



SNU
MEDICINE

Seoul National University
College of Medicine

Deep learning-based fully automated grading system for dry eye disease severity

(Presentation No.99692)

Gahye Lee, MD¹, Seonghwan Kim^{1,2}, Daseul Park^{3,4}, Youmin Shin^{3,4}, Mee Kum Kim^{1,2}, Hyun Sun Jeon^{1,5}, Young-Gon Kim³, Chang Ho Yoon^{1,2}

Department of Ophthalmology, Seoul National University College of Medicine, Seoul, the Republic of Korea¹;
Laboratory of Ocular Regenerative Medicine and Immunology, Seoul Artificial Eye Center, Seoul National University Hospital Biomedical Research Institute, Seoul, the Republic of Korea²;
Transdisciplinary Department of Medicine & Advanced Technology, Seoul National University Hospital, Seoul, the Republic of Korea³;
Interdisciplinary Program in Bioengineering, Graduate School, Seoul National University, Seoul, the Republic of Korea⁴;
Department of Ophthalmology, Seoul National University Bundang Hospital, Seongnam, the Republic of Korea⁵

The authors have no financial disclosures

Introduction

- Dry eye disease (DED)
 - The prevalence of symptomatic dry eye: 5-33% in adult populations
 - In the US, the cost of managing and treating dry eye patients: \$700,000 per million patients
 - Global dry eye products market: 5.22 billion USD in 2019
 - The severity of DED grading
 - The degree of corneal fluorescein staining (CFS) with various methods
 - A. NEI scale is one of the commonly used scales in clinical trials
 - Most CFS scales including National Eye Institute (NEI) scale: subjective, observer-dependent
- A reproducible, reliable and objective method is needed to minimize the subjective bias of human observers.

Purpose

- To develop a clinically applicable fully automated deep learning-based system for the assessment of dry eye severity according to the NEI scale to minimize subjective bias from human observers

Methods

- Cobalt blue filtered anterior segment images (IRB No. 2205-162-1328)
 - Seoul National University Hospital (SNUH): model establishment (1300 cases), serial data (100 cases)
 - Seoul National University Bundang Hospital (SNUBH): external validation (94 cases)
 - The grading system recommended by the NEI
 - Graded by three experts and the median value was used as ground truth

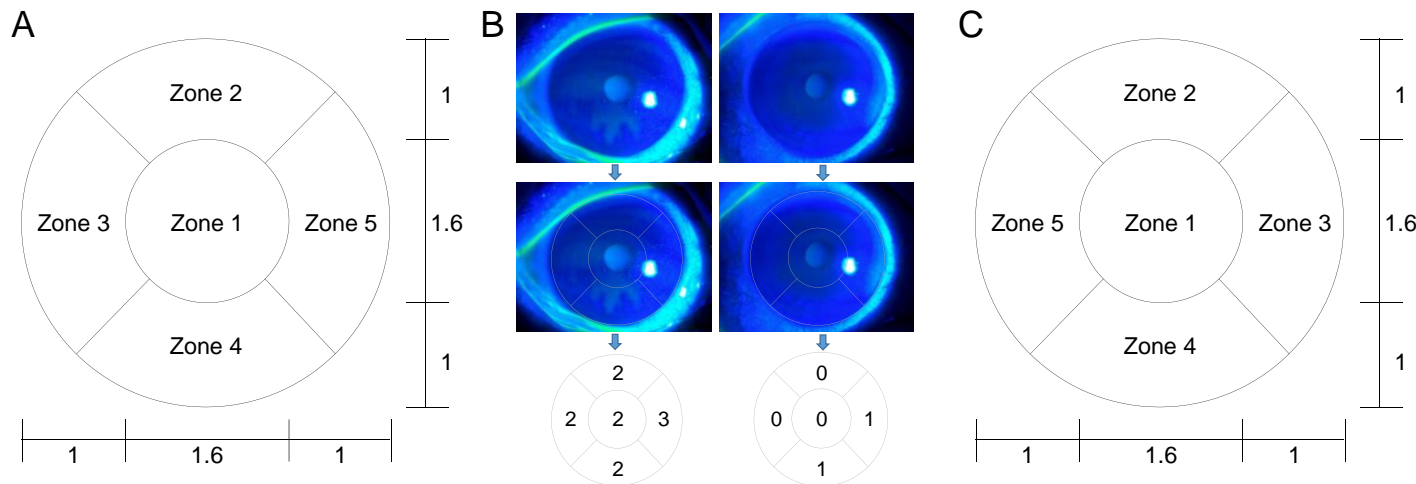


Figure. Corneal segmentation and scoring method.
A. Corneal segmentation grid and proportion (right eye).
The horizontal and vertical ratios of each zone of the grid are the same at 1:1.6:1 as in Amparo et al.
B. Two examples of NEI scale evaluation. PEE of the five zones is assessed and scored using the NEI scale.
C. Corneal segmentation grid and proportion (left eye).
NEI=National Eye Institute; PEE=punctate epithelial erosion

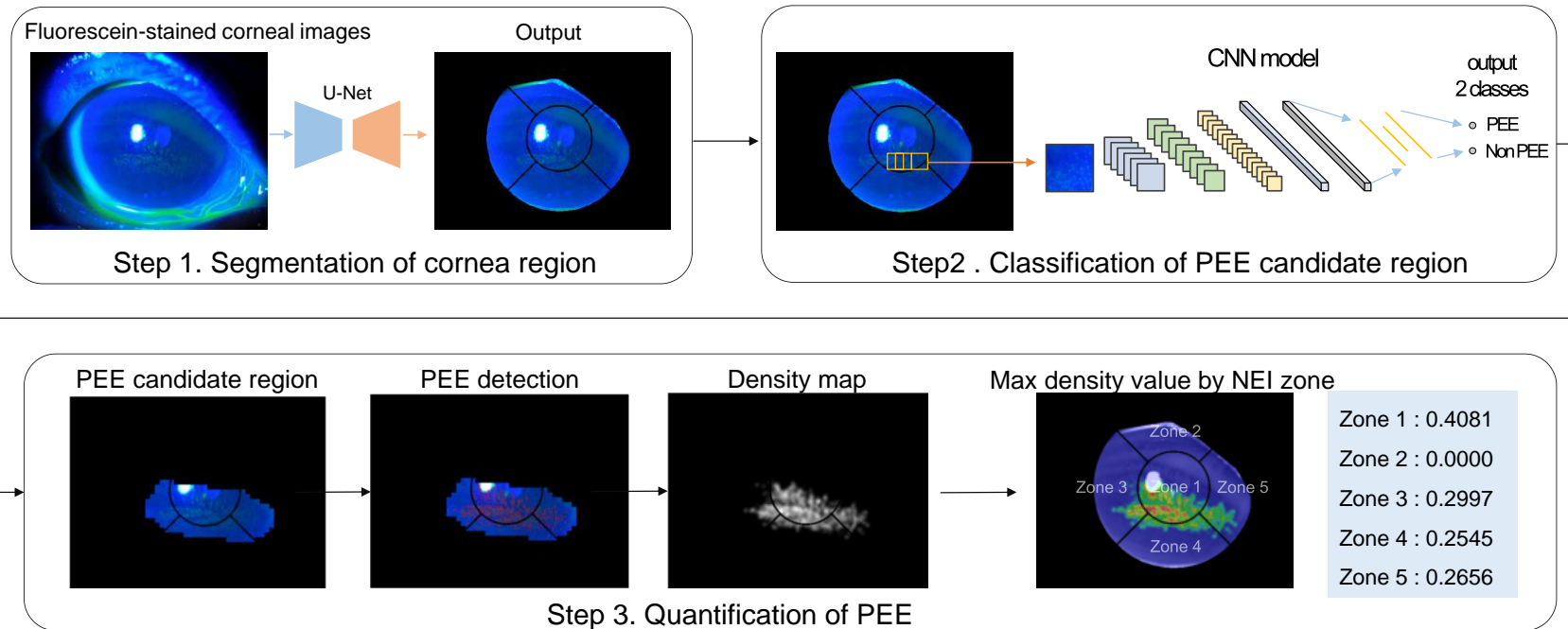
Methods

- AI system development

- (1) Corneal segmentation using U-Net architecture

- (2) CFS candidate region classification

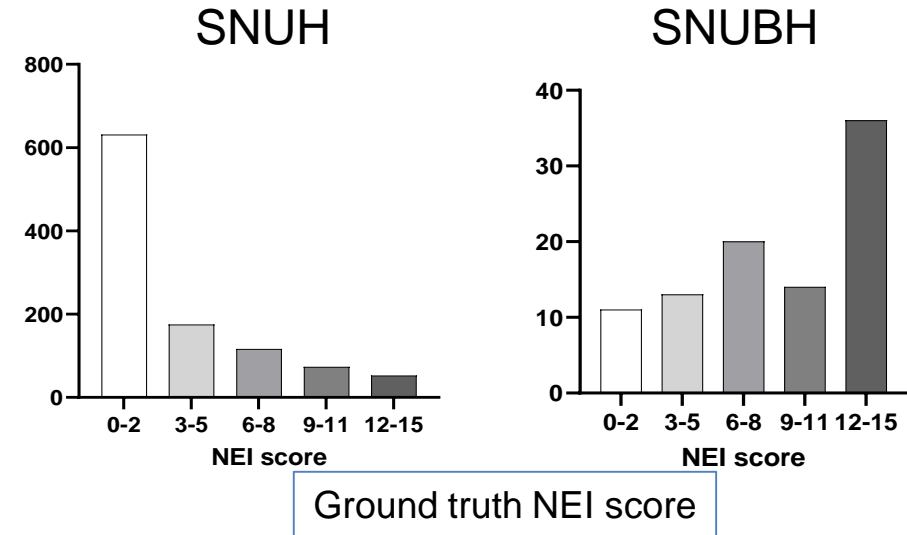
- (3) Quantification of punctate epithelial erosion (PEE) using PEE density map and presentation of max density value (MDV)



Results

- Patient and image characteristics

Diagnosis	SNUH (N= 1100)	SNUBH (N=94)
Sjögren syndrome	113 (10.3%)	65 (69.1%)
Ocular GVHD	15 (1.4%)	0 (0%)
Others	972 (88.4%)	29 (30.9%)



- The clinical score agreement among the three investigators (INVs)

CFS score by NEI scale	INV 1 vs. INV 2	INV 3 vs. INV 3	INV 3 vs. INV 3
Pearson's correlation coefficient (95% CI)	0.905 (0.840-0.944)	0.903 (0.838-0.943)	0.934 (0.888-0.961)
Bland-Altman mean difference (P value)	-0.111 (0.135)	-0.370 (<0.001)*	-0.259 (<0.001)*
SD Difference	0.538	0.469	0.483
Concordance correlation coefficient (95% CI)	0.895 (0.828-0.936)	0.874 (0.806-0.919)	0.876 (0.800-0.925)

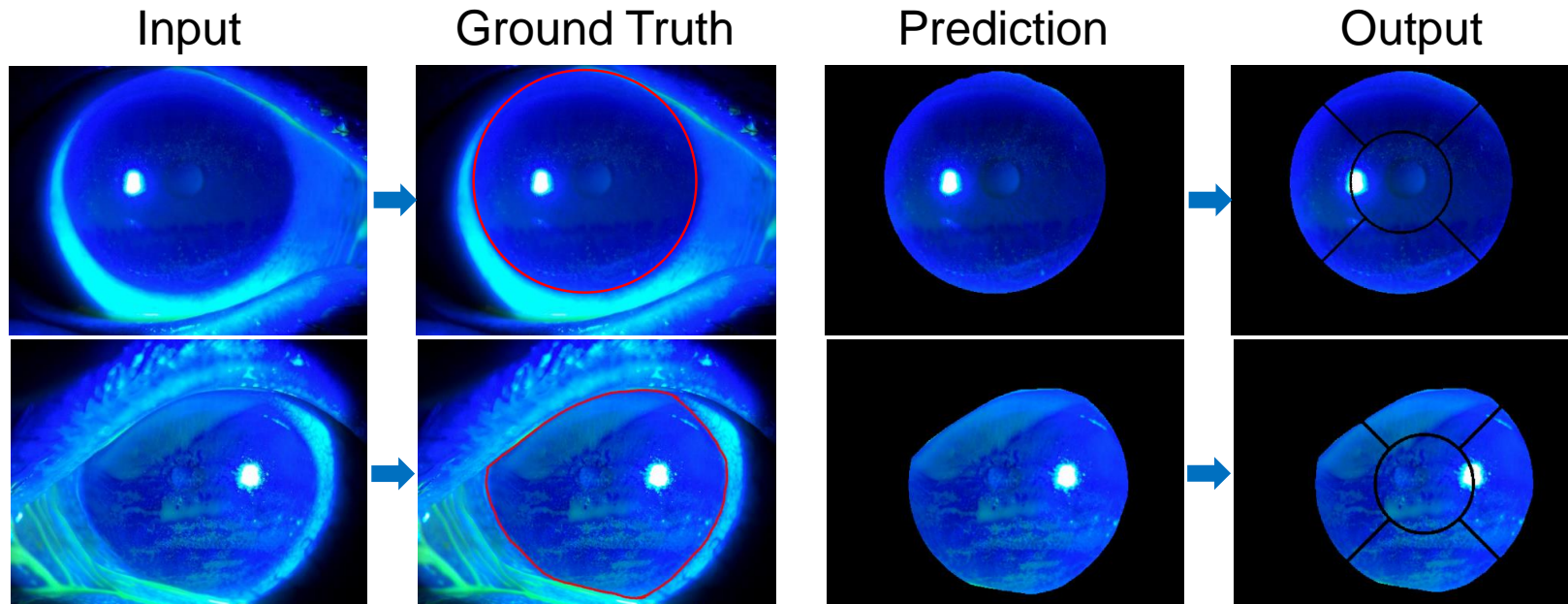
* Differences between investigators are significantly greater than zero ($P < 0.05$).

The correlation of the CFS score among the investigators is higher than 0.9 by Pearson's correlation coefficient.

CI, confidence interval; SD, standard deviation.

Results: Corneal segmentation

- The Dice coefficient
 - An index used to evaluate the similarity between two areas
 - Calculated by doubling the size of the shared area and dividing by the sum of the sizes of the two areas
 - Our model: 0.962



Results: PEE classification model

- To reduce false positives mimicking PEE
 - At a threshold of 0.98: accuracy = 0.89, sensitivity = 0.82, specificity = 0.96, AUC = 0.97
 - Indicating its robust performance

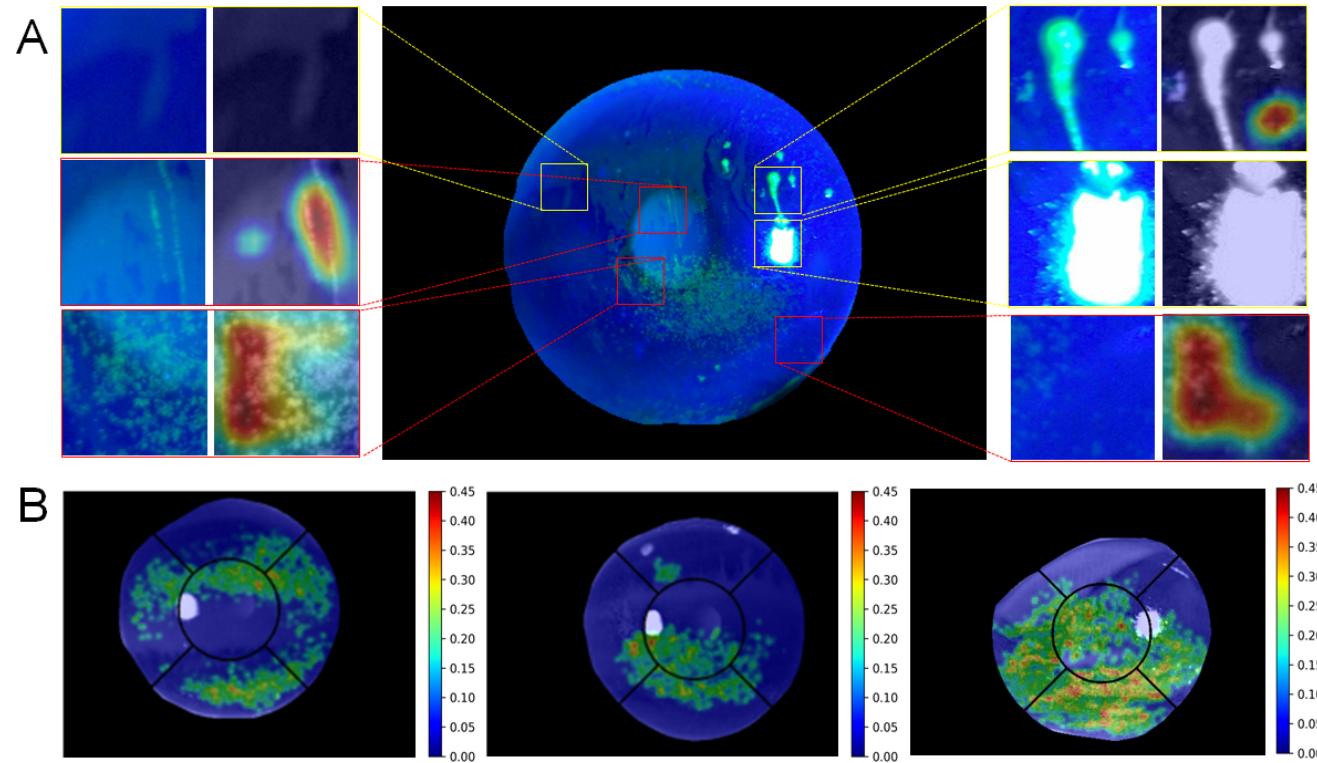
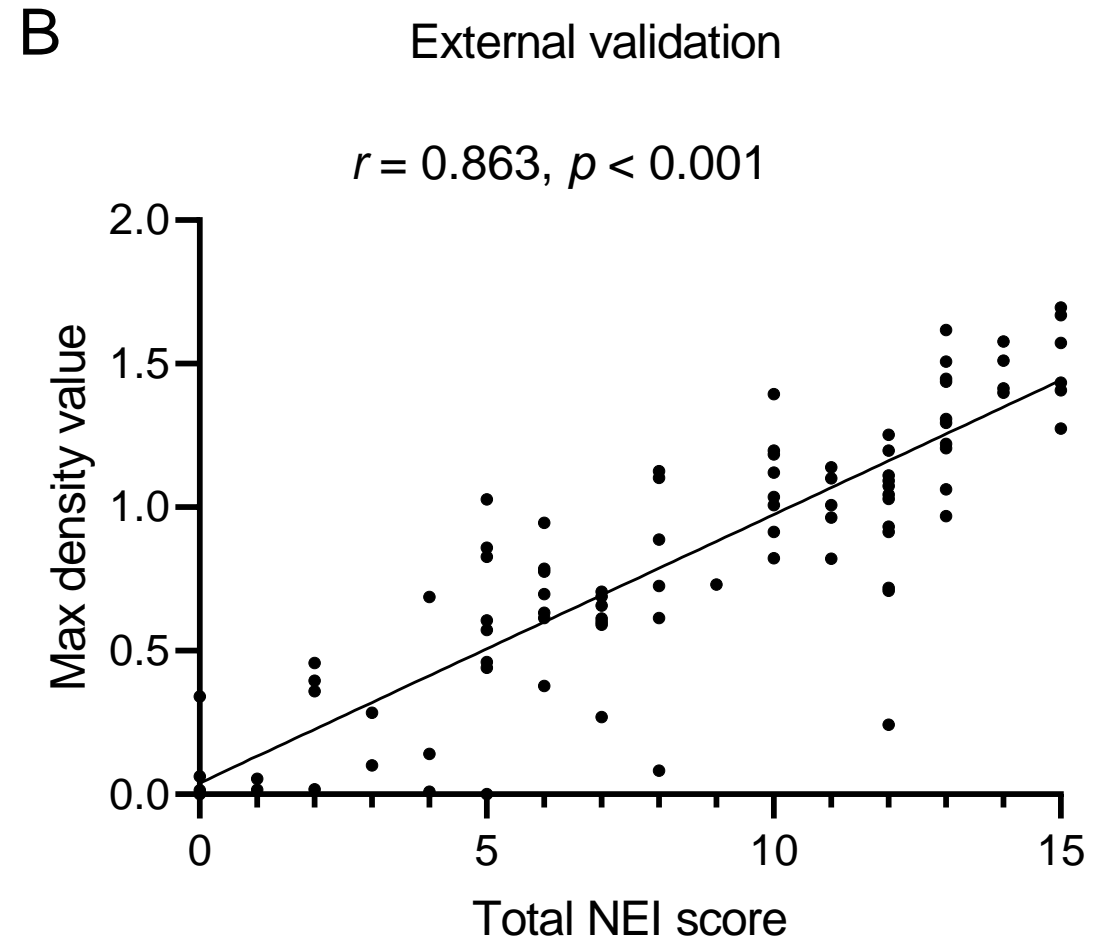
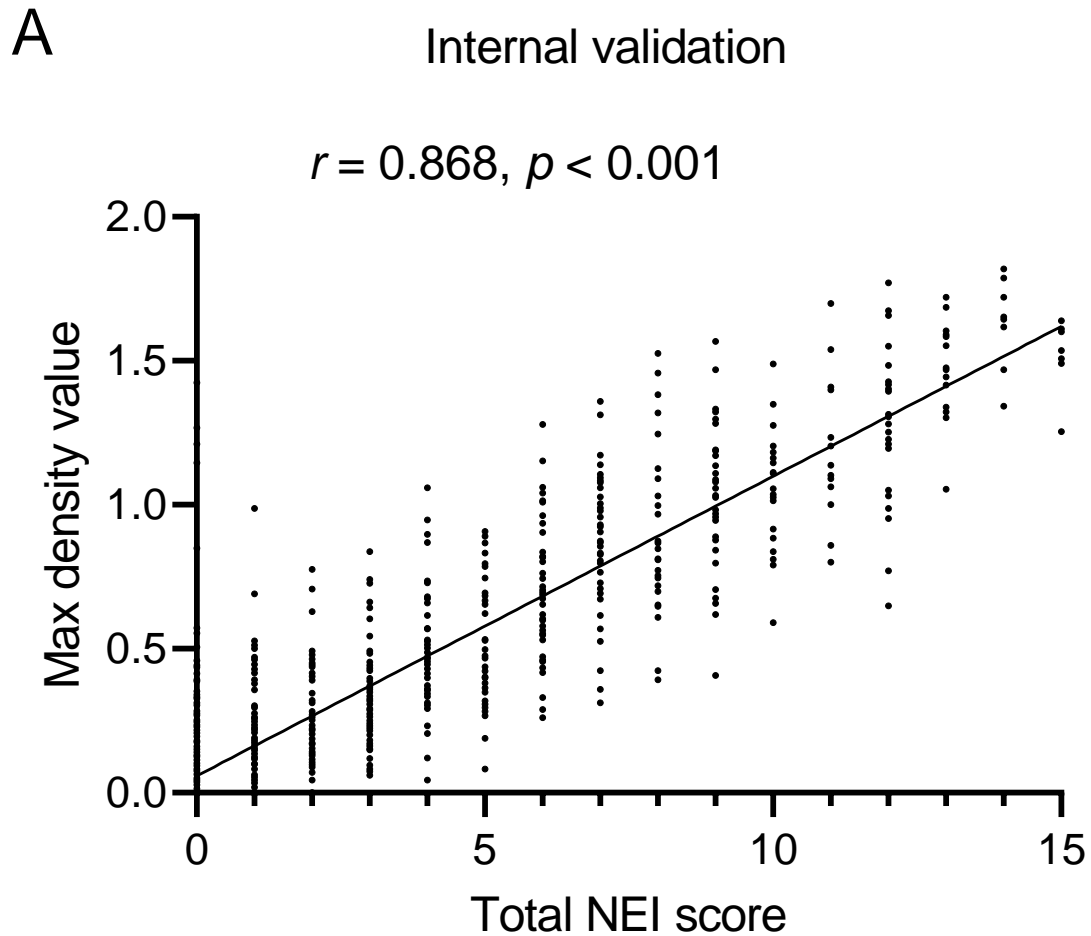


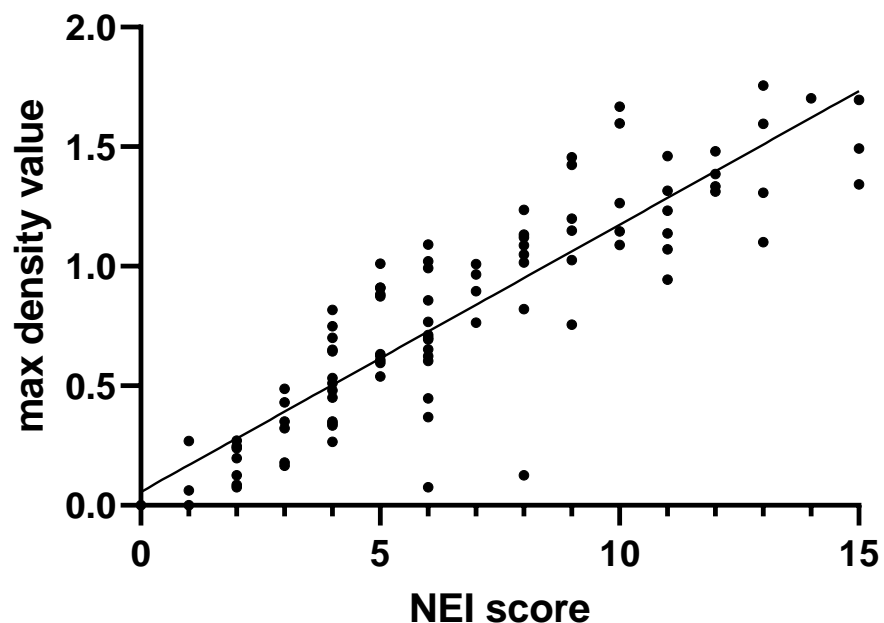
Illustration of classification model and density map results. A. CAM of PEE candidate region classification. The yellow and red boxes represent TP and TN of the PEE classification model, respectively. B. Density map results. Blue indicates low PEE density, and red indicates high PEE density. CAM=class activation map; TP=true positive; TN=true negative; PEE=punctate epithelial erosion

Results: Internal and external validation



The Spearman correlation between MDV and ground truth NEI score was 0.868 in the internal validation datasets (Figure A). The Spearman correlation between MDV and ground truth NEI score was 0.863 in the external validation dataset (Figure B).

Results: Improvement or deterioration evaluation



Total PCC result : **0.9018**

	Δ Prediction (negative)	Δ Prediction (positive)
Δ GT (negative)	19	2
Δ GT (positive)	4	25

The agreement rate for improvement or deterioration between the proposed model and ground truth was 88% (44/50 patients).

Conclusion

- Fully automated deep learning-based dry eye severity evaluation system
 - Automated corneal region segmentation (step 1)
 - PEE candidate region classification (step 2)
 - PEE detection & quantification (step 3)
- evaluate CFS score based on NEI scale with high accuracy
- ⇒ expected to be applied in clinical settings or animal experiment in the future.