Abstract 42

REAL WORLD PERFORMANCE OF AI, HUMAN AND HYBRID SCREENING SYSTEMS FOR DIABETIC RETINOPATHY

Oral

Leng T.*

Byers Eye Institute at Stanford, Stanford University School of Medicine ~ Palo Alto ~ United States of America

Purpose:

To evaluate a commercial artificial intelligence (AI) system and human-based grading, as well as a hybrid model on their ability to detect more than moderate non-proliferative diabetic retinopathy (MTMDR) in a primary care clinic setting.

Methods:

Primary care patients were screened with remote teleophthalmology at seven sites over 2.5 years. During the teleophthalmology phase, single-reader evaluation was used to classify images. During the AI phase, an FDA-cleared device (IDx-DR, Digital Diagnostics), cloud-based AI models assessed image quality prior to clinical evaluation. If the image was gradable, the AI indicated the presence or absence of MTMDR, and/or DME, in at least one eye. Patients who received a MTMDR or ungradable result were referred to an ophthalmologist for in-person examination. Some patients participated in in-person examinations and those exams were used to validate the performance of both workflows.

Results:

2,012 exams were performed (790 teleophthalmology, 1222 AI). Images were gradable in 90.3% of teleophthalmology and 62.5% of AI. Gradable encounters resulted in MTMDR in 5.6% of teleophthalmology and 19% of AI. Overreads of AI images compared to in-person had 91.8% accuracy with 69.5% sensitivity and 96.9% specificity. Moreover, AI had a 70.0% accuracy with a 95.5% sensitivity and 60.3% specificity when compared to in-person. When the DR stages were expanded beyond MTMDR-positive or -negative, the agreement between retina specialist overread and in-person examination was 83.3%, and 96.5% of encounter diagnoses were within one DR stage of each other.

Conclusions:

While both teleophthalmology and AI-based screening approaches demonstrate high accuracy, the sensitivity of AI exceeded that of the retina specialists, whereas the specificity of retina specialists exceeded the AI, possibly caused by imaging artifacts.