Med-X

RESEARCH ARTICLE

Noninvasive and affordable type 2 diabetes screening by deep learning-based risk assessment and detection using ophthalmic images inspired by traditional Chinese medicine

Zhi Zhang[1,2] · Han Wang[3] · Lei Chen[4] · Chensi Cao[1,2] · Tengwen Liu[5] · Ruifang Ren[1,2] · Ruixing Zhou[1,2] · Rudan Huang[1,2] · Dan Hu[4] · Chenxing Zhu[4] · Chong Lu[6] · Yunsheng Xu[7] · Zhaohui Fang[8] · Fuer Lu[9] · Huimin Pan[10] · Yanjin Su[11] · Nanlin Fu[12] · Huixia Zhan[13] · Qin Si[14] · Chenze Bai[14] · Ri Le Ge[14] · Hongmei Cao[1,2] · Wei Dong[1,2] · Guohui Yang[1,2] · Lan Wu[15] · Jiao Guo[16,17,18,19] · Jing Cheng[1,2,3]

Abstract

The global prevalence of diabetes is steadily increasing, with a high percentage of patients unaware of their disease status. Screening for diabetes is of great significance in preventive medicine and may benefit from deep learning technology. In traditional Chinese medicine, specific features on the ocular surface have been explored as diagnostic indicators for systemic diseases. Here we explore the feasibility of using features from the entire ocular surface to construct deep learning models for risk assessment and detection of type 2 diabetes (T2DM). We performed an observational, multicenter study using ophthalmic images of the ocular surface to develop a deep convolutional network, OcularSurfaceNet. The deep learning system was trained and validated with a multicenter dataset of 416580 images from 67151 participants and tested independently using an additional 91422 images from 12544 participants, and can be used to identify individuals at high risk of T2DM with areas under the receiver operating characteristic curve (AUROC) of 0.89–0.92 and T2DM with AUROC of 0.70–0.82. Our study demonstrated a qualitative relationship between ocular surface images and T2DM risk level, which provided new insights for the potential utility of ocular surface images in T2DM screening. Overall, our findings suggest that the deep learning framework using ocular surface images can serve as an opportunistic screening toolkit for noninvasive and low-cost large-scale screening of the general population in risk assessment and early identification of T2DM patients.

Highlights

- Phenotypes of ocular surface can be used for accurate, non-invasive, affordable T2DM risk assessment.
- Ocular surface phenotypes associated with T2DM risk preliminarily elucidated by the neural network OcularSurfaceNet.
- It has potential for generalized high-impact application in T2DM screening for large-scale population.

Zhi Zhang, Han Wang and Lei Chen contributed equally to this work.

Extended author information available on the last page of the article

Introduction

Diabetes mellitus is recognized as a leading cause of premature death and disability worldwide, with the global prevalence estimated to increase from 9.3% in 2019 to 10.9% by 2045, which imposes huge personal and societal health care burdens [1, 2]. Among the different types of diabetes, type

Check for updates

∮ Springer

2 diabetes mellitus (T2DM) accounts for approximately 90%–95% of all cases [3]. However, as T2DM develops over years, it may not be noticed until severe complications have already arisen and damage to the body has taken place. In 2021, almost one in two adults between the ages of 20–79 years old with diabetes were unaware of their diabetes status (44.7%, 239.7 million globally) [4], calling for effective measures for large-scale screening to identify unknown T2DM patients. The current gold standard method for the diagnosis of diabetes by intravenous or fingerstick blood tests is invasive, nonuser-friendly, and costly [5]; unsuitable for screening of the general population. Other risk assessment methods, such as polygenics [6–8] and serum metabolomics profiles [9, 10], still require laboratory tests and are time-consuming and costly [11].

Human eyes may provide a noninvasive and easily accessible observation window of systemic health. Specific features of ophthalmic images, regarded as imaging biomarkers, have been explored as manifestations of eye diseases and systemic diseases, such as diabetic retinopathy [12, 13], cardiovascular diseases [14, 15], stroke [16–18], and other diseases [19, 20]. Slit-lamp images have also been utilized to facilitate early identification of hepatobiliary diseases [21]. Moreover, features on the ocular surface have also been explored as diagnostic indicators of systemic diseases in traditional Chinese medicine (TCM) [22]. In TCM theory, each region of the ocular surface corresponds to a different viscus in the body. By diagnosing the patterns of specific features of a particular region on the ocular surface, the corresponding diseased viscus and pathological changes can be differentiated.

In recent years, deep learning methods have gained popularity due to their capability of extracting predictive features from raw samples. In particular, convolutional neural networks (CNNs) have demonstrated great potential in image-based diagnosis, especially diabetes prediction using retinal fundus images [23]. However, prior studies on deep learning-based diagnosis using ophthalmic images usually rely on prelabeled images for feature learning, which are subject to the availability of high-quality clinical metadata.

In this work, we constructed a deep convolutional network, OcularSurfaceNet, for risk assessment and detection of T2DM by analyzing ophthalmic images of the ocular surface without prior labeling. A two-stage refinement strategy by transfer learning that employed both global and local features of ocular surface images in model development was established. Furthermore, inclusion of unhealthy subjects who carried diseases other than T2DM in the control group during model training significantly improved the accuracy of our AI models. We further conducted multiple assays to elucidate T2DM-relevant regions and features of the ocular surface. As demonstrated using internal and external test datasets, our models showed superior performance and robustness in both risk assessment and detection of T2DM in community screening in 12 cities nationwide.

Methods

Study design and participants

In this observational, multicenter study, patient cohorts from multiple national and regional investigations (Supplementary Methods) were obtained and divided randomly into the development, validation, internal testing, and external testing datasets, respectively. The demographic distribution of the cohorts (Supplementary Table 1) varied largely and covered multiple regions of the whole nation, facilitating accurate and unbiased evaluation of the proposed AI models. These studies were approved by the institutional review boards and ethics committees of the corresponding institutes, and written informed consent was obtained from each participant.

In this study, T2DM and prediabetes were defined according to the international guidelines by the World Health Organization (WHO) [3, 24]. High-risk cases of T2DM assessed by questionnaire scores were defined as HR-QS for short [25]. Both HR-QS and prediabetes were at high risk of T2DM and defined as HR. Datasets with T2DM and HR excluded were defined as non-HR.

Data collection and preprocessing

Eight ocular surface images were acquired using a benchtop eye examination instrument for each participant by looking at the top, bottom, left, and right directions with both eyes while keeping the sclera adequately exposed (Supplementary Methods) [26]. This portable instrument was approved as a second-class medical device of the People's Republic of China in 2018 and has been widely deployed, covering over 200,000 people. Ophthalmic images of the ocular surface were checked automatically and manually, where unqualified images (blurred focus, lack of clarity, light leakage, occlusion, etc.) were excluded. As a result, 70% of the images were found to be qualified, and 90% of the samples were defined as qualified samples with more than four corresponding images being qualified.

We performed white balance on the qualified raw images and then segmented the ocular surface, iris, and sclera areas using U-Net (Supplementary Methods) [27]. We also performed image augmentation to improve the performance of the AI models. Only the training samples were augmented to avoid data leakage. We preprocessed the ocular surface images and their horizontally flipped images randomly with three transformations: rotation (-15° ~ 15°), cropping (70% ~ 100%), and zooming (80% ~ 105%) (Fig. 1), where
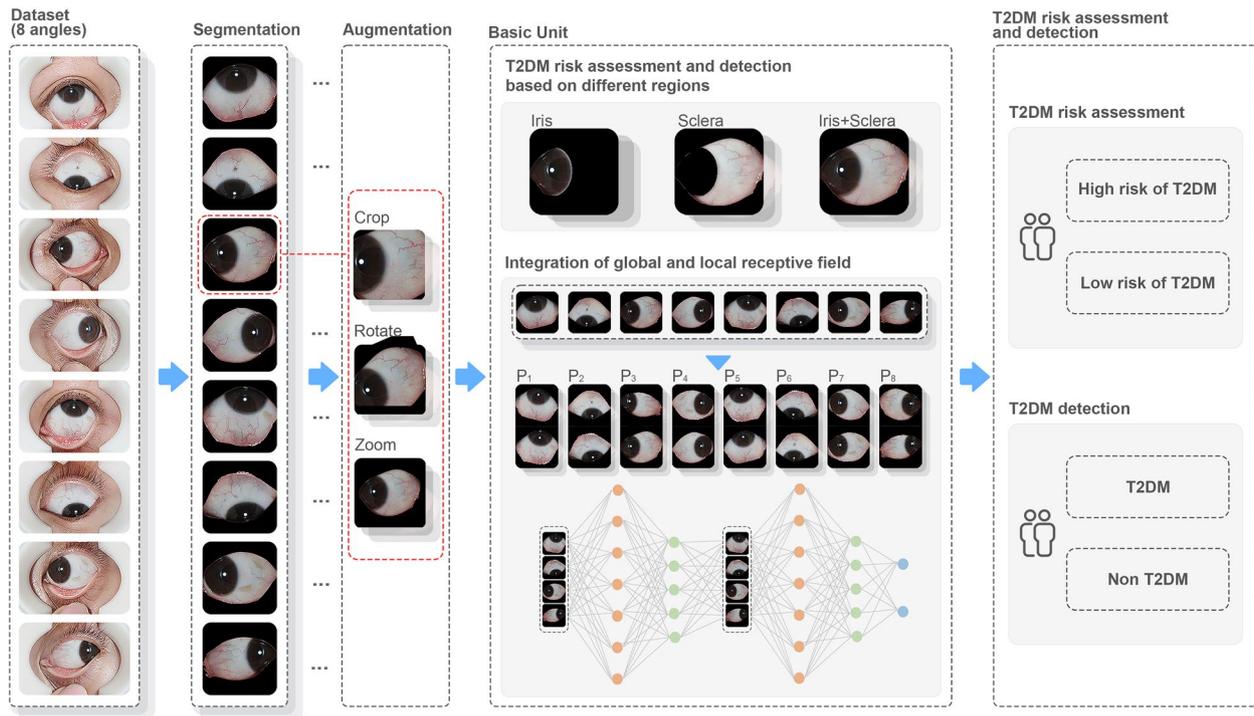
Fig. 1 Overview of the study. Ophthalmic images of the ocular surface were acquired at eight different angles. After segmentation and augmentation of the ophthalmic images, the AI model was developed and trained to recognize distinguished features in different regions of the ocular surface, followed by tuning the attention of the AI model using both merged and individual images at different angles (Basic unit). The average of image-level inference was used to predict participants' status

the degree of transformation was randomly determined. As a result, the ocular surface images were tuned to 3,332,640 in the training phase. All processes were implemented with PyTorch (version 1.4.0) and/or Python (version 3.7.7) software [28].

Development of the T2DM risk assessment and detection models

With ResNet-18 [29] as the backbone (Supplementary Table 2), we first developed an image-level classification model to evaluate the probability of whether an image comes from a high-risk case of T2DM. Then, the final determination of a sample was performed by comparing the average probability and the threshold (which was tuned according to the Matthews correlation coefficient (MCC)) [30]. We developed image-level classification models using different regions of ophthalmic images to assess the contribution of different ocular surface regions in T2DM risk assessment and using groups of ophthalmic images with different angles to integrate local and global ophthalmic features. Models were trained and validated using the same data partition strategy (80% for training, 10% for tuning, and 10% for validation). In T2DM prediction, T2DM samples were regarded as 'hits', and others were included in the control set. In HR

prediction, T2DM and HR samples were included in the 'hit' set, and non-HR samples were included in the control set.

In detail, the region-specific models IrisNet, ScleraNet, and OcularSurfaceNet were trained with the iris, sclera, and whole ocular surface regions of the ophthalmic images, respectively (Fig. 1). Images were grouped into 8 categories according to their angles. Considering the integration of local and global ophthalmic characteristics in T2DM risk assessment, we trained the models with ungrouped images (G) and refined with angle-specific images (G2L) and trained the models with angle-specific images (L) and further refined with ungrouped images (L2G).

Elucidation of T2DM risk assessment based on ophthalmic images

Grad-CAM maps and ΔMCC maps were constructed to understand T2DM-relevant regions on the ocular surface, and the distribution of significant features was tested to explore T2DM-relevant ocular features (Supplementary Methods) [31]. Based on regions in the ΔMCC map, we further tested the distribution differences of 90 ocular features among all qualified 20,221 T2DM/high-risk samples and 7224 T2DM low-risk samples. One hundred and eight ocular features included 6 coarse-grained morphological features

and 102 fine-grained morphological color features (Supplementary Fig. 2). Specifically, we selected T2DM-relevant ocular features with a higher frequency of enrichment in T2DM/high-risk samples and then categorized them into different groups to explore their association with T2DM.

Evaluation of the T2DM risk assessment and detection models

We computed the accuracy (ACC), sensitivity (SEN), specificity (SPE), F1-measure (F1), MCC, Kappa score, and AUROC as evaluation metrics for the T2DM risk assessment and detection models.

Statistical analysis

The 95% CIs of metrics were estimated with the nonparametric bootstrap method (1000 random resamplings with replacement). A t test was used to evaluate the distribution of RGB color values, hue, saturation, and value channels. Chi-square tests were used to test the distribution of shape features such as dark punctate, patchy, and halo-shaped features. All comparisons were two-sided and calculated using Python (version 3.7.7) and/or R (version 4.0.4) software.

Results

In the evaluation of the contribution of different ocular surface regions in T2DM risk assessment, we trained IrisNet, ScleraNet, and OcularSurfaceNet with iris, sclera, and whole ocular surface regions of the ungrouped images. IrisNet, ScleraNet, the ensemble of IrisNet and ScleraNet, and OcularSurfaceNet achieved AUROCs of 0.8587 (95%

CI, 0.8452–0.8715), 0.8728 (95% CI, 0.8599–0.885), 0.8823 (95% CI, 0.8697–0.8941), and 0.8869 (95% CI, 0.8751–0.8988), respectively (Fig. 2a, Supplementary Table 3).

We also tested IrisNet, ScleraNet, the ensemble of IrisNet and ScleraNet, and OcularSurfaceNet on internal and external test sets (Fig. 2b-c, Supplementary Table 3). In T2DM risk assessment, they achieved AUROCs of 0.8627 (95% CI, 0.8533–0.8717), 0.8766 (95% CI, 0.8682–0.8853), 0.8824 (95% CI, 0.8744–0.8905), and 0.8874 (95% CI, 0.8794–0.8951) on the internal test set and AUROCs of 0.9011 (95% CI, 0.8847–0.9196), 0.9104 (95% CI, 0.8922–0.9269), 0.9131 (95% CI, 0.8968–0.9287), and 0.915 (95% CI, 0.8986–0.9299) on external test set 1, respectively. OcularSurfaceNet has achieved higher MCC and AUROC (Supplementary Table 3) than other models on internal and external test sets. The results show that OcularSurfaceNet outperformed the other models, which suggests that T2DM-relevant features may be region-specific and that distinct ocular characteristics may be present in different regions of the iris and sclera. Thus, we used ophthalmic images of the entire ocular surface to develop the following models.

When evaluated on the internal test set and external test set 1, OcularSurfaceNet achieved a sensitivity of 0.9512 (95% CI, 0.9369–0.9637) and 0.8316 (95% CI, 0.7908–0.8673) on T2DM samples, sensitivity of 0.9211 (95% CI, 0.8553–0.9737) and 0.7 (95% CI, 0.55–0.825) on prediabetes samples, sensitivity of 0.9338 (95% CI, 0.9274–0.9401) and 0.756 (95% CI, 0.719–0.793) on HR-QS samples, and specificity of 0.5964 (95% CI, 0.575–0.6173) and 0.8847 (95% CI, 0.8553–0.914) on non-HR samples (Supplementary Table 4).

We constructed a two-stage refining strategy by transfer learning that takes into consideration both local and
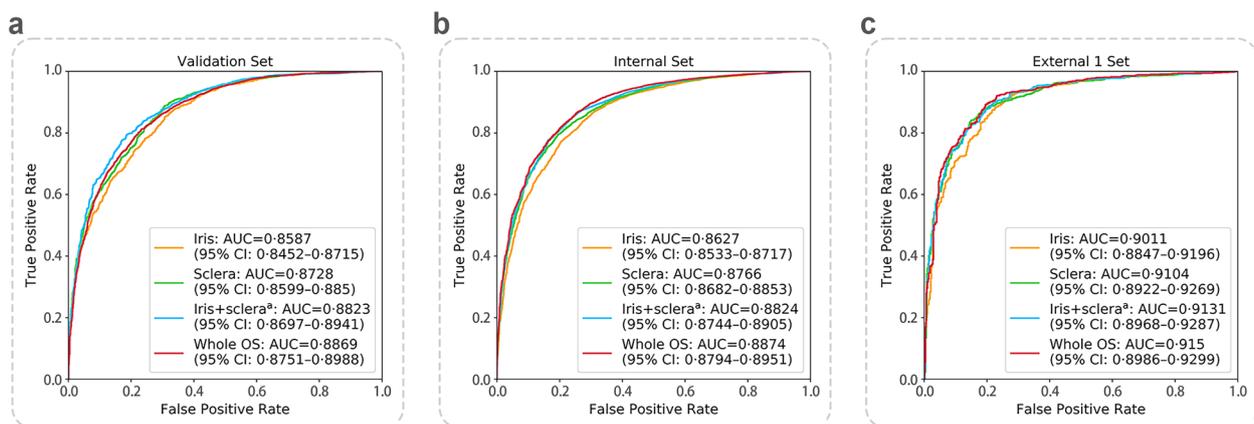
Fig. 2 Performance of the AI system in the prediction of high-risk cases of T2DM based on ophthalmic images. ROC curves showing the performance of T2DM risk assessment using AI models developed with ophthalmic images from different ocular surface regions. a-c, ROC curves of IrisNet (orange), ScleraNet (green), the ensemble of IrisNet and ScleraNet (Iris + Sclera.[a], blue), and OcularSurfaceNet (Whole OS, red) on the validation set (**a**), internal test set (**b**), and external test set 1 (**c**)
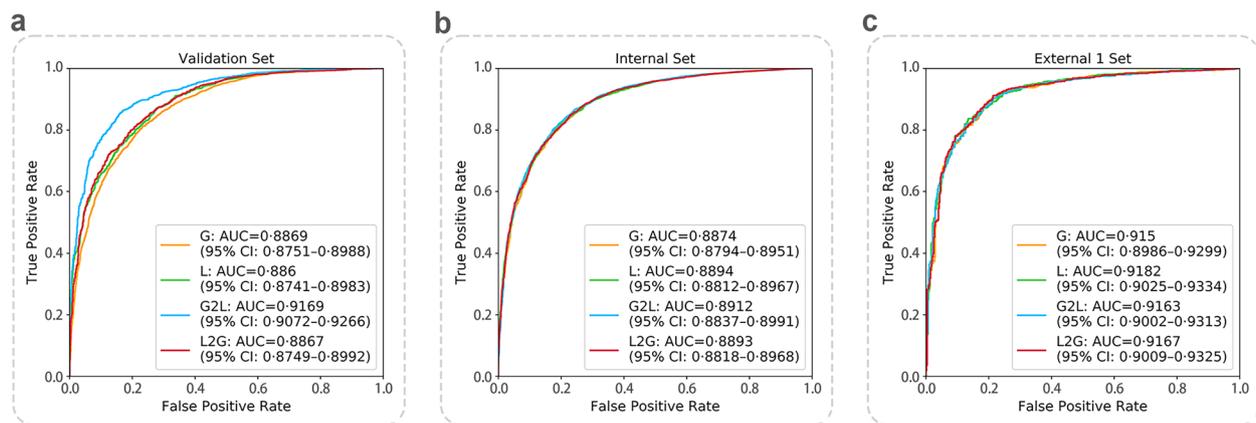
Fig. 3 Performance of the AI system in the prediction of high-risk cases of T2DM based on ophthalmic images by integrating global and local ophthalmic features. ROC curves show the performance of T2DM risk assessment using AI models developed with ophthalmic images from separate or all angles. We trained models with ocular surface images at 8 separate angles (L, green) and all angles (G, orange), then fine-tuned G with images at 8 separate angles (G2L, blue) and fine-tuned L with merged images (L2G, red). **a–c**, ROC curves show their performance on the validation set (**a**), internal test set (**b**), and external test set 1 (**c**)

global characteristics of ophthalmic images for T2DM risk assessment (Supplementary Fig. 3). This strategy improved the model performance and robustness for T2DM risk assessment (Fig. 3, Supplementary Table 5). On the validation set, the G, L, G2L, and L2G models achieved AUROCs of 0.8869 (95% CI, 0.8751–0.8988), 0.886 (95% CI, 0.8741–0.8983), 0.9169 (95% CI, 0.9072–0.9266), and 0.8867 (95% CI, 0.8749–0.8992), respectively. On the internal test set and external test set 1, G achieved AUROCs of 0.8874 (95% CI, 0.8794–0.8951) and 0.915 (95% CI, 0.8986–0.9299), respectively, while L achieved AUROCs of 0.8894 (95% CI, 0.8812–0.8967) and 0.9182 (95% CI, 0.9025–0.9334), respectively, and G2L achieved AUROCs of 0.8912 (95% CI, 0.8837–0.8991) and 0.9163 (95% CI, 0.9002–0.9313), respectively. Moreover, L2G achieved AUROCs of 0.8893 (95% CI, 0.8818–0.8968) and 0.9167 (95% CI, 0.9009–0.9325), respectively. G2L achieved sensitivity of 0.935 (95% CI, 0.9207–0.9484), 0.9474 (95% CI, 0.8816–0.9868), and 0.9045 (95% CI, 0.8971–0.9123) on T2DM, prediabetes, and HR-QS samples, respectively, and specificity of 0.6806 (0.6597–0.6996) on non-HR samples of the internal test set, sensitivity of 0.7908 (95% CI, 0.75–0.8291), 0.65 (95% CI, 0.5–0.775), and 0.6928 (95% CI, 0.6536–0.7386) on T2DM, prediabetes, and HR-QS samples, respectively, and specificity of 0.9182 (0.8931–0.9413) on non-HR samples of external test set 1 (Supplementary Table 6). Therefore, the two-stage fine-tuned models, especially G2L, exhibited higher MCC and AUROC values than the one-stage models on most test sets, demonstrating the utility of the transfer learning method. The two-stage strategy improved model performance by two possible factors: 1) T2DM-relevant ocular characteristics are region specific, so the fine-tuning process endowed G2L/L2G models with the

ability to learn both global and local features; 2) compared to models pretrained with ImageNet [32], refined models provided better source parameters for transfer learning in T2DM risk assessment.

We established the T2DM detection model in a similar manner to the development of OcularSurfaceNet-G2L. The T2DM detection model achieved AUROCs of 0.6957 (95% CI, 0.681–0.7117), 0.7961 (95% CI, 0.7696–0.8214), 0.7922 (95% CI, 0.7256–0.864), and 0.7122 (95% CI, 0.6655–0.7564) on internal and external test sets 1–3, showing the potential of the AI-assisted detection of T2DM (Supplementary Table 7).

However, similar to the detection of chronic kidney diseases based on fundus and slit-lamp images [23], our models achieved better performance in the 50- to 60-year-old age group (SEN of 0.6403–0.73, SPE of 95% CI, 0.6585–0.6927), with lower specificity in elderly groups (age > 60, SEN > 0.9, SPE < 0.4) and lower sensitivity in younger groups (age < 50, SEN < 0.4, SPE > 0.9) (Supplementary Table 8). The results show that age is an important factor that affects the model performance; in particular, the low specificity of the abundant elderly group samples overwhelmed the accuracy of the model in community screening of T2DM. Thus, we used elderly group (> 60) samples to train an independent elderly group sample-specific T2DM detection model as an ensemble factor of the T2DM detection model.

The age-specific T2DM detection model demonstrated higher specificity and F1 score in elderly group samples and a comparable AUROC to the original model (Fig. 4, Supplementary Table 9). On internal and external test sets 1–3, the age-specific T2DM detection model achieved AUROCs of 0.6855 (95% CI, 0.6698–0.7001), 0.7768 (95% CI,

Fig. 4 Performance of the age-specific AI System in the prediction of T2DM based on ophthalmic images by integrating global and local ophthalmic features. ROC curves of the age-specific model on the internal test set (orange), external test set 1 (green), external test set 2 (blue), and external test set 3 (red)
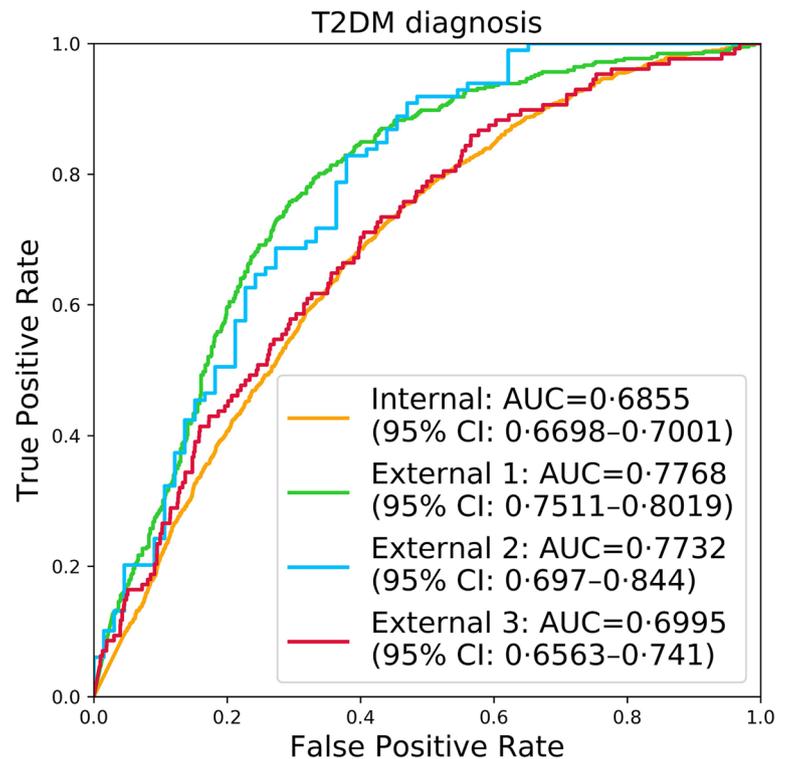
0.7511–0.8019), 0.7732 (95% CI, 0.697–0.844), and 0.6995 (95% CI, 0.6563–0.741), respectively, showing significant improvement in elderly groups over the original model.

Elderly people who are susceptible to T2DM are also prone to acquire other chronic diseases, such as hypertension, as well as complications from T2DM [2, 32]. However, in previous studies on AI-assisted T2DM detection, only healthy samples were included in the control group, and unhealthy samples with diseases other than T2DM were excluded, which may lead to a potentially high false-positive rate when applied to clinical samples [23]. In our study, we developed and tested AI models using both mixed controls and only healthy controls, namely, MC and HC models, respectively. On the one hand, we tested MC and HC models on test sets with unhealthy controls excluded. In this scenario, both the MC and HC models achieved state-of-the-art performance (Supplementary Table 10). On the other hand, we evaluated MC and HC models on more realistic test sets without exclusion of unhealthy samples. HC achieved specificities of 0.0114 (95% CI, 0.009–0.0138), 0.2383 (95% CI, 0.2131–0.2663), 0.1364(95% CI, 0.058–0.2206), and 0.099 (95% CI, 0.086–0.1124), respectively, using the internal test set and external test sets 1–3, even though the sensitivity was high (Supplementary Table 11). MC achieved specificities of 0.4204 (95% CI, 0.4092–0.4313), 0.7797 (95% CI, 0.7525–0.804), 0.7424 (95% CI, 0.6429–0.8471), and 0.7201 (95% CI, 0.7002–0.7412), respectively, using the same datasets. In addition, MC also achieved

AUROCs of 0.6896 (95% CI, 0.6731–0.7059), 0.7842 (95% CI, 0.7575–0.8091), 0.7922 (95% CI, 0.7256–0.864), 0.7122 (95% CI, 0.6655–0.7564), and 0.6938 (95% CI, 0.6814–0.7065) using the internal test set, external test sets 1–3, and the total independent test set, respectively.

From the comparison, it can be clearly seen that healthy and unhealthy control samples all possess distinct features that significantly affect the detection of target T2DM. Therefore, neglecting unhealthy subjects with diseases other than T2DM would hinder learning T2DM-specific features in model development. Moreover, using only healthy samples as controls leads to discrimination between unhealthy and healthy subjects, lacking the specificity of T2DM detection. In the following part of this study, we included both healthy and unhealthy samples with diseases other than T2DM as controls in model development.

Finally, we conducted visualization assays to improve the interpretability of the AI models and shed light on their diagnostic mechanism in T2DM detection. The Grad-CAM maps indicated that the T2DM detection model focused on dark patchy features in regions around the iris and inferior regions of the sclera (Fig. 5a). For a total of 139278 images (32351 and 106927 from T2DM and non-T2DM participants, respectively), we performed an occlusion test according to Cartesian coordinates oriented from the iris centers. The ΔMCC maps also demonstrated that regions around the iris and inferior regions of the sclera of the ocular surface images contributed to T2DM detection (Fig. 5b).

Fig. 5 Relevant regions and features. **a**, the figure shows the original image (left) and saliency map (right) from Grad-CAM. **b**, the figure shows the ΔMCC map (left) from the occlusion test and features (right) in the saliency regions
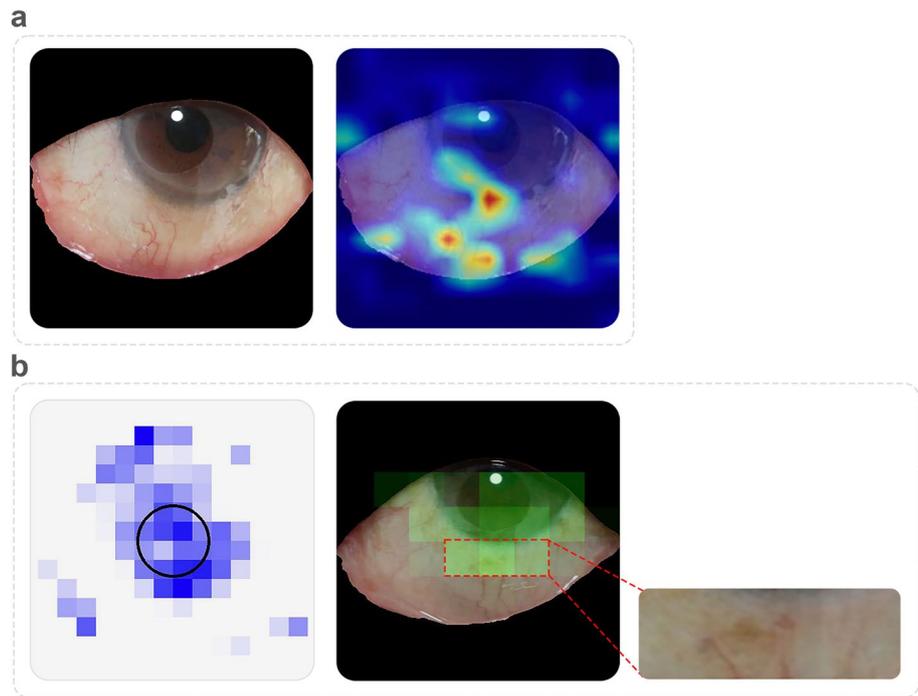
In addition, based on regions in the ΔMCC map (Supplementary Fig. 4), we examined whether the 90 ocular features are enriched in T2DM/high-risk participants. As a result, 36 features are highly enriched in T2DM/high-risk T2DM samples (Supplementary Methods, Supplementary Fig. 5). Among them, 16 ocular features belonged to the syndrome patterns of Yin deficiency, blood deficiency, and blood stasis and heat, 11 ocular features belonged to the syndrome patterns of damp heat, phlegm dampness, and Qi stagnation and heat, and 4 ocular features belonged to the syndrome patterns of Qi deficiency, blood deficiency and Yang deficiency (Supplementary Table 12) [22], while TCM holds that Yin deficiency is the root cause of T2DM, and dryness and heat are the manifestation of T2DM. The syndrome patterns represented by these ocular features were in accord with the TCM syndrome patterns of T2DM, indicating these ocular features might be strongly associated with T2DM.

Discussion

The low awareness rate of T2DM hinders timely intervention and adjustment of personal lifestyle at early stages of disease development, calling for easily accessible means for large-scale screening of T2DM. The results of this study suggest that AI-assisted analysis of ophthalmic images of the ocular surface could be an effective measure for low-cost and scalable screening of the general population.

Previous studies on AI-based ophthalmic image analysis focused on the examination of retinal fundus images and slit-lamp images. However, distinct features are also present at the ocular surface, which are interconnected with the vascular system at the retinal fundus but easier to observe and more sensitive to changes in the microcirculation. In this work, we evaluated the feasibility of using ocular surface images to construct AI models for the risk assessment of T2DM. To avoid temporal effects of subconjunctival hemorrhage by environmental factors, certain contraindications, such as late sleeping, spicy diet, and rough rubbing of the eyes, were established for collecting ocular surface image data in this study.

Compared to previous studies, this work differs mainly in three aspects. First, we established a T2DM risk assessment model along with a detection model, which is of great significance to preventive medicine with early intervention in high-risk patients. Second, qualitative analysis of hyperglycemia using a development dataset that has no labels on the status of T2DM has been performed in this study, which overcomes the limitation of using prelabeled clinical metadata for model development. Third, we have demonstrated the feasibility of ocular surface images as an alternative approach compared to conventional retina fundus images and slit-lamp images for AI-based ophthalmic disease detection. This approach offers new imaging modalities to explore the eyes as opportunistic diagnostic windows for eye and systemic diseases. Moreover, in this study, the proposed deep convolutional network OcularSurfaceNet works in conjunction with a portable imaging instrument to facilitate on-site ophthalmic image acquisition and analysis.

This study has some limitations to be solved in the future. First of all, our model has only been applied to a single ethnicity, and it would be beneficial to evaluate the model with diverse ethnicity and geographical distribution. Secondly, the temporal response of ocular surface features according to changing blood glucose levels needs to be established. Thirdly, distraction from acute microvascular abnormalities of eyes, such as subconjunctival hemorrhage, needs to be minimized by refining algorithms. Furthermore, the performance varied with the age distribution in different samples (Supplementary Table 5). The participants in external test set 1 are much younger than those in the internal test set (Supplementary Table 1). Accordingly, the risk assessment model showed higher specificity and lower sensitivity in external test set 1 and lower specificity and higher sensitivity in the internal test set. However, in terms of early intervention for T2DM, higher sensitivity is preferred for elderly people, and a slight reduction in sensitivity could be tolerated for younger people since elderly people are more susceptible to T2DM and other chronic diseases.

In summary, we constructed a deep convolutional network for risk assessment and detection of T2DM in community screening by analyzing ophthalmic images of the ocular surface. We believe the proposed T2DM risk assessment and detection frameworks could be widely utilized in point-of-care settings and facilitate large-scale screening for early identification and intervention of T2DM.

Authors' contributions Z.Z., H.W., and L.C. contributed to the study design, data analysis, and manuscript drafting. J.C. conceived the original idea of detection of type 2 diabetes by ophthalmic images and, with L.W. and J.G., cosupervised this work and manuscript drafting. C.S.C., R.F.R. and R.X.Z. contributed to the development of the AI models. D.H., C.X.Z., C.L., Y.S.X., Z.H.F., F.E.L., H.M.P., Y.J.S., N.L.F., H.X.Z., Q.S., C.Z.B., and R.L.G. contributed to the data collection. W.D. and G.H.Y. contributed to the data preprocessing. R.D.H. and H.M.C. verified the data supporting this study. The authors read and approved the final manuscript.

Availability of data and materials Data archiving is not mandated, but the de-identified data may be available for research purposes from the corresponding authors on reasonable request.

Code availability The custom code is currently available only on request because it is under a patent examination process.

Declarations

Competing interests Author Jing Cheng is a member of the Editorial Board for Med-X. The paper was handled by another Editor and has undergone a rigorous peer review process. Author Jing Cheng was not involved in the journal's peer review of, or decisions related to, this manuscript.

Patents: ZZ, CN112686855A, CN113378794A, CN113889267A, CN115526888A, CN115775410A; CSC, CN112686855A, CN113378794A, CN113889267A, CN115526888A, CN115775410A; RXZ, CN113889267A; JC, CN112686855A, CN113378794A, CN113889267A, CN115526888A, CN115775410A.

References

1. World Health Organization. Global report on diabetes. https://apps.who.int/iris/rest/bitstreams/909883/retrieve. Accessed 21 Apr 2016.
2. Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the international diabetes federation diabetes Atlas, 9th edition. Diabetes Res Clin Pract. 2019;157:107843.
3. World Health Organization. Classification of diabetes mellitus. https://apps.who.int/iris/rest/bitstreams/1233344/retrieve. Accessed 21 Apr 2019.
4. Ogurtsova K, Guariguata L, Barengo NC, et al. IDF diabetes Atlas: Global estimates of undiagnosed diabetes in adults for 2021. Diabetes Res Clin Pract. 2022;183:109118.
5. Hu PL, Koh YLE, Tan NC. The utility of diabetes risk score items as predictors of incident type 2 diabetes in Asian populations: an evidence-based review. Diabetes Res Clin Pract. 2016;122:179–89.
6. Small KS, Todorčević M, Civelek M, et al. Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. Nat Genet. 2018;50(4):572–80.
7. Thomsen SK, Raimondo A, Hastoy B, et al. Type 2 diabetes risk alleles in PAM impact insulin release from human pancreatic β-cells. Nat Genet. 2018;50(8):1122–31.

8. Udler MS, McCarthy MI, Florez JC, et al. Genetic risk scores for diabetes diagnosis and precision medicine. Endocr Rev. 2019;40(6):1500–20.

9. Wang-Sattler R, Yu ZH, Herder C, et al. Novel biomarkers for pre-diabetes identified by metabolomics. Mol Syst Biol. 2012;8:615.

10. Rebholz CM, Yu B, Zheng ZH, et al. Serum metabolomic profile of incident diabetes. Diabetologia. 2018;61(5):1046–54.

11. Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. Diabet Med. 1998;15(7):539–53.

12. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402–10.

13. Ting DSW, Cheung CYL, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA. 2017;318(22):2211–23.

14. Cheung CY, Xu DJ, Cheng CY, et al. A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. Nat Biomed Eng. 2021;5(6):498–508.

15. Cheung N, Bluemke DA, Klein R, et al. Retinal arteriolar narrowing and left ventricular remodeling. The multi-ethnic study of atherosclerosis. J Am Coll Cardiol. 2007;50(1):48–55.

16. McGeechan K, Liew G, Macaskill P, et al. Prediction of incident stroke events based on retinal vessel caliber: a systematic review and individual-participant meta-analysis. Am J Epidemiol. 2009;170(11):1323–32.

17. Kawasaki R, Xie J, Cheung N, et al. Retinal Microvascular Signs and Risk of Stroke: The Multi-Ethnic Study of Atherosclerosis (MESA). Stroke. 2012;43(12):3245–51.

18. Ikram MK, Witteman JCM, Vingerling JR, et al. Retinal vessel diameters and risk of hypertension: the Rotterdam study. Hypertension. 2006;47(2):189–94.

19. Wong TY, Klein R, Nieto FJ, et al. Retinal microvascular abnormalities and 10-year cardiovascular mortality: a population-based case-control study. Ophthalmology. 2003;110(5):933–40.

20. Seidelmann SB, Claggett B, Bravo PE, et al. Retinal vessel calibers in predicting long-term cardiovascular outcomes: the atherosclerosis risk in communities study. Circulation. 2016;134(18):1328–38.

21. Xiao W, Huang X, Wang JH, et al. Screening and identifying hepatobiliary diseases through deep learning using ocular images: a prospective, multicentre study. Lancet Digit Health. 2021;3(2):e88–97.

22. Wang JJ, editor. Ophthalmology-based syndrome differentiation in diagnostics. Beijing: China Press of Traditional Chinese Medicine; 2013.

23. Zhang K, Liu XH, Xu J, et al. Deep-learning models for the detection and incidence prediction of chronic kidney disease and type 2 diabetes from retinal fundus images. Nat Biomed Eng. 2021;5(6):533–45.

24. Tabák AG, Herder C, Rathmann W, et al. Prediabetes: a high-risk state for diabetes development. Lancet. 2012;379(9833):2279–90.

25. Chinese Diabetes Society, Guideline for the prevention and treatment of type 2 diabetes mellitus in China (2020 edition). Chin J Diabetes Mellitus. 2021;13:317–411.

26. Xue N, Jiang K, Li Q, et al. Original askiatic imaging used in Chinese medicine eye-feature diagnosis of visceral diseases. J Innov Opt Health Sci. 2018;11(04):1850023.

27. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv. Accessed 18 May 2015 https://doi.org/10.48550/arXiv.1505.04597.

28. Paszke A, Gross S, Massa F, et al. An Imperative Style, High-Performance Deep Learning Library. Vancouver, Canada: NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems; 2019. Dec 8–14.

29. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016; 2016-Decem: 770–778.

30. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics. 2020;21(1):1–13.

31. Selvaraju RR, Cogswell M, Das A, et al. Visual Explanations from Deep Networks via Gradient-Based Localization. Venice, Italy: 2017 IEEE International Conference on Computer Vision (ICCV); 2017. Oct 22–29.

32. Deng J, Dong W, Socher R, et al. mageNet: A Large-Scale Hierarchical Image Database. Miami, FL, USA: 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009. June 20–25.

Authors and Affiliations

Zhi Zhang[1,2] · Han Wang[3] · Lei Chen[4] · Chensi Cao[1,2] · Tengwen Liu[5] · Ruifang Ren[1,2] · Ruixing Zhou[1,2] · Rudan Huang[1,2] · Dan Hu[4] · Chenxing Zhu[4] · Chong Lu[6] · Yunsheng Xu[7] · Zhaohui Fang[8] · Fuer Lu[9] · Huimin Pan[10] · Yanjin Su[11] · Nanlin Fu[12] · Huixia Zhan[13] · Qin Si[14] · Chenze Bai[14] · Ri Le Ge[14] · Hongmei Cao[1,2] · Wei Dong[1,2] · Guohui Yang[1,2] · Lan Wu[15] · Jiao Guo[16,17,18,19] · Jing Cheng[1,2,3]

✉ Lan Wu
   wulan@immu.edu.cn

✉ Jiao Guo
   gyguoyz@163.com

✉ Jing Cheng
   jcheng@tsinghua.edu.cn

1   National Engineering Research Center for Beijing Biochip Technology, Beijing, China

2   CapitalBio Corporation, Beijing, China

3   Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China

4   Department of Neurology & Department of Clinical Research Management, West China Hospital, Sichuan University, Chengdu, China

5   School of Basic Medical Sciences, Chengdu University of Traditional Chinese Medicine, Chengdu, China

6   Department of Physical Examination, The First Affiliated Hospital/School of Clinical Medicine of Guangdong Pharmaceutical University, Guangzhou, China

7   Department of Endocrinology, The Second Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, China

8   Department of Endocrinology, The First Affiliated Hospital of Anhui University of Chinese Medicine, Hefei, China

9   Department of Integrative Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

10 Health Management Center, Foshan Hospital Affiliated to Shunde Hospital of Guangzhou University of Chinese Medicine, Guangzhou, China

11 Department of Endocrinology, Affiliated Hospital of Shaanxi University of Chinese Medicine, Xi'an, China

12 Famous Doctor Clinic, The Second Affiliated Hospital/School of Clinical Medicine of Guangdong Pharmaceutical University, Guangzhou, China

13 Famous Doctor Studio, The First Affiliated Hospital/School of Clinical Medicine of Guangdong Pharmaceutical University, Guangzhou, China

14 International Mongolia Hospital of Inner Mongolia, Hohhot, China

15 Mongolian Medicine School, Inner Mongolia Medical University, Hohhot, China

16 Guangdong Metabolic Diseases Research Center of Integrated Chinese and Western Medicine, Guangdong Pharmaceutical University, Guangzhou, China

17 Key Laboratory of Glucolipid Metabolic Disorder, Ministry of Education of China, Guangzhou, China

18 Institute of Chinese Medicine, Guangdong Pharmaceutical University, Guangzhou, China

19 Guangdong TCM Key Laboratory for Metabolic Diseases, Guangzhou, China