Investigating Determinants and Evaluating Deep Learning Training Approaches for Visual Acuity in Foveal Hypoplasia

Volha V. Malechka, MD,[1], Dat Duong, PhD,[2],* Keyla D. Bordonada, MD,[1] Amy Turriff, MS,[1] Delphine Blain, MS, MBA,[1] Elizabeth Murphy, PhD,[1] Wendy J. Introne, MD,[2,3] Bernadette R. Gochuico, MD,[2,3] David R. Adams, MD, PhD,[2,3] Wadih M. Zein, MD,[1] Brian P. Brooks, MD, PhD,[1] Laryssa A. Huryn, MD,[1] Benjamin D. Solomon, MD,[2] Robert B. Hufnagel, MD, PhD[1]*

Purpose: To describe the relationships between foveal structure and visual function in a cohort of individuals with foveal hypoplasia (FH) and to estimate FH grade and visual acuity using a deep learning classifier.

Design: Retrospective cohort study and experimental study.

Participants: A total of 201 patients with FH were evaluated at the National Eye Institute from 2004 to 2018.

Methods: Structural components of foveal OCT scans and corresponding clinical data were analyzed to assess their contributions to visual acuity. To automate FH scoring and visual acuity correlations, we evaluated the following 3 inputs for training a neural network predictor: (1) OCT scans, (2) OCT scans and metadata, and (3) real OCT scans and fake OCT scans created from a generative adversarial network.

Main Outcome Measures: The relationships between visual acuity outcomes and determinants, such as foveal morphology, nystagmus, and refractive error.

Results: The mean subject age was 24.4 years (range, 1−73 years; standard deviation = 18.25 years) at the time of OCT imaging. The mean best-corrected visual acuity (n = 398 eyes) was equivalent to a logarithm of the minimal angle of resolution (LogMAR) value of 0.75 (Snellen 20/115). Spherical equivalent refractive error (SER) ranged from −20.25 diopters (D) to +13.63 D with a median of +0.50 D. The presence of nystagmus and a high-LogMAR value showed a statistically significant relationship ($P < 0.0001$). The participants whose SER values were farther from plano demonstrated higher LogMAR values (n = 382 eyes). The proportion of patients with nystagmus increased with a higher FH grade. Variability in SER with grade 4 (range, −20.25 D to +13.00 D) compared with grade 1 (range, −8.88 D to +8.50 D) was statistically significant ($P < 0.0001$). Our neural network predictors reliably estimated the FH grading and visual acuity (correlation to true value > 0.85 and > 0.70, respectively) for a test cohort of 37 individuals (98 OCT scans). Training the predictor on real OCT scans with metadata and fake OCT scans improved the accuracy over the model trained on real OCT scans alone.

Conclusions: Nystagmus and foveal anatomy impact visual outcomes in patients with FH, and computational algorithms reliably estimate FH grading and visual acuity. *Ophthalmology Science 2023;3:100225 Published by Elsevier on behalf of the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).*

Supplemental material available at www.ophthalmologyscience.org.

Foveal hypoplasia (FH) refers to the incomplete development of foveal architecture, including the lack of a foveal depression, lengthening of the photoreceptor outer segment, and widening of an outer nuclear layer with continuity of inner retinal layers.[1] Foveal hypoplasia may occur as an isolated condition[2−6] or in a syndromic context[7,8] and is frequently observed in individuals with albinism,[9,10] albinism-spectrum syndromes (e.g., Chediak−Higashi syndrome [CHS][11] or Hermansky−Pudlak syndrome [HPS]),[12,13] and aniridia-related diseases (e.g., Wilms tumor-aniridia-genital anomalies-retardation syndrome [WAGR] syndrome and other conditions involving the Paired Box 6 gene [*PAX6*]).[14−16] Foveal architecture can also be affected in conditions involving congenital and early childhood retinal degeneration[17] and retinal dystrophies, such as achromatopsia.[18−21]

The fovea confers high acuity vision in humans and nonhuman primates because of the increased density of cone photoreceptors with the exclusion of other cell types.[22−24] As such, best-corrected visual acuity (BCVA) is often impaired in individuals with FH.[25] However, in some conditions, high visual acuity is preserved despite the underdevelopment of the fovea in normal retinas[26] or in patients with retinal degeneration.[27] To provide a prognostic indicator of visual acuity, Thomas et al[28] developed a structural grading system for FH that combines qualitative descriptions of these architectural features. Because FH is a retinal structural anomaly, OCT

permits analysis and quantification of foveal structures, including the foveal pit contour, extrusion of the inner plexiform layer, lengthening of the outer segments, and widening of the outer nuclear layer.[29,30] Thus, OCT imaging can be used to correlate FH grade with visual outcome.

The current approach of using OCT for visual acuity prediction using the grading scale developed by Thomas et al[28] involves manual assignment of a discrete score from 1 to 4 by a trained expert (although FH degradation is continuous because FH grading changes from one stage to the next one). In albinism cohort studies, FH grading correlates with BCVA, whereas iris translucency, fundus pigmentation, and misrouting do not significantly predict visual acuity.[10] Compared with quantitative measures (e.g., photoreceptor length, outer segment length, and foveal developmental index), structural grading is the strongest predictor of future visual acuity.[31] Other clinical findings in the setting of FH may contribute to vision loss. For example, many individuals with FH may have nystagmus, which can affect vision.[25,29,32,33] Because FH, nystagmus, and other ocular features may each impact visual acuity (although the relationship among these entities is complex), it is important to understand their relative contributions to visual maturation, which may provide prognostic information.

Here, we characterized a cohort of patients with FH to explore the relationships between visual acuity outcomes and determinants, such as foveal morphology, nystagmus, and refractive error. Because visual acuity is often a key clinical question, the aim of our study was to estimate the logarithm of the minimal angle of resolution (LogMAR) visual acuity based on OCT findings. In this paper, while estimating LogMAR, we also tested a new automated algorithm to assign an FH score based on the gold standard grading system by Thomas et al.[28] To predict the FH grading and visual acuity, we built a neural network classifier for a small set of OCT scans (n = 368), along with nystagmus status and spherical equivalents. For predicting the FH grading, we evaluated the effectiveness of training the classifier with real images and images created from a generative adversarial network (GAN). We found that extra clinical data, like nystagmus status and spherical equivalents, and fake OCT images created by a GAN improve classifier accuracy for predicting visual acuity from OCT scans.

Methods

Individuals with FH were retrospectively identified using the National Eye Institute (NEI) electronic health record. A total of 201 individuals were identified through 2 searches. First, we searched for the term "FH" (as well as variants of that term) in any text field for patients seen from 2004 to 2018. Then, to identify potentially missed cases, we performed a second electronic health record search of the same period for one of the following search terms: oculocutaneous albinism, Albinism, Aniridia, Microphthalmia, Microphthalmos, Nanophthalmia, Nanophthalmos, *PAX6*, *PAX2*, WAGR, Chediak, CHS, Pudlak, and HPS. We collected 1 horizontal OCT B-scan from each subject for FH structural grading and analyzed data on foveal imaging, visual acuity, genetic testing, refractive error, and nystagmus.

Best-corrected visual acuity was measured using a Snellen eye chart, and values were converted using standardized LogMAR scale. The Teller Acuity Cards II were used to assess visual acuity in children aged < 3 years old. OCT scans of the macula were obtained on Cirrus (Cirrus HD-OCT; Carl Zeiss Meditec, Inc) (n = 156) or Spectralis (Heidelberg Engineering Inc) (n = 8) machines. OCTs were analyzed and graded by 2 independent human graders using the scale of Thomas et al.[28] Nystagmus was recorded from a clinical examination for individuals who had repetitive, uncontrolled eye movements (because the presence of nystagmus can be correlated with other variables, see below for a description of how analyses were done, without including nystagmus as a variable). Spherical equivalent refractive errors (SERs) were used to measure the refraction. The axial length was not included in the calculation of refractive error; cylinder and sphere component values were used for this purpose. All manifest refraction measurements were performed by certified vision examiners at the NEI, and values were recorded in the electronic health record to be later computationally converted into SERs.

This study was approved by the Institutional Review Board of the NEI. Informed consent was obtained from all participants.

Neural Network Predictor

Our classifier was built based on the EfficientNet-B4 model, which has obtained high performance on the ImageNet data (images of general objects) with relatively few parameters.[34] The pretrained weights from ImageNet were loaded and then trained end-to-end. Pretraining on a larger closely related dataset often helps in predicting accuracy of smaller datasets. Because EfficientNet-B4 was originally trained on ImageNet, which is unrelated to OCT scans, we first retrained EfficientNet-B4 on a large publicly available OCT dataset,[35] which contains 100 000 scans with 3 types of age-related macular degeneration, as well as normal scans. Next, we continued training this classifier specifically on just our datasets to predict FH gradings and LogMAR. Here, we modified just the last classification layer in EfficientNet-B4 to predict 4 choices of FH grading or a single LogMAR value and did not change the component that returned the vector representation of an image. During training and testing, all OCT scans were rescaled into the resolution 448 × 448 pixels, chosen to maximize graphical processing unit usage (one NVIDIA P100, training batch size 32). Supplementary Figure 5 shows a few examples where our classifier analyzes relevant regions of the OCT scans (see the Supplementary material).

Metadata, like nystagmus and spherical equivalents, were represented as a numerical vector input. A discrete variable (e.g., nystagmus) was represented as 0 or 1, denoting its absence or presence, whereas a continuous variable (e.g., spherical equivalents) was kept unchanged. For example, the metadata vector [1, 0.5] would represent the presence of nystagmus and a 0.5 spherical equivalent. This metadata vector was concatenated with the vector representation of an image and then passed the combined information into the last classification layer. Experiments without nystagmus as the metadata input were also performed to avoid unintentional bias (see "Discussion" section).

For FH gradings, the classifier was trained with cross-entropy loss, in which true labels are 1-hot encodings; for example, the label of an image with an FH grading score 1 out of 4 would be represented as the vector [1, 0, 0, 0]. The classifier for predicting LogMAR, which is a continuous value, was trained with smooth-L1-loss.

For each label type (e.g., FH grading or LogMAR), we trained 5 classifiers via fivefold cross-validation (one for each fold). For fair comparison, the same fold partitions were used when predicting FH gradings and LogMAR. We then created an ensemble predictor by averaging the predicted label of an image from these 5 classifiers. When averaging the FH grading predictions, we considered only the classifiers that produced a maximum predicted probability (overall labels) of at least 0.5. Although FH gradings are discrete values in the Thomas et al[28] grading scheme, disease progression is continuous. We computed a continuous FH grading by taking the average score weighted by the prediction probabilities.

Generative Adversarial Network

We trained a generative adversarial network to generate fake images based on FH gradings. A GAN was trained for each data partition from the fivefold cross-validation in the section "Neural Network Classifier." We describe the GAN training and image generation for a data partition p, which also will apply to the other partitions.

We trained StyleGAN2-Ada to generate images using 2 inputs: FH gradings and eye position (left vs. right).[36] We made the following key modification to StyleGAN2-Ada. Its default label embedding is L $\times$ 512, in which L is the number of labels and produces a vector of length 512 for each label. In our case, training this embedding requires many OCTs from 8 unique label combinations (4 possible FH gradings and 2 eye positions). For this reason, the default label embedding was replaced with 2 smaller 4 $\times$ 256 and 2 $\times$ 256 matrices to represent the 4 FH gradings and eye positions (left vs. right). Training the 4 $\times$ 256 disease embedding then used all OCTs from both eyes. Likewise, training the 2 $\times$ 256 eye location embedding uses all images from every FH grading. The outputs of these 2 components are concatenated to a vector of size 512 to match the rest of the StyleGAN2-Ada architecture. We initialized all the StyleGAN2-Ada weights with the pretrained values on Flickr-Faces-High-Quality dataset at resolution 256 $\times$ 256 pixels[36]; all fake images were generated at 256 $\times$ 256-pixel resolution. The image resolution was chosen to maximize graphical processing unit usage (2 NVIDIA P100). Our code and trained models are available at https://github.com/datduong/stylegan2-ada-FovealHypoplasia.

After training the GAN on a data partition p, we generated the following 2 types of fake images: (1) unrelated images for a given label and (2) images between 2 consecutive labels. We generated images for every combination (g,p) from FH grading $g \in \{1, 2, 3, 4\}$ and eye position $p \in \{left, right\}$.[4,21] Conditioned on the eye position p, the number of generated images for each pair (g,p) is equal to the average count with respect to the FH gradings in the data partition. Thus, we created a fake dataset twice the size of our original dataset. We also made more and fewer images for the uncommon and common combinations of FH gradings and eye positions, respectively. In total, we created a fake dataset twice the size of our original dataset.

For type 1, we generated a fake OCT scan i by concatenating the random vector r_{igp} with the FH grading label and eye position embedding e_g and e_p, denoted as $[r_{ige}, e_g, e_p]$, and then passing this new vector to our GAN image generator. Each grading g and eye e combination has images generated from their own unique random vector r_{igp} so that, theoretically speaking, all the fake images are unique (Figure 1).

For type 2, with a random vector r_{igp}, we generated 2 images: the first one shared characteristics of label g and $g\prime$ and the second image with characteristics of label g and $g\prime\prime$. For example, suppose $g = 2$ (e.g., FH grading score 2); we would generate fake images with both characteristics of FH scores 1 and 2 and then fake images with characteristics of scores 2 and 3 (Figures 2 and 3). To do this, we passed into the GAN image generator the inputs $[r_{igp}, ce_g + (1-c)e_{g\prime}, e_p]$ and $[r_{igp}, ce_g + (1-c)e_{g,\prime\prime}, e_p]$, where c is a predefined fraction between 0 and 1. The true labels of these images are soft labels; for example, the image created from the vector $[r_{igp}, ce_g + (1-c)e_{g\prime}, e_p]$, in which $g = 1$ and $g = 2$, would have the soft label encoding $[c, 1-c, 0, 0]$ instead of the traditional 1-hot encoding. We emphasize that g, $g\prime$, and $g\prime\prime$ are consecutive; we do not generate fake images with mixed characteristics of FH grading 1 and 3. When g is 1 or 4, then all the fake images will be a mixture of FH grading 1 and 2, or 4 and 3, respectively. We tested c at 55%, 75%, and 90%. Figure 2 shows examples of fake OCT scans with $c = 55$%.

Next, we created 2 new larger datasets by combining partition p with each of the 2 fake image types. We trained EfficientNet-B4 on each of these new larger datasets. For each type of new dataset, we created the ensemble predictor over all data partitions after the approach described in the section "Neural Network Classifier."

Statistical Analysis

One-way analysis of variance test was used to test the difference in the following: (1) LogMAR visual acuity between different grades of FH, (2) LogMAR visual acuity between different groups of refractive errors, and (3) LogMAR visual acuity between grades of individual foveal features. An unpaired 2-tailed t test was applied to assess the difference in LogMAR between the groups with and without nystagmus. The Levene test was used to evaluate the equality of variances for a variable calculated for 4 FH groups. Agreement between 2 graders was reported to assess the validity of the grading system.

Results

Participant Characteristics

A total of 201 individuals with FH were evaluated at the NEI from 2004 to 2018 and included in this study. The mean age was 24.4 $\pm$ 18.25 years (range, 1−73 years) at the time of OCT imaging. The mean BCVA (n = 398 eyes) was equivalent to a LogMAR value of 0.75 (Snellen 20/115). Of the entire cohort, 88% (n = 177) had nystagmus clinically on examination. Spherical equivalent refractive errors ranged from −20.25 diopters (D) to +13.63 D with a median of +0.50 D in the entire study cohort.

A majority (83%) of subjects (n = 167) had an albinism-spectrum disorder, including albinism, CHS, and HPS. Specifically, 54.7% of patients (n = 110) presented with oculocutaneous albinism, and 3% of the cohort (n = 6) had a clinical diagnosis of ocular albinism, whereas 25.3% of patients (n = 51) had syndromic forms of albinism, with 20.3% of the total cohort having a diagnosis of HPS (n = 41) and 3% with CHS (n = 10). Four percent (n = 8) had aniridia, and of these, 5 patients had mutations in *PAX6*, whereas 4.5% of patients had type 2A Waardenburg syndrome (n = 9). Seven percent (n = 14) presented with an apparently isolated form of FH, 1.5% (n = 1) had achromatopsia, and 1.5% (n = 1) had Erdheim−Chester disease. This finding of FH in Erdheim−Chester disease might be coincidental to Erdheim−Chester disease. The composition of the entire FH cohort is represented in Figure 4.
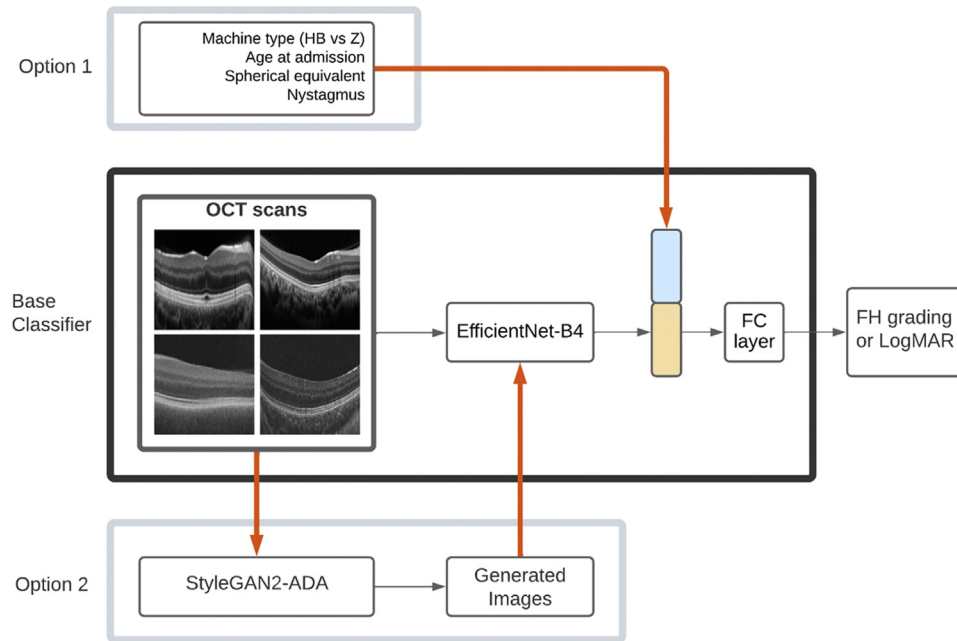
Figure 1. Outline of algorithm training approaches. Our base classifier used EfficientNet-B4 to embed OCT scans as vectors (yellow), which were then passed through a fully connected layer (FC layer) to predict the foveal hypoplasia (FH) grading or logarithm of the minimal angle of resolution (LogMAR) value. Option 1 included the metadata represented as vectors (blue). These vectors were concatenated with the image embeddings (yellow), and the combined outputs were passed through the FC layer. Option 2 used real OCT scans to train a StyleGAN2-Ada image generator. Fake images were then created and jointly trained with real images. We applied option 1 or 2 separately (not simultaneously) to the base classifier. HB = Heidelberg; Z = Zeiss.

Exploratory Data Analysis between Key Visual Determinants and Visual Acuity
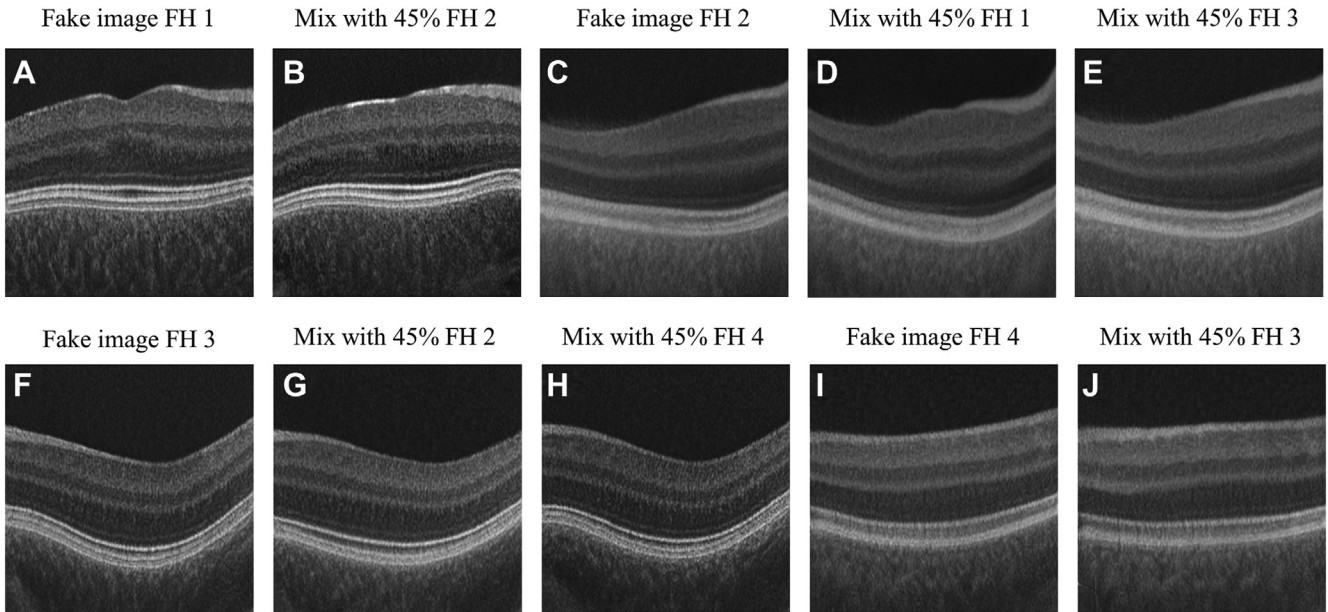
We analyzed the relationships between the visual acuity LogMAR score and the following key visual determinants: FH score, nystagmus status, and SER. These variables were used later as inputs for the neural network model to estimate LogMAR.

OCT scans from the right and left eyes of every participant at the initial visit were assessed for whether visual determinants, such as nystagmus, spherical equivalent, and FH grade influenced the visual outcome. Values from both
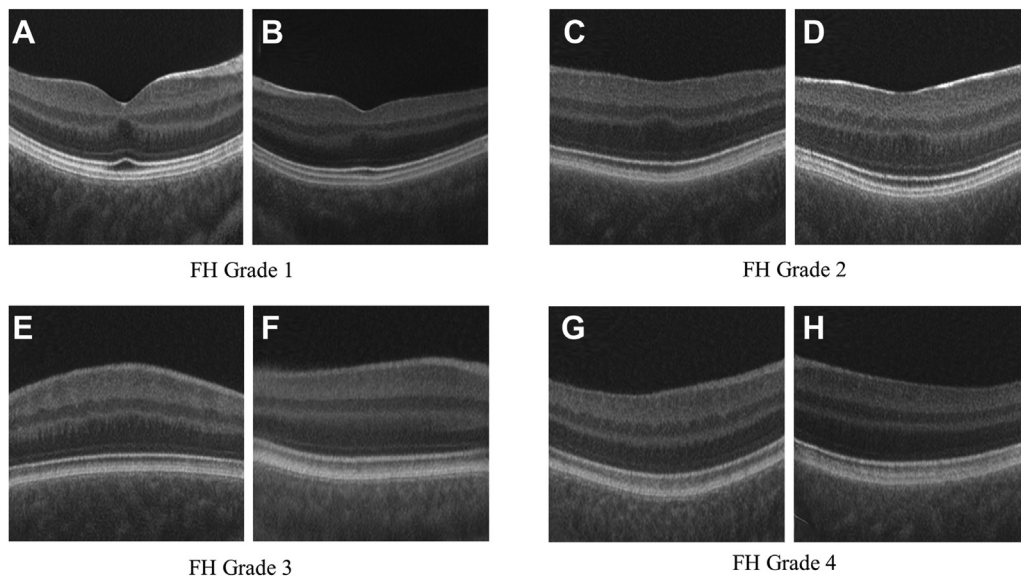
Figure 2. Examples of type 1 fake OCT scans, one for each eye having foveal hypoplasia (FH) grading from 1 to 4 (**A−H**). Fake OCTs were generated at resolution 256 × 256 pixels.

Fake image FH 1	Mix with 45% FH 2	Fake image FH 2	Mix with 45% FH 1	Mix with 45% FH 3

Fake image FH 3	Mix with 45% FH 2	Mix with 45% FH 4	Fake image FH 4	Mix with 45% FH 3

Figure 3. Examples of type 2 fake OCT scans (**A**−**J**). Images with mixed characteristics of 2 foveal hypoplasia (FH) gradings. Only mixed labels were used for training (**B, D, E, G, H, J**); the unmixed images are shown as reference (**A, C, F, I**). Fake images were generated at resolution 256 × 256 pixels.

eyes of every participant were included in the analyses, with a very high concordance of FH grade between the right and left eyes in the same individual. Two trained ophthalmic geneticists discussed and provided a single FH score for each OCT scan using the scale of Thomas et al,[28] currently considered the gold standard for FH grading. For individuals who had scans on both OCT systems (n = 37), we found 100% agreement in OCT grading. Figure 5A shows a positive correlation between a higher FH grade and a higher LogMAR value, which was statistically significant when comparing values of FH grades 1, 2, 3, and 4 ($P <$ 0.05). Although the ranges of LogMAR are highly overlapping between FH scores, 1 subject with FH grade 4 was found to have better vision (LogMAR = 0.4) than the mean of those with FH grade 1 (LogMAR = 0.4714). In the entire cohort (n = 398 eyes), we found a statistically significant difference between the presence of nystagmus and a high-LogMAR value, corresponding to a positive correlation between the presence of nystagmus and reduced visual acuity ($P <$ 0.0001) (Figure 5B). For those individuals with SER and LogMAR recorded, we observed that participants whose SERs were farther from plano demonstrated higher LogMAR values (n = 382 eyes) (Figure 5C).

In summary, a higher LogMAR value is a characteristic feature of subjects with nystagmus, higher SERs, and higher FH grade.

Next, we compared the effects among the visual determinants. First, we examined the OCT-graded FH individuals with (n = 177) and without (n = 24) nystagmus. The proportion of patients with nystagmus increases with a higher FH grade from 70% in grade 1 to 98% in grade 4 (Figure 6A).

Subsequently, we examined the distribution of refractive error among FH grades. We observed a higher variability in SER with grade 4 (range, −20.25 D to +13.00 D) than with grade 1 (range, −8.88 D to +8.50 D) ($P <$ 0.0001) (Figure 6B).

We performed similar analyses of clinical diagnosis subgroups. For analyses of visual determinants with FH grade among the entire cohort, individuals with an albinism-spectrum disorder and individuals with aniridia or another *PAX6* spectrum disorder show a similar pattern in the data distribution between FH grades (Figs S1, Figs S2, Figs S3, Figs S4, available at www.ophthalmologyscience.org).

In summary, in this cohort, FH grade, nystagmus, and refractive error correlate with worse visual acuity. Furthermore, there seem to be similar relationships between FH grade, nystagmus, and refractive error. Effects on visual outcome and between visual determinants do not seem to be dependent on clinical diagnosis.

Neural Network Predictor for FH Grading

We briefly describe our neural network training process in this section; more details are explained in the methods. The scale described by Thomas et al[28] is manual and can be highly subjective because of human judgment. To systematize this process, we evaluated a neural network classifier's ability to automate the FH grading for the OCT scans. In our dataset of 201 individuals, there were OCT scans from both eyes of a patient taken by 2 machine types; hence, some people might have had 4 scans taken. We trained the model on 164 individuals (368 images) and tested the model on 37 individuals (98 images). The test set has the following characteristics: individuals were

Total cohort (N = 201)
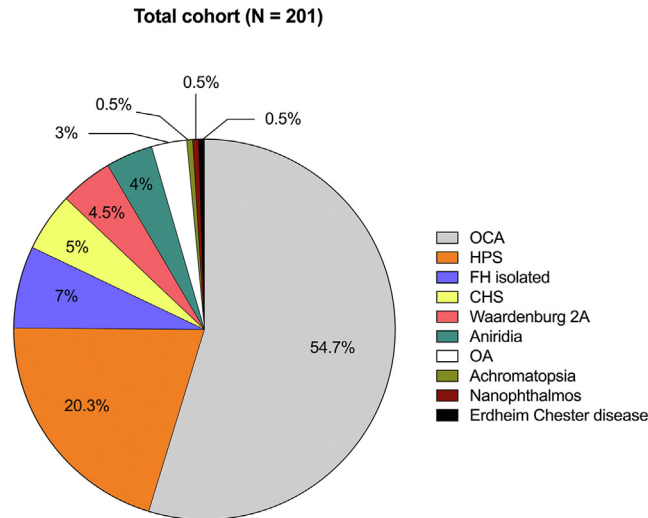
Figure 4. Composition of the study cohort. CHS = Chediak−Higashi syndrome; FH = foveal hypoplasia; HPS = Hermansky−Pudlak syndrome; OA = ocular albinism; OCA = ocular cutaneous albinism.

with the age range at the time of OCT scanning between 5.1 and 65 years, with 31 of 37 individuals having nystagmus; LogMAR ranged from −0.1 to 1.4; and spherical equivalent values were between −12.50 D and +11.13 D. Only high-quality OCT scans from 2 machine types of the same individual were used for analysis.

We evaluated 3 different training approaches for our neural network predictor, which is based on the EfficientNet-B4 architecture.[34] Figure 1 outlines these approaches. First, we pretrained EfficientNet-B4 with a large publicly available OCT dataset and then continued training this model on our dataset (see our base classifier in Figure 1).[35] Second, we represented visual determinants, such as nystagmus and spherical equivalent, and other factors like patient age at the time of OCT scanning as a

vector. We concatenated this metadata vector with the vector representation of an image and then passed the combined information into the last classification layer (option 1 in Figure 1).

Third, because our dataset is small compared with many other computer vision datasets, we generated fake OCTs based on our real OCTs to increase the training sample size. Intuitively, because of the sample size difference, a classifier trained on fake and real images should outperform the one trained on just real images. To this end, the GAN, StyleGAN2-Ada,[36] was applied to create fake OCT scans, which were then used with real OCT scans to train EfficientNet-B4 (option 2 in Figure 1). The following 2 types of fake images were created: (1) unrelated images for a given label (Figure 2); and (2) OCT scans that are between 2 consecutive labels
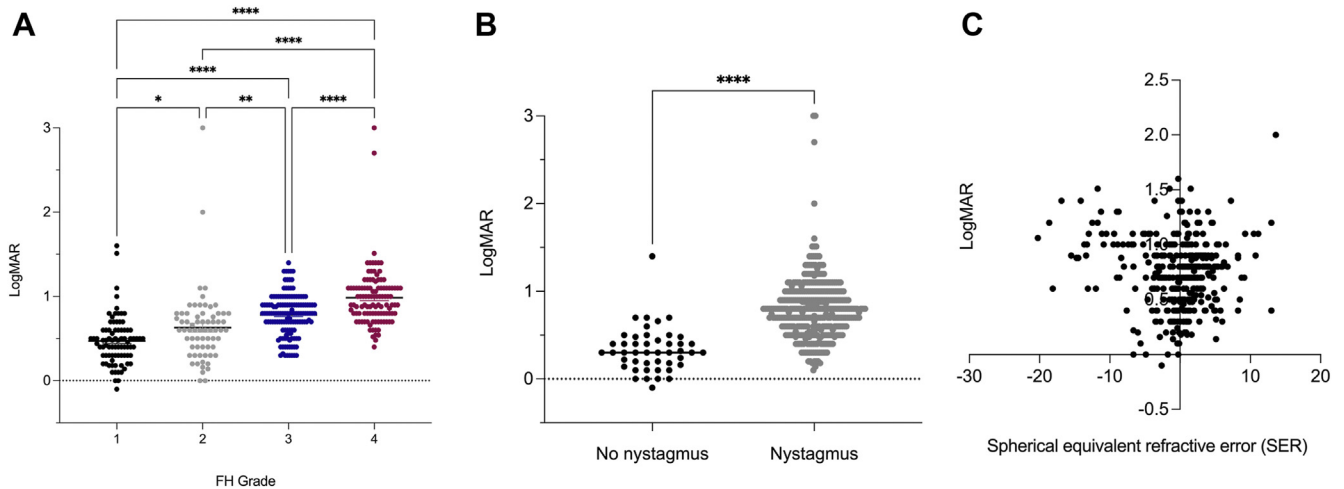
Figure 5. Visual acuity by foveal hypoplasia (FH) grade (A), nystagmus (B), and spherical equivalent refractive errors (SERs) (C). *$P \leq 0.05$; **$P \leq 0.01$; ****$P \leq 0.0001$. LogMAR = logarithm of the minimal angle of resolution.
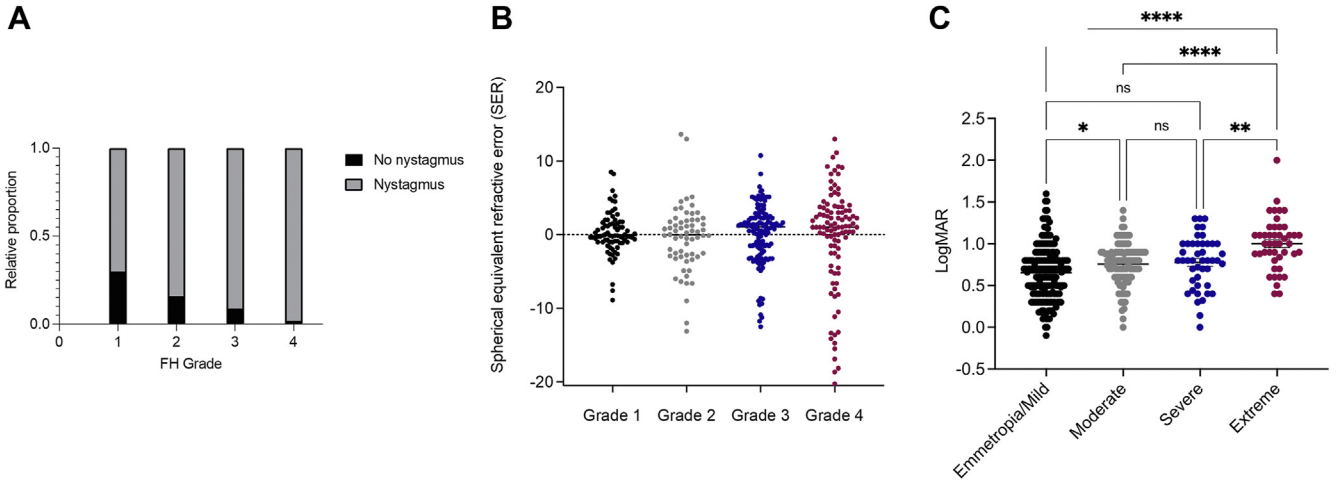
Figure 6. Spherical equivalent refractive error by FH grade (**A**), relative proportion of number of individuals with versus without nystagmus by foveal hypoplasia (FH) grade (**B**), and the distribution of refractive error classes within the cohort (**C**). *$P \leq 0.05$; **$P \leq 0.01$; ****$P \leq 0.0001$. LogMAR = logarithm of the minimal angle of resolution; ns = not significant.

([Figure 3](#)). We combined the real OCTs with each of these 2 fake image types, making 2 new larger datasets, and then trained EfficientNet-B4 on each of them.

For each of the 3 training approaches, we created the ensemble predictor via fivefold cross-validation. We computed a continuous FH prediction by taking the weighted average of predicted probabilities and the corresponding FH gradings. [Supplementary Table 1](#) compares training the model with and without fake OCT scans when assuming the predicted FH grading must be discrete. The following 2 accuracy metrics were considered: (1) Spearman correlation, in which a high value indicates that severe cases are assigned higher scores, and (2) linear regression coefficient (r^2), in which a high value indicates that predicted scores closely match their true values.

[Table 1](#) compares random forest classifiers trained on metadata and neural network classifiers trained according to the 3 training procedures mentioned above. Unsurprisingly, the lowest accuracy ([Table 1](#), row 1) was obtained when ignoring OCT scans and using random forest to predict FH grading based on these variables (denoted as symbol V in [Table 1](#)): machine type (Heidelberg vs. Zeiss), age at admission, spherical equivalent. Including nystagmus into the random forest ([Table 1](#), row 2) increases the accuracy, in concordance with the results shown in [Figure 6](#)A, B, in which nystagmus is correlated with FH, whereas SER is not.

When training neural network predictors, including metadata yielded modest improvement ([Table 1](#), rows 3–5), with a minor drop without modeling nystagmus. Random forest results and these 3 rows imply that (1) without OCT scans, considering nystagmus is useful and (2) with OCT scans, nystagmus provides a slight advantage. There are minor differences among the neural networks trained on real and fake OCT scans; however, the trend is that using both real and fake images improves the accuracy ([Table 1](#), rows 3 and 6–9).

In summary, neural network classifiers reliably estimated FH gradings using OCT scans and related metadata for patients in our cohort.

Table 1. Neural Network Predictor Accuracy When Treating FH Grading as a Continuous Value

		Correlation	R^2
Random forest	V*	0.1460	0.0344
	V + nystagmus	0.3800	0.1589
Neural network	Image	0.8574	0.7114
	Image + V	0.8777	0.7642
	Image + V + nystagmus	0.8945	0.7775
	Image + GAN base[†]	0.8828	0.7543
	Image + GAN55[‡]	0.8632	0.7402
	Image + GAN75[§]	0.8771	0.7614
	Image + GAN90[‖]	0.8942	0.7775

FH = foveal hypoplasia; GAN = generative adversarial network; HB = Heidelberg; Z = Zeiss.
*V: Set of the following variables: machine type (HB vs Z), age at admission, and spherical equivalent.
[†]GAN base: Type 1 fake OCT scans in which each fake image is generated for a specific given FH grading.
[‡]GAN55: Type 2 fake OCT scans in which mix ratio is 55%; for example, a fake image would have 55% and 45% characteristics of FH gradings 1 and 2 as in [Figure 2](#).
[§]GAN75: Type 2 fake OCT scans in which mix ratio is 75%.
[‖]GAN90: Type 2 fake OCT scans in which mix ratio is 90%.

Neural Network Predictor for Visual Acuity

We modified the first 2 neural network training approaches, previously described for predicting FH to estimate Log-MAR. In our dataset, we considered visual acuity as a numerical value measured as LogMAR. We did not consider the third neural network training approach involving fake OCTs because LogMAR does not have a discrete label set, which is required to train StyleGAN2-Ada. Also, we included the true FH gradings based on the scale of Thomas

Table 2. Neural Network Predictor Accuracy When Predicting LogMAR

		Correlation	R^2
ElasticNet	Nystagmus	0.4585	0.2474
	V*	0.0591	0.0229
	V + nystagmus	0.4281	0.2938
	V + FH†	0.7258	0.4948
	V + nystagmus + FH	0.7644	0.5718
Neural network	Image	0.7080	0.4700
	Image + V	0.7081	0.4565
	Image + V + nystagmus	0.7032	0.4967
	Image + V + FH	0.7563	0.5392
	Image + V + nystagmus + FH	0.7765	0.5745

FH = foveal hypoplasia; HB = Heidelberg; LogMAR = logarithm of the minimal angle of resolution; Z = Zeiss.
*V: Set of the following variables: machine type (HB vs. Z), age at admission, and spherical equivalent
†FH: True FH gradings for the OCTs based on Thomas et al, provided by 2 experts.

et al[28] along with the previous metadata: machine type, age at admission, spherical equivalent, and nystagmus.

Among the metadata, elastic net indicates that (1) nystagmus and FH grading are the 2 best predictors and (2) SER does not perform well, agreeing with Figure 5 (Table 2, rows 1−5). With the true FH gradings, OCT scans do not contribute much more to the prediction outcome (Table 2, row 5 vs. 10 and row 4 vs. 9). This is expected because, in our dataset, FH is highly correlated with visual acuity (Figure 5A). Without the true FH gradings, OCT scans alone sufficiently estimated the visual acuity (Table 2, rows 6−8). In summary, our neural network classifiers reasonably estimated LogMAR from just OCT scans because our correlation coefficient 0.7080 falls near the lower end of the range 0.7 to 1, which is considered "strong."[37]

Discussion

Humans possess high acuity vision because of a layerless specialization in the central retina, the fovea.[38] The fovea centralis is an avascular zone located within the macula, presenting as a small anatomic depression rich in cone photoreceptors with high connectivity between cones, bipolar, and ganglion cells.[24,39]

Early morphologic changes associated with the development of human fovea take place from 12 to 24 weeks of gestation,[40] and the fovea remains immature until 4 years of age, when maturation is completed.[41] Foveal hypoplasia results from an interruption of the developmental process and is formally defined as a lack of foveal depression with continuity of neuronal layers through the central retina and a reduced avascular zone.[30,42−44]

Here, based on the scale of Thomas et al[28], we observed that the impact of these features on visual acuity is likely conserved. The clinical diagnosis associated with FH may

be less important than factors, such as nystagmus, spherical equivalent, and OCT scans, in building a neural network classifier to predict FH gradings and visual outcomes. Moreover, individuals with visual differences with increasing refractive error, especially with extreme ametropia ($> +6$ or $< −9$ D) have worse visual acuity than those with emmetropia/mild, moderate, or severe ametropia within our cohort (Figure 6C).

In this paper, we evaluated how well a neural network model estimates the FH gradings and the visual acuity (as LogMAR) for a cohort in which most subjects (84% of the test set) had foveal developmental anomalies and nystagmus. We explored the following 3 approaches to train the neural network model: (1) training the model on just OCT scans to predict the true label, (2) training on OCT scans and metadata, and (3) training on real OCT scans and fake OCT scans created from StyleGAN2-Ada. When estimating LogMAR, the third option was not implemented because StyleGAN2-Ada needs a discrete label set to generate fake images.

All 3 training options reliably estimated the FH gradings and LogMAR values; however, the latter 2 training strategies yield better accuracy. First, metadata provide more information per input sample, which helps model performance. Second, because of sample size difference, a classifier trained on fake and real images should outperform the one trained only on real images. Interestingly, along with studies in other disciplines (e.g., analyses of x-rays or skin lesions), we found that joint training of real and fake images only slightly improves the prediction outcome.[45−47] Our result also shows different types of fake OCT scans; the best outcome is seen for images with mixed characteristics of 2 discrete FH gradings. We suspect that these mixed images may better reflect the continuous progression of FH degradation. Each GAN experiment in Table 1 used a constant mixing fraction between every 2 discrete FH gradings. For example, the experiment Image + GAN75 had fake images containing 75% and 25% characteristics of FH gradings 1 and 2, 75% and 25% characteristics of FH grading 2 and 3, and so forth. Future work will include evaluating different combinations of mixing fractions between FH gradings and exploring other GAN architectures.[48]

Tables 1 and 2 show that the OCT scans, not nystagmus and SER, are the main factor behind the performance of a neural network classifier. Figure 7 depicts selected OCT scans with misclassified FH gradings. Figure 5 and Figure 6 suggest that nystagmus, not the spherical equivalent, is an important predictor of visual acuity. We note that nystagmus, visual acuity, and FH grading may exhibit a causality dilemma; that is, nystagmus and high refractive errors can affect visual acuity, and these manifestations may correlate. We attempted to address this in part by conducting analyses without considering nystagmus but emphasize that correlated variables can present challenges, including in situations in which multiple types of data are analyzed (e.g., images and metadata). Related to this, one limitation is that our dataset does include outliers (including those unrecognized manifestations that may affect our results), such as a single patient with achromatopsia. Although rare outliers would be unlikely
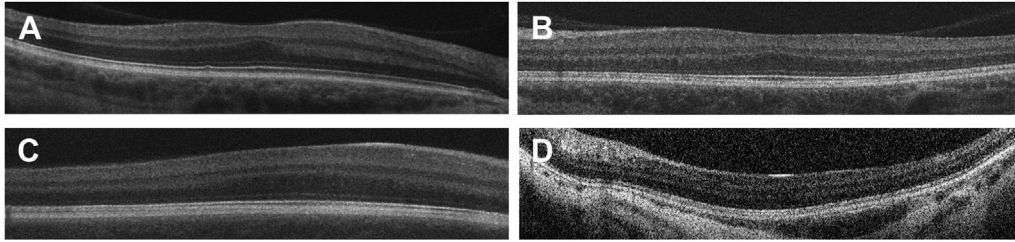
Figure 7. Representative OCT scans showing discrepancy of foveal hypoplasia (FH) grading between deep learning (DL) predictor and trained geneticists. **A**, OCT scan determined as FH grade 1 but estimated as FH grade 2 by DL predictor. Note that all features of complete foveal development are present. **B**, OCT scan determined as FH grade 2 but estimated as FH grade 1 by DL predictor. Note that the extrusion of plexiform layers and foveal pit are absent. **C**, OCT scan determined as FH grade 3 but estimated as FH grade 2 by DL predictor. Note that the extrusion of plexiform layers, foveal pit, and lengthening of the photoreceptor outer segment are absent. **D**, OCT scan determined as FH grade 4 but estimated as FH grade 2 by DL predictor. Note that extrusion of plexiform layers, foveal pit, lengthening of photoreceptor outer segment, and widening of an outer nuclear layer are absent.

to affect the results, it is possible that we were not able to identify all such instances. Similarly, failures in emmetropization in individuals with genetic eye conditions can affect visual acuity in complex ways, which may not be captured by the data types used in these analyses. After this, OCT scans might be the least biased source to estimate FH grading and visual acuity, although they do not represent all potentially relevant information, and further study is needed. In Tables 1 and 2, neural network models without nystagmus did not greatly underperform, indicating that, in our dataset, the OCT scans were sufficient predictors of FH grading and visual acuity.

It is also important to acknowledge that nystagmus and spherical equivalent (among other metadata like age) are clinically observed and can be suboptimal at estimating FH grading and visual acuity. Better inputs would include quantitative continuous structural parameters of the fovea like outer nuclear layer widening and inner retinal layer extrusion values. However, obtaining these parameters can be difficult (and, although of interest for future work, it was unfortunately not possible to gather these data in this study); moreover, the measurements can be noisy because of biases in human judgment. In this paper, we opted to have the neural network classifier holistically analyze the OCT scans; hence, the models may implicitly use the foveal structural parameters. In future work, we hope to explore image segmentation methods to automatically separate each foveal component, measure the relevant structural parameters, and use these values to predict FH grading and visual acuity.

In albinism, central macular thickness was thought to correlate with visual acuity.[49,50] However, later studies were unable to clearly establish this relationship.[30] Seo et al[51] developed a grading system for FH in patients with albinism predominantly considering pigmentation defects because the grading was determined from OCT images and based on foveal hyporeflectivity, the degree of choroidal transillumination, the presence of the tram-tract sign, and foveal depression. The results suggest that the prognostic value of FH gradings was superior to grading scores based on iris transillumination or macular

transparency in young patients with albinism. In 2011, Thomas et al[28] developed a structural grading system for FH that combines qualitative characteristics of the fovea based on the step at which foveal development was arrested and that helps predict visual acuity in disorders with FH. Kruijt et al[10] investigated the phenotype of a cohort of individuals with albinism and found that the FH grading correlated best with the visual acuity, whereas iris translucency, fundus pigmentation, and misrouting were not as useful. Rufai et al[31] evaluated structural grading and quantitative segmentation of FH using handheld OCT versus preferential looking technique as a prognostic indicator of the visual acuity in preverbal children with infantile nystagmus. They found that the application of a structural grading system can predict future visual acuity compared with quantitative measures (e.g., photoreceptor length, outer segment length, and foveal developmental index). Exploring associations between FH and visual acuity in a cohort of individuals with and without albinism, with infantile nystagmus syndrome, Healey et al[25] demonstrated a relation between higher FH grade and poorer visual acuity. They also observed an association between FH grade and hyperopia degree.

Building a classifying algorithm as a predictor of FH gradings and visual outcomes in individuals with foveal developmental anomalies is a potentially useful approach when considering prognostic information in the clinical care of rare conditions. Neural network classifier as an automated technique may be a useful method for identifying the most efficient strategy in the management of FH. This algorithm can be modified, retrained, and applied to other rare eye conditions.

Nystagmus makes it difficult to assess visual acuity in patients with foveal anomalies. When OCT scans can be obtained in these individuals, the development of a method for visual acuity prediction based on OCT results in individuals with FH could help vision researchers and clinicians as well as patients.

In practice, visual acuity measurements can be obtained for older children and individuals with nystagmus but with some difficulty. Previous studies showed that most but not all individuals with disorders associated with FH exhibited

nystagmus.[10,31] Generally, this could be the case in individuals with a lower FH grade and lower LogMAR who have a more normally developed fovea.[10] Because our LogMAR predictions correlated well with true measurements, we plan to evaluate our current model with a larger cohort and design a similar but more sophisticated approach to handle longitudinal data.[31]

In our study, the entire cohort of individuals with an albinism-spectrum disorder and individuals with aniridia or another *PAX6* spectrum disorder demonstrated a correlation among LogMAR, presence of nystagmus, refractive error, and FH grade. Individuals with a higher FH grade showed a higher LogMAR and a higher variability in refractive error. Nystagmus was present in a larger number of individuals who had a higher FH grade.

In summary, our results provide novel insight into the application of computational techniques in the study of FH and foveal developmental anomalies. The ability to extract meaningful prognostic data from OCT alone may be beneficial.

Footnotes and Disclosures

[1] Ophthalmic Genetics and Visual Function Branch, National Eye Institute, National Institutes of Health, Bethesda, Maryland.

[2] Medical Genomics Unit, Medical Genetics Branch, National Human Genome Research Institute, Bethesda, Maryland.

[3] Human Biochemical Genetics Section, National Human Genome Research Institute, Bethesda, Maryland.

*V.V.M. and D.D. contributed equally to this work.

Disclosures:

All authors have completed and submitted the ICMJE disclosures form.

The authors made the following disclosures: B.D.S.: Support − National Human Genome Research Institute; (NHGRI); Payment − Wiley, Inc.; Participation − FDNA; Stock or stock; Options − Opko Health.

The other authors have no proprietary or commercial interest in any materials discussed in this article.

HUMAN SUBJECTS: Human subjects were included in this study. The study was approved by the Institutional Review Board of the NEI. Informed consent was obtained from all participants.

No animal subjects were used in this study.

Abbreviations and Acronyms:

BCVA = best-corrected visual acuity; **CHS** = Chediak−Higashi syndrome; **D** = diopters; **FH** = foveal hypoplasia; **GAN** = generative adversarial network; **HPS** = Hermansky−Pudlak syndrome; **LogMAR** = logarithm of the minimal angle of resolution; **NEI** = National Eye Institute; *PAX6* = Paired Box 6 gene; **SER** = spherical equivalent refractive error; **WAGR** = Wilms tumor-aniridia-genital anomalies-retardation syndrome.

Correspondence:

Robert B. Hufnagel, MD, PhD, Bethesda, MD 20892. E-mail: robert.hufnagel@nih.gov.

References

1. Kondo H. Foveal hypoplasia and optical coherence tomographic imaging. *Taiwan J Ophthalmol*. 2018;8:181−188.
2. Curran RE, Robb RM. Isolated foveal hypoplasia. *Arch Ophthalmol*. 1976;94:48−50.
3. Oliver MD, Dotan SA, Chemke J, Abraham FA. Isolated foveal hypoplasia. *Br J Ophthalmol*. 1987;71:926−930.
4. Poulter JA, Al-Araimi M, Conte I, et al. Recessive mutations in *SLC38A8* cause foveal hypoplasia and optic nerve misrouting without albinism. *Am J Hum Genet*. 2013;93:1143−1150.
5. Perez Y, Gradstein L, Flusser H, et al. Isolated foveal hypoplasia with secondary nystagmus and low vision is associated with a homozygous *SLC38A8* mutation. *Eur J Hum Genet*. 2014;22:703−706.
6. Karaca EE, Cubuk MO, Ekici F, et al. Isolated foveal hypoplasia: clinical presentation and imaging findings. *Optom Vis Sci*. 2014;91(4 suppl 1):S61−S65.
7. Matsushita I, Nagata T, Hayashi T, et al. Foveal hypoplasia in patients with stickler syndrome. *Ophthalmology*. 2017;124:896−902.
8. Gale MJ, Titus HE, Harman GA, et al. Longitudinal ophthalmic findings in a child with Helsmoortel-Van der Aa syndrome. *Am J Ophthalmol Case Rep*. 2018;10:244−248.
9. Mietz H, Green WR, Wolff SM, Abundo GP. Foveal hypoplasia in complete oculocutaneous albinism. A histopathologic study. *Retina*. 1992;12:254−260.
10. Kruijt CC, de Wit GC, Bergen AA, et al. The phenotypic spectrum of albinism. *Ophthalmology*. 2018;125:1953−1960.

11. Liu T, Aguilera N, Bower AJ, et al. High-resolution imaging of cone photoreceptors and retinal pigment epithelial cells in Chediak−Higashi syndrome. *Invest Ophthalmol Vis Sci.* 2021;62:1903.

12. Gradstein L, FitzGibbon EJ, Tsilou ET, et al. Eye movement abnormalities in Hermansky−Pudlak syndrome. *J AAPOS.* 2005;9:369−378.

13. Zhou M, Gradstein L, Gonzales JA, et al. Ocular pathologic features of Hermansky−Pudlak syndrome type 1 in an adult. *Arch Ophthalmol.* 2006;124:1048−1051.

14. Azuma N, Nishina S, Yanagisawa H, et al. *PAX6* missense mutation in isolated foveal hypoplasia. *Nat Genet.* 1996;13:141−142.

15. Chauhan BK, Yang Y, Cveklova K, Cvekl A. Functional properties of natural human *PAX6* and *PAX6(5a)* mutants. *Invest Ophthalmol Vis Sci.* 2004;45:385−392.

16. Thomas S, Thomas MG, Andrews C, et al. Autosomal-dominant nystagmus, foveal hypoplasia and presenile cataract associated with a novel *PAX6* mutation. *Eur J Hum Genet.* 2014;22:344−349.

17. Langlo CS, Trotter A, Reddi HV, et al. Long-term retinal imaging of a case of suspected congenital rubella infection. *Am J Ophthalmol Case Rep.* 2022;25:101241.

18. Thiadens AA, Somervuo V, van den Born LI, et al. Progressive loss of cones in achromatopsia: an imaging study using spectral-domain optical coherence tomography. *Invest Ophthalmol Vis Sci.* 2010;51:5952−5957.

19. Thomas MG, Kumar A, Kohl S, et al. High-resolution in vivo imaging in achromatopsia. *Ophthalmology.* 2011;118:882−887.

20. Sundaram V, Wilde C, Aboshiha J, et al. Retinal structure and function in achromatopsia: implications for gene therapy. *Ophthalmology.* 2014;121:234−245.

21. Aydin R, Ozbek M, Karaman ES, Senturk F. Foveal hypoplasia in a patient with achromatopsia. *J Fr Ophtalmol.* 2018;41:e211−e214.

22. Hendrickson AE, Yuodelis C. The morphological development of the human fovea. *Ophthalmology.* 1984;91:603−612.

23. Hendrickson A, Possin D, Vajzovic L, Toth CA. Histologic development of the human fovea from midgestation to maturity. *Am J Ophthalmol.* 2012;154:767−778.e2.

24. Bringmann A, Syrbe S, Gorner K, et al. The primate fovea: structure, function and development. *Prog Retin Eye Res.* 2018;66:49−84.

25. Healey N, McLoone E, Mahon G, et al. Investigating the relationship between foveal morphology and refractive error in a population with infantile nystagmus syndrome. *Invest Ophthalmol Vis Sci.* 2013;54:2934−2939.

26. Marmor MF, Choi SS, Zawadzki RJ, Werner JS. Visual insignificance of the foveal pit: reassessment of foveal hypoplasia as fovea plana. *Arch Ophthalmol.* 2008;126:907−913.

27. Aleman TS, Han G, Serrano LW, et al. Natural history of the central structural abnormalities in choroideremia: a prospective cross-sectional study. *Ophthalmology.* 2017;124:359−373.

28. Thomas MG, Kumar A, Mohammad S, et al. Structural grading of foveal hypoplasia using spectral-domain optical coherence tomography a predictor of visual acuity? *Ophthalmology.* 2011;118:1653−1660.

29. Recchia FM, Carvalho-Recchia CA, Trese MT. Optical coherence tomography in the diagnosis of foveal hypoplasia. *Arch Ophthalmol.* 2002;120:1587−1588.

30. Holmstrom G, Eriksson U, Hellgren K, Larsson E. Optical coherence tomography is helpful in the diagnosis of foveal hypoplasia. *Acta Ophthalmol.* 2010;88:439−442.

31. Rufai SR, Thomas MG, Purohit R, et al. Can structural grading of foveal hypoplasia predict future vision in infantile nystagmus?: A longitudinal study. *Ophthalmology.* 2020;127:492−500.

32. Weiner C, Hecht I, Rotenstreich Y, et al. The pathogenicity of *SLC38A8* in five families with foveal hypoplasia and congenital nystagmus. *Exp Eye Res.* 2020;193:107958.

33. Papageorgiou E, McLean RJ, Gottlob I. Nystagmus in childhood. *Pediatr Neonatol.* 2014;55:341−351.

34. Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. In: *International Conference on Machine Learning.* PMLR; 2019.

35. Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell.* 2018;172:1122−1131. e9.

36. Karras T, Aittala M, Hellsten J, et al. Training generative adversarial networks with limited data. *Adv Neural Inf Process Syst.* 2020;33:12104−12114.

37. Ratner B. The correlation coefficient: its values range between +1/−1, or do they? *J Target Meas Anal Mark.* 2009;17:139−142.

38. Provis JM, Dubis AM, Maddess T, Carroll J. Adaptation of the central retina for high acuity vision: cones, the fovea and the avascular zone. *Prog Retin Eye Res.* 2013;35:63−81.

39. Iwasaki M, Inomata H. Relation between superficial capillaries and foveal structures in the human retina. *Invest Ophthalmol Vis Sci.* 1986;27:1698−1705.

40. Dubis AM, Costakos DM, Subramaniam CD, et al. Evaluation of normal human foveal development using optical coherence tomography and histologic examination. *Arch Ophthalmol.* 2012;130:1291−1300.

41. Lee H, Purohit R, Patel A, et al. In vivo foveal development using optical coherence tomography. *Invest Ophthalmol Vis Sci.* 2015;56:4537−4545.

42. Charbel IP, Foerl M, Helb HM, et al. Multimodal fundus imaging in foveal hypoplasia: combined scanning laser ophthalmoscope imaging and spectral-domain optical coherence tomography. *Arch Ophthalmol.* 2008;126:1463−1465.

43. Vincent A, Kemmanu V, Shetty R, et al. Variable expressivity of ocular associations of foveal hypoplasia in a family. *Eye (Lond).* 2009;23:1735−1739.

44. Querques G, Bux AV, Iaculli C, Delle NN. Isolated foveal hypoplasia. *Retina.* 2008;28:1552−1553.

45. Qin Z, Liu Z, Zhu P, Xue Y. A GAN-based image synthesis method for skin lesion classification. *Comput Methods Programs Biomed.* 2020;195:105568.

46. Finlayson SG, Lee H, Kohane IS, Oakden-Rayner L. Towards generative adversarial networks as a new paradigm for radiology education. arXiv preprint arXiv:181201547. 2018.

47. Duong D, Hu P, Tekendo-Ngongang C, et al. Neural networks for classification and image generation of aging in genetic syndromes. *Front Genet.* 2022;13:864092.

48. Chen JS, Coyner AS, Chan RP, et al. Deepfakes in ophthalmology: applications and realism of synthetic retinal images from generative adversarial networks. *Ophthalmol Sci.* 2021;1:100079.

49. Harvey PS, King RA, Summers CG. Spectrum of foveal development in albinism detected with optical coherence tomography. *J AAPOS.* 2006;10:237−242.

50. Harvey PS, King RA, Summers CS. Foveal depression and albinism. *Ophthalmology.* 2008;115:756.

51. Seo JH, Yu YS, Kim JH, et al. Correlation of visual acuity with foveal hypoplasia grading by optical coherence tomography in albinism. *Ophthalmology.* 2007;114:1547−1551.