.Specifically, we performed spatial dimension reductionmethod used in the second pathway to select the mostimportant 64 b-scans from a total of 200 b-scans (i.e.rows with the highest weights along this dimension).
This means that more than two-third of the voxels ofinput volumes are discarded. For example, given a cubeof size 1024_200_200, the extracted attention-croppedcube size is 1024_64_200, while the size of the regionwhich is taken away is 1024_136_200. The attentioncroppedcube is also downsampled to size 256_64_64and is used as an input for the third pathway.D. Training LossThe three pathways of the proposed model are trainedconcurrently, so that the attention maps are learnedjointly. Each pathway has its prediction vector and lossas shown in Figure 1. The objective of the training isto minimize the total loss described in Equation 1, thatis the sum of the three pathway losses, in addition toa regularization loss term to avoid overfitting duringtraining.
$L = L(Y0p1, Y) + L(Y0p2, Y) + L(Y0p3, Y) + k\mathbf{W}k22(1)$
Where Y0p 1 , Y0p2 , Y0p3 are the predictions of pathway #1,2, and 3 in order. Y is the ground truth vector and Lis the loss and k**W**k22is the regularization loss term forconvolutional layers.

III. GLAUCOMA DETECTION

In this section, we explain how the proposed AGOCTmodel is used for glaucoma detection. The model is trained to classify an OCT volume as healthy orglaucoma. The output has a value of [1,0] for healthyclass and [0,1] for glaucoma class.
DATASET. The dataset contains 3782 OCT scans fromboth eyes of 555 individuals, acquired on a Cirrus SDOCTScanner (Zeiss; Dublin, CA, USA) over multiplevisits. The dataset has 427 healthy scans from 109 individualsand 3355 glaucoma scans from 446 individualswith primary open angle glaucoma (POAG). The clinicaldefinition of healthy/glaucoma is made based on the visualfield test results. The scans are centered on the ONHand has 200_200_1024 (a-scans_b-scans_depth) voxelsper cube covering an area of 6_6_2 mm3. This study isan observational study that is conducted in accordancewith the tenets of the Declaration of Helsinki and theHealthy Insurance Portability and Accountability Act.The Institutional Review Board of New York Universityand the University of Pittsburgh approved the study, andall subjects give written consent before participation.
TRAINING AND TESTING. We use a fully-connected softmaxlayer with 2 units for FC prediction layer (see Figure1). The 3782 OCT volumes are split into a training,validation and testing subsets, containing 3031 (healthy:325, POAG: 2706), 379 (healthy: 47, POAG: 332) and 372(healthy: 55, POAG: 317) scans, respectively. OCT scansbelonging to the same patient are included in only oneof the three splits. The proposed model is trained usingAdam optimizer with a learning rate of 1e□4. We alsouse weighted cross entropy loss [46] to avoid biasedtraining due to the class size imbalance in the data.Training is performed with a batch of size 12 through100 epochs. To avoid overfitting during training, we useL2 regularization loss with $l = 0.0001$ and drop out layerwith probability of 0.3. After each epoch, the area underthe curve (AUC) was computed for the validation set,and the network is saved if an improvement in the AUCis observed.
EVALUATION. For the evaluation of the proposed model,five statistical performance measures are used, namely,AUC, accuracy, Matthews correlation coefficient (MCC),recall, precision and F1-score. Performance measures arecomputed using predictions from each pathway separatelyas well as the fusion of 3-pathway predictionsusing min, max and average operations. For reliable andstable results, we repeat the training 5 times and reportthe average performance measures for the five folds.

IV. STRUCTURAL-FUNCTIONAL CORRELATION

The purpose of this section is to explore the generalizabilityand robustness of our attention-guided model(AG-OCT). To do this, we train the AG-OCT model to estimatethe VFI parameter from structural data (i.e. OCTvolumes), which is very important since clinicians useboth structural and functional data to monitor glaucomaprogression. The output is a value between 0 and 100.**Dataset.** We use a large dataset consisting of 10,370ONH OCT volumes and their corresponding visual fieldtest results. Structural OCT scans are captured from1678 individuals across multiple visits using Cirrus SDOCTscanners. Scans with signal strength less than 6 arediscarded. The visual field test is performed using theSwedish interactive thresholding algorithm 24-2 perimetry(SITA standard; Humphrey Field Analyzer; Zeiss).The VFI can range from 0% (perimetrically blind field)to 100% (normal visual field). Similar to the previouscohort, all subjects give written consent before participation

TRAINING AND TESTING. We trained and evaluated theregression model using the same architecture and experimentalsetup as the glaucoma detection experiment withthree main changes. Firstly, we replace the last softmaxlayer with one unit fully connected layer with linearactivation. Secondly, the mean squared error loss is usedduring training instead of weighted cross entropy. Lastly,we employ polynomial regression [47] for combining the3-pathway predictions that is trained using same trainingdata as the AG-OCT model (i.e. 70% of data) and the restis used for testing (i.e. 30%, 3100 scans). Also, hyperparameterselection for polynomial degree is performedusing Grid-search with 10-fold cross validation.

Evaluation. For the evaluation, five evaluation metricsare computed namely, root mean squared error(RMSE), mean absolute error (MAE), median absoluteerror (MDAE), Pearson's correlation coefficient (r) andSpearman's rank correlation coefficient (r).

V. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed attention-guided DL model is implementedusing Python and TensorFlow [48] on a singleV100 GPU. We divide our results and experiments intofive sections. In section V-A, we report the performancemeasures for glaucoma detection. In section V-B, we conductdifferent experiments to analyze the performance ofour AG-OCT model. This followed by detailed clinicalanalysis of grad-CAM attention maps in section V-C. Insection V-D, we provide comparative results with stateof-art approaches including classical ML techniques. Finally,in section V-E, we present the regression modelresults for VFI estimation.

A. GLAUCOMA DETECTION

Table I has the performance measures for glaucomadetection using the proposed attention-guided model.The table shows that the average of the 3-pathwaypredictions has the best performance with an AUC of93.8%. Interestingly, the second best performance comesfrom pathway #2, i.e. the occluded cube with an AUCof 93.0%. This means that hiding some of the least importantregions in 3D cubes improves the performance,because those regions might refer to noisy or redundantareas in 3D volume. This is followed by predictionsfrom max-fusion, min-fusion, pathway #3, and pathway#1 with AUCs of 92.8, 92.2, 91.2, and 86.7% in order.Further, to examine the computational complexity ofthe proposed framework, we compute the average andstandard deviation for execution time of the validationand test sets with 23.4_ 2.2 and 22.7_3.8 (in seconds),respectively, for a batch of size 12. This means that thenetwork takes less than 2 seconds for one cube to beprocessed.

B. QUANTITATIVE ANALYSIS RESULTS

To demonstrate the influence of attention maps onthe performance of the AG-OCT model, we trained one

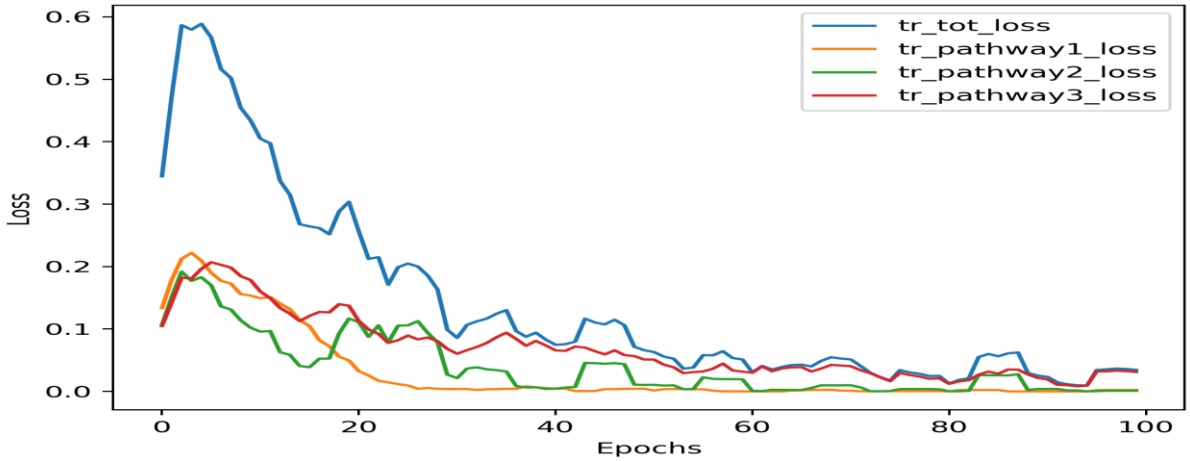


Fig. 2. Training losses for glaucoma detection using AG-OCT model

TABLE I 5-FOLD AVERAGE PERFORMANCE MEASURES FOR THE ATTENTION GUIDED AND BASELINE MODELS

	Accuracy	MCC	Recall	precision	F1-scorew	AUC
Pathway#1	**85.184**	**0.373**	**89.425**	**93.576**	**91.371**	**86.741**
Pathway#2	**91.300**	**0.553**	**94.838**	**95.418**	**95.102**	**93.007**
Pathway#3	**90.300**	**0.513**	**94.826**	**94.115**	**94.426**	**91.189**
Fusion-average	**91.073**	**0.557**	**95.119**	**94.730**	**94.882**	**93.769**
Fusion-min	**90.452**	**0.430**	**98.630**	**91.182**	**94.736**	**92.243**
Fusion-max	**85.319**	**0.528**	**85.216**	**97.873**	**91.056**	**92.785**
Baseline DL model	**86.315**	**0.399**	**90.928**	**93.409**	**92.088**	**86.803**

TABLE II
SHARING WEIGHTS IMPACT ON THE PERFORMANCE THE PROPOSED ATTENTION-GUIDED MODEL

	Accuracy	MCC	Recall	Precision	F1-score	AUC
No weight sharing: each pathway learns its weights						
Pathway#1	90.349	0.496	94.311	94.880	94.595	90.198
Pathway#2	91.413	0.553	94.838	95.418	95.102	93.007
Pathway#3	92.225	0.580	95.808	95.522	95.665	88.579
Fusion - average	91.421	0.521	95.808	94.675	95.238	91.392
Fusion - min	92.225	0.484	99.701	92.244	95.827	90.095
Fusion - max	89.008	0.533	91.018	96.508	93.683	90.922
Full weight sharing: Layers of all pathways share weights						
Pathway#1	90.349	0.585	91.916	97.152	94.462	90.527
Pathway#2	91.153	0.586	93.413	96.594	94.977	91.528
Pathway#3	92.493	0.599	95.808	95.808	95.808	91.475
Fusion - average	91.689	0.611	93.713	96.904	95.282	92.210
Fusion - min	93.298	0.601	97.904	94.783	96.318	91.481
Fusion - max	89.008	0.588	89.521	98.033	93.584	91.137
Partial sharing: Only pathway 1 and 2 share weights						
Pathway#1	89.544	0.520	92.216	95.950	94.046	90.970
Pathway#2	90.080	0.534	92.814	95.975	94.368	93.054
Pathway#3	92.225	0.590	95.509	95.796	95.652	93.432
Fusion - average	90.349	0.553	92.814	96.273	94.512	94.405
Fusion - min	93.298	0.584	98.802	94.017	96.350	92.647
Fusion - max	88.204	0.559	88.922	97.697	93.103	94.505
FusionAvgPath.1&2	89.812	0.539	92.216	96.250	94.190	92.930
FusionAvgPath.1&3	92.493	0.627	94.910	96.646	95.770	94.528
FusionAvgPath.2&3	94.102	0.699	96.108	97.273	96.687	94.271

branch of our proposed framework, where the input isthe downsampled original volumes and the output is theprediction label (i.e. glaucoma or healthy). This meansthat we trained pathway #1 only without the attentionguidedbranches, which we used as our baseline model.Table I reported an AUC

of 86.8% for the baseline modelthat is very close to the performance of pathway #1 whenit is learned jointly with Pathway #2 and 3. This confirmsthat training the network without the guidance of grad-CAM heatmaps has approximately 7% and 3% drop inthe AUC and F1-score measures espectively.Further, Figure 2 displays the training loss curves foreach pathway separately as well as the total training loss.Loss of pathway#1 reached its minimum at epoch 20while other losses decreased further until epoch 40. Also,the total loss was highly influenced by pathway#2 loss.This also clarifies why pathway#2 predictions has thebest performance for glaucoma detection. This also suggeststhat weighting losses of 3-pathways might guidethe network to pay more attention to the branch withslow convergence.We also examined the impact of sharing weightsacross ranches/pathways by running 3 experimentsthat use same data split and different settings for sharingweights, colorbluenamely: i) full sharing: the 3-pathwaysshare the same weights, ii) no sharing: each pathwaylearned its own weights, and iii) partial sharing: onlypathway#1 and pathway#2 share the same weights, whilepathway #3 had its own weights. Table II reports theperformance measures for the 3 experiments, whereseparate weights for each pathway has the lowest performance,while partial sharing recorded the highestAUCs. The results confirm our hypothesis for sharingweights between first two pathways since they share thesame field of view, where the inputs are centered onthe ONH and cover an area of 6_6_2 mm3. While thethird pathway has a smaller field of view, i.e. differentcoverage area, as it focuses on a small region of theoriginally scanned area.

Table II also shows the average fusion results forall possible pairs of pathways namely, pathways# 1&2,1&3 and 2&3. From the table, there is a very slightperformance difference for average fusion of 3-pathwaysversus 2-pathways. More importantly, dropping the attentionmap pathway from the fusion (i.e. pathway# 3)resulted in approximately 2% decrease in the recordedperformance measures. While, fusion of pathway# 2&3had the highest performance measures.C. Clinical Analysis of Attention Maps for GlaucomaIn this section, we present detailed analysis of grad-CAM heatmaps to give insights about the clinicalbiomarkers that our AG-OCT model relies on for glaucomadetection. This is very essential not only to understandthe network decision, but also to increase reliabilityof DL approaches for direct analysis of 3DOCT scans by showing agreement of decision makingprocess between DL approaches and clinicians. Figure3 visualizes the important retinal structures for bothhealthy and glaucoma cases by overlaying grad-CAMheatmap on the original volumes. The figure displays theoverlaid heatmaps for both the enface/top view as wellas the b-scans/side view. It is clear from the figure thatthe AG-OCT model depends on the OCT retinal layersregion for detecting glaucoma.Further, to show which retinal structure/layer hasthe greatest impact, we quantify the presence of eachretinal layer in the generated grad-CAM heatmaps. Ifthe presence of a specific retinal layer, i.e bio-marker, ishigh, then this means that our model depends on thisretinal structure for detecting glaucoma. Visualizationand abbreviation of retinal layers are presented in Figure4. In this regard, we adopt the OCT retinal layerssegmentation method described in [49], to classify each2D b-scan slice into 9 classes, namely: background and8 different retinal layers, namely RNFL, GCL+IPL, INL,OPL, ONL, IS, OS, and RPE as shown in Figure 5.To run this experiment, we perform the followingsteps. We select 80 3D OCT scans from the test set(40: healthy and 40: POAG) and apply the segmentationmethod on each b-scan separately to generate 9 binarymasks: one for each of the eight retinal layers in additionto the background mask. We also extracted theforeground mask to assess the influence of the wholeretina area. Then, we computed the average heatmapweights for each b-scan in each generated mask. In total,this resulted in 16000 average heatmap values for eachbinary mask (200 b-scans _ 80 cubes). The validationprocess was done for each set of healthy and POAGcases, separately.To report the contribution of each retinal layer, we usebox-plots to represent the average computed heatmapvalues for each layer separately, as shown in the firstrow of Figure 6. From the figure, RPE, OS, IS, RNFL andGCL+IPL have the highest correlation with grad-CAMheatmaps, in order, where the median heatmap valuelies between 0.3 and 0.5 for those layers. While ONL,INL, and ONL have shown less influence on the networkdecision where median value lies between 0.1 and 0.3.These findings are non-intuitive for clinicians becausethese layers are not traditionally associated with glaucoma.

However, these findings show that DL approachesmight depend on information from unknown features inthe tissue such as thinning of the inner retina layers.Also, the figure demonstrates that RNFL has higherheatmap values in healthy cases versus POAG cases withmedian values of 0.3 and 0.4 respectively. As expected,background area has the least influence on the networkdecision.D. Comparative ResultsFor comparative study, we follow Maetschke et al.[38] and compare our AG-OCT model against feature

TABLE III
5-FOLD AVERAGE PERFORMANCE MEASURES FOR GLAUCOMA DETECTION USING FEATURE-BASED MACHINE LEARNING METHODS

	Accuracy	MCC	Recall	Precision	F1-score	AUG
Extra Trees	**85.563**	**0.436**	**87.367**	**96.160**	**91.498**	**89.556**
Gradient Boosting	**90.318**	**0.395**	**95.660**	**93.69**	**94.586**	**90.380**
Logistic Regression	**82.908**	**0.475**	**83.223**	**97.464**	**89.660**	**91.499**
Naive Bayes	**81.583**	**0.448**	**82.370**	**96.806**	**88.776**	**89.689**
Random forest	**88.128**	**0.469**	**91.026**	**95.588**	**93.176**	**89.098**
SVM(Linear)	**81.120**	**0.454**	**81.202**	**97.477**	**88.450**	**91.550**
SVM(poly)	**88.008**	**0.447**	**91.166**	**95.244**	**93.117**	**91.204**
SVM(RBF)	**80.749**	**0.442**	**81.126**	**97.140**	**88.254**	**90.602**
Proposed AG-OCT	**91.073**	**0.557**	**95.119**	**94.730**	**94.882**	**93.769**

based ML approaches, where we use Cirrus OCT scannercomputed measurements for training classical ML algorithms.Specifically, we use 22 measurements includingperipapillary RNFL thickness at 12 clock-hours, peripapillaryRNFL thickness in the four quadrants, averageRNFL thickness, rim area, disc area, average cup-todiscratio, vertical cup-to-disc ratio and cup volume.We normalize all features by subtracting the mean andscaling to unit variance.Six ML classifiers are trained, namely, Support VectorMachine (SVM), Logistic Regression ,Na¨ıve Bayes,Gradient Boosting, Extra Trees and Random Forest. Weused the same dataset used for training AG-OCT modeland same split (i.e. train, validation and test). We usedthe validation set to select the best hyper-parameters foreach classifier using grid-search. We also computed thesame performance measures. For reliability, we perform5-fold cross-validation and report the average performancemeasures for the test set as shown in Table V-D.From the table, the best feature based ML model thathas the highest AUC and a good balance between recalland precision is gradient boosting with an AUC of 91.5%,that is 2.3% less than our AUC (i.e. AG-OCT). This isfollowed by SVM with polynomial kernel with AUC of91.2%, with higher measure for precision than recall. Allother feature based ML classifiers either showed an AUCless than 89% or strongly biased towards one class (i.e.significant difference between recall and precision). Ina nutshell, not only does this experiment confirm theeffectiveness of our proposed approach but it also showsthat DL approaches have the potential to learn directlyfrom raw volumes with better performance than the oneachieved by relying on scanner extracted features.Etructural-Functional CorrelationTable V-E reports the performance measures for theattention-guided regression model. It is revealed fromthe table that polynomial regression using 3-pathwaypredictions has significantly outperformed the other predictionswith Pearson correlation (r) of 0.75 and MAE of8 for a test set of size 3100 cubes. The table also showsthat predictions from pathway#2 has slightly better performance
es than the other two pathways with

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 9, No.3, September : 2024

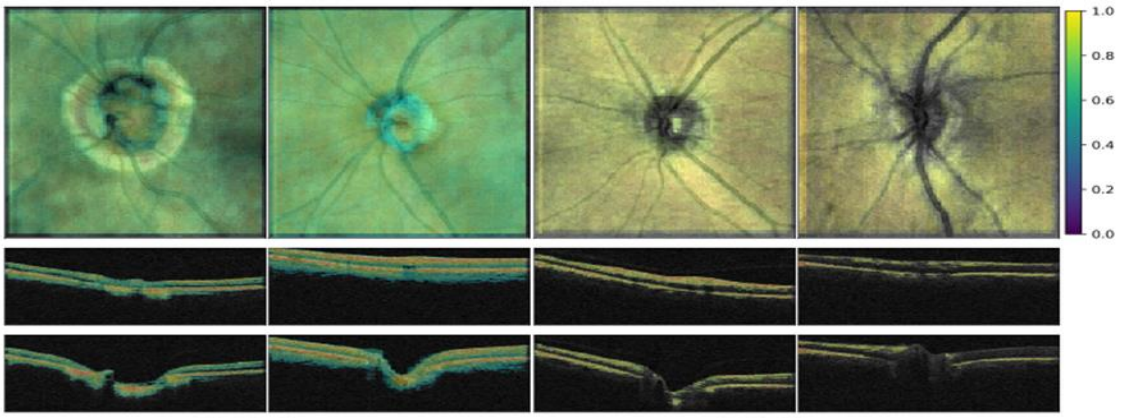
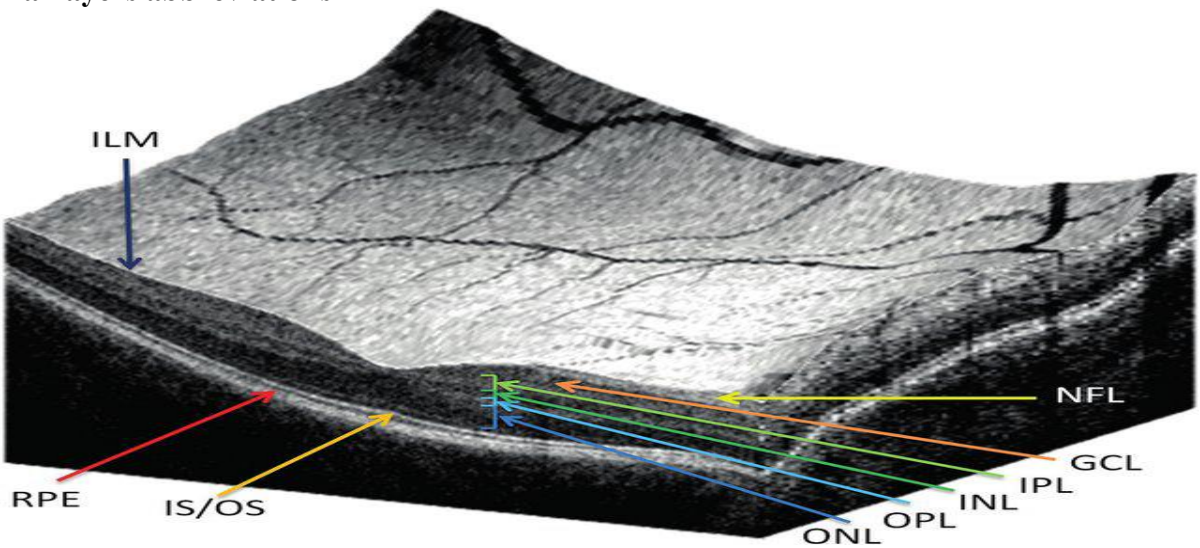


Fig. 3. Grad-CAM attention maps. First row shows overlaid grad-CAM heatmap for enface view while second and third rows show b-scan slices#50 and 100 in order

Retinal layers abbreviations



RNFL: retinal nerve fiber layer
GCL: ganglion cell layer
IPL: inner plexiform layer
INL: inner nuclear layer
OPL: outer plexiform layer
ONL: outer nuclear layer
IS: photoreceptor inner segments
OS: photoreceptor outer segments
RPE retinal pigment epithelium

Fig. 4. Visualization and abbreviation of different retinal layers in OCTscan. The left image is taken from this paper [50]

TABLE IV PERFORMANCE MEASURES FOR VISUAL FIELD INDEX ESTIMATION EXPERIMENT

predications	RMSE	MAE	MADE	r	ρ
Pathway#1	17.139	12.362	9.071	0.497	0.648
Pathway#2	16.783	11.628	8.000	0.491	0.649
Pathway#3	19.371	14.060	10.701	0.459	0.582
Fusion-regression (3-Pathways)	13.403	7.954	3.615	0.582	0.750
Regression(Pathway#1)	14.806	9.103	4.441	0.497	0.680
Regression(Pathway#2)	14.834	9.035	3.729	0.451	0.681

Regression(Pathway#3)	**15.874**	**10.078**	**4.602**	**0.451**	**0.620**

Pearson correlation (*r*) of 0.65 and MAE of 11.6. Also,refining the predictions from the individual pathwayshad decreased the MAE by at least a values of 2.5 (i.e.MAE = 9.1).

VI. CONCLUSION AND FUTURE WORK

We present an end-to-end 3D attention-guided modelthat can be used for multiple tasks including classificationand regression through direct analysis of 3Draw volumes that outperformed the scanner computedmeasurements. The model leverages the rich structuralinformation embedded in the high resolution 3D OCTcubes by the guidance of grad-CAM attention map,which resulted in better performance compared withbaseline models and feature based ML approaches. Importantly,we showed that using the attention-guidedframework we can identify the important regions inthe OCT volumes, whereby redundant regions of thescan can be excluded from the analysis. Also, grad-CAM allowed for a qualitative clinical analysis and understandingof the DL network. In particular, we quantitativelymeasured the importance of different retinallayers in 3D OCT cubes which the network relied on fordetecting glaucoma. Further, the glaucoma detection andVFI estimation experiments confirmed the effectiveness,robustness and generalization of the proposed modelthat is able to learn from high resolution 3D volumes.Both tasks showed that the fusion of predictions fromthe three pathways (i.e. attention-guided) had the bestperformance. In the future, we will apply this approachfor estimating other functional parameters and detectingother ocular diseases. We also plan to improve the performanceof the model by enhancing the attention map.

VII. ACKNOWLEDGMENTS

This work is supported by a grant from the NationalInstitutes of Health National Eye Institute R01EY013178.

REFERENCES

[1] S. R. Flaxman, R. R. Bourne, S. Resnikoff, P. Ackland, T. Braithwaite,M. V. Cicinelli, A. Das, J. B. Jonas, J. Keeffe, J. H. Kempenet al., “Global causes of blindness and distance vision impairment1990–2020: a systematic review and meta-analysis,” The LancetGlobal Health, vol. 5, no. 12, pp. e1221–e1234, 2017.

[2] B. M. Davis, L. Crawley, M. Pahlitzsch, F. Javaid, and M. F.Cordeiro, “Glaucoma: the retina and beyond,” Acta neuropathologica,vol. 132, no. 6, pp. 807–826, 2016.

[3] K. A. Lucy and G. Wollstein, “Structural and functional evaluationsfor the early detection of glaucoma,” Expert review ofophthalmology, vol. 11, no. 5, pp. 367–376, 2016.[4] D. C. Broadway, “Visual field testing for glaucoma–a practicalguide,” Community eye health, vol. 25, no. 79-80, p. 66, 2012.

[5] M. Peracha, B. Hughes, J. Tannir, R. Momi, A. Goyal, M. Juzych,C. Kim, M. McQueen, A. Eby, and F. Fatima, “Assessing the reliabilityof humphrey visual field testing in an urban population,”Investigative Ophthalmology & Visual Science, vol. 54, no. 15, pp.3920–3920, 2013.

[6] J. Ho, S. Ameen, L. Crawley, E. Normando, M. F. Cordeiro,P. Bloom, and F. Ahmed, “Key predictors of visual field testreliability,” Investigative Ophthalmology & Visual Science, vol. 54,no. 15, pp. 2632–2632, 2013.

[7] M. T. Leite, L. M. Zangwill, R. N. Weinreb, H. L. Rao, L. M.Alencar, and F. A. Medeiros, “Structure-function relationshipsusing the cirrus spectral domain optical coherence tomograph andstandard automated perimetry,” Journal of glaucoma, vol. 21, no. 1,p. 49, 2012.

[8] N. Nilforushan, N. Nassiri, S. Moghimi, S. K. Law, J. Giaconi,A. L. Coleman, J. Caprioli, and K. Nouri-Mahdavi, “Structure–function relationships between spectral-domain oct and standardachromatic perimetry,” Investigative ophthalmology & visual science,vol. 53, no. 6, pp. 2740–2748, 2012.

[9] R. Malik, W. H. Swanson, and D. F. Garway-Heath, “Structure–function relationship in glaucoma: past thinking and currentconcepts,” Clinical & experimental ophthalmology, vol. 40, no. 4, pp.369–380, 2012.
[10] T. M. Kuang, C. Zhang, L. M. Zangwill, R. N. Weinreb, andF. A. Medeiros, “Estimating lead time gained by optical coherencetomography in detecting glaucoma before development of visualfield defects,” Ophthalmology, vol. 122, no. 10, pp. 2002–2009, 2015.
[11] K. R. Sung, S. Kim, Y. Lee, S.-C. Yun, and J. H. Na, “Retinalnerve fiber layer normative lassification by optical coherencetomography for prediction of future visual field loss,” Investigativeophthalmology & visual science, vol. 52, no. 5, pp. 2634–2639, 2011.
[12] F. A. Medeiros, L. M. Zangwill, C. Bowd, K. Mansouri, andR. N. Weinreb, “The structure and function relationship in glaucoma:implications for detection of progression and measurementof rates of change,” Investigative ophthalmology & visual science,vol. 53, no. 11, pp. 6939–6946, 2012.
[13] M. Norouzifard, A. Nemati, A. Abdul-Rahman, H. GholamHosseini,and R. Klette, “A comparison of transfer learning techniques,deep convolutional neural network and multilayer neuralnetwork methods for the diagnosis of glaucomatous optic neuropathy,”in International Computer Symposium. Springer, 2018,pp. 627–635.
[14] E. G. P. S. E. Group et al., “Results of the european glaucomaprevention study,” Ophthalmology, vol. 112, no. 3, pp. 366–375,2005.
[15] N. Panwar, P. Huang, J. Lee, P. A. Keane, T. S. Chuan, A. Richhariya,S. Teoh, T. H. Lim, and R. Agrawal, “Fundus photographyin the 21st century—a review of recent technological advancesand their implications for worldwide healthcare,” Telemedicine ande-Health, vol. 22, no. 3, pp. 198–208, 2016.
[16] P. J. Saine and M. E. Tyler, Ophthalmic photography: retinal photography,angiography, and electronic imaging. Butterworth-HeinemannBoston, 2002, vol. 132.
[17] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, “A crossmodalitylearning approach for vessel segmentation in retinalimages,” IEEE transactions on medical imaging, vol. 35, no. 1, pp.109–118, 2016.
[18] K.-K. Maninis, J. Pont-Tuset, P. Arbel´aez, and L. Van Gool, “Deepretinal image understanding,” in International conference on medicalimage computing and computer-assisted intervention. Springer, 2016,pp. 140–148.
[19] A. Dasgupta and S. Singh, “A fully convolutional neural networkbased structured prediction approach towards the retinal vesselsegmentation,” in *2017* IEEE *14*th International Symposium onBiomedical Imaging (ISBI *2017*). IEEE, 2017, pp. 248–251.
[20] H. Fu, Y. Xu, D. W. K. Wong, and J. Liu, “Retinal vessel segmentationvia deep learning network and fully-connected conditionalrandom fields,” in *2016* IEEE *13*th international symposium onbiomedical imaging (ISBI). IEEE, 2016, pp. 698–701.
[21] H. Fu, J. Cheng, Y. Xu, D. W. K. Wong, J. Liu, and X. Cao,“Joint optic disc and cup segmentation based on multi-label deepnetwork and polar transformation,” IEEE transactions on medicalimaging, vol. 37, no. 7, pp. 1597–1605, 2018.
[22] S. Sadhukhan, G. K. Ghorai, S. Maiti, V. A. Karale, G. Sarkar,and A. K. Dhara, “Optic disc segmentation in retinal fundusimages using fully convolutional network and removal of falsepositivesbased on shape features,” in Deep Learning in MedicalImage Analysis and Multimodal Learning for Clinical Decision Support.Springer, 2018, pp. 369–376.
[23] A. Sevastopolsky, “Optic disc and cup segmentation methodsfor glaucoma detection with modification of u-net convolutionalneural network,” Pattern Recognition and Image Analysis, vol. 27,no. 3, pp. 618–624, 2017.
[24] Z. Li, Y. He, S. Keel, W. Meng, R. T. Chang, and M. He, “Efficacyof a deep learning system for detecting glaucomatous opticneuropathy based on color fundus photographs,” Ophthalmology,vol. 125, no. 8, pp. 1199–1206, 2018.

[25] U. Raghavendra, H. Fujita, S. V. Bhandary, A. Gudigar, J. H.Tan, and U. R. Acharya, "Deep convolution neural network foraccurate diagnosis of glaucoma using digital fundus images,"Information Sciences, vol. 441, pp. 41–49, 2018.
[26] S. Phene, R. C. Dunn, N. Hammel, Y. Liu, J. Krause, N. Kitade,M. Schaekermann, R. Sayres, D. J. Wu, A. Bora et al., "Deeplearning to assess glaucoma risk and associated features in fundusimages," arXiv preprint arXiv:*1812.08911*, 2018.
[27] M. Kim, O. Janssens, H.-m. Park, J. Zuallaert, S. Van Hoecke,and W. De Neve, "Web applicable computer-aided diagnosis ofglaucoma using deep learning," arXiv preprint arXiv:*1812.02405*, 2018.
[28] J. M. Ahn, S. Kim, K.-S. Ahn, S.-H. Cho, K. B. Lee, and U. S.Kim, "A deep learning model for the detection of both advancedand early glaucoma using fundus photography," PloS one, vol. 13,no. 11, p. e0207982, 2018.
[29] D. Mahapatra, B. Antony, S. Sedai, and R. Garnavi, "Deformablemedical image registration using generative adversarial networks,"in *2018* IEEE *15*th International Symposium on BiomedicalImaging (ISBI *2018*). IEEE, 2018, pp. 1449–1453.
[30] J. Fujimoto and E. Swanson, "The development, commercialization,and impact of optical coherence tomography," Investigativeophthalmology & visual science, vol. 57, no. 9, pp. OCT1–OCT13,2016.
[31] F. Lavinsky, G.Wollstein, J. Tauber, and J. S. Schuman, "The futureof imaging in detecting glaucoma progression," Ophthalmology,vol. 124, no. 12, pp. S76–S82, 2017.
[32] R. Asaoka, H. Murata, K. Hirasawa, Y. Fujino, M. Matsuura,A. Miki, T. Kanamoto, Y. Ikeda, K. Mori, A. Iwase et al., "Usingdeep learning and transfer learning to accurately diagnose earlyonsetglaucoma from macular optical coherence tomographyimages," American journal of ophthalmology, vol. 198, pp. 136–145,2019.
[33] H. Muhammad, T. J. Fuchs, N. C. De, C. M. De, D. M. Blumberg,J. M. Liebmann, R. Ritch, and D. C. Hood, "Hybrid deep learningon single wide-field optical coherence tomography scans accuratelyclassifies glaucoma suspects." Journal of glaucoma, vol. 26,no. 12, pp. 1086–1094, 2017.
[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classificationwith deep convolutional neural networks," in Advances inneural information processing systems, 2012, pp. 1097–1105.
[35] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp.5–32, 2001.
[36] G. An, K. Omodaka, K. Hashimoto, S. Tsuda, Y. Shiga, N. Takada,T. Kikawa, H. Yokota, M. Akiba, and T. Nakazawa, "Glaucomadiagnosis with machine learning based on optical coherencetomography and color fundus images," Journal of Healthcare Engineering,vol. 2019, 2019.
[37] J. Wang, C. Chen, F. Li, Z. Wang, G. Qu, Y. Qiao, H. Lv, andX. Zhang, "Sd net: Joint segmentation and diagnosis revealing thediagnostic significance of using entire rnfl thickness in glaucoma,"in Conference on Medical Imaging with Deep Learning (MIDL), 2018.
[38] S. Maetschke, B. Antony, H. Ishikawa, and R. Garvani, "A featureagnostic approach for glaucoma detection in oct volumes," arXivpreprint arXiv:*1807.04855*, 2018.
[39] Y. George, B. Antony, H. Ishikawa, G. Wollstein, J. Schuman,and R. Garnavi, "3d-cnn for glaucoma detection using opticalcoherence tomography," in International Workshop on OphthalmicMedical Image Analysis. Springer, 2019, pp. 52–59.
[40] S. Maetschke, B. Antony, H. Ishikawa, G. Wollstein, J. Schuman,and R. Garnav, "Inference of visual field test performance fromoct volumes using deep learning," arXivpreprint rXiv:*1908.01428*,2019.
[41] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medicalimage processing: Overview, challenges and the future," in Classificationin BioApps. Springer, 2018, pp. 323–350.
[42] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba,"Learning deep features for discriminative localization," in Proceedingsof the IEEE conference on computer vision and pattern recognition,2016, pp. 2921–2929.

[43] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh,and D. Batra, “Grad-cam: Visual explanations from deep networksvia gradient-based localization,” in Proceedings of the IEEE InternationalConference on Computer Vision, 2017, pp. 618–626.
[44] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neuralnetworks,” in Proceedings of the fourteenth international conferenceon artificial intelligence and statistics, 2011, pp. 315–323.
[45] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deepnetwork training by reducing internal covariate shift,” arXivpreprint arXiv:*1502.03167*, 2015.
[46] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MITpress, 2016.
[47] S. A. Bini, “Artificial intelligence, machine learning, deep learning,and cognitive computing: what do these terms mean and howwill they impact health care?” The Journal of arthroplasty, vol. 33,no. 8, pp. 2358–2361, 2018.
[48] A. Gulli and S. Pal, Deep Learning with Keras. Packt PublishingLtd, 2017.
[49] S. Sedai, B. Antony, D. Mahapatra, and R. Garnavi, “Joint segmentationand uncertainty visualization of retinal layers in opticalcoherence tomography images using bayesian deep learning,”in Computational Pathology and Ophthalmic Medical Image Analysis.Springer, 2018, pp. 219–227.
[50] A. Stankiewicz, T. Marciniak, A. Dabrowski, M. Stopa, P. Rakowicz,and E. Marciniak, “Denoising methods for improving automaticsegmentation in oct images of human eye,” Bulletin of thePolish Academy of Sciences Technical Sciences, vol. 65, no. 1, pp. 71–78, 2017.