Artificial Intelligence in Dry Eye Disease: Benefits, Challenges and Future Directions

Carolyn YTW[1,2,3]*, Timing L[1,2], Tin LW[3], Raina JX[4] and Henry HWL[5]

[1]Institute of Ophthalmology, University College London, United Kingdom

[2]Moorfields Eye Hospital NHS Foundation Trust, United Kingdom

[3]Faculty of Medicine, The Chinese University of Hong Kong, China

[4]Department of Ophthalmology, The University of Cambridge, United Kingdom

[5]Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, China

*Corresponding author: Carolyn Yu Tung Wong, Prince of Wales Hospital, 30-32 Ngan Shing Street, Shatin, New Territories, Hong Kong, Email: carolcarol.carolyn@gmail.com

Abstract

Current Diagnostic Challenges in Dry Eye Disease

Dry eye disease (DED) is a challenging condition to pin down, given the various probable aetiologies, signs, and symptoms. DED is characterised by its loss of tear-film homeostasis, ocular surface inflammation, hyperosmolarity, eye discomfort, and visual abnormalities [1]. However, the signs of DED are occasionally inconsistent with the symptoms stated by patients [1]. There is presently no one clinical test that can uniformly pinpoint DED [1]. DED is diagnosed using a variety of subjective tests and symptom questions, including tear breakup time (TBUT), Schirmer's test (ST), fluorescein and lissamine green staining of the corneal (CSS1) and conjunctival surface (CSS2), and the ocular surface disease index (OSDI) [1]. Furthermore, the differentiation of different tear film breakup patterns is thought to be at the centre of a tear film-oriented diagnosis, which helps elucidate the pathophysiology of DED (i.e., identify the insufficient component of the tear film or of the corneal surface epithelium responsible for TFBU), sub classify DED, and select the optimal topical therapy (decide on the most appropriate treatment) [2]. Additionally, although meibomian gland dysfunction (MGD) is the leading cause of evaporative DED and one of the most common conditions encountered in DED, diagnosing MGD can be difficult due to the non-specific nature of the symptoms and great inter-examiner variability in grading clinical variables associated with MGD [3]. As a result, standardised and universal diagnostic and decision-making tools in DED are highly valued.

Artificial intelligence (AI) through machine learning (ML) and deep learning (DL) has garnered attention in the ophthalmological field, particularly in the screening and diagnosis of retinal and optic nerve conditions [4]. These AI algorithms perform image-intensive analyses on fundus or optical coherence tomography (OCT) images [4]. Similarly, in the current practices of DED, AI is expected to facilitate the data-intensive analysis of DED signs and symptoms when diagnosing, triaging, and managing DED patients.

Keywords: Artificial Intelligence; Optical Coherence Tomography; Eye; Cornea

Abbreviations: OCT: Optical Coherence Tomography; ST: Schirmer's Test; AI: Artificial Intelligence; ML: Machine Learning; DL: Deep Learning.

AI in the Automated Diagnosis of DED

Videography and photography of the ocular surface collected during a slit lamp examination can assist in the diagnosis of DED because they provide information on tear film stability (TBUT) and volume (tear meniscus height (TMH)) [5]. However, the evaluation of videos and photographs frequently exhibits poor reproducibility and repeatability due to a lack of tools for consistent, objective, and quantitative analysis, and the incorporation of AI will hopefully aid in the establishment of precise data interpretation tools and reliable DED diagnostics [5].

Shimizu E, et al. [4] devised a DL model that uniformly diagnoses DED from the TBUT information presented by slit lamp videos of a portable device (Smart Eye Camera) and the OSDI reported by patients. The model interpreted a patient as having DED when it estimated a TBUT ≤5 seconds and OSDI input >13. After being trained on 16,440 fluorescence-enhanced blue light ocular video frames annotated for TBUT, the model was able to estimate TBUT with high accuracy (0.789) and area under curve (AUC) (0.877), and diagnose DED with high sensitivity (0.778), specificity (0.857), and AUC (0.813). Chase C, et al. [6], on the other hand, has leveraged TMH information captured by anterior segment optical coherence tomography (AS-OCT) images to diagnose DED. Having been trained and tested on 27,180 AS-OCT images, the model exhibited a high accuracy, sensitivity and specificity of 84.62%, 86.36%, and 82.35% respectively in the diagnosis of DED. The model even outperformed the diagnostic accuracy of CSS1, CSS2, and ST1 (P < 0.05), which are standard clinical tests used to diagnose DED.

Similarly, utilising the tear meniscus information captured on AS-OCT images, two DL models that adopted a direct tear meniscus segmentation approach (DSA) and a region of interest localization followed by segmentation approach (LSA) have been constructed to diagnose DED [7]. The DSA and LSA models segmented the tear meniscus for DED diagnosis at a sensitivity of 96.36% and 96.43%, a specificity of 99.98% and 99.86%, and a Jaccard index of 93.24% and 93.16%, respectively. Keratography videos capturing patients' TBUT and TMH have also been exploited to develop a DED diagnostic model via the transfer learning approach [8]. The model trained on 244 videos demonstrated a high diagnostic accuracy of 0.98 in the detection of DED. The lower paracentral cornea was reported by the network activation maps to be the most significant region for DED detection.

AI in the Understanding of DED Pathophysiology

Tear film interferometry is a promising technique for scrutinizing the tear film dynamics and tracing the pathophysiological cause of DED. Tear inferometric colour and fringe patterns reflect the balance between the aqueous and lipid layers of the tear film and can be used to trace the aqueous deficient (AD-DED) or evaporative nature (EDED) of the DED [9]. The knowledge of the DED's nature type may facilitate targeted treatment actions to tackle the root causes of DED for optimal symptom and sign relief [9,10]. For instance, applying lipid-containing eye drops to improve the lipid layer in EDED [10] and administering aqueous-based artificial tears to improve aqueous deficiency in AD-DED [11].

Da CLB, et al. [12,13] derived two ML models that classified the inferometry images into five fringe patterns (debris, fine fringes, coalescing fine fringes, strong fringes, and coalescing strong fringes). Using 106 interferometry images retrieved from the VOPTICAL_GCU dataset to train the two ML classifiers separately, the best-performing Random Forest-based classifiers differentiated the fringe patterns at high area under the receiver operating characteristic curves (AUROCs) of 0.99, Kappa indexes of 0.995-0.96 and F-Measures of 0.996-0.97. Kikukawa Y, et al. [14], on the other hand, harnessed the power of ML to classify tear inferometric patterns into nine categories for the fine-grained analysis of tear film dynamics in different DEDs. With the help of data augmentation and transfer learning, the model re-trained on 9,089 inferometric image patches achieved an AUC of 0.898, sensitivity of 84.3%, and a specificity of 83.3%.

Furthermore, Arita R, et al. [15] designed an algorithm for the classification of AD-DED and EDED subtypes based on interferometric fringe patterns. The algorithm corresponded the three fringe patterns: pearl-like appearance, jupiter-like appearance, and crystal-like appearance to the diagnosis of normal tear condition, AD-DED, and EDED, respectively. After training with 138 inference images of each fringe pattern type, the algorithm demonstrated high F-scores of 0.954, 0.806 and 0.762 for AD-DED, EDED and normal condition diagnoses, respectively.

AI in the Interpretation of Meibomian Gland Dysfunction

Infrared meibography provides insights into the two-dimensional silhouette of meibomian glands (MGs) and delivers information such as the amount of dropout, the area of the MG's acini, and the length of the MG's duct [5]. Such information are vital for evaluating the degree of

Carolyn YTW, et al. Artificial Intelligence in Dry Eye Disease: Benefits, Challenges and Future Directions. J Ophthalmol 2024, 9(1): 000310.

Copyright© Carolyn YTW, et al.

MGD and the DED associated with it, hence encouraging the formation of individualised treatment plans [5]. The manual interpretation of infrared meibography is vulnerable to diagnostic quality fluctuations caused by image quality and inter-oberver variability [16]. As a result, an automated segmentation approach for infrared meibography assures diagnostic consistency, quality, and neutrality in the deciphering of MGs and MGD-related DED.

Yu Y, et al. [17] developed a meibography image grading model using 1878 annotated meibography images. The model achieved good performance in terms of MG area detection and segmentation as reflected by the small validation loss values (validation loss < 0.35-1.0) and high mean value of average precisions (mAP) (mAP > 0.92-0.976). Similarly, Setu MAK, et al. [16] utilized 728 clinical infrared meibography images to build a DL model for MG segmentation and morphology assessment. The model achieved an average precision, recall, F1 score, AUROC, and dice coefficient of 83%, 81%, 84%, 0.96, and 84%, respectively, for MG segmentation.

In another study by Saha RK, et al. [18] 1600 meibography images were employed to develop a MG and eyelid segmentation model for the detection of individual MGs and quantification of meibomian gland area and area ratio. The model achieved 73.01% accuracy for meiboscore classification on the validation set and 59.17% accuracy when tested on images from an independent center, which were superior to the 53.44% validation accuracy by MGD experts. Koh YW, et al. [19] constructed a ML algorithm that differentiates healthy and unhealthy infrared meibography images by automatically detecting the length and width of MGs. The model successfully attained a specificity of 96.1% and a sensitivity of 97.9% in detecting healthy and unhealthy meibography images. The user-free computational method

was claimed to be fast and did not suffer from inter-observer variability.

Benefits, Challenges and Future Directions of AI in DED

AI is projected to improve consistency, impartiality, and efficiency in DED diagnostics. It may be possible to expand DED screening coverage by implementing AI screening tools in primary care settings and on individual smartphones. DED patients can also be triaged more effectively into distinct pathophysiology or severity groups by leveraging AI's ability to aggregate multimodal data for fine-grained sub classification of DED type. Individualised therapy plans can be better developed to address the underlying causes of patients' complaints. Despite the reported success of AI, problems and obstacles remain. Before the broad implementation of AI in DED practice, critical technical and clinical constraints must be addressed.

The quantity and quality of data are critical to the robustness of an AI model. The majority of the aforementioned AI applications in DED relied on raw data acquired by ophthalmologists during their clinical practice, which were often confined to one or a few demographic groups and of limited quantity. As a result of spectrum bias, the algorithm's generalizability may suffer, as may its performance during model deployment. Additionally, the majority of AI in DED employs a fully supervised learning strategy, which necessitates high-quality data and annotations provided by ophthalmologists. The labor-intensive and time-consuming nature of manual data labelling, the discrepancy in image processing protocols, and the disparity in labelling standards and levels across labellers, make it challenging to assure DED data quality in AI training (Table 1).

Study	Task	Data Modality	Dataset	Artificial Intelligence Model/ Network	Significant Findings	Limitations
Shimizu E, et al. [4]	Automated diagnosis of dry eye disease (DED)	Slit-lamp captured videography	16440 fluorescence-enhanced blue light ocular video frames	Deep learning (DL) model (Swin Transformer)	-The model estimate tear breakup time (TBUT) at an accuracy of 0.789 and area under curve (AUC) of 0.877, and diagnosed DED with a sensitivity of 0.778, specificity of 0.857, and AUC of 0.813.	Limited sample number and selection bias
						model reliance on DED diagnostic criteria by Asia dry eye society (ADES)
						Absence of inclusion of other data modalities e.g. Schirmer's test (ST1)

Author	Aim	Image Type	Dataset	Model	Results	Limitations
Chase C, et al. [6]	Automated diagnosis of DED	Anterior segment optical coherence tomography (AS-OCT) images	27,180 AS-OCT images	DL model (VGG19)	-The model showed an accuracy, sensitivity and specificity of 84.62%, 86.36%, and 82.35% respectively in the diagnosis of DED.	Limited sample number Lack of gold standard in dry eye testing Absence of quality control phase during model testing
Stegmann H, et al. [7]	Automated diagnosis of DED	AS-OCT images	6658 images	Two DL models where one utilized a direct tear meniscus segmentation approach (DSA) and the other used a region of interest localization followed by segmentation approach (LSA)	-The DSA and LSA models segmented the tear meniscus for DED diagnosis at a sensitivity of 96.36% and 96.43%, a specificity of 99.98% and 99.86%, and a Jaccard index of 93.24% and 93.16%, respectively.	Limited images Single device acquisition of images
Abdelmotaal H, et al. [8]	Automated diagnosis of DED	Keratography videos	244 videos	DL model (transfer learning approach)	-The model achieved a diagnostic accuracy of 0.98 in the detection of DED.	Limited number of eyes and videos High degree of similarity between some of the extracted video frames Single institution acquisition of data
Da CLB, et al. [12]	Classification of tear film fringe patterns	Inferometry images	106 images	DL networks (random forest (RF), Support Vector Machine)	-The best-performing RF model differentiated the fringe patterns at an area under the receiver operating characteristic curve (AUROC) of 0.99, Kappa index of 0.995 and F-Measure of 0.996.	NA
Da CLB, et al. [13]	Classification of tear film fringe	Inferometry images	106 images	DL networks (random forest, Support Vector Machine)	-The best-performing RF model differentiated the fringe patterns at an area under the receiver operating characteristic curve (AUROC) of 0.99, Kappa index of 0.96 and F-Measure of 0.97.	Removal of certain relevant features Other regions of interest segmentation technique not explored

Carolyn YTW, et al. Artificial Intelligence in Dry Eye Disease: Benefits, Challenges and Future Directions. J Ophthalmol 2024, 9(1): 000310.

Copyright© Carolyn YTW, et al.

Kikukawa Y, et al. [14]	Classification of tear film fringe patterns and breakup patterns	Inferometry images	9,089 image patches	DL model (ResNet50)	-The model classified the fringe patterns at achieved an AUC of 0.898, sensitivity of 84.3%, and a specificity of 83.3%.	Limited data quantity
Arita R, et al. [15]	Classification of tear film fringe patterns	Inferometry images	138 images	DL model	-The model attained F-scores of 0.954, 0.806 and 0.762 for AD-DED, EDED and normal condition diagnoses, respectively.	NA
Yu Y, et al. [17]	Meibomian gland (MG) area detection and segmentation	Meibography images	1878 images	Mask R-CNN DL model	The model achieved small validation loss values (validation loss < 0.35-1.0) and high mean value of average precisions (mAP) (mAP > 0.92-0.976) for MG area detection and segmentation.	Small dataset
						Single center database
Setu MAK, et al. [16]	MG segmentation and morphology assessment	Meibography images	728 images	Dl model (Inception-ResNet-v2)	The model achieved an average precision, recall, F1 score, AUROC, and dice coefficient of 83%, 81%, 84%, 0.96, and 84%, respectively, for MG segmentation.	Device specificity of DL model
						Inter observer variability in ground truth masks
						Limited sample number
Saha RK, et al. [18]	MG and eyelid segmentation	Meibography images	1600 images	DL method	The model achieved 73.01% accuracy for meiboscore classification on the validation set and 59.17% accuracy when tested on images from an independent center.	NA
Koh YW, et al. [19]	Differentiates healthy and unhealthy meibography images by automatically detecting the length and width of MGs	Meibography images	Meibography images of 55 patients	ML method (Scale invariant feature transform)	The model attained a specificity of 96.1% and a sensitivity of 97.9% in detecting healthy and unhealthy meibography images.	-Limited image quantity
						Inter observer variability in ground truths

Table 1: Artificial intelligence applications in dry eye disease.

In response, publicly accessible benchmark datasets are essential, especially because they may provide an equitable platform for comparing the outcomes of AI models in DED.

Currently, public DED databases are sparse. Future AI research in DED should consider the establishment of a large-scale public DED dataset. A generative adversarial network

can generate an immense amount of random and diverse DED images, while federated learning can help with data privacy issues through its decentralised data management mechanism.

Furthermore, transparent AI reporting is required at both the model development and prospective clinical testing levels in order for AI to be implemented clinically effectively and reliably. Although established AI reporting protocols such as CONSORT-AI, STARD-AI, SPIRIT-AI, and TRIPOD have emerged charalambides M, et al. [20], report transparency has varied significantly between DED models, and rigorous adherence to standardised AI reporting is often inadequate. Many models neglected to include extensive demographic information in their training and validation data. Most models' appropriate usage scopes were also ill-defined (e.g. principal users), making contextualising many of these DED models challenging. Future DED AI research will require reduced reporting selectivity via strengthened adoptions of standardised AI reporting approaches. Poor transparency was also seen in the AI models' decision-making. The majority of the DED models were black boxes, making it hard for ophthalmologists to comprehend how they produced their predictions in the first place. To fully trust AI's clinical reasonableness, future DED AI research must embrace the explainable AI approach, in which the AI system is dissected into multiple modules (e.g., pre-diagnosis module, image segmentation module, and final diagnosis module) for ophthalmologists' thorough visualisation.

Conclusion

Finally, concerns about the attribution of responsibility for particular harms caused by the use or misuse of AI should be addressed. Ethical frameworks that define the legal obligations of various parties (e.g. AI firms and clinicians) in ensuring AI operates in a specific manner and taking necessary compensation steps when harm occurs.

References

1. Stapleton F, Alves M, Bunya VY, Jalbert I, Lekhanont K, et al. (2017) TFOS DEWS II Epidemiology Report. Ocul Surf 15(3): 334-365.

2. Tsubota K, Pflugfelder SC, Liu Z, Baudouin C, Kim HM, et al. (2020) Defining Dry Eye from a Clinical Perspective. Int J Mol Sci 21(23): 9271.

3. Geerling G, Baudouin C, Aragona P, Rolando M, Boboridis KG, et al. (2017) Emerging strategies for the diagnosis and treatment of meibomian gland dysfunction: Proceedings of the OCEAN group meeting. Ocul Surf 15(2): 179-192.

4. Shimizu E, Ishikawa T, Tanji M, Agata N, Nakayama S, et al. (2023) Artificial intelligence to estimate the tear film breakup time and diagnose dry eye disease. Sci Rep 13(1): 5822.

5. Yang HK, Che SA, Hyon JY, Han SB (2022) Integration of Artificial Intelligence into the Approach for Diagnosis and Monitoring of Dry Eye Disease. Diagnostics (Basel) 12(12): 3167.

6. Chase C, Elsawy A, Eleiwa T, Ozcan E, Tolba M, et al. (2021) Comparison of Autonomous AS-OCT Deep Learning Algorithm and Clinical Dry Eye Tests in Diagnosis of Dry Eye Disease. Clin Ophthalmol 15: 4281-4289.

7. Stegmann H, Werkmeister RM, Pfister M, Garhofer G, Schmetterer L, et al. (2020) Deep learning segmentation for optical coherence tomography measurements of the lower tear meniscus. Biomed Opt Express 11(3): 1539-1554.

8. Abdelmotaal H, Hazarbasanov R, Taneri S, Al-Timemy A, Lavric A, et al. (2023) Detecting dry eye from ocular surface videos based on deep learning. Ocul Surf 28: 90-98.

9. Arita R, Morishige N, Fujii T, Fukuoka S, Chung JL, et al. (2016) Tear Interferometric Patterns Reflect Clinical Tear Dynamics in Dry Eye Patients. Invest Ophthalmol Vis Sci 57(8): 3928-3934.

10. Rolando M, Merayo-Lloves J (2022) Management Strategies for Evaporative Dry Eye Disease and Future Perspective. Curr Eye Res 47(6): 813-823.

11. Semp DA, Beeson D, Sheppard AL, Dutta D, Wolffsohn JS (2023) Artificial Tears: A Systematic Review. Clin Optom (Auckl) 15: 9-27.

12. Cruz LBD, Souza JC, Sousa JAD, Santos AM, Paiva ACD, et al. (2020) Interferometer eye image classification for dry eye categorization using phylogenetic diversity indexes for texture analysis. Comput Methods Programs Biomed 188: 105269.

13. Cruz LBD, Souza JC, Paiva ACD, Almeida JDSD, Junior GB, et al. (2020) Tear Film Classification in Interferometry Eye Images Using Phylogenetic Diversity Indexes and Ripley's K Function. IEEE J Biomed Health Inform 24(12): 3491-3498.

14. Kikukawa Y, Tanaka S, Kosugi T, Pflugfelder SC (2023) Non-invasive and objective tear film breakup detection on interference color images using convolutional neural networks PLoS 18(3): e0282973.

15. Arita R, Yabusaki K, Yamauchi T, Ichihashi T, Morishige N (2018) Diagnosis of dry eye subtype by artificial

intelligence software based on the interferometric fringe pattern of the tear film obtained with the Kowa DR-1α instrument. Invest Ophthalmol Vis Sci 59(9): 1965-1965.

16. Setu MAK, Horstmann J, Schmidt S, Stern ME, Steven P (2021) Deep learning-based automatic meibomian gland segmentation and morphology assessment in infrared meibography. Sci Rep 11(1): 7649.

17. Yu Y, Zhou Y, Tian M, Zhou Y, Tan Y, et al. (2022) Automatic identification of meibomian gland dysfunction with meibography images using deep learning. Int Ophthalmol 42(11): 3275-3284.

18. Saha RK, Chowdhury AMM, Na KS, Hwang GD, Eom Y, et al. (2022) Automated quantification of meibomian gland dropout in infrared meibography using deep learning. Ocul Surf 26: 283-294.

19. Koh YW, Celik T, Lee HK, Petznick A, Tong L (2012) Detection of meibomian glands and classification of meibography images. J Biomed Opt 17(8): 086008.

20. Charalambides M, Flohr C, Bahadoran P, Matin RN (2021) New international reporting guidelines for clinical trials evaluating effectiveness of artificial intelligence interventions in dermatology: strengthening the SPIRIT of robust trial reporting. Br J Dermatol 184(3): 381-383.

Carolyn YTW, et al. Artificial Intelligence in Dry Eye Disease: Benefits, Challenges and Future Directions. J Ophthalmol 2024, 9(1): 000310.

Copyright© Carolyn YTW, et al.