



OPEN

## Assessing the efficacy of 2D and 3D CNN algorithms in OCT-based glaucoma detection

Rafiq Karim Rasel<sup>1</sup>, Fengze Wu<sup>1,2</sup>, Marion Chiariglione<sup>1</sup>, Stacey S. Choi<sup>1,4</sup>, Nathan Doble<sup>1,4</sup> & Xiaoyi Raymond Gao<sup>1,2,3✉</sup>

Glaucoma is a progressive neurodegenerative disease characterized by the gradual degeneration of retinal ganglion cells, leading to irreversible blindness worldwide. Therefore, timely and accurate diagnosis of glaucoma is crucial, enabling early intervention and facilitating effective disease management to mitigate further vision deterioration. The advent of optical coherence tomography (OCT) has marked a transformative era in ophthalmology, offering detailed visualization of the macula and optic nerve head (ONH) regions. In recent years, both 2D and 3D convolutional neural network (CNN) algorithms have been applied to OCT image analysis. While 2D CNNs rely on post-prediction aggregation of all B-scans within OCT volumes, 3D CNNs allow for direct glaucoma prediction from the OCT data. However, in the absence of extensively pre-trained 3D models, the comparative efficacy of 2D and 3D-CNN algorithms in detecting glaucoma from volumetric OCT images remains unclear. Therefore, this study explores the efficacy of glaucoma detection through volumetric OCT images using select state-of-the-art (SOTA) 2D-CNN models, 3D adaptations of these 2D-CNN models with specific weight transfer techniques, and a custom 5-layer 3D-CNN-Encoder algorithm. The performance across two distinct datasets is evaluated, each focusing on the macula and the ONH, to provide a comprehensive understanding of the models' capabilities in identifying glaucoma. Our findings demonstrate that the 2D-CNN algorithm consistently provided robust results compared to their 3D counterparts tested in this study for glaucoma detection, achieving AUC values of 0.960 and 0.943 for the macular and ONH OCT test images, respectively. Given the scarcity of pre-trained 3D models trained on extensive datasets, this comparative analysis underscores the overall utility of 2D and 3D-CNN algorithms in advancing glaucoma diagnostic systems in ophthalmology and highlights the potential of 2D algorithms for volumetric OCT image-based glaucoma detection.

Glaucoma, a neurodegenerative disease affecting nearly 80 million people globally<sup>1</sup>, is characterized by the progressive degeneration of retinal ganglion cells, leading to vision impairment<sup>1-5</sup>. It stands as a leading contributor to irreversible vision loss, often going undetected until considerable damage has occurred<sup>6-8</sup>. In developed countries, almost half of the individuals with glaucoma remain undetected<sup>9,10</sup>, while in developing countries, the number of undetected cases can reach as high as 90%<sup>11</sup>. Therefore, early, and accurate diagnosis is essential to reduce further visual deterioration. In this regard, advances in optical coherence tomography (OCT) present a promising avenue, offering a viable means for early glaucoma detection.

By providing detailed visualization of crucial ocular structures, such as the macula and optic nerve head (ONH)<sup>12-14</sup>, OCT offers an exceptional opportunity to study the effects of vision-related diseases such as glaucoma. While automated algorithms have significantly streamlined the analysis of intricate details in OCT, the interpretation of data from these automated systems still requires expert input for accurate predictions. Consequently, deep learning (DL) based algorithms have garnered considerable popularity in streamlining OCT image analysis, providing substantial insights for deriving meaningful conclusions<sup>15-19</sup>.

Previous reports on DL-based algorithms for predicting glaucoma from OCT images showed a predominant use of the 2D convolutional neural network (CNN) compared to 3D-CNN models for glaucoma detection. For instance, Mehta et al.<sup>15</sup> proposed a multimodal model leveraging various ocular and clinical data for glaucoma detection. Their OCT-based model employs a 2D-CNN-based DL framework with a Densenet201<sup>15</sup> backbone to

<sup>1</sup>Department of Ophthalmology and Visual Sciences, The Ohio State University, Columbus, OH 43212, USA. <sup>2</sup>Department of Biomedical Informatics, The Ohio State University, Columbus, OH 43210, USA. <sup>3</sup>Division of Human Genetics, The Ohio State University, Columbus, OH 43210, USA. <sup>4</sup>College of Optometry, The Ohio State University, Columbus, OH 43210, USA. ✉email: raymond.gao@osumc.edu

train on each B-scan from OCT volumes obtained from the UK Biobank (UKB) Supplementary Materials. The model provides final glaucoma predictions by collectively analyzing each B-scan of the OCT volume, achieving a notable AUC score of 0.950.

Additionally, Christopher et al.<sup>20</sup> introduced a 2D-CNN-based approach that capitalizes on the wide-angle swept-source OCT images to extract relevant RNFL features for improved glaucoma classification. Their method has been shown to surpass glaucoma detection based on conventional metrics, such as circumpapillary RNFL thickness, standard automated perimetry, and frequency-doubling technology visual field tests. Furthermore, Garcia et al.<sup>21</sup> introduced an innovative algorithm that combines manually crafted features with a 2D-CNN algorithm, incorporating tailored residual and attention modules to achieve robust glaucoma detection. This combination enhances the discrimination between healthy, early, and advanced glaucoma samples, resulting in superior classification accuracy.

In contrast to the 2D-CNN-based methodology, Maetschke et al.<sup>22</sup> proposed a 3D-CNN algorithm for glaucoma detection. Their 3D-CNN-based model directly classifies healthy and glaucomatous eyes from unsegmented OCT volumes of the ONH region, attaining a robust AUC of 0.940 on the test dataset. Following a similar trajectory, Yasmeen et al.<sup>23</sup> introduced an attention-guided 3D DL model for analyzing OCT images to detect glaucoma. The proposed model operates through three pathways with different inputs but the same model architecture and, when validated, achieved an AUC of 0.983, outperforming traditional models and other machine learning methods. Despite the promising outcomes of 2D and 3D-CNN algorithms in glaucoma detection, our review highlights a research gap. Specifically, research examining the 2D and 3D-CNN algorithms for glaucoma detection, particularly through weight transfer techniques to enhance 3D models' performance across diverse OCT datasets, is notably limited.

Considering that macular and ONH OCT inherently offers volumetric imaging data and given the scarcity of pre-trained 3D models trained on extensive datasets, the importance of examining the strengths of current 2D and 3D-CNN models in detecting glaucoma from OCT volumes cannot be underestimated. While 2D-CNN models may not utilize volumetric information available in OCT volumes, they benefit from a wealth of pre-trained models, offering better generalization in smaller datasets. In contrast, although 3D-CNN models can leverage the volumetric information, they are often limited by a notable scarcity of pre-trained models and large volume datasets. Even though there are ways to incorporate transfer learning, i.e., through dimensionality expansion of 2D-CNN models<sup>24–26</sup>, such techniques remain approximate and possibly not fully optimized.

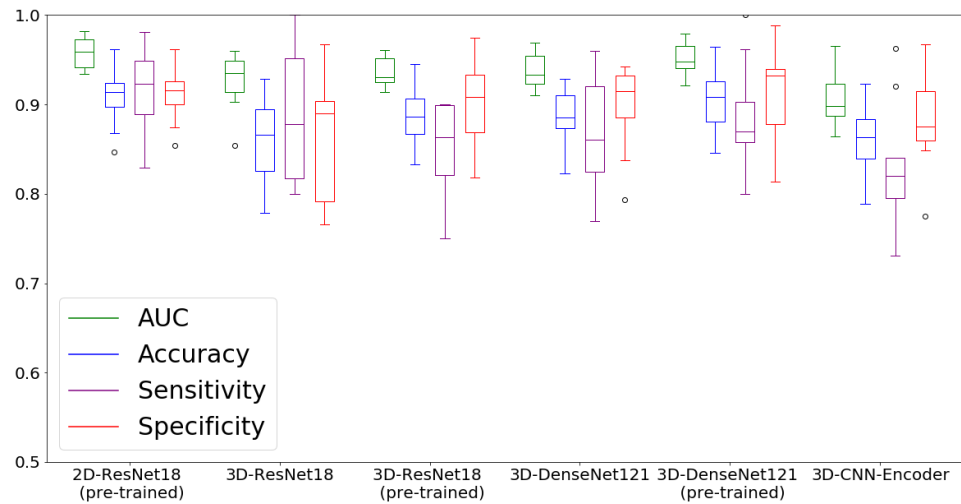
Therefore, our study investigates the effectiveness of select 2D and 3D-CNN algorithms for OCT-based glaucoma detection by leveraging DL models, including 2D and 3D versions of ResNet18<sup>27,28</sup> and DenseNet121<sup>29</sup>, and a 5-layer 3D-CNN-Encoder model<sup>22</sup>. In this study, we utilized dimensionality expansion to adapt 2D models for 3D applications, enabling weight transfer from an ImageNet1k-trained 2D model to improve a 3D model's performance. Through extensive evaluations using two publicly available volumetric OCT image datasets, the first available through the UKB<sup>30</sup> focusing on the macula and the other published by Maetschke et al.<sup>22</sup> focusing on ONH—this study underscores the robust performance of 2D-CNN algorithms in detecting glaucoma compared to their 3D counterparts and the weight transfer method tested in this study. These findings underscore the strengths of 2D-CNN in glaucoma detection via volumetric OCT images and point out the limitations of current weight transfer techniques for 3D-CNN models. They also highlight the potential of 3D algorithms and the necessity for future research to improve 3D-CNNs' performance in glaucoma detection. Therefore, the findings of this study have the potential to transform diagnostic approaches in the field of ophthalmology using OCT-based datasets.

## Results

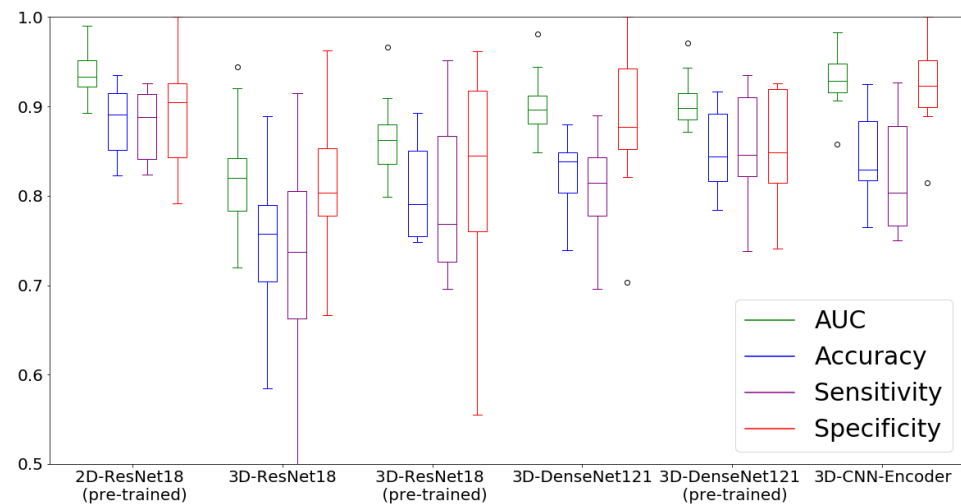
Table 1 presents the comparative performance of various models trained on the macular-OCT and ONH-OCT test datasets. The evaluation metrics include tenfold cross-validation (CV) average values for the AUC, accuracy, sensitivity, and specificity. Accuracy, sensitivity, and specificity scores, calculated based on the Youden Index, are specific to datasets and models, thus providing a qualitative assessment of the model's performance. Meanwhile, Fig. 1 displays the box plots of the evaluation metrics obtained from the macular-OCT and ONH-OCT test datasets during the tenfold CV study.

Models	Pre-trained	macular-OCT				ONH-OCT			
		AUC	ACC	Sen	Spe	AUC	ACC	Sen	Spe
2D-ResNet18	Yes	<b>0.960</b>	0.901	<b>0.891</b>	0.913	<b>0.943</b>	<b>0.890</b>	<b>0.917</b>	0.910
3D-ResNet18	No	0.928	0.863	0.885	0.861	0.823	0.750	0.730	0.820
3D-ResNet18	Yes	0.937	0.890	0.850	0.900	0.863	0.810	0.810	0.820
3D-DenseNet121	No	0.938	0.901	0.848	<b>0.923</b>	0.889	0.818	0.806	0.857
3D-DenseNet121	Yes	0.945	<b>0.905</b>	0.878	0.915	0.906	0.840	0.829	0.880
3D-CNN-Encoder	No	0.910	0.852	0.820	0.873	0.931	0.844	0.821	<b>0.921</b>

**Table 1.** Comparative performance of models on macular-OCT and ONH-OCT test datasets: average AUC, accuracy, sensitivity, and specificity values from tenfold CV. CV cross-validation, CNN convolutional neural network, OCT optical coherence tomography, ONH-OCT optic nerve head OCT, Acc accuracy, Sen sensitivity, Spe specificity.



(a) Box plot of the evaluation metrics obtained during 10-fold CV for macular-OCT dataset.



(b) Box plot of the evaluation metrics obtained during 10-fold CV for ONH-OCT dataset.

**Figure 1.** Box plots of evaluation metrics AUC, accuracy, sensitivity, and specificity obtained through tenfold cross-validation on the test data. This figure presents the box plot of the evaluation metrics AUC, accuracy, sensitivity, and specificity obtained through tenfold cross-validation on the test data. Subfigure (a) illustrates the results for the Macular-OCT model, while subfigure (b) displays the results for the ONH-OCT model. The AUC values box plot reveals that, for both macular-OCT and ONH-OCT datasets, the 2D CNN model delivers superior overall results. Additionally, the 2D model yields robust results for other essential metrics such as accuracy and sensitivity. Among the 3D CNN models, the pre-trained 3D-ResNet18 outperformed the 3D-ResNet18 model trained from scratch. Notably, the 3D-CNN-Encoder exhibits superior performance compared to the 3D-ResNet18 models when tested on the ONH-OCT dataset, while delivering subpar performance on the macular-OCT dataset. AUC area under the receiver operating characteristic curve, OCT optical coherence tomography, *macular-OCT* macular-OCT, *ONH-OCT* optic nerve head OCT.

For the macular-OCT dataset, the pre-trained 2D-ResNet18 model emerged as the top performer, achieving an AUC of 0.960, and showcasing an accuracy, sensitivity, and specificity of 0.910, 0.891, and 0.913, respectively. These results were closely followed by the pre-trained 3D-ResNet18 and 3D-DenseNet121, which posted an AUC of 0.937 and 0.945, respectively. Conversely, the 3D-ResNet18, 3D-DenseNet121, and 3D-CNN-Encoder models, trained from scratch on this dataset, recorded AUCs of 0.928, 0.938, and 0.910, respectively, highlighting the advantage of utilizing weight transfer technique to improve 3D models.

Regarding the ONH-OCT dataset, the 2D ResNet18 model once again surpassed its 3D counterparts tested in this study, achieving an AUC of 0.943, an accuracy of 0.890, a sensitivity of 0.917, and a specificity of 0.803. Among the 3D models utilized for this dataset, the 3D-CNN-Encoder model outperformed the other 3D models by achieving an AUC of 0.931. Meanwhile, the 3D-ResNet18 and its pre-trained version achieved AUCs of 0.823 and 0.863, respectively, while the 3D-DenseNet121 and its pre-trained counterpart recorded AUCs of 0.889 and 0.906, respectively. The diminished AUCs observed in 3D-ResNet18 and 3D-DenseNet121 models

on the ONH-OCT dataset might potentially be attributed to a predisposition for overfitting. We suspect that this overfitting is due to the inherent complexity of the 3D models employed, compounded by the smaller size of the dataset and lower resolution of the images utilized in this study.

Moreover, the box plot analysis (as depicted in Fig. 1) of the AUCs obtained from the tenfold CV emphasizes the steadfast performance of the 2D model. This illustration also demonstrates an overall strong performance achieved by the pre-trained 3D CNN models compared to their counterparts trained from scratch. Therefore, a significantly larger dataset could greatly enhance the 3D model's performance. Overall, as indicated by the results presented in Table 1 and Fig. 1 for the models tested in this study, the 2D-ResNet18 model consistently exhibited higher performance across both volumetric OCT datasets in the tenfold CV.

## Discussion

Glaucoma, a prevalent neurodegenerative disease affecting millions worldwide, necessitates early and precise diagnosis to mitigate irreversible vision loss. OCT has emerged as a powerful tool in ophthalmology, offering detailed insights into ocular structures. In this study, we investigated the effectiveness of select 2D and 3D CNN algorithms for glaucoma detection using both macular and ONH OCT images.

Our findings demonstrate that the pre-trained 2D-ResNet18 model provides robust results when applied to volumetric OCT datasets. The consistent performance of the 2D-ResNet18 model on both macular-OCT and ONH-OCT datasets indicates its potential to improve diagnostic practices in ophthalmology, especially considering the scarcity of large glaucoma datasets or the lack of 3D-CNN models trained on extensive datasets. Notably, compared to the 3D-CNN models tested in this study, the 2D-ResNet18 model achieved the highest AUC scores of 0.960 and 0.943 on macular-OCT and ONH-OCT, respectively.

Furthermore, when considering the comparison between 2D and 3D CNN algorithms, it is beneficial to highlight several advantages of 2D-CNN models. 2D-CNN models encompass a wider array of available pre-trained models, decreased computational complexity in model training, easier to interpret and visualize intermediate model layers due to lower dimensionality, simplified data augmentation due to existing libraries and lower dimensionality, and enhanced scalability. Moreover, in light of data scarcity and appropriate pre-trained 3D models, the application of 2D-CNNs is not only limited to glaucoma detection, it can be extended to other medical fields with reasonable accuracy and performance<sup>31–33</sup>. Consequently, considering the results presented and the evident benefits of 2D algorithms, their significance cannot be underestimated in glaucoma detection.

Despite yielding robust results, this study is not devoid of limitations. For instance, our analysis indicated that the 3D-ResNet18 and 3D-DenseNet121 models exhibited severe tendencies towards overfitting when applied to the ONH-OCT dataset. These findings suggest that the inherent complexity of the model may not be ideally suited for this dataset, featuring a resolution of  $64 \times 64 \times 128$ . Therefore, the study could have benefited from a higher resolution and larger ONH-OCT dataset. These outcomes reiterate the importance of model selection in achieving optimal performance.

Another limitation is the scarcity of the large OCT volume dataset and pre-trained state-of-the-art (SOTA) 3D-CNN models. While it is plausible that a 3D model pre-trained on a comprehensive dataset could significantly enhance the model's accuracy, further research is required. Therefore, in future studies, we intend to explore additional models and datasets as they become available. Additionally, we also plan to pre-train a 3D-CNN model from scratch by curating a substantial corpus of medical and other volumetric image datasets to improve models' performance in smaller glaucoma-specific volumetric datasets. These will provide more comprehensive insights into the comparative performance between 2D and 3D-CNN counterparts. Nevertheless, based on the publicly available glaucoma datasets and the ML resources readily available at our disposal, it is evident that 2D-CNN algorithms consistently deliver better results in glaucoma detection compared to the 3D-CNN algorithms and the weight transfer method utilized in this study for these datasets.

In conclusion, our investigation provides a thorough understanding of the strengths of 2D and 3D-CNN architectures in OCT-based glaucoma detection. Despite the loss of volumetric information in each 2D B-scan derived from the OCT volume; the 2D-CNN model still provides robust results by aggregating the prediction of all B-scans. The higher accuracy of 2D-CNN models underscores their potential to drive advancements in glaucoma diagnosis and management, especially considering the scarcity of pre-trained 3D CNN models trained on adequately large datasets and the constraints posed by the smaller sizes of glaucoma-specific datasets. It is also essential to emphasize that while the 2D approach yields excellent results, the selections of 2D and 3D ML algorithms are inherently problem-specific, demanding continuous efforts for ongoing progress. Therefore, by shedding light on the comparative performance of select 2D and 3D-CNN models on these smaller glaucoma datasets, our research aids in refining diagnostic tools in ophthalmology by highlighting the robust performance of 2D-CNN models and providing probable future direction to improve the performance of 3D-CNN models, thus, making a significant stride towards improving glaucoma detection techniques.

## Materials and methods

### Macular OCT data

For the macular-OCT dataset, we utilized the UKB dataset<sup>30</sup>, an ongoing project comprising health records of over half a million individuals aged between 40 and 70 years. Within this extensive dataset, a subset of spectral domain macular-OCT images was released for approximately 87,000 participants, captured between 2009 to 2013 using the TOPCON 3D OCT 1000 Mk2 device. Each OCT volume consists of 128 B-scans, of which each represents a  $512 \times 650$  pixel grayscale image. As the B-scan index advances from 0 to 127, the OCT B-scans progressively shift from the superior region down to the inferior region. Our access and use of the UKB data were approved under application #23424 in accordance with their Access Procedures and Ethics regulations. We obtained fully de-identified data. Informed consent was obtained from the participants for their participation

in the study by the UKB Committee upon recruitment. The study protocol was approved by The North West Multi-centre Research Ethics Committee.

In our cohort selection, we extracted participants with OCT scans from the UKB dataset consisting of both glaucoma cases and healthy individuals. Glaucoma cases were identified by the ICD-10 code H40.1, indicative of primary open-angle glaucoma (POAG). The healthy participants in our study had not reported any glaucoma or other eye-related conditions. Building upon previous research, we further refined our selection of healthy participants by excluding those who reported secondary health issues, such as high blood pressure or obesity, and those whose visual acuity was reported worse than 20/30 on the logMAR chart. After these exclusions, our dataset comprised of 255 individuals with glaucoma (~448 eyes) and 2,812 healthy participants (~5,619 eyes). For our research, we randomly chose 765 healthy individuals (about three times the number of cases encompassing ~1,530 eyes) from the healthy group and incorporated all 255 participants diagnosed with POAG. Figure 2 displays the age and sex distribution of the study sample.

### Optic nerve head OCT data

For the ONH-OCT, we utilized the publicly available OCT image dataset provided by Maetschke et al.<sup>22</sup>. This ONH-OCT dataset contains a collection of 1110 OCT volumes derived from 624 patients, imaged using a Cirrus SD-OCT Scanner (Zeiss, Dublin, CA, USA), focusing on the ONH region. Each publicly available OCT volume corresponds to a down-sampled version of  $64 \times 64 \times 128$  voxels, originating from the  $200 \times 200 \times 1024$  voxel dimensions. The dataset includes 263 scans representing healthy individuals and 847 cases were attributed to primary open-angle glaucoma (POAG) based on the provided labels. Demographic information, such as gender and race distribution, as well as mean values and standard deviations for patient age, intraocular pressure, and visual field test results, have previously been reported<sup>22</sup>.

### Train, validation, and test splits

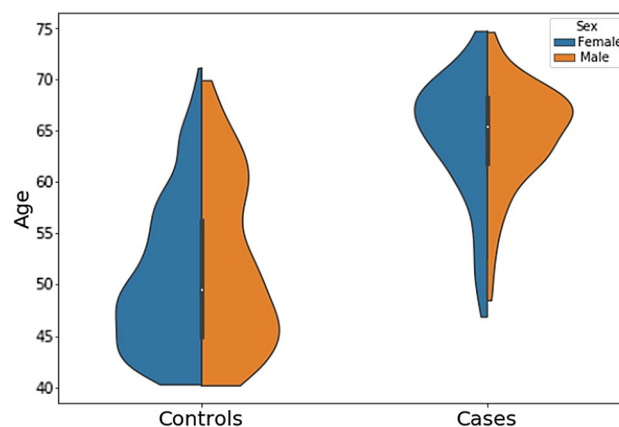
We employed a tenfold cross-validation (CV) strategy to train and assess the performance of 2D and 3D-CNN algorithms on the macular-OCT and ONH-OCT datasets. Within each fold, we used stratified splitting to ensure a balanced representation of healthy and glaucomatous individuals and partitioned the dataset into 80% for training, 10% for validation, and 10% for testing. Special precautions were taken to ensure that data from eye images of a single individual were allocated to the same data split.

### Data preprocessing

Before model training, each B-scan and OCT volume underwent a series of preprocessing steps. Initially, images were cropped and resized to  $224 \times 224$  pixels to align with the specifications of models pretrained on the ImageNet1k dataset. This was followed by standardizing the images to ensure their intensity ranges matched those of the images used for pretraining. To enhance the models' ability to generalize, data augmentation techniques were employed, such as random horizontal and vertical flips, rotations, and translations. While the data augmentation and preprocessing pipeline was similar for both 2D and 3D models, special attention was given to ensure that all slices within a given volume underwent identical augmentations during the training of 3D models.

### 2D Models

The primary DL architecture chosen for our 2D model is ResNet18, a convolutional neural network consisting of 18 deep layers. ResNet18 is part of the ResNet (Residual Network) family, known for its utilization of residual blocks<sup>27,28</sup> to address the vanishing gradient problem common in deep models, thereby facilitating the training of very deep networks. The selection of ResNet18 is driven by its relative simplicity and efficiency in model training



**Figure 2.** Age and sex distribution of the UKB study sample. The figure demonstrates the age and sex distribution present within the glaucoma (cases) and non-glaucoma (controls) individual in the UKB dataset.

and previously demonstrated robust performance in disease detection<sup>34,35</sup>. This choice enhances its resistance to overfitting when working with smaller datasets.

Our model begins with training on all B-scans (i.e., 128 for macular-OCT and 64 for ONH-OCT) collectively to generate per-B-scan glaucoma predictions. To achieve this, we consolidate each volume to match the corresponding number of B-scans and labeled each B-scan accordingly based on the labels of the OCT volume. As a result, after model training, the model can provide glaucoma predictions on a per-B-scan basis, rather than at the volume level. Therefore, these predictions are then aggregated using XGBoost<sup>36</sup> to obtain final glaucoma prediction. We implemented the DL model using the PyTorch framework in Python and utilized a pre-trained ResNet18 model trained on the ImageNet1K dataset as the foundation for our architecture. The DL framework used in this work is depicted in Fig. 3.

During the fine-tuning process, we adopted a batch size of 64 and employed the Adam optimizer with an initial learning rate of 0.001 and an epsilon value of 0.1. The model underwent 50 epochs of training, with the best-performing model being determined based on the lowest loss observed in the training phase. For loss calculation, we employed the binary cross-entropy (BCE) loss function.

### XGBoost

To aggregate the results from all available OCT B-scan predictions obtained from the 2D model, XGBoost was trained and validated with a split between training and validation data. The hyperparameters for the macular-OCT-based XGBoost model were set as the following learning rate 0.01, maximum depth 3, subsample 0.7, alpha 0.1, and lambda 0.5. Similarly, the hyperparameters for the ONH-OCT-based XGBoost model were learning rate 0.01, maximum depth 3, subsample 0.7, colsample\_bylevel 0.2, alpha 0.1, and lambda 0.5. Each XGBoost model was trained for 2000 epochs, and the best model was selected based on the lowest BCE loss.

### 3D-models

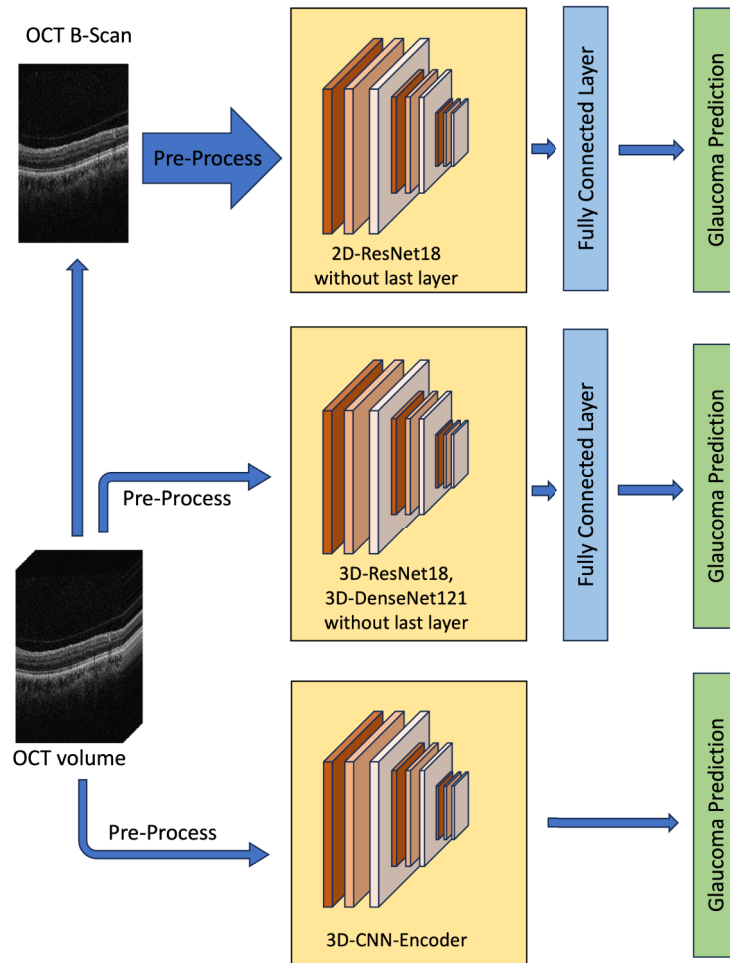
In this study, we explored two distinct 3D architectures: the 3D-ResNet18, 3D-DenseNet121<sup>29</sup>, and the 3D-CNN-Encoder. The 3D-ResNet18 and 3D-DenseNet121 are three-dimensional adaptations of their 2D counterparts, specifically designed to leverage the unique strengths of each architecture in volumetric contexts. In contrast to ResNet18, which utilizes residual connection, DenseNet121 employs a dense connectivity approach, directly linking each layer to every other, ensuring maximum information flow and excelling in detail-rich feature recognition due to its 121-layer depth<sup>29</sup>. To develop the 3D version, we replaced the 2D layers of the original model with 3D counterparts, maintaining the original model's essence but increasing the dimensionality. Transfer learning was also utilized; by employing dimensionality expansion, we adapted pretrained weights from the 2D CNN model to fit the 3D counterparts<sup>24–26</sup>. The weight transfer technique employed in this study mirrors a strategy elucidated by Ebrahimi et al.<sup>24</sup>, wherein they leveraged dimensionality expansion to iteratively replicate 2D CNN weights into their 3D equivalents. Xue et al.<sup>25</sup> also utilized a similar weight transfer technique to enhance the performance of their 3D model. In both works, comprehensive training of all layers was undertaken. Given the inherent approximations associated with the adaptation of 2D weights to a 3D model, it is recommended to train all layers for optimal results. Moreover, Ebrahimi et al.<sup>24</sup> and Xue et al.<sup>25</sup> trained their models on 264 and 263 volumetric images, respectively, achieving robust results. This illustrates the weight transfer technique's success in training full models effectively on smaller datasets.

The 3D-CNN-Encoder architecture, depicted in Fig. 3b, is based on the foundational framework described in Maetschke et al.'s study<sup>22</sup>. This model consists of five convolutional layers, each incorporating a batch normalization layer, with the first layer including both a max pool and a batch normalization layer. Kernel sizes for these convolutional layers are set at 7 for the first, 5 for the second, and 3 for the others. The max pooling layer follows with a kernel size of 1 and a stride of 2. Each batch normalization is succeeded by a ReLU activation layer. The architecture concludes with a global average pooling layer feeding 32 features into the final fully connected layer, which is then used for predicting glaucoma.

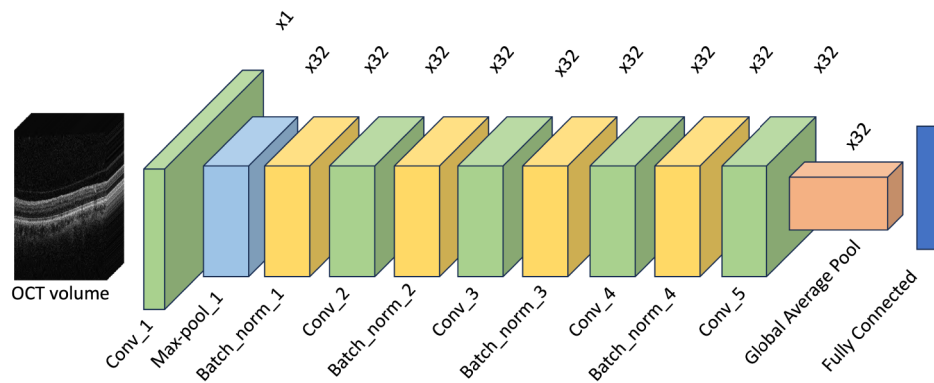
During the fine-tuning process, we noticed that the hyperparameters for model training on both datasets converged to similar values. Consequently, we employed a batch size of 16 for both the 3D-ResNet18 and 3D-DenseNet121 models and a batch size of 64 for the 3D-CNN-Encoder model, all utilizing the Adam optimizer. The learning rates were set at 1e-5 for both the 3D-ResNet18 and 3D-DenseNet121 models and at 0.001 for the 3D-CNN-Encoder model, respectively. Each of these 3D models underwent 50 epochs of training, and we selected the best model based on achieving the lowest BCE loss.

### Performance metrics

The performance of the model was evaluated using a comprehensive set of metrics, including accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC-ROC). To generate ROC curves, we used the 'roc\_curve' function from scikit-learn to plot the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings. To determine the accuracy, sensitivity, and specificity of the model, we utilized the Youden Index<sup>37</sup>, which identifies the best balance between sensitivity and specificity for each model and each particular dataset. However, the Youden Index does not generalize effectively across different datasets and models. For this reason, it was utilized in this study primarily to support the results qualitatively, rather than serving as a definitive tool for validating the model's performance.



(a) Deep learning framework for 2D and 3D CNN models.



(b) Architecture of the 3D-CNN-Encoder model used in this study.

**Figure 3.** Deep learning framework utilized with the 2D and 3D CNN models. This figure illustrates the deep learning (DL) framework utilized with the 2D and 3D CNN models. Subfigure (a) displays the DL framework employed in this study, encompassing a data preprocessing pipeline, responsible for both data preprocessing and augmentation before the images are input into the model. Following this, a DL backbone (specifically, ResNet18 sourced from PyTorch’s torchvision library) is used for feature extraction, culminating in a final fully connected layer dedicated to glaucoma prediction. In contrast, subfigure (b) showcases the 3D-CNN-Encoder model, which is built from scratch and structured around five convolutional layers and a fully connected layer for the final prediction. *DL* deep learning, *CNN* convolutional neural network, *OCT* optical coherence tomography.

## Data availability

The data used in this paper is publicly available except the UKB data which was obtained via contract using application ID #23424. Applications to access the data can be completed at: <https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access>. Informed consent was obtained from the participants for their participation in the study by the UKB Committee upon recruitment. The study protocol was approved by The North West Multi-centre Research Ethics Committee.

Received: 9 October 2023; Accepted: 16 May 2024

Published online: 23 May 2024

## References

1. Tribble, J. R. *et al.* Neuroprotection in glaucoma: Mechanisms beyond intraocular pressure lowering. *Mol. Asp. Med.* **92**, 101193. <https://doi.org/10.1016/j.mam.2023.101193> (2023).
2. Munemasa, Y. & Kitaoka, Y. Molecular mechanisms of retinal ganglion cell degeneration in glaucoma and future prospects for cell body and axonal protection. *Front. Cell. Neurosci.* **6**, 60. <https://doi.org/10.3389/fncel.2012.00060> (2012).
3. Dervisevic, E., Pavljasevic, S., Dervisevic, A. & Kasumovic, S. S. Challenges In early glaucoma detection. *Med. Arch.* **70**, 203–207. <https://doi.org/10.5455/medarh.2016.70.203-207> (2016).
4. Agarwal, R., Gupta, S. K., Agarwal, P., Saxena, R. & Agrawal, S. S. Current concepts in the pathophysiology of glaucoma. *Indian J. Ophthalmol.* **57**, 257–266. <https://doi.org/10.4103/0301-4738.53049> (2009).
5. Beykin, G., Norcia, A. M., Srinivasan, V. J., Dubra, A. & Goldberg, J. L. Discovery and clinical translation of novel glaucoma biomarkers. *Prog. Retin. Eye Res.* **80**, 100875. <https://doi.org/10.1016/j.preteyeres.2020.100875> (2021).
6. Chang, R. T. & Singh, K. Myopia and glaucoma: Diagnostic and therapeutic challenges. *Curr. Opin. Ophthalmol.* **24**, 96–101. <https://doi.org/10.1097/ICU.0b013e32835cef31> (2013).
7. Butt, N. H., Ayub, M. H. & Ali, M. H. Challenges in the management of glaucoma in developing countries. *Taiwan J. Ophthalmol.* **6**, 119–122. <https://doi.org/10.1016/j.tjo.2016.01.004> (2016).
8. Tham, Y. C. *et al.* Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis. *Ophthalmology* **121**, 2081–2090. <https://doi.org/10.1016/j.ophtha.2014.05.013> (2014).
9. Heijl, A., Bengtsson, B. & Oskarsdottir, S. E. Prevalence and severity of undetected manifest glaucoma: Results from the early manifest glaucoma trial screening. *Ophthalmology* **120**, 1541–1545. <https://doi.org/10.1016/j.ophtha.2013.01.043> (2013).
10. Soh, Z. *et al.* The global extent of undetected glaucoma in adults: A systematic review and meta-analysis. *Ophthalmology* **128**, 1393–1404. <https://doi.org/10.1016/j.ophtha.2021.04.009> (2021).
11. Olawoye, O., Fawole, O. I., Teng, C. C. & Ritch, R. Evaluation of community eye outreach programs for early glaucoma detection in Nigeria. *Clin. Ophthalmol.* **7**, 1753–1759. <https://doi.org/10.2147/OPTH.S46823> (2013).
12. Dong, Z. M., Wollstein, G. & Schuman, J. S. Clinical utility of optical coherence tomography in glaucoma. *Invest. Ophthalmol. Vis. Sci.* **57**, 556–567. <https://doi.org/10.1167/iovs.16-19933> (2016).
13. Savini, G. *et al.* Correlation between retinal nerve fibre layer thickness and optic nerve head size: An optical coherence tomography study. *Br. J. Ophthalmol.* **89**, 489–492. <https://doi.org/10.1136/bjo.2004.052498> (2005).
14. Kim, M., Choung, H. K., Lee, K. M., Oh, S. & Kim, S. H. Longitudinal changes of optic nerve head and peripapillary structure during childhood myopia progression on OCT: Boramae myopia cohort study report 1. *Ophthalmology* **125**, 1215–1223. <https://doi.org/10.1016/j.ophtha.2018.01.026> (2018).
15. Mehta, P. *et al.* Automated detection of glaucoma with interpretable machine learning using clinical data and multimodal retinal images. *Am. J. Ophthalmol.* **231**, 154–169. <https://doi.org/10.1016/j.ajo.2021.04.021> (2021).
16. Barros, D. M. S. *et al.* Machine learning applied to retinal image processing for glaucoma detection: Review and perspective. *Biomed. Eng. Online* **19**, 20. <https://doi.org/10.1186/s12938-020-00767-2> (2020).
17. Kim, S. J., Cho, K. J. & Oh, S. Development of machine learning models for diagnosis of glaucoma. *PLoS ONE* **12**, e0177726. <https://doi.org/10.1371/journal.pone.0177726> (2017).
18. Hood, D. C. *et al.* Detecting glaucoma with only OCT: Implications for the clinic, research, screening, and AI development. *Prog. Retin. Eye Res.* **90**, 101052. <https://doi.org/10.1016/j.preteyeres.2022.101052> (2022).
19. Wu, F., Chiariglione, M. & Gao, X. R. In 2022 *International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* 567–570 (IEEE, 2022).
20. Christopher, M. *et al.* Retinal nerve fiber layer features identified by unsupervised machine learning on optical coherence tomography scans predict glaucoma progression. *Invest. Ophthalmol. Vis. Sci.* **59**, 2748–2756. <https://doi.org/10.1167/iovs.17-23387> (2018).
21. Garcia, G. *et al.* Circumpapillary OCT-focused hybrid learning for glaucoma grading using tailored prototypical neural networks. *Artif. Intell. Med.* **118**, 102132. <https://doi.org/10.1016/j.artmed.2021.102132> (2021).
22. Maetschke, S. *et al.* A feature agnostic approach for glaucoma detection in OCT volumes. *PLoS One* **14**, e0219126. <https://doi.org/10.1371/journal.pone.0219126> (2019).
23. George, Y. *et al.* Attention-guided 3D-CNN framework for glaucoma detection and structural-functional association using volumetric images. *IEEE J. Biomed. Health Inform.* **24**, 3421–3430. <https://doi.org/10.1109/JBHI.2020.3001019> (2020).
24. Ebrahimi, A., Luo, S. H., Chiong, R. & Initia, A. D. N. Introducing transfer learning to 3D ResNet-18 for Alzheimer's Disease detection on MRI images. 2020 *35th International Conference on Image and Vision Computing New Zealand (Ivcnz)*. (2020).
25. Xue, S. H. & Abhayaratne, C. Covid-19 diagnostic using 3d deep transfer learning for classification of volumetric computerised tomography chest scans. 2021 *Ieee International Conference on Acoustics, Speech and Signal Processing (Icassp 2021)*. 8573–8577 (2021).
26. Solovyev, R., Kalinin, A. A. & Gabruseva, T. 3D convolutional neural networks for stalled brain capillary detection. *Comput. Biol. Med.* **141**, 105089. <https://doi.org/10.1016/j.compbiomed.2021.105089> (2022).
27. He, K. M., Zhang, X. Y., Ren, S. Q. & Sun, J. Identity mappings in deep residual networks. *Lect. Notes Comput. Sci.* **9908**, 630–645. [https://doi.org/10.1007/978-3-319-46493-0\\_38](https://doi.org/10.1007/978-3-319-46493-0_38) (2016).
28. He, K. M., Zhang, X. Y., Ren, S. Q. & Sun, J. Deep residual learning for image recognition. 2016 *Ieee Conference on Computer Vision and Pattern Recognition (Cvpr)*. 770–778 (2016).
29. Iandola, F. *et al.* DenseNet: Implementing Efficient ConvNet Descriptor Pyramids. *arXiv* (2014).
30. Sudlow, C. *et al.* UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779. <https://doi.org/10.1371/journal.pmed.1001779> (2015).
31. Saeedi, S., Rezayi, S., Keshavarz, H. R. & Niakan Kalhori, S. MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. *BMC Med. Inform. Decis. Mak.* **23**, 16. <https://doi.org/10.1186/s12911-023-02114-6> (2023).
32. Sahoo, S. K., Das, S. & Panda, S. K. *International Conference on Communications and Cyber Physical Engineering 2018*. (Springer Nature Singapore, 2024)



33. Shah, A. A., Malik, H. A. M., Muhammad, A., Alourani, A. & Butt, Z. A. Deep learning ensemble 2D CNN approach towards the detection of lung cancer. *Sci. Rep.* <https://doi.org/10.1038/s41598-023-29656-z> (2023).
34. Akter, N. *et al.* Glaucoma diagnosis using multi-feature analysis and a deep learning technique. *Sci. Rep.* **12**, 8064. <https://doi.org/10.1038/s41598-022-12147-y> (2022).
35. Odusami, M., Maskeliunas, R., Damasevicius, R. & Krilavicius, T. Analysis of features of alzheimer's disease: Detection of early stage from functional brain changes in magnetic resonance images using a finetuned ResNet18 network. *Diagnostics* <https://doi.org/10.3390/diagnostics11061071> (2021).
36. Chen, T. Q. & Guestrin, C. XGBoost: A scalable tree boosting system. *Kdd'16: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*. 785–794 (2016).
37. Ruopp, M. D., Perkins, N. J., Whitcomb, B. W. & Schisterman, E. F. Youden index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biom. J.* **50**, 419–430. <https://doi.org/10.1002/bimj.200710415> (2008).

## Acknowledgements

This study was supported in part by National Institutes of Health (NIH; Bethesda, MD, USA) grant P30EY032857 and Research to Prevent Blindness New Chair Challenge Grant. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We would like to thank the study participants from the UK Biobank as well as the staff who aided in data collection and processing.

## Author contributions

R.K.R. prepared the dataset, trained the deep learning models, and analyzed the data presented in the manuscript. X.R.G. conceived the study, organized the dataset, and oversaw its execution. F.W. processed the dataset used in this study. Both R.K.R. and X.R.G. wrote the manuscript and edited it to its final version. F.W. and M.C. contributed to the overall draft and organization of the manuscript. S.S.C. and N.D. reviewed and edited the final version.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-62411-6>.

**Correspondence** and requests for materials should be addressed to X.R.G.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024