Deep Learning Using Images of the Retina for Assessment of Severity of Neurological Dysfunction in Parkinson Disease

Janan Arslan, Daniel Racoceanu, Kurt K Benke

Deep Learning Using Images of the Retina for Assessment of Severity of Neurological Dysfunction in Parkinson Disease

Janan Arslan, Daniel Racoceanu, Kurt K Benke

Deep Learning Using Images of the Retina for Assessment of Severity of Neurological Dysfunction in Parkinson Disease

Janan Arslan, MSc, MBiostat, PhD; Daniel Racoceanu, PhD; Kurt K. Benke, PhD

The deep learning (DL) approach is a recent development in research on the automation of medical diagnostics using machine learning. This type of statistical learning is not based on biophysical (mechanistic) models but on training (for parameter estimation) using a sample of population data. There are issues about robustness in performance when testing on new data sets in uncontrolled conditions separate from validation experiments. The use of DL in diagnostic modeling has been demonstrated in recent years with very encouraging results in the assessment of diabetic retinopathy, geographic atrophy in age-related macular degeneration, and cardiovascular disease when using images of the retina.[1]

Related article

The study by Ahn et al[1] in this issue of *JAMA Ophthalmology* provides insights and useful information on the application of DL to explore possible associations between changes in the retina and brain function. The reported work is interesting as it explores the use of DL approaches for assessment of cognitive function scores for disease severity in the case of Parkinson disease (PD), which is projected to be an increasing economic burden.[2] In the study, fundus photographs were collected from 615 participants who were outpatients at the neurology department of Kangbuk Samsung Hospital in South Korea (August 2020 to April 2021). To compensate for the typically small sample sizes often encountered in ophthalmology research, the authors applied a range of image augmentation techniques, which included random rotation and random flip, to increase the sample size.

Five common convolutional neural network (CNN) models were investigated to predict the Hoehn and Yahr Scale (H-Y) and Unified Parkinson's Disease Rating Scale part III (UPDRS) scores for PD,[3,4] with the training based on fundus photographs and patient demographic characteristics. The study echoed other works that have investigated the idea that changes manifested in the retina may provide indicators of neurological impairment due to PD.[5] The prospective case-control hospital-based study revealed that, for the internal validation data, the DL system had high sensitivity that was above 80% for both the H-Y and UPDRS scores. The specificity ranged from 66% to 67% for H-Y and UPDRS scores. For the external validation data, the sensitivity and specificity were 70.73% and 66.66%. The area under the receiver operating characteristic curve was 0.67 and accuracy was 70.5%.

An advantage of this method for severity assessment is that it is noninvasive and based on readily available fundus retina images, rather than using images of the brain, such as radiological images like computed tomography and magnetic resonance imaging scans. Comparatively, these images are a little more invasive and time consuming in their acquisition. An additional salient feature of this study was that it suggests the possibility of augmenting or even eventually replacing the 2 standard tests for severity of PD dysfunction, which are rating scales with a subjective component, with a more objective approach that reduces risk and uncertainty in diagnosis and therefore improves reliability. An approach based on DL may be used as a fast screening method for cognitive impairment during routine eye examinations. A feature of the diagnostic method is that finding an association between the algorithm prediction and the results for the 2 standard tests provides a degree of "explainability." This is because the DL algorithm used to develop the predictive model produces results that are associated with a disease severity scale. Furthermore, the study revealed that the CNN uses the macular area for finding associations with PD cognitive impairment, as was demonstrated using the class activation map (CAM)—a heatmap-based feature visualization tool. CAM is useful for explainability, given it highlights the regions of interests and the degree of importance of these regions as estimated by the CNN.[6]

To create a CAM, a DL architecture must apply global average pooling (GAP) to the final convolutional feature maps. The GAP will be followed by a layer, fully connected, designed to produce the predictions. While this study has used CAM, a limitation of this approach is that CNNs that do not have the GAP layer cannot take advantage of CAM-based visualization. To solve this problem, Grad-CAM has been designed as a generalization of CAM. While the CAM weights features are based on the final fully connected layer, Grad-CAM is not an architecture-dependent method but creates an importance score based on the gradients of any trained CNN (which is a form of post hoc attention, meaning that it is a method that produces heatmaps from an already trained neural network). The idea is to exploit the spatial information that is preserved by the convolution layers and to understand which parts of an input image were important for a classification decision.

While a stepping stone above CAM, Grad-CAM does have its own disadvantages. For example, it cannot locate objects in instances where an image has multiple occurrences of the same class. Another consequence is that the localization often does not correspond to the whole object, but to parts of it. To overcome this problem, GradCAM++ and its variant XGrad-CAM were designed, with the former using second-order gradients to obtain activations that are independent of the object size, and the latter scaling the gradients by the normalized activation maps.[7] Thus, in this study, the results and application of DL, as well as explainable steps, are encouraging. However, there needs to be further consolidation in future experiments to establish the potential of this fast, noninvasive tool for routine applications in a clinical context as well as in-depth analysis to verify the explainable features as determined by the DL model using techniques such as GradCAM++ or XGradCAM.

This study, along with its predecessors, has demonstrated the importance of DL in health care solutions. There are several benefits attached to medical DL, including automating previously human-based and subjective processes, time and cost efficiency, and expedited steps toward development of new and novel therapies, all of which will lead to increasing patient life expectancy and improved quality of life. However, a limitation of DL uptake in clinical care has been the lack of transparency or explainability, given that DL methods are typically black box and opaque. Through the use of various DL algorithms for medical conditions, we can clearly see good quantitative results in terms of improving diagnostic time and accuracy. However, the associations between DL features and clinical outputs or how DL methods reach conclusions is not immediately apparent, and this connection is currently missing in the breadth of ophthalmic DL literature. Through the integration of explainable tools into the DL pipeline, such as the CAM, GradCAM, GradCAM++, and XGradCAM, we can deconstruct the solution given by a DL-trained model and pinpoint which regions, locations, or spatial information within medical images correlate with DL-predicted outputs. Not only will explainability assist in elucidating imaging features associated with outcomes but extracted features could further our current clinical understanding of many diseases and their progression. Furthermore, by incorporating an explainable component within a DL framework, we can expedite the uptake and adoption of DL in health care.

ARTICLE INFORMATION

Author Affiliations: Sorbonne Université, Institut du Cerveau–Paris Brain Institute–ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France (Arslan, Racoceanu); School of Engineering, University of Melbourne, Parkville, Victoria, Australia (Benke).

Corresponding Author: Janan Arslan, MSc, MBiostat, PhD, Paris Brain Institute, 47 Bd de l'Hôpital, 75013 Paris, France (janan.arslan@icm-institute.org).

Published Online: February 9, 2023. doi:10.1001/jamaophthalmol.2022.6036

Conflict of Interest Disclosures: None reported.

REFERENCES

1. Ahn S, Shin J, Song SJ, et al. Neurologic dysfunction assessment in Parkinson disease based on fundus photographs using deep learning. *JAMA Ophthalmol*. Published February 9, 2023. doi:10.1001/jamaophthalmol.2022.5928

2. Yang W, Hamilton JL, Kopil C, et al. Current and projected future economic burden of Parkinson's disease in the U.S. *NPJ Parkinsons Dis*. 2020;6(15):15. doi:10.1038/s41531-020-0117-1

3. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. *Neurology*. 1967;17(5):427-442. doi:10.1212/WNL.17.5.427

4. Goetz CG, Tilley BC, Shaftman SR, et al; Movement Disorder Society UPDRS Revision Task Force. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord*. 2008;23(15):2129-2170. doi:10.1002/mds.22340

5. Ahn J, Shin JY, Lee JY. Retinal microvascular and choroidal changes in parkinson disease. *JAMA Ophthalmol*. 2021;139(8):921-922. doi:10.1001/jamaophthalmol.2021.1728

6. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2016:2921-2929. .

7. Racoceanu D, Ounissi M, Kergosien Y. Explicabilité en Intelligence Artificielle; vers une IA Responsable. Instantiation dans le domaine de la santé. *Techniques de l'Ingénieur*. Published online December 10, 2022. doi:10.51257/a-v1-h5030