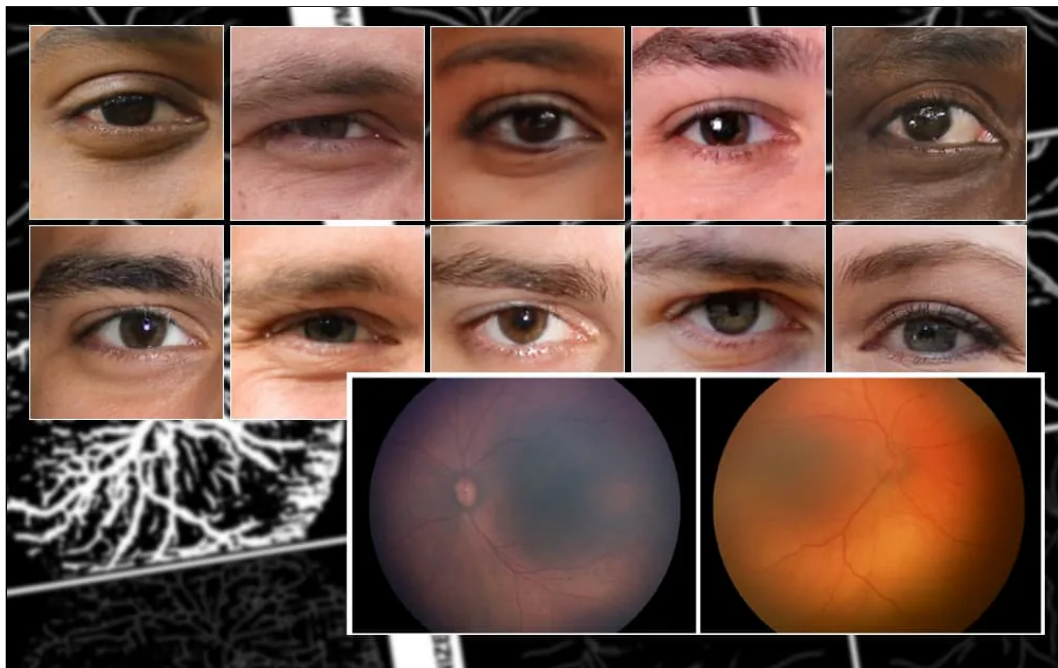


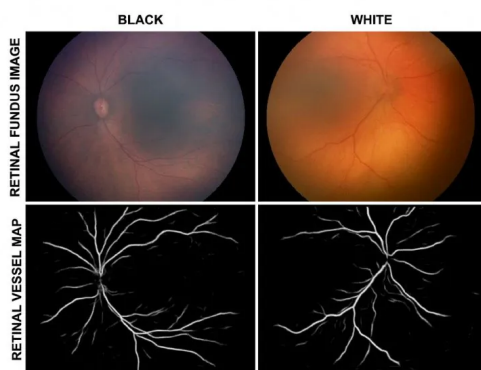
Retinal Veins Reveal Race, Widening Scope for AI Race Bias

Originally published at <https://www.unite.ai/retinal-veins-reveal-race-widening-scope-for-ai-race-bias/>



Inspired by recent revelations that medical AI imaging [can disclose race](#), a research consortium in the US and UK has conducted a study into whether retinal vein patterns are indicative of race, and has concluded that this is indeed the case, with AI able to predict parent-reported race in babies from retinal images – images that would not reveal racial identity to a human physician studying them, and which were previously thought to contain no potential for racial disclosure.

The group has expressed concern that this additional vector of racial stratification in medical imaging opens up the possibility for increased bias in the use of artificial intelligence systems in healthcare.



Retinal images from a black and a white subject. Above, a full-color Retinal Fundus Image from each race already has race-distinguishing features related to color disposition. Below, the Retinal Vessel Maps derived from these images, which are supposed to 'level' these racial anchor-marks, do actually contain race-identifying characteristics, according to the new report.

Source: <https://arxiv.org/ftp/arxiv/papers/2109/2109.13845.pdf>

The authors further note the possibility that [U-Net](#), the machine learning framework that has come to define this sector of AI-based healthcare, and which was trained on [predominantly white subjects](#)*, may have an influence on this observed phenomena. However, the authors assert that they are 'as of yet unable to fully explain these findings based on the U-Net hypothesis alone'.

Commenting on the findings at the project's associated [GitHub repository](#), the authors state:

'AI can detect race from grayscale RVMS [Retinal Vessel Maps] that were not thought to contain racial information. Two potential explanations for these findings are that: retinal vessels physiologically differ between Black and White babies or the U-Net segments the retinal vasculature differently for various fundus pigmentations.

'...Either way, the implications remain the same: AI algorithms have potential to demonstrate racial bias in practice, even when preliminary attempts to remove such information from the underlying images appear to be successful.'

The [paper](#) is titled *Not Color Blind: AI Predicts Racial Identity from Black and White Retinal Vessel Segmentations*, and is an equal collaboration between doctors and researchers from five institutions and research departments in the US, and one in the UK.

Medical doctors participating in the research consortium include [R.V. Paul Chan](#), MD, MSc, FACS, board-certified in ophthalmology, and a fellow of the American College of Surgeons; [Michael F. Chiang](#), M.D., Director of the National Eye Institute at the National Institutes of Health in Bethesda, Maryland; and [J. Peter Campbell M.D.](#), M.P.H., Associate Professor of Ophthalmology at the School of Medicine at Oregon Health & Science University in Portland.

The Eyes Have It

The paper notes the previously-proven potential for human-originated bias to propagate into AI medical systems, not least in [the study of eyes](#)*. Retinal Fundus Images (RFIs, see image comparison above), used in evaluating ocular disease, are full-color images that contain enough pigmentation information to identify race.

Greyscale Retina Vessel Maps (RVMs) discard most of this information in order to extract the underlying pattern of capillaries that are likely to define many disease conditions. It has always been assumed, at this level of distillation, that no racial characteristics remain in such reductive medical images.

The authors have tested this assumption with the use of a dataset of RFIs (full color retinal images) obtained from infants screened for a potentially blinding disease. Screening for such images, the authors note, is increasingly featured outside of personal consultations, in telemedicine and other remote diagnosis contexts, and is becoming more and more the subject of machine learning analysis.

The new study examines whether various types of reductionist versions of the race-identifying full-color images retain racial information, as reported by the parents of the babies, and has found that even the most information-destructive distillations of RFIs (thresholded, skeletonized and binarized) enable some level of racial identification.

Data and Methodology

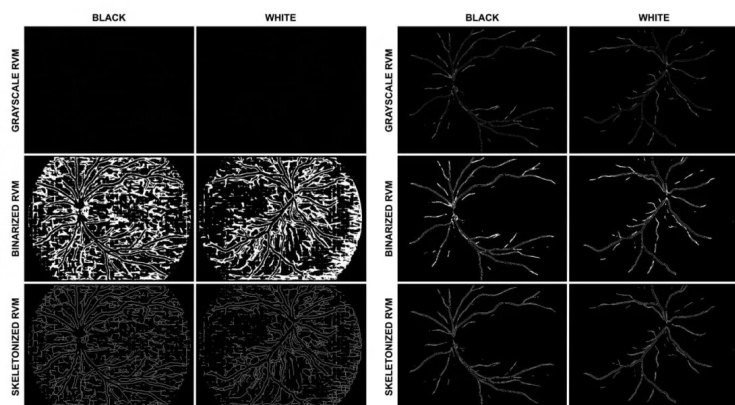
Data from 245 infants, gathered between January 2012 and July 2020 as part of a multicenter i-ROP cohort study, were divided into training, validation and test datasets on a 50/20/30 basis, respectively, with a natural distribution of races preserved as best the source data allowed.

Color RFIs were reduced down to the three aforementioned reductive styles of imaging, so that 'obvious' racial markers should technically have been removed from the data.

Multiple Convolutional Neural Networks (CNNs) were trained to achieve binary classification ('black'/'white', based on reported race from parents) using PyTorch. The CNNs ran the data across all versions of the images, from RFIs down to skeletonized versions, applying the usual random flips and rotations, with derived images having a resolution of 224×244 pixels.

The models were trained with stochastic gradient descent for up to ten epochs with a constant learning rate of 0.001, and early stopping implemented and training ceased where perceived convergence was identified after five epochs (i.e. the model was not going to get any more accurate with further training).

Since there was a demographically natural imbalance between white and black subjects, compensation was applied to ensure that minority sources were not systematically discounted as outliers, and the results were cross-checked to verify that no data leakage occurred across the experiments.



Samples of retinal maps from the study at varying levels of threshold tolerance.

Results

RVMs, which extract veins and capillaries from the full-color RFI images, should not theoretically be race-discernible by a CNN, according to the authors. However, results have shown that a higher number of major arteries are segmented by U-Net for white eyes than for black eyes.

In concluding remarks, the researchers observe 'We found that AI was easily able to predict the race of babies from retinal vessel segmentations that contain no visible information regarding pigmentation', and that 'even images that appeared devoid of information to the naked eye retained predictive information of the race of the original baby'. The researchers further offer the possibility that the retinal vessels of black vs. white babies differ 'in some way that AI can appreciate, but humans cannot'.

Prior Bias?

The authors also suggest that the discrimination could be a function of the white-predominant data on which U-Net was originally trained. Though they describe this as their 'leading theory', they also admit that the capabilities of capture sensors may be a factor in the phenomena, if it should turn out that the discovered bias is a corollary of the technical aspects of retinal imaging practices, or of data bias in U-Net which is perpetuating itself over the years. Addressing these

possibilities, the paper concedes:

'However, the U-Net was trained on RFIs that were first converted to grayscale images and subjected to contrast adjustment — specifically, contrast limited histogram equalisation (CLAHE) — and was therefore never actually trained on color RFIs. Thus, we are as of yet unable to fully explain these findings based on the U-Net hypothesis alone.'

However, the authors assert that the cause is less alarming than the effect, stating that the ability of AI models to discern race entails a possible *'risk of bias in medical AI algorithms that use them as input'*.

The authors point out the high-contrast nature of the races studied, and postulate that 'intermediate' racial groups may be more difficult to identify by similar means, and that this is an aspect that they intend to study in ongoing and related works.

** All supporting links provided by the paper that are included in this article have been converted from limited-access PaperPile links to publicly available online versions, where possible.*

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.

Portfolio site: martinanderson.ai

Contact: martin@martinanderson.ai



Discover More