npj | digital medicine

Article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-025-01577-3

Development and validation of a deep learning model for diagnosing neuropathic corneal pain via in vivo confocal microscopy

Neslihan Dilruba Koseoglu[1,7], Eric Chen[2,7], Rudraksh Tuwani[3], Benjamin Kompa[4], Stephanie M. Cox[1], M. Cuneyt Ozmen[1], Mina Massaro-Giordano[5], Andrew L. Beam[3,4,7] ✉ & Pedram Hamrah[1,6,7] ✉

Neuropathic corneal pain (NCP) is an underdiagnosed ocular disorder caused by aberrant nociception and hypersensitivity of corneal nerves, often resulting in chronic pain and discomfort even in the absence of noxious stimuli. Recently, microneuromas (aberrant growth and swelling of the corneal nerve endings) detected using in vivo *confocal microscopy* (IVCM) have emerged as a promising biomarker for NCP. However, this process is time-intensive and error-prone, limiting its clinical use and availability. In this work, we present a new NCP screening system based on a deep learning model trained to detect microneuromas using a multisite dataset with a combined total of 103,168 IVCM images. Our model showed excellent discriminative ability detecting microneuromas (AuROC: 0.97) and the ability to generalize to data from a new institution (AuROC: 0.90). Additionally, our pipeline provides an uncertainty quantification mechanism that allows it to communicate when its predictions are reliable, further increasing its clinical relevance.

Neuropathic corneal pain (NCP) is an underdiagnosed ocular surface disorder caused by aberrant nociception and hypersensitivity of corneal nerves. Given that the cornea is the most densely innervated tissue in the human body, patients who develop NCP experience chronic, severe pain and discomfort, even in the absence of noxious stimuli[1,2]. NCP is an ocular form of neuropathic pain, which is defined by The International Association for the Study of Pain as "pain initiated or caused by a primary lesion or dysfunction of the nervous system"[2,3]. Current understanding of the pathophysiology suggests that damaged corneal nerves initiate chronic ectopic activity in response to both noxious and innocuous stimuli following inflammation and sensitization[3]. The clinical manifestation varies in frequency and severity but commonly presents as severe, unremitting pain, burning, dryness, discomfort, and photophobia[4]. Dry or windy environments can exacerbate NCP, and patients report significant morbidity with daily activities like using computer screens or going outside during the daytime due to persistent pain[4].

At present, NCP is a diagnosis of exclusion based on clinical history, symptoms, and ophthalmologic examination, by utilizing the proparacaine challenge test or by identifying microneuromas by IVCM[3]. Patients are often first misdiagnosed with dry eye disease (DED), given both the similar clinical presentation and the fairly recent recognition of NCP[1]. Unsurprisingly, patients often fail to respond when presumptively treated for DED and continue to experience refractory pain of neuropathic origin. Lack of response to conventional DED treatment, discordant signs and symptoms, and pain out of proportion to exam all result in suspicion for NCP. While an estimated $3.84 billion is spent on DED annually in the U.S., only 37.3% of patients report being satisfied with their treatment[5]. Recent work suggests that many of these patients who receive treatment for DED instead may have NCP[3,6,7].

Accurate diagnosis has significant treatment implications. While topical treatments for DED focus on preserving and/or adding lubricants, and anti-inflammatory therapies, the main therapeutic

[1]Center for Translational Ocular Immunology and Cornea Service, New England Eye Center, Department of Ophthalmology, Tufts Medical Center, Tufts University School of Medicine, Boston, MA, USA. [2]Harvard-MIT Program in Health Sciences and Technology, Boston, MA, USA. [3]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [4]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [5]Scheie Eye Institute, Department of Ophthalmology, University of Pennsylvania, Philadelphia, PA, USA. [6]Department of Neuroscience, Tufts University School of Medicine, Boston, MA, USA. [7]These authors contributed equally: Neslihan Dilruba Koseoglu, Eric Chen, Andrew L. Beam, Pedram Hamrah. ✉e-mail: Andrew_Beam@hms.harvard.edu; pedram.hamrah@tufts.edu; p_hamrah@yahoo.com

approaches for NCP include combination of lubrication, anti-inflammatory therapies, and neuro-regenerative therapies, with peripheral or central neuromodulation[3]. In cases with peripheral NCP[3], topical anti-inflammatory therapy and topical neuro-regenerative therapies, such as autologous serum tears[8,9], cryopreserved amniotic membrane transplantation[10], and/or peripheral neuromodulators[11,12] have been shown to be successful in managing this condition. However, when patients present with centralized NCP[3], the first-line therapy shifts to oral neuromodulators, such as nortriptyline[13], gabapentin[14], low-dose naltrexone[15], and duloxetine[3].

NCP remains particularly difficult to diagnose because the abnormal nerves that drive the aberrant nociception cannot be seen on traditional slit-lamp examination[16]. The condition has therefore earned several other names, such as corneal "pain without stain"[17], "phantom cornea"[18], "corneal neuralgia"[19], or "keratoneuralgia"[18]. Many other corneal pathologies, on the other hand, typically have objective exam findings that can be visualized directly under traditional ophthalmologic slit-lamp examination. The lack of specific exam findings in NCP thus leads to improper referrals, misdiagnoses, and under-treatment. Furthermore, since these patients present with pain out of proportion to clinical examination, they are at risk of being dismissed or suspected of malingering[20]. This results in significant impact on the patients' quality of life and unnecessary suffering[13,15,21]. Thus, there is a clear clinical need for an objective and accurate method to diagnose NCP.

An avenue for accurate and objective diagnosis of NCP is via direct visualization of corneal nerves at the subbasal epithelium level. The ciliary nerves arising from the ophthalmic branch of the trigeminal ganglion enter the cornea in a radial fashion after forming a circumferential ring around the limbus and lose their myelin sheath upon entry. These unmyelinated nerve fascicles move 90° anteriorly and enter through the Bowman's layer to form the subbasal nerve plexus between Bowman's layer and basal epithelium[22]. These nerves are particularly susceptible to damage due to their proximity to the ocular surface, and damage is implicated in the pathogenesis of NCP[1,3]. NCP can be caused either by direct injury to the corneal nerves, such as chronic ocular surface diseases, surgery and abnormal post-surgical healing, infections, trigeminal neuralgia, and UV exposure, or as a result of systemic diseases, including small fiber neuropathy, fibromyalgia, and Sjögren's syndrome[1,3,23,24]. NCP can be referred to as peripheral when symptoms are a result of corneal nerve damage and associated inflammation. Over time, this pain can be centralized due to increased sensitivity of these damaged nerves, causing pain that is disconnected from the peripheral signs[3,10]. In many cases, damaged nerves can regenerate and heal properly, but ineffective and unregulated nerve regeneration can result in aberrant regeneration and swelling of the nerve endings known as microneuromas[7]. Microneuromas are often observed in the context of other nerve alterations, including changes in nerve density, tortuosity, beading, and thickness. A recent study of 71 patients demonstrated that the microneuromas can serve as an objective biomarker for the presence of NCP – microneuromas were present in all patients with NCP, but not in healthy controls or in patients with DED[7]. More recently, the FDA has accepted the presence of microneuromas as a diagnostic criterion for differentiation of NCP from DED in a randomized, placebo-controlled clinical trial (ClinicalTrials.gov ID: NCT06637527)[25].

The association of microneuromas with NCP lays the groundwork for objective diagnosis of NCP. In vivo confocal microscopy (IVCM) is a non-invasive diagnostic imaging modality that can capture the dense network of nerve fibers in the subbasal epithelium (typically in 50–80 μm depth) and visualize microneuromas[6]. IVCM is a powerful tool for the diagnosis of ocular surface diseases because of its efficiency in providing 600- to 800-fold magnification (due to 0.9 lens aperture and 63x magnification achieved via an exchangeable lens), and its ability to generate a large number of images per scan at various depths of the cornea. These images are generated in black and white, covering an area of 0.16 mm^2 (400 μm × 400 μm)[22]. Therefore, multiple scans with high volumes of images are required to adequately image the entire central cornea. As such, IVCM generates optical biopsies that visualize individual neurons at quasi-histological levels for clinical studies and diagnostics[7,20]. The application of IVCM to diagnose NCP is analogous to performing skin biopsies for systemic neuropathies[26].

IVCM has been utilized for corneal visualization and disease evaluation in other conditions, such as diabetic peripheral neuropathy, DED, and inflammatory keratitis. Previous studies have shown that metrics derived from IVCM can lead to sensitive disease detection, as in the detection of diabetic peripheral neuropathy from corneal nerve fiber length. Additionally, subbasal nerve plexus properties (count, density, length) have clinical utility in HIV, Parkinson's disease, multiple sclerosis, and diabetes—conditions that have systemic and ocular manifestations.

One of the obstacles for adoption of IVCM as a diagnostic tool is that it is difficult for ophthalmologists to review the large number of images in a robust and efficient manner[27]. A single IVCM scan produces hundreds of images during each imaging sequence, and clinical studies can require multiple scans per patient. While semi-automated software can help with basic tasks like nerve tracing, the more complex task of identifying and labeling images with microneuromas has yet to be automated. Thus, there is a clear opportunity for automated image analysis to aid in diagnosis, a task that is optimal for a deep-learning-based approach.

Analysis of IVCM images is currently slow and cumbersome. Technologies like Neuron J have accelerated the characterization of some image features like nerve density quantification[28], but microneuromas still have to be assessed manually—a process that is both time-consuming and requires significant training for human graders. Thus, resource constraints reduce the availability of IVCM-based diagnosis of NCP. Machine learning techniques are uniquely suited to address this challenge by enabling robust and rapid detection of microneuromas from a high volume of images. In particular, deep learning techniques[29,30] have demonstrated great utility in discovering complex patterns in large sets of data that cannot be efficiently evaluated by an individual or by traditional statistical modeling[31–35].

The primary innovation of this work lies in the integration of two valuable—but currently separate—technologies, to provide a new standard for the diagnosis of NCP. Current diagnostic methods for NCP rely on basic clinical history and ophthalmologic examination, and do not yet widely involve IVCM[1]. Here we present the first integration of IVCM imaging with convolutional neural networks to automatically analyze the corneal subbasal nerve plexus for nerve alterations and to identify microneuromas, which may serve as surrogate biomarkers for NCP[6]. We combine deep, convolutional neural network training to detect microneuromas with an uncertainty quantification mechanism[36] that allows the model to communicate when it is unsure about a given image to increase safety and reliability when deployed[37]. We developed the model using a cohort of patients from Tufts Medical Center (Tufts) and validated it on an external cohort from the University of Pennsylvania (UPenn). Images from both cohorts were processed by multiple expert graders from each site. As a result, the model developed in this work will allow clinicians to focus on the images with the highest diagnostic yield and provide a standardized assessment of the likelihood of NCP for each patient. Taken together, this work creates a physician-in-the-loop workflow by providing a deep learning clinical decision support system to aid ophthalmologists in screening patients at risk for NCP.

Results

Study cohort

A demographic summary of the study cohorts from Tufts Medical Center and the University of Pennsylvania is presented in Table 1.

In total, the training dataset contained 82,359 IVCM images from 51 patients seen at Tufts Medical Center with a mean visit number of 2.36 ± 2.15 [range 1–11], and 1610 (2.0%) of these images were found to contain microneuromas. A separate external validation set of 20,809 IVCM images was acquired from 100 patients (from a single visit each) seen at UPenn, and 512 (2.5%) of these images contained microneuromas.

Table 1 | Demographic summary of the Tufts and UPenn patient cohorts

	Total	Tufts	UPenn
Age, *N* (%)			
18–44 years	38	17 (33)	21 (21)
45–64 years	58	18 (35)	40 (40)
65–74 years	36	9 (18)	27 (27)
>74 years	19	7 (14)	12 (12)
Total	151	51	100
Gender, *N* (%)			
Female	127	41 (80)	86 (86)
Male	24	10 (20)	14 (14)
Total	151	51	100
Race/Ethnicity, *N* (%)			
Black non-latino	20	5 (10)	15 (15)
White non-latino	105	39 (76)	66 (66)
Other	26	7 (14)	19 (19)
Total	151	51	100
Diagnosis			
DED	61	11 (22)	50 (50)
NCP	90	40 (78)	50 (50)

Table 2 | Model performance on the internal (Tufts) and external validation data (UPenn)

	Tufts (cross-val)	UPenn (external-val)
N	82359	20809
AuROC	0.966 (0.963–0.969)	0.907 (0.893–0.92)
AuPRC	0.42 (0.393–0.444)	0.341 (0.298–0.386)
F1-score	0.351 (0.337–0.364)	0.262 (0.243–0.284)
Specificity	0.939 (0.938–0.941)	0.901 (0.897–0.905)
Sensitivity	0.862 (0.844–0.879)	0.746 (0.709–0.783)
PPV	0.22 (0.21–0.23)	0.159 (0.146–0.174)
Accuracy	0.938 (0.936–0.939)	0.897 (0.893–0.901)

Model performance results

We first evaluated the model's ability to detect the presence or absence of a microneuroma in individual IVCM images. The model was evaluated on the training data (Tufts) using a nested cross-validation procedure (Methods) and validated on an external cohort (UPenn). Performance was assessed using three metrics: area under the precision-recall curve (AuPRC), area under the receiver-operator curve (AuROC), and F1-score, and is summarized in Table 2. Notably, on the Tufts data, the model achieved an AuROC of 0.966 (95% CI: 0.963–0.969), which closely matched performance on the external validation data from UPenn (AuROC of 0.907, 95% CI: 0.893–0.920) for image-level classification, indicating strong generalization.

Visualization of the model's predictions and qualitative error analysis

Next, we provide a visualization of the model's predictions to offer insight into how the model processes the images in various scenarios. We used the integrated gradients method[38] to produce a *saliency map* for each image in the external validation dataset. A saliency map highlights regions of the image that are important to the model's prediction process. However, we note that these methods offer only qualitative insight into how a model works at a high level and can be misleading in some instances; thus, caution is needed when interpreting these kinds of saliency maps[39]. Figure 1 contains saliency maps for four prediction scenarios: true negatives, false negatives, true positives, and false positives. Inspection of the true positive saliency maps indicates reasonable behavior as the model appears to be focusing on the microneuromas present in the images. The false positive maps reveal an interesting failure case of the model as in the 2D image, it does appear that a microneuroma is present. However, these are likely dendritiform immune cells, and this distinction could only be resolved if a 3-dimensional reconstruction of the patient's cornea were performed using the full IVCM sequence.

Model performance relative to visit number

Images in the Tufts dataset were collected across a number of clinic visits for each patient, so we measured AuROC and AuPRC at each visit to assess if model performance varied over time (Fig. 2). AuROC across all visits was fairly consistent, ranging from a minimum of 0.950 to a maximum of 0.987 (Supplementary Table 1). AuPRC, on the other hand, generally improved for later clinic visits as compared with earlier visits. The initial AuPRC for the first visit was 0.457, and a maximum AuPRC of 0.78 was achieved during visit 9. Accompanying estimates and confidence intervals for AuROC, AuPRC, and other metrics calculated for each visit are found in Supplementary Table 1.

Model uncertainty evaluation

We next tested the ability of our model to automatically "abstain" from providing a diagnosis when it had sufficiently high uncertainty. In Fig. 3, model performance at the image-level (AuROC and AuPRC) was displayed for different levels of model uncertainty. A monotonically decreasing relationship between model performance and model uncertainty was observed in the Tufts dataset. On the validation dataset, the same trend generally held, though there was a decrease in AuROC and AuPRC associated with the images in the top 5th percentile of predicted probability. These results demonstrate that model uncertainty could be used to initiate human review, since high uncertainty on a given image indicates that the model is more likely to be wrong.

Discussion

In this work we demonstrate the capacity of deep convolutional neural networks to detect microneuromas from in vivo confocal images of the cornea, a key step toward the standardized and widely available diagnosis and treatment of NCP. Thus far, IVCM has provided clinicians and researchers with the capability of acquiring high-resolution images of the ocular surface at a cellular level. These images serve as "optical biopsies" with quasi-histological visualization of corneal nerve architecture and underlying abnormalities such as microneuromas, which are currently understood to play a key role in the pathophysiology of NCP. IVCM has also been utilized for corneal visualization and disease evaluation in other conditions such as diabetic peripheral neuropathy, DED, and inflammatory keratitis[22]. Previous studies have shown that metrics derived from IVCM can lead to sensitive disease detection, as in the detection of diabetic peripheral neuropathy from corneal nerve fiber length[40,41]. Additionally, subbasal plexus nerve properties (count, density, length), have clinical utility in HIV, Parkinson's disease, multiple sclerosis, and diabetes—conditions that have systemic and ocular manifestations[42,43]. The application of deep learning models to IVCM images bridges the gap between a powerful imaging modality and a diagnostic tool with clinical feasibility.

Our deep learning model achieved high performance on both internal and external validation datasets with image-level AuROCs of 0.966 and 0.904, respectively. Similarly, we observed excellent performance and generalization across a wide range of metrics and scenarios, including detection of microneuromas across subsequent patient visits. Additionally, the model is able to automatically "flag" images that could warrant human review using our proposed uncertainty quantification method. The proposed pipeline aligns well with clinical workflows, as images of low uncertainty could be

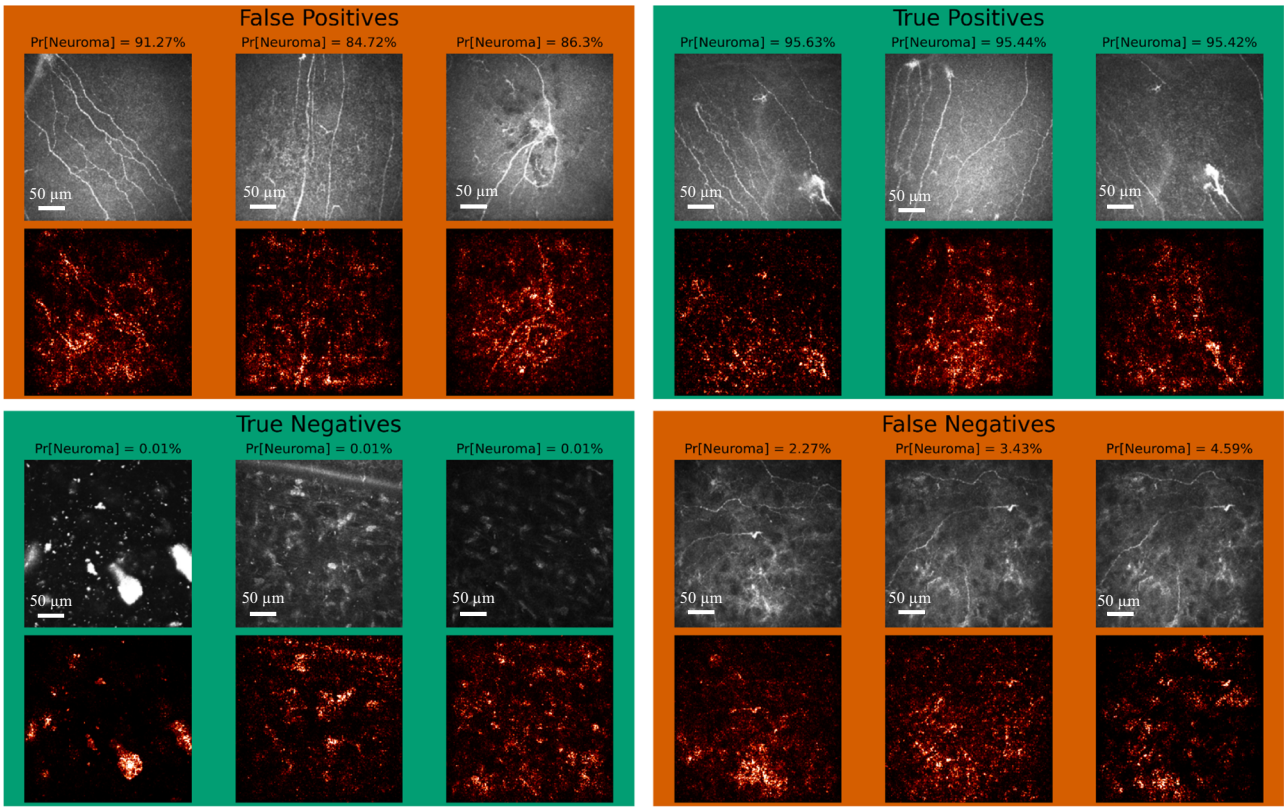

Fig. 1 | Representative saliency maps in four prediction scenarios: true negatives, false positives, false negatives, and true positives. The top row in each panel shows the original image, while the second row highlights regions of high importance to the model (brighter colors indicate increased importance). The probability that a microneuroma is present is shown above each image.

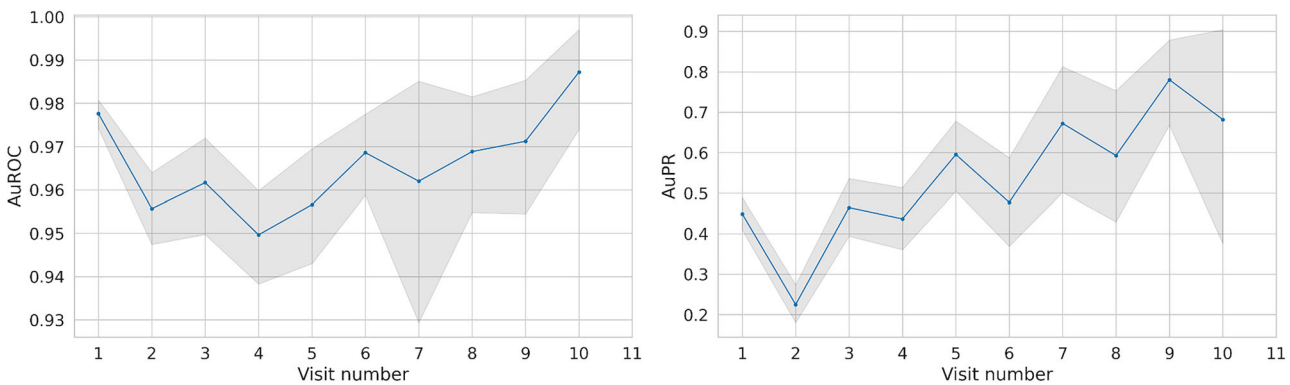

Fig. 2 | Performance of the model in predicting microneuromas. The model is able to achieve good performance across all visits. AuROC (left) and AuPRC (right) and vs. visit number in the Tufts cohort. 95% confidence intervals are shown as shaded regions in gray.

automatically handled by the model, while the few images that are difficult for the model could be triaged to a human reviewer. However, this integration into the clinical workflow will be most beneficial once the model's diagnostic performance of NCP patients is established. This clinical validation is currently underway in a multicenter, prospective clinical trial (ClinicalTrials.gov ID: NCT05653921)[44], in which clinical metrics are collected using standardized protocols and subjects are classified as DED and NCP based on these metrics by expert clinicians, in order to assess the model's patient-level diagnostic performance. More recently, the FDA has accepted the presence of microneuromas as a diagnostic criterion for differentiation of NCP from DED in a randomized, placebo-controlled clinical trial (ClinicalTrials.gov ID: NCT06637527)[25]. The work presented herein and our ongoing clinical trial lay the foundation for clinicians to be able to provide patients with rapid and accurate diagnoses of NCP, to improve our management for this underdiagnosed condition.

From a technical perspective, our deep learning model leverages several key architectural and training decisions that warrant discussion. We chose EfficientNet B3 as our base architecture due to its strong performance-to-parameter ratio and used ImageNet pre-training to improve generalization. To address the significant class imbalance in our data (only 2.0% of images containing microneuromas), we employed balanced mixup during training and implemented class weighting in our loss function. The model's uncertainty quantification mechanism, based on predicted probability thresholds, provides an important safety feature for clinical deployment by allowing automatic triage of low-confidence predictions for human review. Our saliency map analysis demonstrated that the model successfully learns

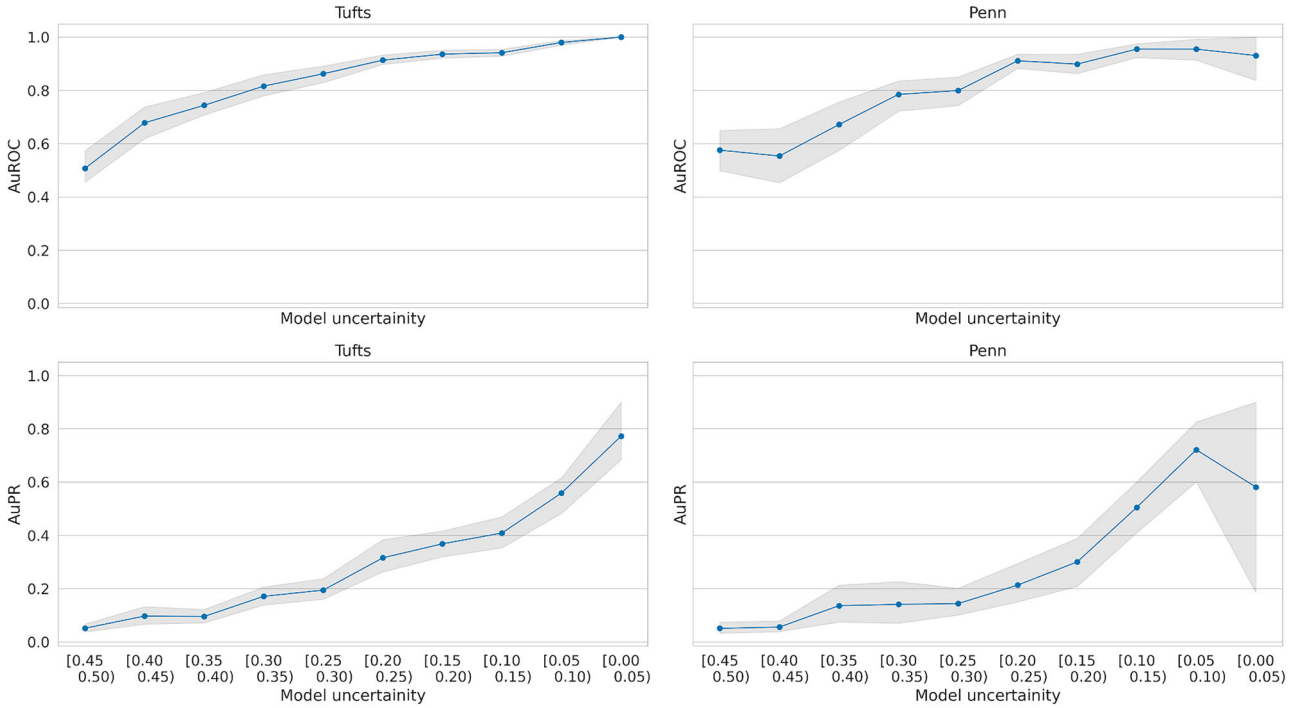

Fig. 3 | Model confidence correlates with accuracy. Model certainty is correlated with model AuROC and AuPRC. Both AuROC (top row) and AuPRC (bottom row) decrease as model uncertainty increases on each dataset.

to focus on relevant morphological features for identifying microneuromas, providing interpretable evidence for its decision-making process. However, like many deep learning models in medical imaging, our approach may be sensitive to domain shift and variations in image acquisition protocols. Future extensions of our model include moving beyond the presence or absence of microneuromas in NCP, and some recent studies indicate that the ultrastructural characteristics are important to NCP diagnosis[17]. Further, future utilization of our model in longitudinal studies of NCP patients could show that our model's performance could provide clinicians with information for prognosis and management of this condition. There are additional future benefits that arise from this work. Since other diseases like diabetes, Parkinson's disease, and HIV show evidence of corneal nerve changes via IVCM, the modeling and prediction approaches detailed in this work may be adapted for the detection and management of other diseases[42].

However, our model is not without limitations. Our cohorts include patients being screened exclusively for DED or NCP, as opposed to patients with other corneal nerve pathologies, which could limit the applicability of our approach to other diagnostic scenarios. In a preliminary study by our group, we have observed that patients with clinical signs of DED (irrespective of diagnosis) can present with microneuromas on IVCM imaging, but that patients with microneuromas suffer with symptoms of discomfort[45]. Given the clinical similarity between these two conditions, this prediction task is arguably more difficult than other corneal nerve diseases with more distinct pathologies. Nonetheless, we believe this approach would also further help in the evaluation of other ocular surface pathologies, specifically affecting the nerves, such as neurotrophic keratitis, herpetic keratitis, and post-surgical changes. Future studies in diverse range of clinical settings and patient populations will be important to facilitate broader adoption.

Another possible limitation is that our model is trained to detect microneuromas as defined by current criteria, which may continue to evolve over time. Additionally, for the purposes of this manuscript, we focused solely on identifying microneuromas at the image level. A prospective, multicentric, patient-level clinical trial is currently underway, funded by the NIH, aimed at validating the presence of microneuromas as a biomarker for NCP[44]. Furthermore, benchmarking against expert clinical assessments remains a limitation, as our model's performance has not been directly compared to comprehensive clinical evaluations by specialists in the field. The current clinical trial aims to differentiate NCP from DED by implementing clinical assessments including: ocular surface and tear film evaluation (Schirmer's test, corneal fluorescein staining and tear break up time), the Ocular Pain Assessment Survey (OPAS), a 27-question survey designed to measure pain severity and its impact of quality of life, and functional tests that assesses the pain response of NCP patients to hyperosmolar and anesthetic drops to determine whether the pain is centralized, peripheral, or mixed. A secondary outcome of this trial is determined as the correct classification of patients into NCP and DED groups using this algorithm, which would facilitate its clinical validation. Finally, prospective studies that combine both the detection of nerve abnormalities and functional tests are warranted to provide stronger evidence and more robust results regarding the clinical applicability of this algorithm. The deep learning approach itself has inherent limitations: the model's performance depends heavily on image quality and standardization of IVCM capture protocols; the 'black box' nature of deep learning makes it difficult to fully interpret how the model makes its decisions, as evidenced by our saliency map analysis showing potential confusion between microneuromas and dendritiform cells; and like all deep learning models, performance may degrade when encountering edge cases or presentations significantly different from the training data.

In summary, we have presented a new approach to aid the diagnosis of NCP that combines two cutting-edge technologies for the first time: IVCM images of microneuromas and deep learning. Our system was shown to be accurate and is capable of grading tens of thousands of images per second. To improve the clinical utility of our system, we have incorporated an uncertainty quantification system which can reliably express how certain it is about a given image. We believe the system in this work represents an important step towards widespread and accurate diagnosis of NCP.

Methods

This retrospective imaging study was conducted in adherence to the ethical standards outlined in the Declaration of Helsinki and conformed to the Health Insurance Portability and Accountability Act (HIPAA). Images were obtained from patients seen at the Cornea Service of the New England Eye

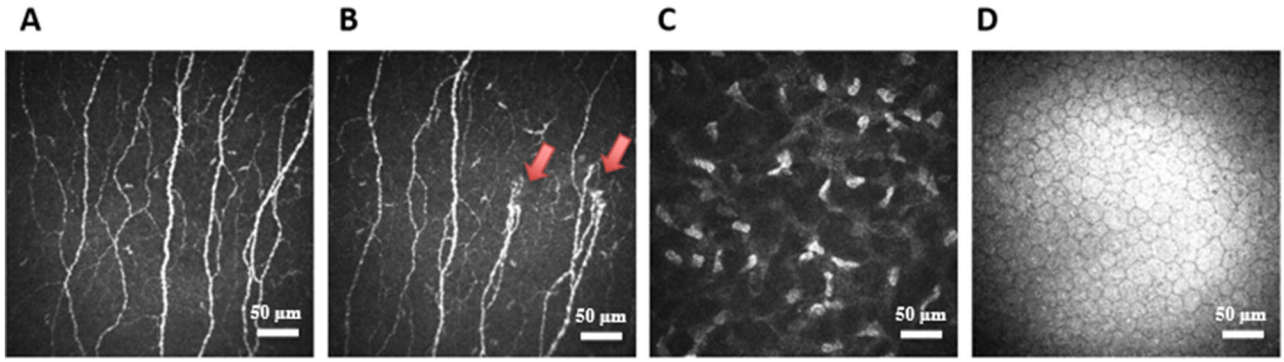

Fig. 4 | Common IVCM image types. a Normal sub-basal nerves. **b** Microneuromas (indicated with arrows) (**c**) Stroma. **d** Endothelium.

Center, Tufts Medical Center, Boston, MA, and the University of Pennsylvania Health System. The protocol for this study was approved by the Institutional Board Review of Tufts Medical Center/Tufts University Health Sciences (IRB protocol number: 12877) and University of Pennsylvania (IRB protocol number: 831744). We adhere to the TRIPOD guidelines for reporting methods and results[46]. All analyses were performed on the Harvard Medical School computing cluster, and model training was performed using 8 Titan-X GPUs in Python using the PyTorch (version 2.0) library.

Patient selection

Diagnosis of NCP was made based on the discordance of clinical signs and symptoms. Patients' primary complaints were neuropathic ocular symptoms, such as burning, stinging, photophobia and pain with an unremarkable clinical slit-lamp examination [tear break-up time (TBUT) ≥ 10 seconds or Schirmer's I score >10 at 5 min, and with or without a maximum of trace corneal fluorescein staining (CFS)] along with the presence of confirmed corneal nerve damage by IVCM. In addition, functional tests, including the use of anesthetic drops, were performed for all patients. Diagnosis of DED was based on the following criteria; symptoms of DED (foreign body sensation, burning, stinging and photophobia) for at least 6 months and 2 or more of objective signs on slit-lamp examination (TBUT < 10 seconds, Schirmer's I score<10 at 5 minutes, CFS > 3/15 and lissamine green staining of nasal and temporal conjunctiva). Exclusion criteria for both groups included any other ocular pathologies that could cause symptoms of DED or ocular pain including corneal infections and abrasions, recurrent erosion syndrome, iridocyclitis or uveitis, and ocular surgery. The patients were identified and verified by one observer at each institution, and both eyes of each eligible patient were included in the final analyses (PH and MMG).

IVCM imaging and data. IVCM images were obtained from a mix of patients with DED and NCP from both Tufts and UPenn. The central cornea was scanned with laser in vivo confocal microscopy, using the Heidelberg Retinal Tomograph (HRT-III) with a Rostock Cornea Module (Heidelberg Engineering GmBH, Heidelberg, Germany). All images were 400 µm x 400 µm in dimension, and all were in grayscale. The total number of images depended on the number of scans the patients received and the number of visits. Two trained ophthalmologists classified each image as subbasal nerve plexus, epithelium, or stroma (Fig. 4), and they also noted whether images contained microneuromas. All images were annotated by multiple clinical experts independently; discordance was adjudicated by another reviewer to produce the final training and validation datasets.

Neural network specifications and pre-processing. Neural network pre-training has demonstrated significant performance gains across an array of computer vision tasks[47]. The EfficientNet B3 neural network pre-trained on the ImageNet dataset[48] augmented with a scalar-valued linear layer (with randomly initialized weights) was selected as the base configuration for the NCP deep learning model. The IVCM images were transformed to be compatible with the pre-trained network. Each pixel value of the grayscale IVCM images was duplicated thrice as the pre-trained network expects RGB images as input. The images were further normalized to have channel-wise mean (0.485, 0.456, 0.406) and standard deviation (0.229, 0.224, 0.225) to match the training distribution statistics of the pre-trained network. The model was trained using AdamW optimizer with a learning rate of 1e-3 and weight decay of 0.01. We employed binary cross-entropy loss with class weighting to handle potential class imbalance. The training pipeline includes validation-based early stopping monitoring AUROC, and comprehensive metrics tracking including precision, recall, F1-score, specificity, AUROC, and average precision. The implementation uses PyTorch Lightning for standardized training procedures and reproducibility.

Data augmentation. Neural networks work best when they train on input that comprises a diversity of data. The standard for machine learning is to leverage a data augmentation strategy, in order to apply various transformations on the original image set, and thereby generate images that are not explicitly specified in the training set. Such data augmentation increases the robustness of deep neural networks, decreases its susceptibility to adversarial attack, and increases its generalizability. The data augmentation strategy for training the neural network consisted of random horizontal flipping of input image (with probability = 0.5), followed by a random affine transformation while keeping the center invariant (with range of degrees from −30 to 30 and maximum absolute fraction for horizontal and vertical translations set to 0.2).

Model training and validation. The pre-trained neural network was fine-tuned for predicting microneuromas by optimizing the binary cross entropy loss with a stochastic gradient descent optimizer (learning rate of 0.005, momentum of 0.9, batch size of 192, and weight decay of 0.01) and a one cycle learning rate scheduler with cosine annealing. More details and the code used to train the model can be found at this repository: https://github.com/beamlab-hsph/ncp_prediction.

As the proportion of images with microneuromas was significantly less than those without, balanced mixup (mixup proportion of 0.1) was used to tackle both the extreme class imbalance and model overfitting of microneuroma images[49]. The neural network was fine-tuned for 3 epochs on an Nvidia A100 GPU.

The in-distribution performance of the neural network model was assessed using leave-one-out cross-validation at the patient level on the Tufts dataset. Specifically, in each iteration of cross-validation, the images from all patients except one were used to train the model, and the images corresponding to the left-out patient were used to assess the model[50]. This resulted in a total of 51 trained models. The exact number of images in each of these 51 held-out folds will differ since each patient has a unique number of images. The out-of-distribution performance was estimated using the

UPenn dataset and the average of probability predictions from the trained models. Model performance was assessed at the patient level by considering the top-9 images with the highest predicted probability for neuroma.

Model testing. The dataset for testing was a separate set of 20,809 IVCM images, obtained from the University of Pennsylvania. This image set was also independently labeled by two expert annotators to identify the images with and without microneuromas. The performance of this trained model was evaluated according to the statistical measures, detailed below. Performance at the patient level was determined using the top-9 image probabilities for each patient. That is, each patient's aggregate probability was defined as the average of the top-9 images with the highest probabilities for the patient, or the top-9 images most likely to contain microneuromas.

Uncertainty quantification mechanisms

In machine learning, the practice of model abstention refers to the possibility of refusing a prediction in cases of uncertainty. In effect, model abstention is important in cases where model performance is expected to differ significantly with respect to model certainty. To implement an abstention mechanism in our model, images were separated into incremental bins of width 5% based on top-label predicted probability, and associated performance metrics were calculated. Intuitively, values closer to 100% for top-label predicted probabilities suggest that the model is confident about the top-label prediction, whereas values closer to 50% indicate that model is not very sure. The performance metrics for each bin can be used to select an abstention threshold for top-label predicted probability above which the desired performance is expected to be maintained at the image level.

Statistical measures

Performance of the deep neural network during training and validation was measured with several standard metrics, including: Specificity: TN/(TN + FP), Sensitivity = TP/(TP + FN), Accuracy = (TP + TN)/Total Population, and Precision = TP/(TP + FP), in addition to other metrics like AuROC, AuROC Precision-Recall, and PPV.

Data availability

The data in this study cannot be shared publicly due to privacy concerns but are available upon request from the investigators.

Code availability

The code used to train the model and produce the results in this manuscript may be found here: https://github.com/beamlab-hsph/ncp_prediction.

Received: 19 April 2024; Accepted: 20 March 2025;
Published online: 14 May 2025

References

1. Goyal, S. & Hamrah, P. Understanding neuropathic corneal pain–gaps and current therapeutic approaches. *Semin Ophthalmol.* **31**, 59–70 (2016).
2. Galor, A. et al. Neuropathic pain and dry eye. *Ocul. Surf.* **16**, 31–44 (2018).
3. Dieckmann, G., Goyal, S. & Hamrah, P. Neuropathic corneal pain: approaches for management. *Ophthalmology* **124**, S34–s47 (2017).
4. Borsook, D. & Rosenthal, P. Chronic (neuropathic) corneal pain and blepharospasm: five case reports. *Pain* **152**, 2427–2431 (2011).
5. Yu, J., Asche, C. V. & Fairchild, C. J. The economic burden of dry eye disease in the United States: a decision tree analysis. *Cornea* **30**, 379–387 (2011).
6. Moein, H. -R. et al. In vivo confocal microscopy demonstrates the presence of microneuromas and may allow differentiation of patients with corneal neuropathic pain from dry eye disease. *Investig. Ophthalmol. Vis. Sci.* **58**, 2656–2656 (2017).
7. Moein, H. R. et al. Visualization of microneuromas by using in vivo confocal microscopy: An objective biomarker for the diagnosis of neuropathic corneal pain?. *Ocul. Surf.* **18**, 651–656 (2020).
8. Aggarwal, S. et al. Autologous serum tears for treatment of photoallodynia in patients with corneal neuropathy: efficacy and evaluation with in vivo confocal microscopy. *Ocul. Surf.* **13**, 250–262 (2015).
9. Aggarwal, S., Colon, C., Kheirkhah, A. & Hamrah, P. Efficacy of autologous serum tears for treatment of neuropathic corneal pain. *Ocul. Surf.* **17**, 532–539 (2019).
10. Morkin, M. I. & Hamrah, P. Efficacy of self-retained cryopreserved amniotic membrane for treatment of neuropathic corneal pain. *Ocul. Surf.* **16**, 132–138 (2018).
11. Olcucu, O., de Leeuw, A., Lamazales, L. L., Mallone, F. & Hamrah, P. Efficacy of extranasal neurostimulation for patients with neuropathic corneal pain: a pilot study. *Cornea*, 10.1097/ICO.0000000000003719 https://doi.org/10.1097/ico.0000000000003719 (9900).
12. Mehra, D. et al. Long-term trigeminal nerve stimulation as a treatment for ocular pain. *Neuromodulation* **24**, 1107–1114 (2021).
13. Ozmen, M. C. et al. Efficacy and tolerability of nortriptyline in the management of neuropathic corneal pain. *Ocul. Surf.* **18**, 814–820 (2020).
14. Yoon, H. J., Kim, J. & Yoon, K. C. Treatment response to gabapentin in neuropathic ocular pain associated with dry eye. *J. Clin. Med.* **9** https://doi.org/10.3390/jcm9113765 (2020).
15. Dieckmann, G., Ozmen, M. C., Cox, S. M., Engert, R. C. & Hamrah, P. Low-dose naltrexone is effective and well-tolerated for modulating symptoms in patients with neuropathic corneal pain. *Ocul. Surf.* **20**, 33–38 (2021).
16. Rosenthal, P., Baran, I. & Jacobs, D. S. Corneal pain without stain: is it real?. *Ocul. Surf.* **7**, 28–40 (2009).
17. Guerrero-Moreno, A. et al. Corneal nerve abnormalities in painful dry eye disease patients. *Biomedicines* **9** https://doi.org/10.3390/biomedicines9101424 (2021).
18. Moshirfar, M. & Benstead, E. E., Sorrentino, P. M. & Tripathy, K. in *StatPearls* (StatPearls Publishing Copyright © 2023, StatPearls Publishing LLC., 2023).
19. Theophanous, C., Jacobs, D. S. & Hamrah, P. Corneal neuralgia after LASIK. *Optom. Vis. Sci.* **92**, e233–e240 (2015).
20. Ross, A. R. et al. Clinical and in vivo confocal microscopic features of neuropathic corneal pain. *Br. J. Ophthalmol.* **104**, 768–775 (2020).
21. Bayraktutar, B. N. et al. Comparison of clinical characteristics of post-refractive surgery-related and post-herpetic neuropathic corneal pain. *Ocul. Surf.* **18**, 641–650 (2020).
22. Cruzat, A., Qazi, Y. & Hamrah, P. In vivo confocal microscopy of corneal nerves in health and disease. *Ocul. Surf.* **15**, 15–47 (2017).
23. D'Souza, S., Khamar, P. & Shetty, R. Fibromyalgia syndrome and the eye-Implications in corneal ultrastructure on confocal microscopy. *Indian J. Ophthalmol.* **71**, 1656–1657 (2023).
24. Rosenthal, P. & Borsook, D. Ocular neuropathic pain. *Br. J. Ophthalmol.* **100**, 128–134 (2016).
25. The Study Assessing the Safety and Efficacy of OK-101 Treatment in Subjects With Neuropathic Corneal Pain, https://clinicaltrials.gov/study/NCT06637527?cond=neuropathic%20corneal%20pain%20&term=OK-101&rank=1.
26. Farhad, K. Current diagnosis and treatment of painful small fiber neuropathy. *Curr. Neurol. Neurosci. Rep.* **19**, 103 (2019).
27. Wei, S., Shi, F., Wang, Y., Chou, Y. & Li, X. A deep learning model for automated sub-basal corneal nerve segmentation and evaluation using in vivo confocal microscopy. *Transl. Vis. Sci. Technol.* **9**, 32 (2020).
28. Meijering, E. et al. Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. *Cytom. A* **58**, 167–176 (2004).

29. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
30. Hinton, G. Deep learning-a technology with the potential to transform health care. *Jama* **320**, 1101–1102 (2018).
31. Ben-David, S. & Shalev-Shwartz, S. *Understanding Machine Learning: From Theory to Algorithms* (Cambridge University Press, 2014).
32. Beam, A. L. & Kohane, I. S. Big data and machine learning in health care. *Jama* **319**, 1317–1318 (2018).
33. Schmaltz, A. & Beam, A. L. Sharpening the resolution on data matters: a brief roadmap for understanding deep learning for medical data. *Spine J.* **21**, 1606–1609 (2021).
34. Beaulieu-Jones, B. et al. Trends and focus of machine learning applications for health research. *JAMA Netw. Open* **2**, e1914051 (2019).
35. Yu, K. H., Beam, A. L. & Kohane, I. S. Artificial intelligence in healthcare. *Nat. Biomed. Eng.* **2**, 719–731 (2018).
36. Kompa, B., Snoek, J. & Beam, A. L. Second opinion needed: communicating uncertainty in medical machine learning. *NPJ Digit. Med.* **4**, 4 (2021).
37. Kompa, B., Snoek, J. & Beam, A. L. Empirical frequentist coverage of deep learning uncertainty quantification procedures. *Entropy* **23** https://doi.org/10.3390/e23121608 (2021).
38. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic attribution for deep networks. In *Proc. of the 34th International Conference on Machine Learning*, Sydney, Australia, PMLR, Vol. 70, 3319–3328 (2017).
39. Ghassemi, M., Oakden-Rayner, L. & Beam, A. L. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit. Health* **3**, e745–e750 (2021).
40. Wang, M. et al. Diagnostic utility of corneal confocal microscopy in type 2 diabetic peripheral neuropathy. *J. Diab. Investig.* **12**, 574–582 (2021).
41. Hossain, P., Sachdev, A. & Malik, R. A. Early detection of diabetic peripheral neuropathy with corneal confocal microscopy. *Lancet* **366**, 1340–1343 (2005).
42. Oakley, J. D. et al. Deep learning-based analysis of macaque corneal sub-basal nerve fibers in confocal microscopy images. *Eye Vis.***7**, 27 (2020).
43. Mikolajczak, J. et al. Patients with multiple sclerosis demonstrate reduced subbasal corneal nerve fibre density. *Mult. Scler.* **23**, 1847–1853 (2017).
44. Prospective Study to Validate the Imaging Biomarker for NCP (R33).
45. Koseoglu, N. D. et al. Patients with clinical signs of dry eye disease demonstrate presence of signs of neuropathic corneal pain. *Investig. Ophthalmol. Vis. Sci.* **60**, 4199–4199 (2019).
46. Collins, G. S. et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open* **11**, e048008 (2021).
47. Girshick, R. B., Donahue, J., Darrell, T. & Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 580–587 (2013).
48. Tan, M. & Le, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. ArXiv abs/1905.11946 (2019).
49. Zhang, H., Cissé, M., Dauphin, Y. & Lopez-Paz, D. mixup: Beyond Empirical Risk Minimization. *ArXiv* abs/1710.09412 (2017).
50. Wong, T. -T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognit.* **48**, 2839–2846 (2015).

Acknowledgements

Funding Sources: NIH R61NS113341. Burton G. Bettingen Corporation Research to Prevent Blindness Challenge Grant to the Department of Ophthalmology Mass. Lions Eye Research Fund Inc. Lions Club International Foundation.

Author contributions

N.D.K.*: Data collection, data analyses, manuscript writing. E.C.*: Model development and training, data analyses, manuscript writing. R.T.: Model development and training, data analyses, manuscript writing. B.K.: Data analyses, manuscript writing. S.C.: Data analyses, manuscript writing. C.O.: Data analyses. M.M.G.: Data collection. A.L.B.*: Model development and training, data analyses, manuscript writing. P.H.*: Data collection, manuscript writing.

Funding

NIH R61NS113341. Burton G. Bettingen Corporation Research to Prevent Blindness Challenge Grant to the Department of Ophthalmology Mass. Lions Eye Research Fund Inc. Lions Club International Foundation.

Competing interests

The authors declare no competing interests. Patent applicant: Harvard College, Tufts Medical Center. Name of inventors: Pedram HAMRAH, Neslihan Dilruba KOSEOGLU, Andrew BEAM. Patent number: US20210113078A1.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41746-025-01577-3.

Correspondence and requests for materials should be addressed to Andrew L. Beam or Pedram Hamrah.