A Deep Learning Model to Identify Homonymous Defects on Automated Perimetry

Aaron Hao Tan[1] MASc, Laura Donaldson[2] MD PhD, Luqmaan Moolla[3] MEng, Austin Pereira[2] MD MEng, Edward Margolin[2,4] MD

[1]Faculty of Applied Science and Engineering University of Toronto, Toronto, Canada
[2]Department of Ophthalmology and Vision Sciences, University of Toronto
[3]University of Toronto Temerty School of Medicine
[4]Department of Medicine, Division of Neurology, University of Toronto

The authors have no conflict of interest to disclose

There are no funding sources to disclose

Main message

- Homonymous visual field (VF) defects are usually an indicator of serious intracranial pathology but can often be subtle and difficult to detect

- AI is a valuable tool for the interpretation of complex datasets and we provide proof of principle for the use of deep learning in detecting homonymous VF defects on automated perimetry.

- The developed deep learning model achieved an overall average accuracy of 87%, making it highly effective at identifying homonymous VF defects on automated perimetry.

Introduction

Homonymous VF defects are usually an indicator of serious intracranial pathology and their presence indicates the need for urgent neuro-imaging. While the most common cause of homonymous VF defect is stroke, other important causes including brain tumours, metastatic cancers, demyelinating disease, and traumatic brain injury can all be culprits.[1,2]

The field of artificial intelligence (AI) technology in medicine is an important avenue of innovation and is becoming increasingly prevalent in the field of ophthalmology for a wide range of clinical applications.[3] Previous studies have explored its use in the analysis of fundus photos for detection of diabetic retinopathy[4] and predicting progression of VF defects in glaucomatous optic neuropathy[5]. The use of AI in VF interpretation to date has been almost exclusively within the field of glaucoma. We identified the identification of homonymous defects as an important goal for an AI-based tool due to 1) the critical importance of detecting these VF defects, and 2) our observations that these defects can be subtle and thus overlooked even by skilled ophthalmologists.

1. Zhang X, Kedar S, Lynn MJ, Newman NJ, Biousse V. Homonymous hemianopia in stroke. *Journal of Neuro-Ophthalmology*. 2006;26(3):180-183.
2. Goodwin D. Homonymous hemianopia: challenges and solutions. *Clinical Ophthalmology (Auckland, NZ)*. 2014;8:1919.
3. Grewal PS, Oloumi F, Rubin U, Tennant MTS. Deep learning in ophthalmology: a review. *Canadian Journal of Ophthalmology*. 2018;53(4):309-313.
4. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *npj Digital Medicine 2018 1:1*. 2018;1(1):1-8.
5. Berchuck SI, Mukherjee S, Medeiros FA. Estimating Rates of Progression and Predicting Future Visual Fields in Glaucoma Using a Deep Variational Autoencoder. *Scientific Reports 2019 9:1*. 2019;9(1):1-12.

Objectives

This study aimed to develop an AI-based tool using a deep learning approach[6] to make accurate classifications of homonymous VF defects.

The proposed model in this paper, which we named the Deep Homonymous Classifier (DHC), utilizes convolutional layers to extract spatial features from 2D Humphrey Visual Field (HVF) images to perform binary classification of either homonymous defect or no homonymous defect.

6. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics*. 2018;19(6):1236-1246.

Methods

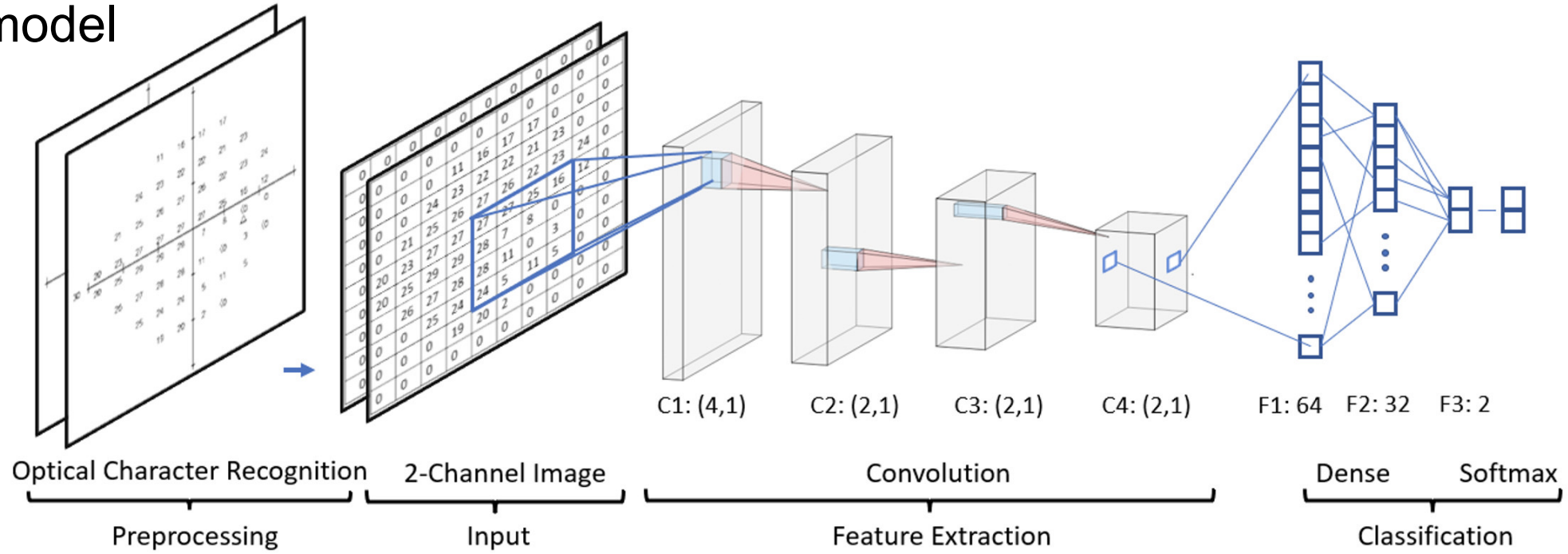
- Retrospective, proof-of-concept study using 24-2 Humphrey VFs from controls and patients with homonymous defects. HVF tests collected were reviewed by two independent reviewers to ensure the identification of both control and homonymous samples was correct

- To extract the visual threshold values from the collected patient PDFs, a custom optical character recognition (OCR) program was developed using an open-source library, pytesseract[7]. A random sample of 50 extracted fields was manually inspected and the results showed a 97% recognition accuracy.

- The total dataset utilized in this study included 1236 VFs, of which 820 were controls and 416 were represented homonymous defects. To attempt to train the proposed deep learning model with a balanced dataset, augmentation techniques were used to increase the number of homonymous defects examples. This augmentation process included random flipping of available homonymous VFs along the vertical or horizontal axis in both eyes to create new training examples for the proposed model. In addition, VFs of patients with bitemporal visual defects were flipped in one eye along the vertical axis to create a homonymous defect pattern. None of the augmented VFs were included in the validation and testing sets of the model and were used strictly for the training of the model.

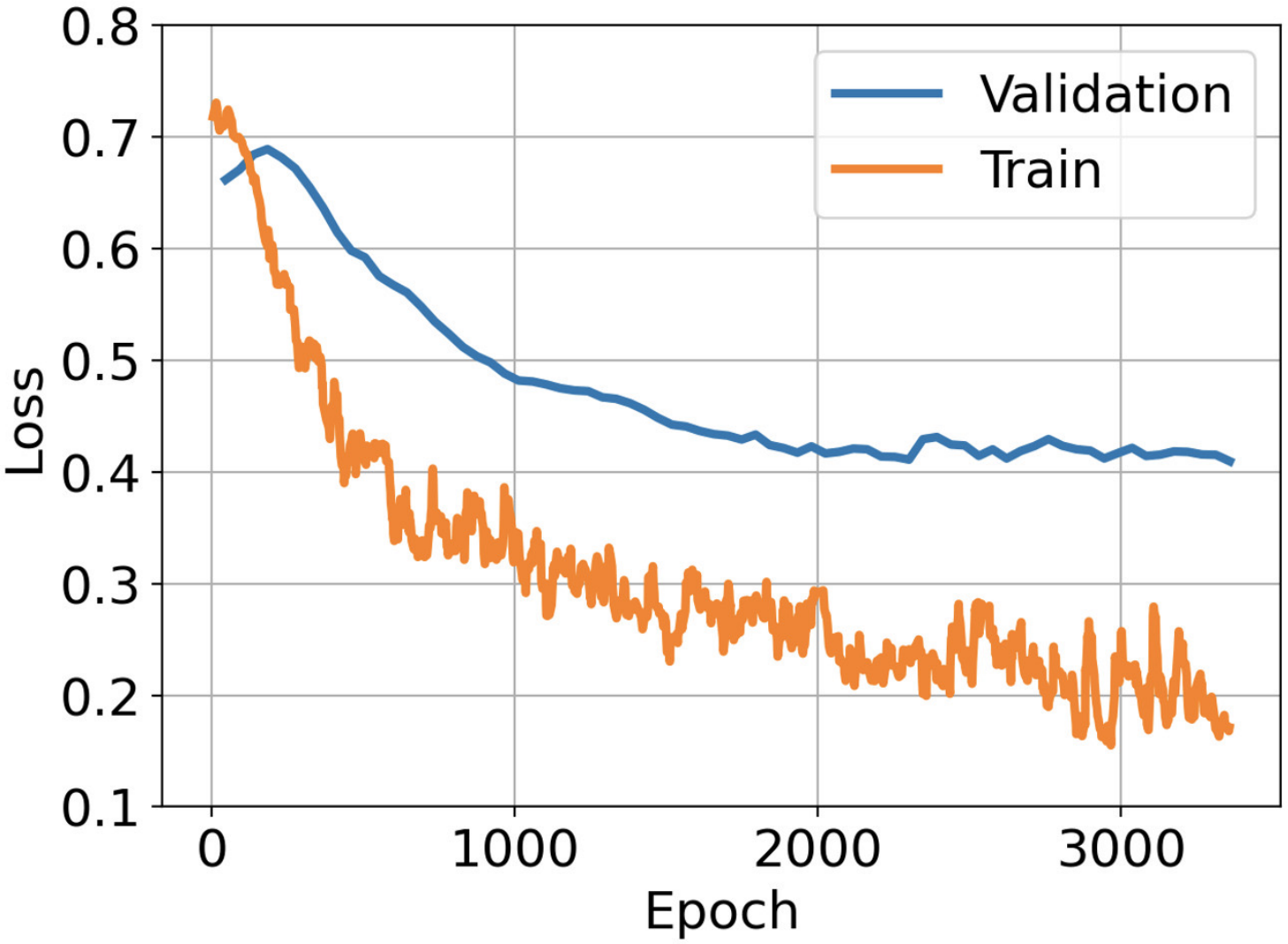
7. pytesseract · PyPI. Accessed October 8, 2021. https://pypi.org/project/pytesseract/

DHC model



The DHC model was designed using convolutional neural networks (CNN) to perform binary classification for patients with homonymous defects on visual field examination. The proposed model was developed using the PyTorch framework.[8] More specifically, each training sample consisted of a 10 x 12 pixel, two-channel image, that contained the left and right eye. Extracted visual threshold values from the patient PDFs are padded with zeros to create input images for the DHC model. In terms of the network architecture, DHC utilized four convolutional layers to extract location invariant features from the input images followed by three dense layers to perform classification. Rectified linear unit functions were adopted to incorporate non-linearity within the model. A softmax function was applied to transform the output vector of the last layer into a vector of probabilities for binary classification. To stabilize training, batch normalization was implemented to reduce the internal covariate shift of the network by rescaling and recentering the layer inputs.[9]

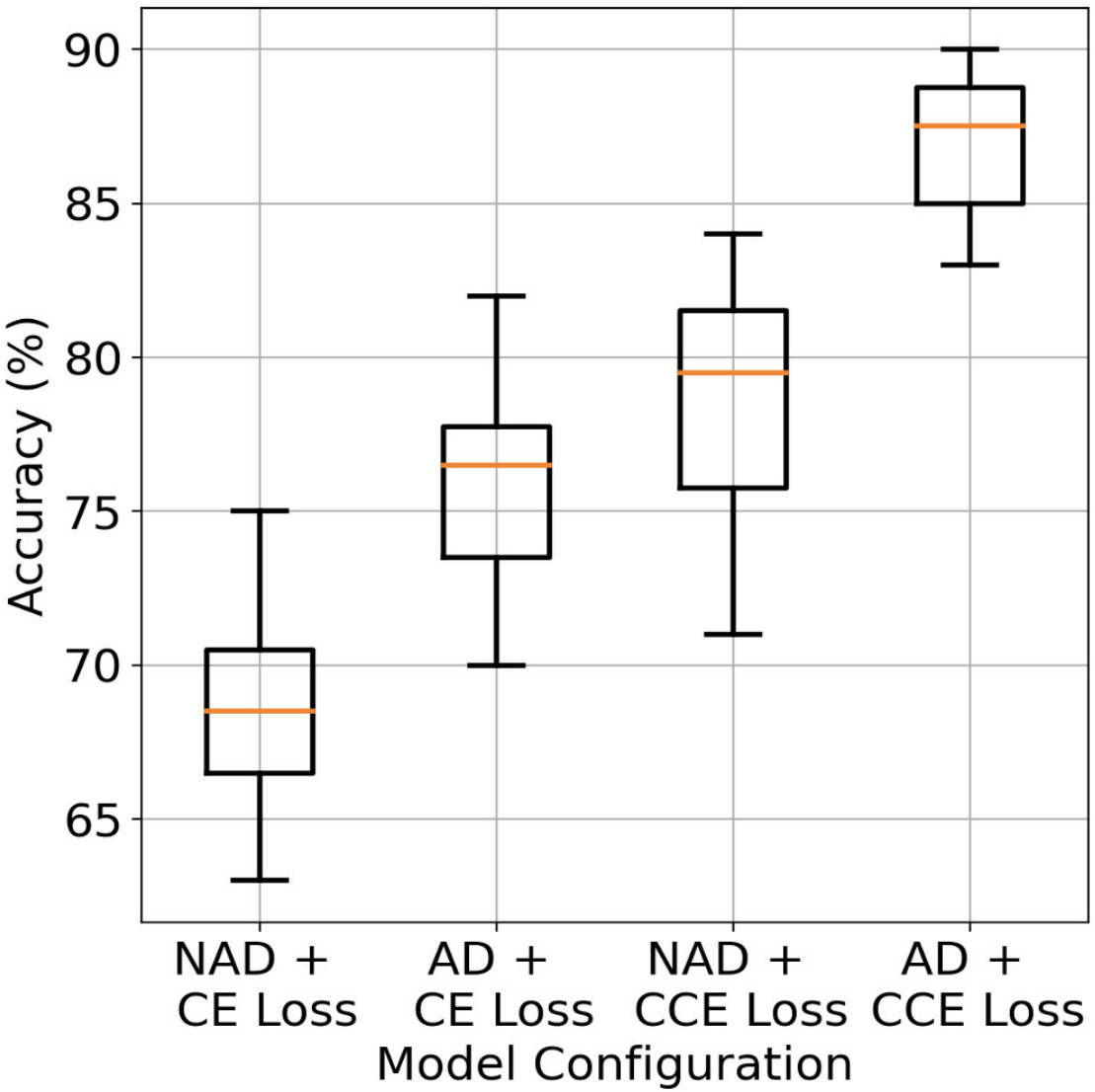
8. PyTorch. Accessed October 8, 2021. https://pytorch.org/
9. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Published online June 1, 2015:448-456. Accessed October 8, 2021. https://proceedings.mlr.press/v37/ioffe15.html



During training, an adaptive optimizer, ADAM, was utilized to update the proposed model via backpropagation.[10] Other hyperparameters included a batch size of 32 and a learning rate of 0.0001. To avoid overtraining, early stopping was implemented with a patience of 18. In terms of the loss function, the recently proposed complement cross-entropy loss (γ set to -1) was implemented for imbalance classification.[11] Successful convergence was achieved after 3000 epoch, where an epoch is defined as a complete cycle of the training dataset.

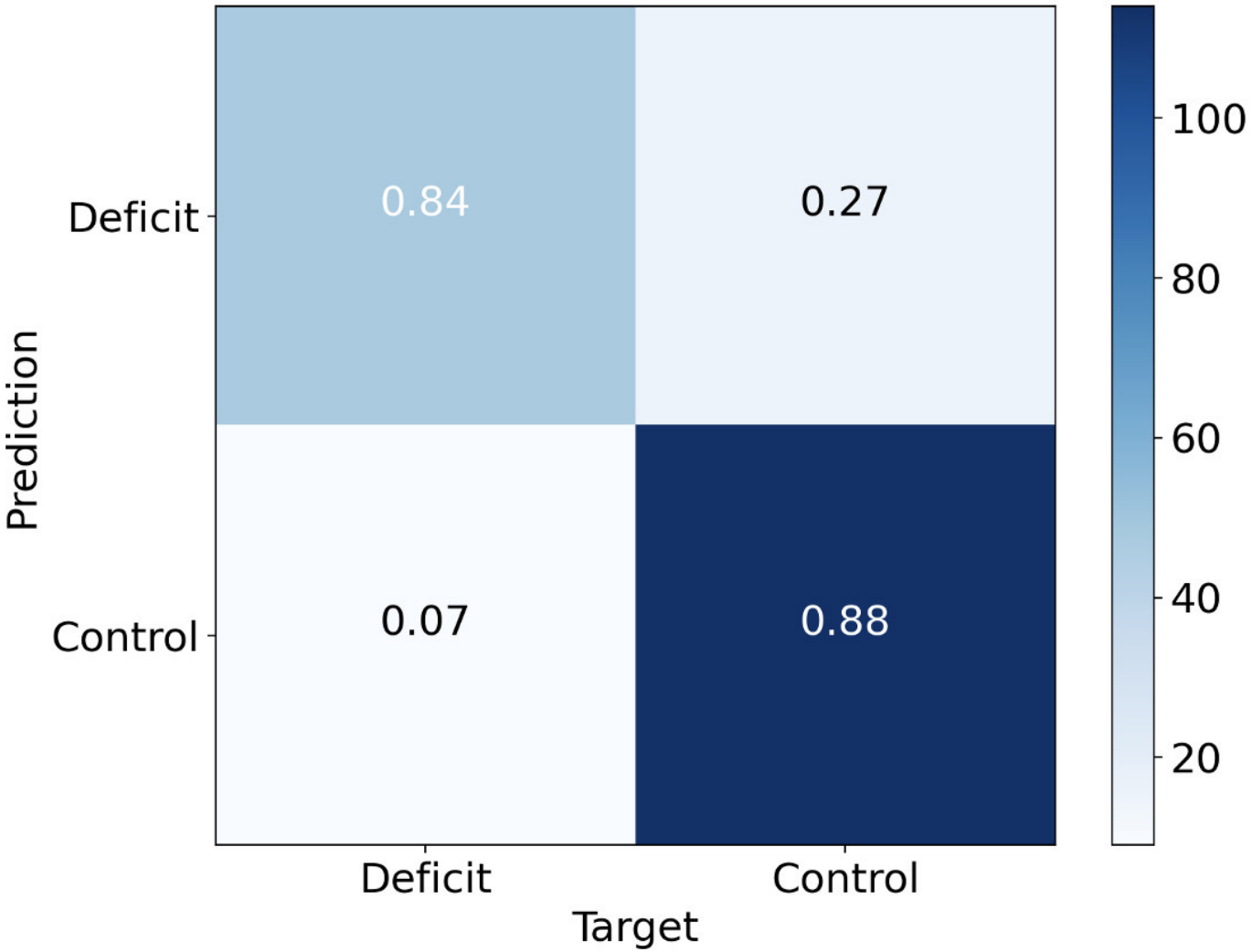
10. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. Published online December 22, 2014. Accessed October 8, 2021. https://arxiv.org/abs/1412.6980v9
11. Kim Y, Lee Y, Jeon M. Imbalanced Image Classification with Complement Cross Entropy. Pattern Recognition Letters. 2020;151:33-40. Accessed October 10, 2021. https://arxiv.org/abs/2009.02189v4

Results



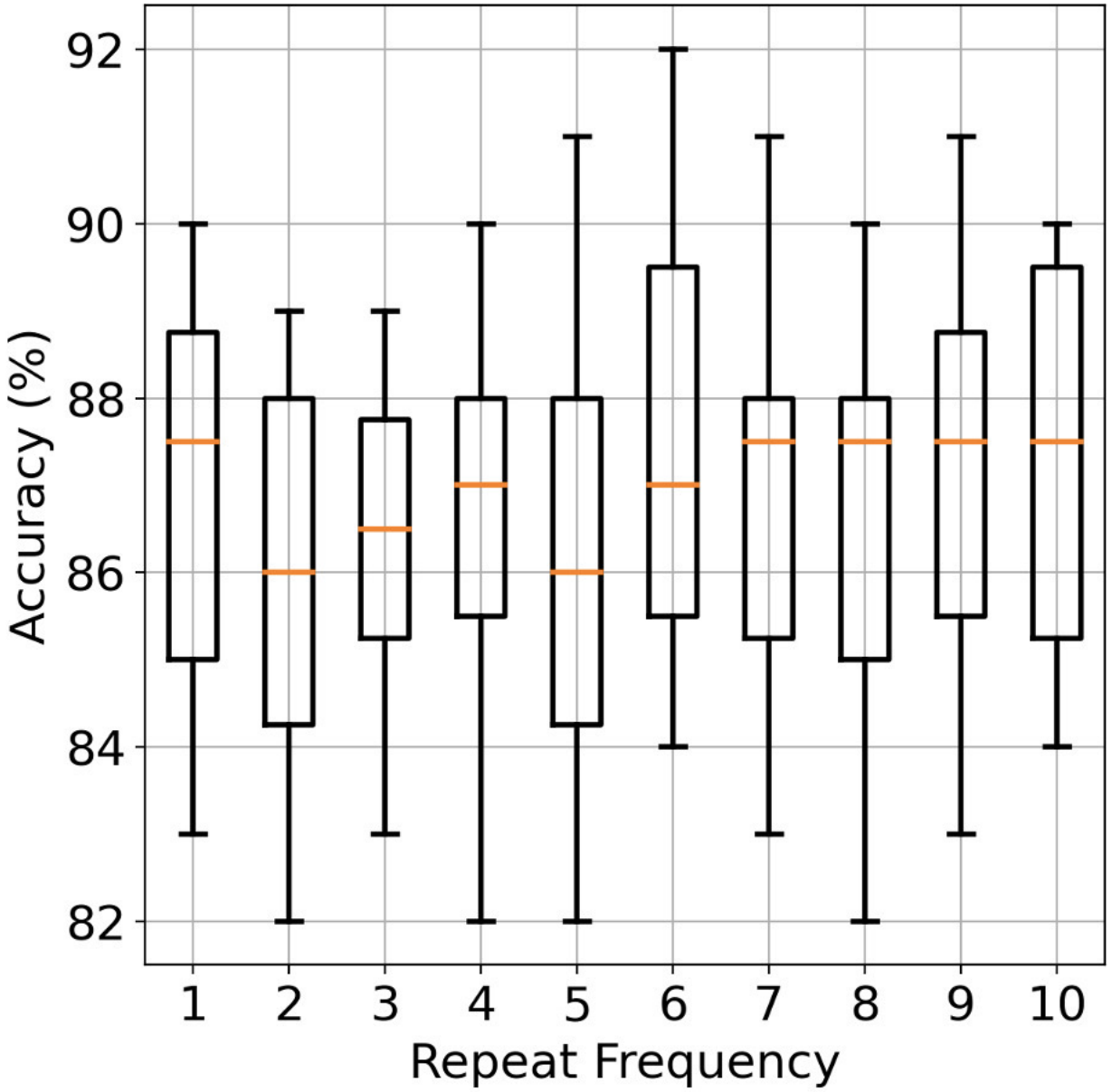
7-fold validation for four different models was conducted to analyze the effects of data augmentation and different loss functions on the model's performance. More specifically, we trained a model with 1) no augmented data (NAD) with cross entropy (CE) loss, 2) augmented data (AD) with cross entropy (CE) loss 3) NAD with complement cross entropy (CCE) loss and 4) AD with CCE loss. Note, for the models trained with augmented data, the augmentation was only applied to the training partitions. The accuracy of each model was averaged across all folds. At each fold, the test set included approximately 176 VFs that contained roughly 70% controls and 30% homonymous defects (the exact numbers of controls and homonymous defects varied as the dataset was randomly shuffled for each partition).

The best model was trained with AD and CCE loss, which achieved an average overall accuracy of 87%.



Test confusion matrix for the DHC model with AD and CCE Loss, where the accuracies are averaged across 7 folds.

Greater emphasis was placed on recall, since the consequence of missing a homonymous defect is more catastrophic than a false positive prediction. Recall, calculated by true positives/ (true positives + false negatives) was found to be an average of 84%. The calculated F2 score for this model was found to be 0.82, with a Cohens Kappa value of 0.74.



To evaluate DHC's training robustness, repeated k-fold cross-validation was conducted up to ten times.

Repeated k-fold cross validation and the average accuracies of the proposed model trained with AD and CCE Loss are shown. Average model accuracies stabilized to approximately 87%,

Discussion

- Automated perimetry is routinely used in ophthalmology clinics for diagnosis and monitoring of ocular disease, most commonly glaucomatous optic neuropathy. Correspondingly, existing methods for the interpretation of VFs were developed to detect the presence and progression of glaucomatous field loss; however, interpretation of VFs is of great importance in the identification of neurological disease.

- The widely used glaucoma hemifield test (GHT) compares symmetry of defined superior and inferior sectors looking for asymmetry along the horizontal axis.[12] As the segregation of nasal and temporal axons occurs at the optic chiasm, the important axis of symmetry for detecting homonymous defects is the vertical axis.A vertical analog to the GHT has been proposed, termed the neurological hemifield test.[13] This approach was effective at discriminating VFs from glaucoma patients versus those with intracranial pathology, though subtle defects were less readily picked up and binasal VF defects in glaucoma patients could thus be misclassified as neurologic VF defects.

- We opted to use a deep learning algorithm rather than a pre-defined rule above for detecting homonymous VF defect as deep learning excels in identifying intricate relationships and patterns within data sets.[14]

12. Åsman P, Heijl A. Glaucoma Hemifield Test: Automated Visual Field Evaluation. *Archives of Ophthalmology*. 1992;110(6):812-819.
13. Boland M v., McCoy AN, Quigley HA, et al. Evaluation of an Algorithm for Detecting Visual Field Defects Due to Chiasmal and Postchiasmal Lesions: The Neurological Hemifield Test. *Investigative Ophthalmology & Visual Science*. 2011;52(11):7959-7965.
14. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*. 2018;15(141).

Discussion

- To the best of our knowledge, this study is the first application of deep learning to the classification of homonymous defects on automated perimetry

- Imbalanced datasets are a major obstacle for many medical applications of AI technology as it can often be difficult to acquire ideal datasets due to various reasons, such as difficulty with data collection, poor uniformity of data, and when dealing with rare diseases. The use of data augmentation and complement cross-entropy loss addressed the issue of our imbalanced dataset and resulted in a significant increase in the accuracy of the DHC model.
- Our model is currently limited by the size of the dataset. As with all deep learning models, the proposed model will improve in generalizability with a larger training set and will be able to further validate the presented recall and accuracy with a larger testing set. Other limitations of the DHC model are related to deep learning and include its low explainability or “black box” nature and incorrect feature attribution in smaller training data sets.

- The model can be further improved by adding recurrent layers to incorporate temporal features during classification. This would enable it to analyze the progression of vision loss to enhance its ability to detect early homonymous defects.

Conclusions

- The overall average accuracy achieved by the DHC was 87% with an average recall of 84% when evaluated with 7-fold cross-validation. The robustness of training was evaluated with repeated k-fold cross-validation, where the average accuracy also converged to approximately 87%. This is an excellent result as homonymous defects are often subtle and can be difficult to identify even by a skilled ophthalmologist.

- As a screening tool, this has the potential for meaningful impact on clinical practice as the consequences of missing these defects can include significant patient morbidity and mortality.

- Patients flagged as having a possible homonymous defect may questioned regarding neurologic symptoms and risk factors for stroke, and undergo repeat perimetry tests with neuroimaging if homonymous defect is confirmed