

RESEARCH

Open Access



# Establishment of an automatic diagnosis system for corneal endothelium diseases using artificial intelligence

Jing-hao Qu<sup>1,2†</sup>, Xiao-ran Qin<sup>3†</sup>, Zi-jun Xie<sup>1,2</sup>, Jia-he Qian<sup>3,4</sup>, Yang Zhang<sup>5</sup>, Xiao-nan Sun<sup>6</sup>, Yu-zhao Sun<sup>7</sup>, Rong-mei Peng<sup>1,2</sup>, Ge-ge Xiao<sup>1,2</sup>, Jing Lin<sup>8</sup>, Xiao-yan Bian<sup>9</sup>, Tie-hong Chen<sup>10</sup>, Yan Cheng<sup>11</sup>, Shao-feng Gu<sup>1,2</sup>, Hai-kun Wang<sup>1,2</sup> and Jing Hong<sup>1,2\*</sup>

<sup>†</sup>Jing-hao Qu, Xiao-ran Qin, Zi-jun Xie and Jia-he Qian contributed equally to this work.

\*Correspondence: hongjing196401@163.com

<sup>1</sup> Department of Ophthalmology, Peking University Third Hospital, Beijing, China  
Full list of author information is available at the end of the article

## Abstract

**Purpose:** To use artificial intelligence to establish an automatic diagnosis system for corneal endothelium diseases (CEDs).

**Methods:** We develop an automatic system for detecting multiple common CEDs involving an enhanced compact convolutional transformer (ECCT). Specifically, we introduce a cross-head relative position encoding scheme into a standard self-attention module to capture contextual information among different regions and employ a token-attention feed-forward network to place greater focus on valuable abnormal regions.

**Results:** A total of 2723 images from CED patients are used to train our system. It achieves an accuracy of 89.53%, and the area under the receiver operating characteristic curve (AUC) is 0.958 (95% CI 0.943–0.971) on images from multiple centres.

**Conclusions:** Our system is the first artificial intelligence-based system for diagnosing CEDs worldwide. Images can be uploaded to a specified website, and automatic diagnoses can be obtained; this system can be particularly helpful under pandemic conditions, such as those seen during the recent COVID-19 pandemic.

**Keywords:** Corneal endothelium disease, In vivo confocal microscopy, Enhanced compact convolutional transformer

## Introduction

The corneal endothelium is responsible for maintaining corneal transparency, and dysfunction of hexagonal corneal endothelial cells leads to corneal opacity and blindness. According to the 2020 statistical report of the Eye Bank Association of America (EBAA), 53.5% of all keratoplasty procedures were performed for corneal endothelium disease (CED), and 76.7% of the indications for endothelial keratoplasty (EK) were for CED [1]; in addition, the number of EK procedures increased 15.3% to 30,098 in 2021 [2]. Similar patterns have been observed in the United States and Europe. In 2016, 58.4% of patients in Germany underwent keratoplasty for CED [3]. These findings suggest that CEDs are

the most common indication for keratoplasty, and their incidence has increased in recent years. Ophthalmologists should be more aware of the incidence of CEDs worldwide.

The corneal endothelium is the innermost layer of the cornea, and CEDs cannot be accurately diagnosed without specific examination equipment, leading to very high rates of clinical missed diagnosis and misdiagnosis. CEDs include Fuchs' endothelial corneal dystrophy (FECD), posterior polymorphous corneal dystrophy (PPCD), bullous keratopathy, iridocorneal endothelial (ICE) syndrome and viral endotheliitis. In the past, diagnosing CEDs was difficult due to the lack of appropriate equipment. In recent years, with the advent of in vivo confocal microscopy (IVCM), the morphology and structure of corneal endothelial cells can be clearly observed and analysed in vivo, and even mild oedema of the cornea can be detected at the corneal endothelial level. This technological progress has great importance for the understanding and diagnosis of CED [4–8]. As the application of IVCM (HRT III) devices in ophthalmology continues to advance and experience accumulates, imaging features of diagnostic significance are constantly being refined and summarized, and the characteristics and diagnostic criteria of these diseases have been clarified. Unfortunately, however, the microscope itself, while providing clear images, cannot supply reports or directly provide a diagnosis, similar to computed tomography (CT) and magnetic resonance imaging (MRI). Therefore, images are sent to doctors for analysis, but diagnosis is often difficult because CEDs are not common in clinical practice; doctors have a very limited understanding of this disease, and no analysis software is provided for the machine. Moreover, China lacks systematic training on IVCM images and a detailed atlas of CEDs. As a result, the ability to read IVCM images remains limited, and IVCM is often not used effectively. Therefore, the ability to diagnose CEDs in China remains inadequate.

Artificial intelligence (AI) has demonstrated rapid advancements in disease diagnosis. In ophthalmology, substantial progress has been made in the diagnosis of fundus diseases using deep neural networks [9–12] and the detection of glaucomatous optic neuropathy with multimodal machine learning [13]. However, AI-based diagnosis of corneal diseases is in its infancy and has focused mainly on corneal endothelial cell (CEC) morphology [14], keratitis [15, 16] and keratoconus [17, 18]. According to a literature review, there is no research on the diagnosis of CEDs using AI technology.

The aim of this study is to develop an automatic diagnostic system to identify FECD, PPCD, owl eye cells in cytomegalovirus (CMV) infection, viral endotheliitis (other CEDs) and normal corneas using AI technology. Through observation and investigation of the imaging characteristics of different endothelial diseases, we identify three aspects that should be considered when designing a proposed model. First, local features and long-range context information are both useful for improving the discriminability of representations, and not all regions in IVCM images equally contribute to helping identify diseases. For example, guttata are markers of FECD, and owl eye cells are markers of CMV infection, both of which occupy limited areas. However, in corneas with PPCD, there is a wide area of abnormal regions, such as craters or ridges, on the corneal endothelium. Second, certain spatial interactions and implicit relationships among abnormal regions or cells may be key to avoiding misdiagnosis. For example, guttata can sometimes be found in corneas with PPCD, which may be confused with FECD. However, guttata appear in the middle of the cornea and spread to other parts

of the cornea in FECD, while they are distributed along the ridges in PPCD. Third, our dataset for endothelial diseases is relatively small compared with public DR grading datasets. Therefore, models suitable for small-scale datasets are more effective for this task.

After considering the above aspects, in this paper, we incorporate a cross-head relation-aware self-attention mechanism and a token-attention feed-forward network (TaFFN) into a compact convolutional transformer (CCT) [19] to enhance its discriminability, termed an enhanced CCT (ECCT), for diagnosing CEDs from IVCN images. The CCT is designed by adding convolutional blocks to generate tokens from input images with the goal of maintaining local information and reducing the computational burden for subsequent transformer blocks. Therefore, the CCT not only combines local features and global representations but is also suitable for small datasets. Based on the CCT, we introduce a cross-head relative position encoding (CHRPE) scheme to the standard multihead self-attention module with the goal of capturing spatial relationships and semantic information among different tokens. Inspired by LocalViT [20], we adopt a TaFFN to adaptively learn the importance of each token for different inputs.

Overall, our contributions are as follows:

- 1) To our knowledge, this is the first study to utilize deep learning methods to automatically diagnose CEDs from IVCN images; these methods can assist ophthalmologists in clinical diagnosis and promote the application of IVCN.
- 2) We propose a CHRPE scheme to aggregate the spatial interactions and contextual information among different regions. To give more weight to valuable abnormal regions, we propose a TaFFN.
- 3) The experimental results show that our ECCT is superior in identifying endothelium diseases compared to certain popular convolutional neural network (CNN)-based methods and transformer-based methods.

## Methods

### Image capture

In this prospective study, images of corneal endothelial cells are acquired using an IVCN system (HRT III Rostock Cornea Module [RCM]; Heidelberg Engineering GmbH, Heidelberg, Germany). The specific IVCN image acquisition steps have been described previously [21]. The images are taken from the focal zone of the cornea using section mode and saved in JPG format with 8-bit grey levels and a size of  $384 \times 384$  pixels ( $400 \times 400 \mu\text{m}$ ). The study is performed according to the tenets of the Declaration of Helsinki and was approved by the institutional review board of Peking University Third Hospital (PUTH) (IRB00006761-M2022834). All participants provided written informed consent to participate in the study.

### Procedure

First, we select IVCN endothelial images from CED patients. CEDs are diagnosed by corneal specialists (J H, GG X and RM P) in the ophthalmology department of PUTH. CMV endotheliitis is confirmed by reverse transcription-polymerase chain reaction (RT-PCR). Next, the images are used to train our automatic diagnosis system. Seven

Chinese hospitals (Beijing Tongren Hospital, Shenyang the Fourth Hospital of People, The First Hospital of China Medical University, The Affiliated Hospital of Qingdao University, Baotou Chaoju Eye Hospital, Liaoning Aier Eye Hospital and The First Affiliated Hospital of Northwest University) supply data to construct the multicentre test set, which is used to test the automatic diagnosis system (Fig. 1). The diagnosis of CED is reviewed by a corneal specialist (J H) from the PUTH Ophthalmology Department. An example IVCM image of a CED is shown in Fig. 2.

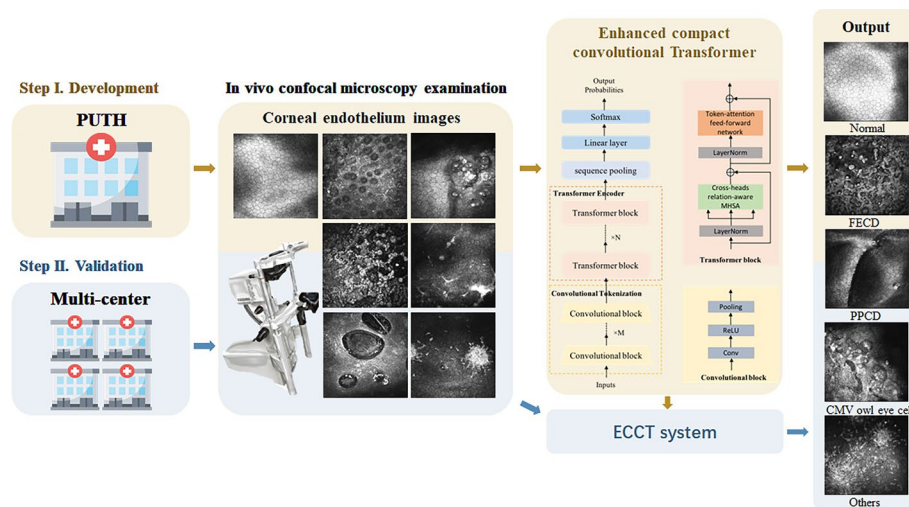
**Datasets**

The datasets include IVCM images from the PUTH ophthalmology department and the multicentre cohort. A total of 3723 images are included in the PUTH dataset, which is divided into a development set and a testing set. The development set (3110 images) is used to train (2723 images) and validate (387 images) the model. The testing set (613 images) is used to test the model. The images are divided so that data from the same patient are not include in both the development set and the testing set. The total number of images for each of the diseases in the PUTH dataset is shown in Table 1.

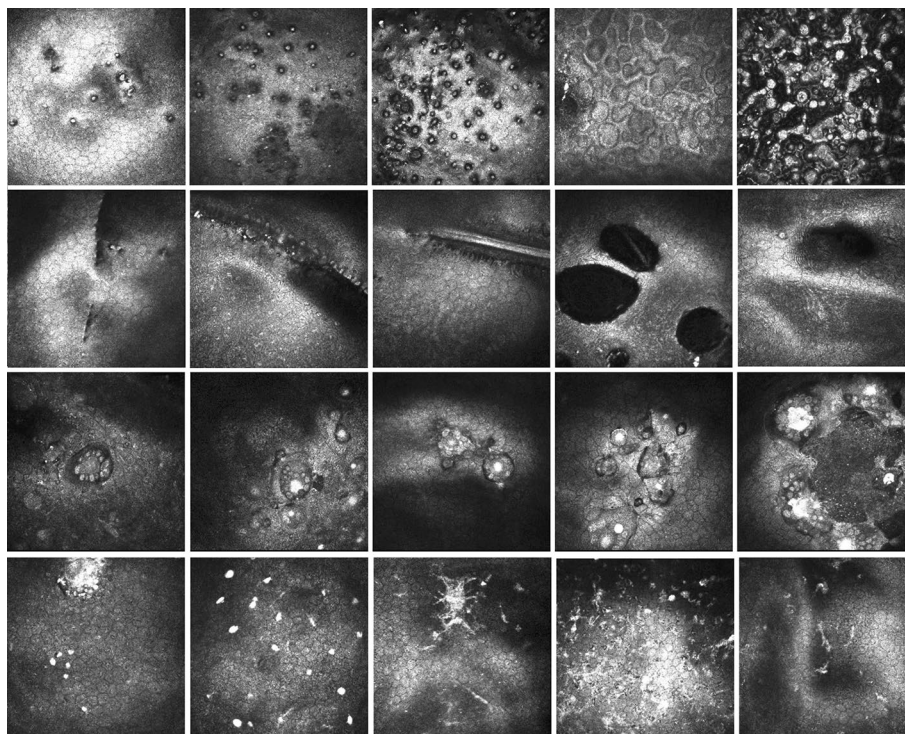
A total of 449 IVCM images from multiple centres are included in the testing set.

**Development of the automated algorithm**

CNNs have achieved great success in various medical image analysis tasks. The convolution operations in CNNs utilize convolution kernels with shared weights to interact with input images, and their limited receptive field cannot establish long-range feature dependencies. Recently, transformers with a self-attention mechanism were employed to capture long-range information and global representations. The transformer was first introduced to solve problems in natural language processing [22], in which it has demonstrated excellent performance. Subsequently, the vision transformer (ViT) first applied a standard transformer for image recognition and achieved great performance [23]. The authors of the ViT argued that transformers, unlike



**Fig. 1** Summary flow chart of our research. The brown arrow shows the training procedure of the automatic diagnosis system. The blue arrow shows the validation procedure using CED images from multiple centres



**Fig. 2** Example IVCM images of CEDs included in our study. The first row shows FECD of different severities; the second line shows different kinds of PPCD; the third line shows different kinds of owl eye cells; and the fourth line shows positive examples for the “others” group

**Table 1** Characteristics of the datasets

Item	PU TH dataset		Multicentre dataset	
Total no. of images	3723		449	
No. of subjects	301		231	
Age, mean/range	51.9/5–88		52.4/7–87	
No. of males	140		110	
No. of females	161		121	
	Training set	Validation set	Test set	Test set
Normal	167 (6%)	33 (9%)	40 (7%)	32 (8%)
FECD	279 (10%)	38 (10%)	100 (16%)	239 (53%)
PPCD	1508 (56%)	213 (55%)	284 (46%)	50 (11%)
CMV owl eye cell	576 (21%)	82 (21%)	124 (20%)	24 (5%)
Others	193 (7%)	21 (5%)	65 (11%)	104 (23%)
Total	2723	387	613	449

CNNs, lack inductive biases and must be trained on large-scale datasets to eliminate inductive biases. Therefore, some studies have attempted to add locality to transformers, producing systems such as the CCT [19]. The CCT combines local features and global representations while reducing the computational burden for the standard transformer, thus making it suitable for small-scale learning in medical research.

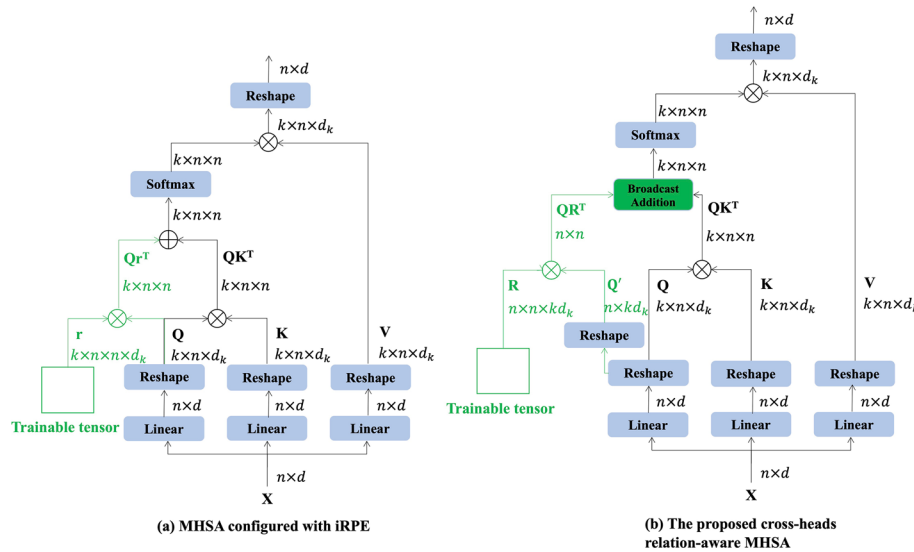
We briefly revisit the CCT as follows. It consists of two parts: a convolutional tokenization and a transformer encoder followed by sequence pooling (SeqPool). Given an input image, several convolutional blocks, each of which contains a convolutional layer, a rectified linear unit (ReLU) activation function, and a max pooling layer, are used to generate tokens (a sequence of vectors) [24]. Then, the transformer encoder takes the tokens as input and utilizes a series of stacked transformer blocks to extract global features. Each transformer block comprises two sublayers: a multihead self-attention (MHSA) module and a feed forward network (FFN). The normalization layer and the residual connection are applied to the two sublayers. Finally, to predict the final class index, the SeqPool module is used to pool the output sequential embeddings of the transformer encoder in a learnable attention scheme and generate probability estimates for different class labels.

We propose a transformer-based model based on the CCT [19] for automatically diagnosing CEDs from IVCN images. The diagnosis task is regarded as a five-class classification problem (normal, FECD, PPCD, CMV and others). To identify the correct CED, two factors are considered: (1) certain spatial interactions and implicit relationships between abnormal regions or cells may be key to avoiding misdiagnosis, and (2) not all regions in the IVCN images contribute equally to disease diagnosis. Based on the above observations, we incorporate a cross-head relation-aware self-attention mechanism and a TaFFN into the CCT to enhance its discriminability, producing an ECCT. Specifically, a novel CHRPE scheme, which utilizes cross-head features to capture spatial relationships and semantic information among different regions, is introduced to the standard MHSA module. The TaFFN employs a token-attention scheme to adaptively learn the importance of each token and substitutes for the conventional FFN.

### Cross-head relative position encoding

Transformers cannot explore the order of sequential tokens. Therefore, position encoding methods, including absolute and relative position encoding, have been studied recently to add token location information. For absolute position encoding [22], the encodings are learnable or generated from sinusoidal functions with different frequencies and then added to the input tokens. The ViT [23] utilizes this approach. Relative position encoding [25] focuses on the pairwise distances between sequential tokens and was further improved by Transformer-XL [26] and image RPE (iRPE) [27]. In this paper, we use relative position encoding to extract implicit relationships among different regions in IVCN images. The authors of iRPE [27] introduced two relative position modes, bias and contextual, where the bias mode represents encodings as learnable scalars that are independent of input tokens, and the contextual mode represents encodings as trainable vectors that need to interact with query embedding. The encodings are applied to each attention head independently of the MHSA module, as shown in Fig. 3a. Specifically, for an input sequence  $X \in R^{n \times d}$ , an MHSA module configured with iRPE runs self-attention  $k$  times (i.e.,  $k$  attention heads) in parallel, which can be formulated as follows:

$$MHSA(Q, K, V) = softmax\left(\frac{QK^T + B}{\sqrt{d_k}}\right)V \quad (1)$$



**Fig. 3** Comparison between **a** the multihead self-attention (MHSA) module configured with image RPE (iRPE) and **b** the proposed cross-head, relation-aware MHSA. The green areas are newly added parts

where the query  $Q$ , the key  $K$  and the value  $V$  are generated by applying projection matrices to  $X$  and reshaping; generally, we have  $Q, K, V \in R^{k \times n \times d_k}$  and  $d_k = d/k$ .  $B \in R^{k \times n \times n}$  is the relative position encoding for  $k$  heads. In contextual mode, each  $B_{k_0ij} \in R$  in the  $B$  matrix is calculated as follows:

$$B_{k_0ij} = Q_{k_0i} r_{k_0ij}^T \tag{2}$$

where  $k_0 \in [0, k)$  is the  $k_0$ -th head,  $i, j \in [0, n)$  are position indices, and  $r_{k_0ij} \in R^{d_k}$  is a trainable vector that interacts with query embedding  $Q_{k_0i}$ .  $r_{k_0ij}$  can also be operated on both query and key embeddings. To represent the relative position on 2D feature maps,  $r_{k_0ij}$  can be defined by multiple mapping methods following the original iRPE [27].

However, employing iRPE at each attention head independently ignores information from other heads, which may cause performance degradation. Especially in contextual mode, representations from multiple heads can help to learn more sufficient semantic information. On the other hand, the pairwise positional relationships between tokens are the same for all attention heads; thus, it is reasonable to maintain consistent relative position encoding among various attention heads. Therefore, we design our CHRPE based on iRPE to utilize cross-head embeddings to obtain richer encodings. Specifically, query  $Q$  is reshaped to  $Q' \in R^{n \times kd_k}$  to aggregate cross-head features, and multiplication of the trainable vectors  $R_{ij} \in R^{kd_k}$  and  $Q_{i'}$  is performed to generate positional encoding. Finally, the relative position encoding matrix  $B$  can be broadcast-added to the attention maps in each head. As illustrated in Fig. 3b, our cross-head relation-aware MHSA can be formulated as follows:

$$CH - MHSA(Q, K, V) = softmax \left( \frac{QK^T \oplus B}{\sqrt{d_k}} \right) V \tag{3}$$

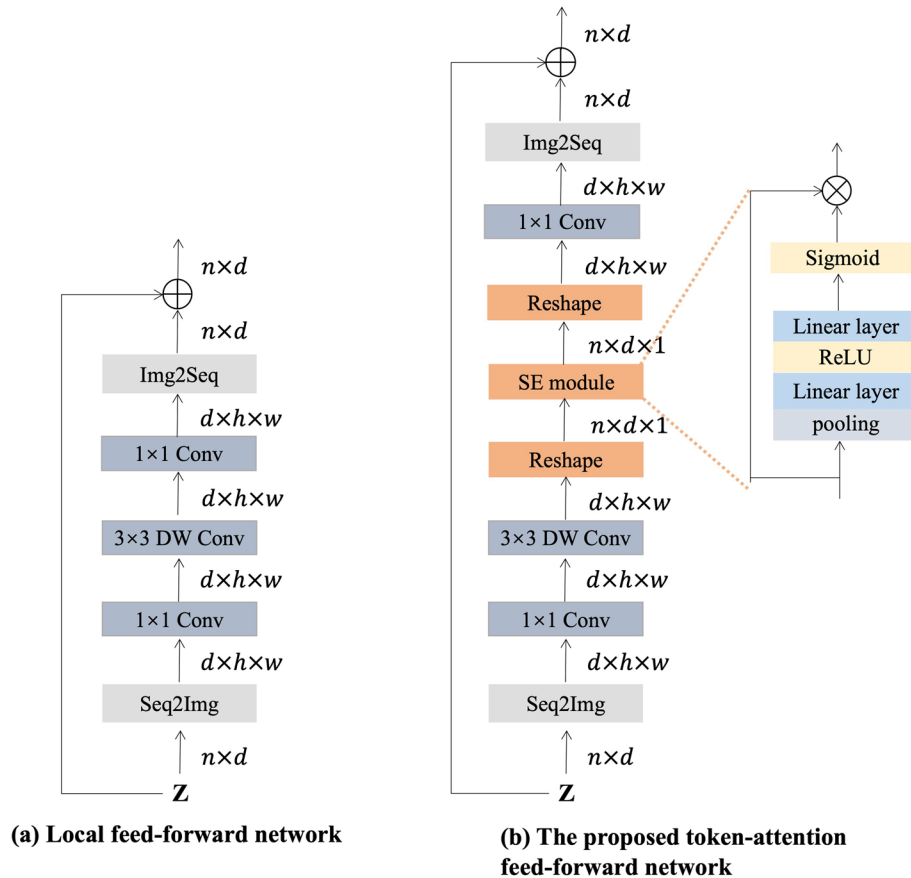
$$B_{ij} = Q_i R_{ij}^T \tag{4}$$

where  $\oplus$  is the broadcast addition. Additionally, the proposed cross-head relation-aware MHSA configured with CHRPE also provides certain interactions of information among different heads.

**Token-attention feed-forward network**

In a standard transformer, the FFN is composed of two fully connected layers that establish global information along the embedding dimension. A nonlinear activation function is applied in the hidden layer. To add more locality, LocalViT [20] incorporates a novel FFN that first converts the input sequence to a 2D feature map, then performs two  $1 \times 1$  convolutions along with a depthwise convolution, and finally converts the feature map back to a sequence, as shown in Fig. 4a.

In this study, not all regions in IVCM images contribute equally to identifying CEDs, and each disease has its own characteristic area that should be given additional attention. To determine the importance of each token, we propose a TaFFN inspired by LocalViT. Specifically, we reshape the 2D feature map so that a



**Fig. 4** Comparison between **a** the local feed-forward network and **b** the proposed token-attention feed-forward network (TaFFN). The orange areas are newly added parts



squeeze-and-excitation (SE) module [28] can be applied on the tokenwise dimension. As shown in Fig. 4b, for an input sequence  $Z \in R^{n \times d}$ , our TaFFN can be formulated as follows:

$$Z^r = Seq2Img(Z) \in R^{d \times h \times w} \quad (5)$$

$$U^r = W_d \sigma(W_1 Z^r) \quad (6)$$

$$V^r = R^{-1}(SE(R(U^r))) \quad (7)$$

$$Y^r = W_2 V^r \quad (8)$$

$$Y = Img2Seq(Y^r) \in R^{n \times d} \quad (9)$$

where  $h$  and  $w$  are the height and width of the 2D feature map, respectively;  $W_1$  and  $W_2$  are the two  $1 \times 1$  convolutions;  $W_d$  is the depthwise convolution; and  $\sigma$  is the nonlinear activation function.  $R$  represents a reshaping operation that converts  $U^r \in R^{d \times h \times w}$  to  $U \in R^{n \times d \times 1}$ . The SE module learns the importance in the token dimension and weights  $U$  to generate  $V$ , and then,  $R^{-1}$  converts  $V \in R^{n \times d \times 1}$  to  $V^r \in R^{d \times h \times w}$ . In this way, the model is configured to focus on regions that contain more information.

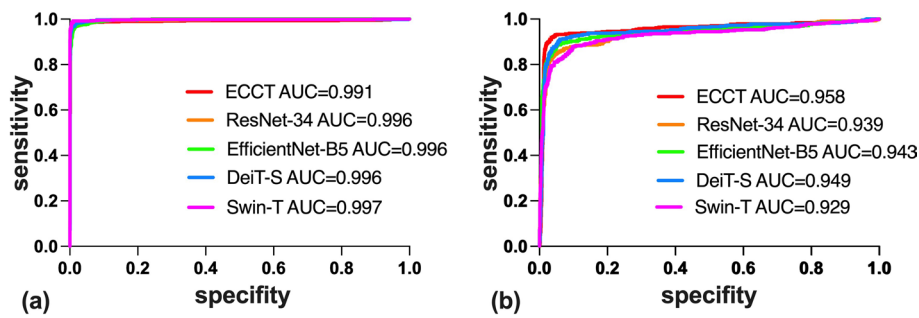
### Implementation details

To train the proposed model, the development set is randomly divided into a training set (2723 images) and a validation set (387 images). All the input images have a resolution of  $384 \times 384$  pixels, and the pixel values are normalized to values of 0–1. To reduce the risk of overfitting, data augmentation strategies, including random cropping, random horizontal flipping, random erasing [29], CutMix [30] and RandAugment [31], are adopted on the training set. For the structure of our proposed network, the number of convolutional blocks in the convolutional tokenization is set to 4 for  $16 \times$  downsampling. For the transformer encoder, the number of transformer blocks and the number of attention heads are set to 10 and 8, respectively; the dimension of sequential embeddings is set to 512. During training, we first train the proposed model on the ImageNet [32] dataset and then fine-tune the pretrained parameters on our own CED training set. For fine-tuning, the AdamW optimizer is used with a weight decay of  $5e^{-2}$  and a batch size of 40 [33]. We train the network for 100 epochs. The learning rate starts at  $5e^{-5}$  and then gradually decreases to  $1e^{-8}$  with the cosine annealing schedule [34].

## Results

### Results with the PUTH testing set

The total accuracy of our ECCT system in the PUTH testing set is 97.06%, and the area under the receiver operating characteristic curve (AUC) is 0.991 (95% CI 0.984–0.997) (Fig. 5a). The sensitivities for normal corneas, FECD, PPCD, owl eye cells in CMV infection and other conditions are 95.000%, 100.000%, 97.183%, 100.000% and 87.692%, respectively, and the corresponding specificities are 99.826%, 97.856%, 100.000%, 98.978% and 99.818%, respectively.



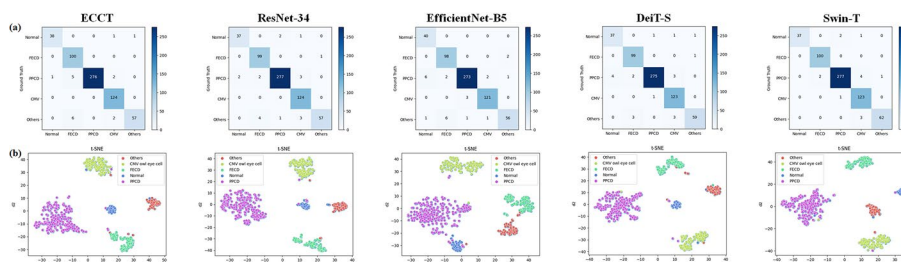
**Fig. 5** The AUCs of the different automatic diagnostic systems on the PUTH and multicentre datasets. **a** The AUCs of the different automatic diagnostic systems on the PUTH dataset. **b** The AUCs of the different automatic diagnostic systems on the multicentre dataset

**Results with a multicentre testing set**

The total accuracy in the multicentre testing set using the ECCT is 89.53%, and the AUC is 0.958 (95% CI 0.943–0.971) (Fig. 5b). The sensitivities for normal corneas and those with FECD, PPCD, owl eye cells in CMV infection and others are 96.875%, 98.326%, 100.000%, 91.667% and 61.539%, respectively, and the corresponding specificities are 98.801%, 94.762%, 96.742%, 96.236% and 99.420%, respectively.

**Comparison with other AI methods on the PUTH testing set**

To verify the performance of our method, we compare the ECCT with other state-of-the-art CNNs and ViTs on the PUTH testing set and the multicentre testing set. Two CNNs, ResNet [35] and EfficientNet [36], and two ViT models, DeiT [37] and Swin Transformer [38], are used for the comparisons. Specifically, we use ResNet-34, EfficientNet-B5, DeiT-S and Swin-T to balance the number of parameters with the proposed ECCT. For fairness, we also utilize ImageNet [32] for pretraining and adopt the same training configuration as the proposed ECCT. The total accuracy in the PUTH testing set using ResNet-34 is 96.90%, and the AUC is 0.996 (95% CI 0.994–0.998) (Fig. 5a). The sensitivities for normal corneas and those with FECD, PPCD, owl eye cells in CMV infection and others are 92.500%, 99.000%, 97.535%, 100.000% and 87.692%, respectively; the corresponding specificities are 99.651%, 98.830%, 99.088%, 98.569% and 99.818, respectively. The total accuracy in the PUTH testing set using EfficientNet-B5 is 95.92%, and the AUC is 0.997 (95% CI 0.995–0.998) (Fig. 5a). The sensitivities for the above corneas are 100.000%, 98.000%, 96.127%, 97.581% and 86.154%, respectively, and the corresponding specificities are 98.778%, 98.440%, 98.784%, 99.387% and 99.453%, respectively. The total accuracy in the PUTH testing set using DeiT-S is 96.74%, and the AUC is 0.996 (95% CI 0.992–0.999) (Fig. 5a). The sensitivities are 92.500%, 99.000%, 96.831%, 99.194% and 90.769%, respectively, and the corresponding specificities are 99.302%, 99.025%, 99.392%, 98.569% and 99.635%, respectively. The total accuracy in the PUTH testing set using Swin-T is 97.72%, and the AUC is 0.997 (95% CI 0.994–0.999) (Fig. 5a). The sensitivities are 92.500%, 100.000%, 97.535%, 99.194% and 95.385%, respectively, and the corresponding specificities are 100.000%, 99.610%, 99.088%, 98.364% and 99.818%, respectively. The confusion matrices of the different algorithms are shown in Fig. 6a. The t-distributed stochastic neighbour embedding (t-SNE) technique indicates that the features of each category learned by the ECCT algorithm are nearly as separable as those



**Fig. 6** Performance of the deep learning algorithms on the PUTH test dataset. **a** Confusion matrices describing the accuracies of the five deep learning algorithms. **b** Visualization by t-distributed stochastic neighbour embedding (t-SNE) of the separability of the features learned by the deep learning algorithms. Different coloured point clouds represent different categories of features

learned by ResNet-34, EfficientNet-B5, DeiT-S and Swin-T (Fig. 6b). The detailed performance of the five AI algorithms on the PUTH test dataset is shown in Tables 2 and 3.

**Comparison with other AI methods on the multicentre testing set**

The total accuracy on the multicentre testing set using ResNet-34 is 85.52%, and the AUC is 0.939 (95% CI 0.923–0.951) (Fig. 5b). The sensitivities for normal corneas and those with FECD, PPCD, owl eye cells in CMV infection and other conditions are 81.250%, 97.908%, 96.000%, 83.333% and 53.846%, respectively, and the specificities are 98.801%, 90.952%, 93.734%, 96.706% and 99.420%, respectively. The total accuracy on the multicentre testing set using EfficientNet-B5 is 85.52%, and the AUC is 0.943 (95% CI 0.925–0.958) (Fig. 5b). The sensitivities for the different types of corneas are 90.603%, 95.816%, 98.000%, 83.333% and 54.808%, respectively, and the specificities are 97.842%, 93.810%, 93.233%, 96.706% and 99.420%, respectively. The total accuracy on the multicentre testing set using DeiT-S is 86.41%, and the AUC is 0.949 (95% CI 0.934–0.961) (Fig. 5b). The sensitivities for the different types of corneas are 93.750%, 94.561%, 98.000%, 87.500% and 59.615%, respectively, and the specificities are 99.041%, 95.238%, 93.484%, 95.529% and 99.420%, respectively. The total accuracy on the multicentre testing set using Swin-T is 80.62%, and the AUC is 0.929 (95% CI 0.911–0.943) (Fig. 5b). The sensitivities for the different types of corneas are 84.375%, 88.285%, 94.000%, 75.000% and 56.731%, respectively, and the specificities are 98.082%, 85.714%, 92.983%, 97.177% and 97.391%, respectively. The confusion matrices of the different algorithms are shown in Fig. 7a. The t-SNE technique indicates that the features of each category learned by the ECCT algorithm are more separable than those learned by ResNet-34, EfficientNet-B5, DeiT-S and Swin-T (Fig. 7b). The detailed performances of the five AI algorithms on the multicentre test dataset are shown in Tables 2 and 3.

**Heatmaps**

To analyse the regions with the greatest contributions to the diagnosis of CEDs using our system, we generate a heatmap that visualizes the attention maps in the transformer blocks of the ECCT by using the attention rollout method [39]. For the CED findings, the heatmaps effectively highlight the regions containing lesions on the corneal

**Table 2** Performance of the five AI algorithms in the PUTH and multicentre test datasets

CED/algorithms	PUTH test dataset		Multicentre enter test dataset	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Normal				
ECCT	95.000	99.826	96.875	98.801
ResNet-34	92.500	99.651	81.250	98.801
EfficientNet-B5	100.000	98.778	90.625	97.842
DeiT-S	92.500	99.302	93.750	99.041
Swin-T	92.500	100.000	84.375	98.082
FECD				
ECCT	100.000	97.856	98.326	94.762
ResNet-34	99.000	98.830	97.908	90.952
EfficientNet-B5	98.000	98.441	95.816	93.810
DeiT-S	99.000	99.025	94.561	95.281
Swin-T	100.000	99.610	88.285	85.714
PPCD				
ECCT	97.183	100.000	100.000	96.742
ResNet-34	97.535	99.088	96.000	93.734
EfficientNet-B5	96.127	98.784	98.000	93.233
DeiT-S	96.831	99.392	98.000	93.484
Swin-T	97.535	99.088	94.000	92.983
Owl eye cells				
ECCT	100.000	98.978	91.667	96.235
ResNet-34	100.000	98.569	83.333	96.706
EfficientNet-B5	97.581	99.387	83.333	96.706
DeiT-S	99.194	98.569	87.500	95.529
Swin-T	99.194	98.364	75.000	97.177
Others				
ECCT	87.692	99.818	61.539	99.420
ResNet-34	87.692	99.818	53.846	99.420
EfficientNet-B5	86.154	99.453	54.808	99.420
DeiT-S	90.769	99.635	59.615	99.420
Swin-T	95.385	99.818	56.731	97.391

FECD: Fuchs' endothelial corneal dystrophy; PPCD: posterior polymorphous corneal dystrophy

endothelium. Typical examples of heatmaps for corneas with FECD, PPCD, owl eye cells in CMV infection, and other CEDs and for normal corneas are presented in Fig. 8.

We also compare the heatmaps from other methods, such as the class activation maps of ResNet-34 and the attention maps of DeiT-S, to our heatmap in Fig. 9. Compared to the class activation maps of ResNet-34, the attention mechanism in DeiT-S and ECCT tends to activate abnormal regions precisely for all cases due to its ability to handle long-range dependencies; compared to DeiT-S, ECCT captures more complete features, which illustrates its greater discriminative capacity for learned features.

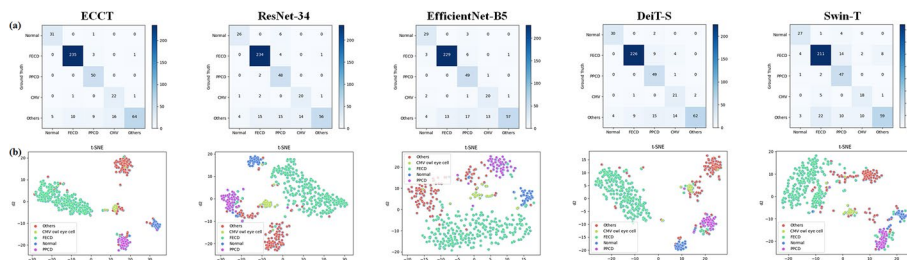
### Misclassified images

In the PUTH testing set, one “normal” image is misclassified as an “owl eye” image, and another “normal” image is misclassified as “others”. One “PPCD” image is misclassified as a “normal” image, five “PPCD” images are misclassified as “FECD”, and two

**Table 3** Overall performance of the five AI algorithms in the PUTH and multicentre test datasets

Algorithms	Pretraining	PUTH testing set		Multicentre testing set		Params (M)	MACs (G)	Throughput (Images/s)
		ACC (%)	AUC	ACC (%)	AUC			
ECCT	–	<u>96.09</u>	<u>0.993</u>	83.29	0.942	18.49	23.48	82
ResNet-34	–	94.94	<u>0.993</u>	81.95	<u>0.946</u>	21.29	10.8	280
EfficientNet-B5	–	87.92	0.979	78.39	0.935	28.35	7.08	65
DeiT-S	–	86.78	0.974	68.15	0.865	21.81	12.43	215
Swin-T	–	89.07	0.977	75.05	0.919	27.57	12.8	145
ECCT	ImageNet	97.06	0.991	<b>89.53</b>	<b>0.958</b>	18.49	23.48	82
ResNet-34	ImageNet	96.90	0.996	85.52	0.939	21.29	10.8	280
EfficientNet-B5	ImageNet	95.92	0.996	85.52	0.943	28.35	7.08	65
DeiT-S	ImageNet	96.73	0.996	86.41	0.949	21.81	12.43	215
Swin-T	ImageNet	<b>97.71</b>	<b>0.997</b>	80.62	0.929	27.57	12.8	145

Underlined numbers indicate the best results when training from scratch; bold numbers indicate the best results when fine-tuning with ImageNet pretraining

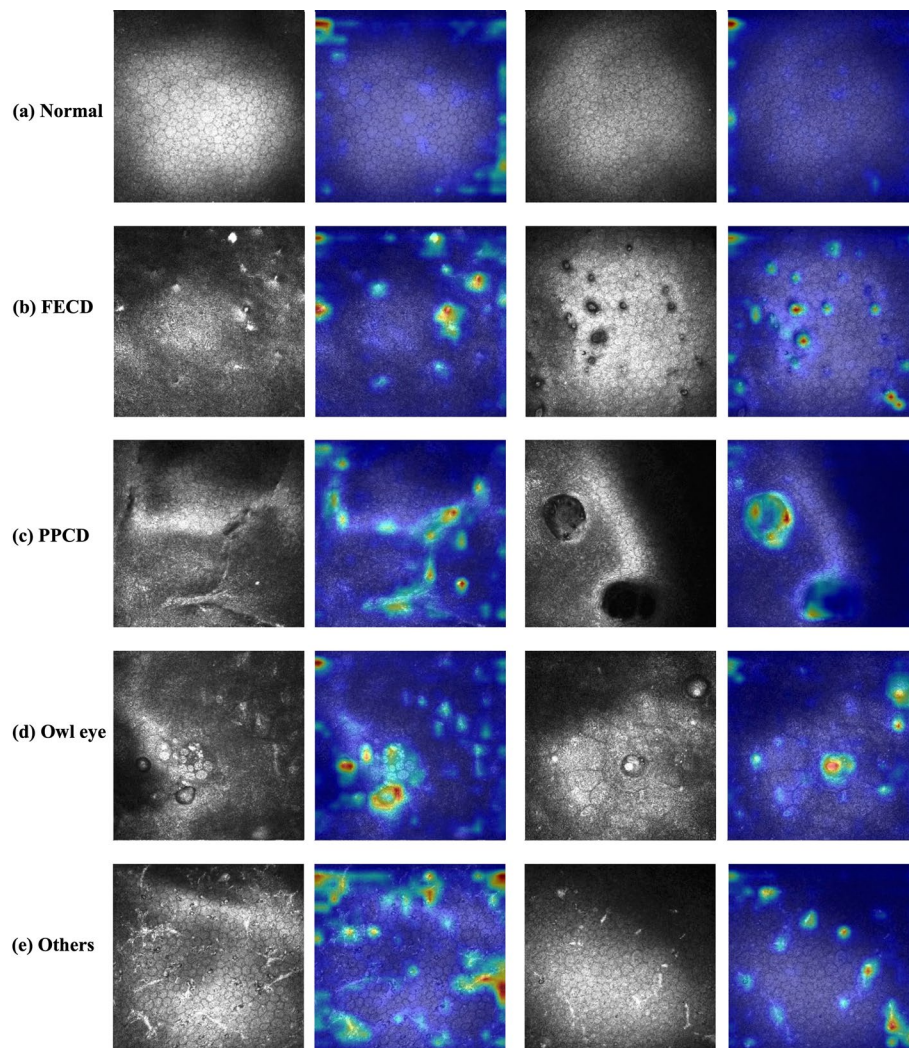


**Fig. 7** Performance of the deep learning algorithms on the multicentre test dataset. **a** Confusion matrices describing the accuracies of the five deep learning algorithms. **b** Visualization by t-distributed stochastic neighbour embedding (t-SNE) of the separability of the features learned by the deep learning algorithms. The differently coloured point clouds represent the different feature categories

“PPCD” images are misclassified as “owl eye”. Six “others” images are misclassified as “FECD”, and two “others” images are misclassified as “owl eye”. In the multicentre testing set, one “normal” image is misclassified as “PPCD”; three “FECD” images are misclassified as “PPCD”, and one “FECD” image is misclassified as “others”; one “owl eye” image is misclassified as “PPCD”, and one “owl eye” image is misclassified as “others”; five “others” images are misclassified as “normal”, ten “others” images are misclassified as “FECD”, nine “others” images are misclassified as “PPCD”, and sixteen “others” images are misclassified as “owl eye”. The details of the classification errors from the ECCT are described in Fig. 10.

**Ablation studies on the internal PUTH testing set**

To verify the effectiveness of each component of our method, we conduct experiments without ImageNet pretraining on the PUTH testing set. Specifically, we separately analyse the effects of CHRPE and TaFFN. To verify the effectiveness of CHRPE, we compare the results of three RPE options: without RPE, iRPE and CHRPE. The findings show that taking relative positional relationships into account is effective in extracting the characteristics of CEDs, and the proposed CHRPE performs better than the iRPE,

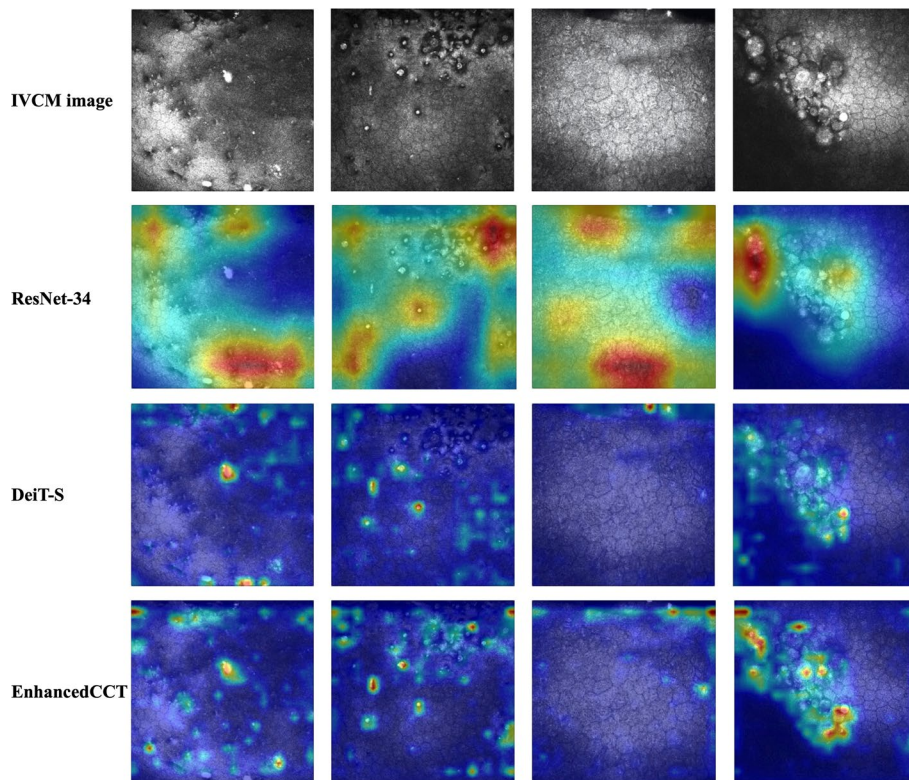


**Fig. 8** Colour heatmaps demonstrating typical findings for different corneas, shown in pairs with the original images (left) and the corresponding heatmaps (right) for each category. **a** Normal. **b**. FECD. **c** PPCD. **d** Owl eye cells in CMV infection. **e** Other CEDs

as shown in Table 4. For the feed-forward network (FFN), we also compare the results of three options: linear FFN, local FFN and TaFFN. A linear FFN is used in the standard transformer and is composed of two fully connected layers. A local FFN is the type of FFN adopted in LocalViT. As shown in Table 5, the performance of TaFFN is better than that of the other FFNs.

### Discussion

The total accuracy of the ECCT on the PUTH testing set is 97.06%, and the AUC is 0.991. Moreover, the total accuracy of the ECCT on the multicentre testing set is 89.53%, and the AUC is 0.958. The t-SNE technique shows that the features of each category learned by the ECCT algorithm are more separable than those of the other four AI algorithms on both the PUTH and multicentre testing datasets, as shown in Fig. 6b and Fig. 7b. As shown in Fig. 5 and Tables 2 and 3, ECCT not only performs well on the PUTH dataset



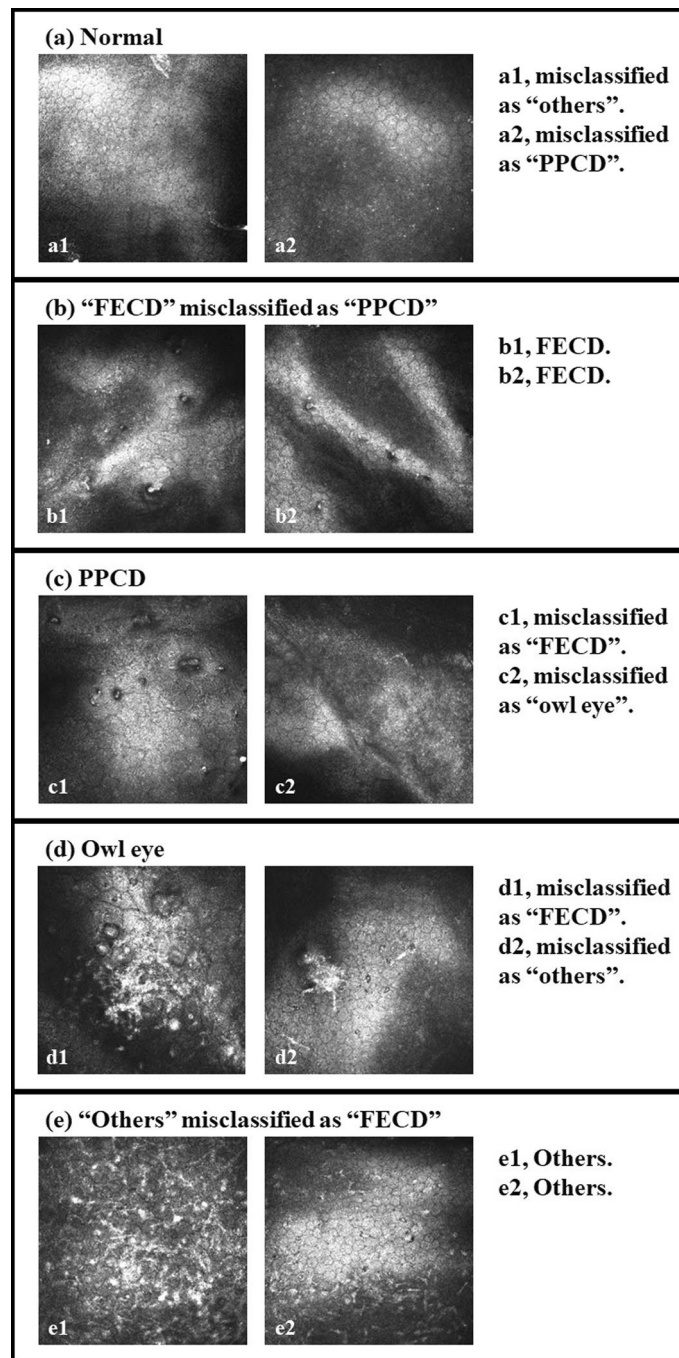
**Fig. 9** Comparison among class activation maps of ResNet-34, attention maps of DeiT-S and attention maps of ECCT. The first row shows the original images

but also achieves the best accuracy and sensitivity and significantly surpasses the other four AI algorithms on the multicentre dataset, which demonstrates the superiority of our system in generalizing to unseen images.

According to the heatmaps, the ECCT effectively highlights the regions containing lesions on the corneal endothelium. This finding suggests that the ECCT can accurately focus on the regions with lesions in CEDs, especially in PPCD and owl eye cell images, which are often ignored or unknown by junior ophthalmologists.

The sensitivity for “others” is relatively lower on the multicentre testing set because the “others” images in the PUTH dataset mainly focus on areas of disturbance that are similar to FECD, PPCD and owl eye cell images; consequently, the “others” images in the PUTH dataset are unable to depict all the alterations seen in endotheliitis. The “others” images from the multicentre dataset contain different diseases that are not found in the PUTH dataset, which is why these images are the most commonly misclassified.

While the automatic diagnosis of several diseases, such as diabetic retinopathy, diabetic macular oedema and keratitis, has been studied in ophthalmology, most of the systems were developed with large datasets (tens of thousands of images). However, for the diagnosis of CEDs based on IVCM images, there are no public datasets, and PPCD and CMV cases are relatively rare; thus, our datasets are relatively small. In the PUTH dataset, we collect multiple images from different corneal positions for each patient to increase the number of PPCD and CMV images. To learn discriminative feature representations on such a small-scale dataset, the proposed ECCT is based on the main



**Fig. 10** Typical examples of misclassified images

**Table 4** Ablation studies of CHRPE

RPE	ACC (%)	AUC
w/o RPE	94.78	0.989
iRPE	95.11	0.991
CHRPE	<b>96.09</b>	<b>0.993</b>



**Table 5** Ablation studies of TaFFN

FFN	ACC (%)	AUC
Linear FFN	93.15	0.991
Local FFN	93.96	0.989
TaFFN	<b>96.09</b>	<b>0.993</b>

architecture of the CCT, utilizing convolutional blocks to avoid overfitting and transformer blocks to capture long-range information. As shown in Table 2, we also conduct experiments without ImageNet pretraining (i.e., training from scratch). The table shows the superiority of the ECCT in both configurations. First, this shows that the ECCT can achieve reasonable performance by training from scratch on our relatively small dataset, while other transformer-based methods (DeiT-S and Swin-T) do not perform well. Second, pretraining on ImageNet significantly boosts the performance of all the methods, which shows that features learned from natural images are also helpful for medical image tasks.

Furthermore, the proposed architecture captures both local and global features for various patterns of endothelial diseases, as implied by the heatmap comparisons (Fig. 9) to other CNN- and transformer-based methods. Moreover, to establish contextual relationships among different lesion regions when designing the model, we integrate a CHRPE scheme into a standard multihead self-attention module by utilizing cross-head features to obtain richer encoding. In addition, a TaFFN is introduced to learn the importance of tokens for all transformer blocks. Ablation studies on the PUTH testing set demonstrate the advantages of adopting both CHRPE and TaFFN (Tables 4 and 5).

The prevalence of FECD is approximately 7.33%, and the total number of people aged > 30 years with FECD is currently estimated to be nearly 300 million. An increase of 41.7% in the number of FECD-affected patients is expected by 2050 [40]. The prevalence of FECD varies by race and geographic location. A study from Iceland (a white population) revealed that the prevalence of cornea guttata was 9.2% [41]. In Asia, the incidence of FECD is lower than that in Europe, with rates of 6.7% [42] and 4.1%, respectively, among Singaporean and Japanese individuals [43]. There are no data on the prevalence of FECD in China, which reflects the lack of diagnostic ability for this disease in the country. Therefore, developing an automatic diagnostic system for this disease is logical.

PPCD is a relatively rare, autosomal dominant disease. Ophthalmologists have a poor understanding of this disease, which can easily lead to missed diagnoses. Many asymptomatic PPCD cases are found and diagnosed during air force/civil aviation physical examinations in China [44]. IVCN reveals hyporeflexive, crater-like, vesicular lesions of different sizes on the corneal endothelium [45].

CMV endotheliitis is defined as corneal endothelium-specific inflammation triggered by CMV infection [46] and has been reported mainly in Asian countries [47]. The Japan Corneal Endotheliitis Study, which included the largest case series of 106 patients, reported that CMV endotheliitis is most common in middle-aged and older men [48]. The features of the owl eye morphology include large cells with nuclei presenting a highly reflective area surrounded by a halo of low reflection [7]. These

cells, which are considered pathognomonic for CMV, can be detected with IVCN, which may be helpful as an adjunct examination method. IVCN can assist in the evaluation of FECD guttae and owl eye cells. Our system can effectively distinguish between these two diseases. Images of other corneal endotheliitis patients were used as disturbance terms in our study, which is important for improving diagnostic accuracy. The characteristics of corneal endotheliitis on IVCN are diverse and might be confused with those of FECD, PPCD and owl eye cells.

IVCN is a very effective methodology for studying corneas and improving the diagnostic ability for CEDs. For the reasons mentioned above, the ability to diagnose CEDs in China remains low. We used an HRT III machine agent to establish five WeChat groups of 500 people each; these groups consisted of ophthalmologists who, every day, consulted on IVCN images within the groups. The development of this system can greatly improve the level of CED diagnosis. Due to the COVID-19 pandemic, free personnel flow between cities is sometimes restricted; with this system, ophthalmologists can upload images to a website and automatically obtain a diagnosis.

Although this study includes a large sample, it is still relatively small compared to that of other AI systems. More images of endotheliitis patients from multiple centres will be used to train and improve our system. The diagnosis of CEDs using the proposed system should be further confirmed through large-scale clinical trials.

## Conclusion

This is the first report of an AI diagnostic system for CEDs, and our results show that this system can achieve excellent diagnosability. IVCN is a reliable and effective diagnostic method for CEDs.

Once an ophthalmologist suspects CEDs after IVCN examination, the obtained image is input into our system, and the system automatically recognizes the image and assists in diagnosis to improve the ophthalmologist's understanding of CED.

However, images of endotheliitis patients are still rarer than those of other CEDs. In the future, additional images of endotheliitis patients from multiple centres will be used to train and improve our system. Moreover, the proposed system was tested on an ordinary computer, after which the system was tested online and run on a web page.

With the increased incidence of CEDs, this AI system will play a key role in the prevention of corneal blindness.

## Acknowledgements

We are very grateful to Dr. Xiao-yong Chen and Hong-yuan Cai (Department of Ophthalmology, Peking University Third Hospital, Beijing, China) for their hard work during the study period.

## Author contributions

Study design (GG Xiao/J Hong); study execution (JH Qu/XR Qin/ZJ Xie/JH Qian); collection and management of the data (JH Qu/RM Peng/ZJ Xie/Y Zhang/XN Sun/YZ Sun/TH Chen/XY Bian/J Lin); image capture (SF Gu/HK Wang); analysis and interpretation of the data (XR Qin/JH Qian); design and experimentation of network structure (JH Qian/XR Qin); writing of the manuscript (JH Qu/XR Qin/JH Qian); and review or approval of the manuscript (J Hong).

## Funding

This work was supported by the Peking University Medicine Sailing Program for Young Scholars' Scientific & Technological Innovation under Grant No. BMU2023YFJHPY018 and by the National Natural Science Foundation of China under Grant Nos. 81970768 and 81800801.

## Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Declarations

### Ethical approval and consent to participate.

The study is performed according to the tenets of the Declaration of Helsinki and was approved by the institutional review board of Peking University Third Hospital (PUTH) (IRB00006761-M2022834). All participants provided written informed consent to participate in the study.

### Consent for publication

Not applicable.

### Competing interests

The authors report no competing interests. The authors alone are responsible for the content and writing of the paper.

### Author details

<sup>1</sup>Department of Ophthalmology, Peking University Third Hospital, Beijing, China. <sup>2</sup>Beijing Key Laboratory of Restoration of Damaged Ocular Nerve, Peking University Third Hospital, Beijing, China. <sup>3</sup>Institute of Automation, Chinese Academy of Sciences, Beijing, China. <sup>4</sup>School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China. <sup>5</sup>Beijing Tongren Eye Center, Beijing Key Laboratory of Ophthalmology and Visual Sciences, Beijing Institute of Ophthalmology, Beijing Tongren Hospital, Capital Medical University, Beijing, China. <sup>6</sup>Shenyang the Fourth People's Hospital, Shenyang, Liaoning, China. <sup>7</sup>Department of Ophthalmology, The First Hospital of China Medical University, Shenyang, Liaoning, China. <sup>8</sup>Department of Ophthalmology, The Affiliated Hospital of Qingdao University, Qingdao, Shandong, China. <sup>9</sup>Baotou Chaoju Eye Hospital, Baotou, Neimenggu, China. <sup>10</sup>Liaoning Aier Eye Hospital, Shenyang, Liaoning, China. <sup>11</sup>Shanxi Provincial Ophthalmology Clinical Disease Research Center, The First Affiliated Hospital of Northwest University, Xi'an, Shanxi, China.

Received: 11 November 2023 Accepted: 7 April 2024

Published online: 08 May 2024

## References

1. Eye Bank Association of America. 2020 EYE BANKING STATISTICAL REPORT. 11 December. [https://restoresight.org/wp-content/uploads/2021/03/2020\\_Statistical\\_Report-Final.pdf](https://restoresight.org/wp-content/uploads/2021/03/2020_Statistical_Report-Final.pdf).
2. Eye Bank Association of America. 2021 EYE BANKING STATISTICAL REPORT. 11 December. <https://restoresight.org/members/publications/statistical-report/>.
3. Flockerzi E, Maier P, Böhringer D, et al. Trends in corneal transplantation from 2001 to 2016 in Germany: a report of the DOG-section cornea and its Keratoplasty registry. *Am J Ophthalmol.* 2018;188:91–8.
4. Aggarwal S, Cavalcanti BM, Regali L, et al. In Vivo confocal microscopy shows alterations in nerve density and dendritiform cell density in fuchs' endothelial corneal dystrophy. *Am J Ophthalmol.* 2018;196:136–44.
5. Guier CP, Patel BC, Stokkermans TJ, Gulani AC. Posterior Polymorphous Corneal Dystrophy. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing. 2022. <http://www.ncbi.nlm.nih.gov/books/NBK430880/>. Accessed 11 Dec 2022.
6. Malhotra C, Seth NG, Pandav SS, et al. Iridocorneal endothelial syndrome: evaluation of patient demographics and endothelial morphology by in vivo confocal microscopy in an Indian cohort. *Indian J Ophthalmol.* 2019;67:604–10.
7. Kobayashi A, Yokogawa H, Higashide T, et al. Clinical significance of owl eye morphologic features by in vivo laser confocal microscopy in patients with cytomegalovirus corneal endotheliitis. *Am J Ophthalmol.* 2012;153:445–53.
8. Peng R-M, Guo Y-X, Xiao G-G, et al. Characteristics of corneal endotheliitis among different viruses by in vivo confocal microscopy. *Ocul Immunol Inflamm.* 2021;29:324–32.
9. Cen L-P, Ji J, Lin J-W, et al. Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. *Nat Commun.* 2021;12:4828.
10. Abdulsahib AA, Mahmoud MA, Mohammed MA, et al. Comprehensive review of retinal blood vessel segmentation and classification techniques: intelligent solutions for green computing in medical images, current challenges, open issues, and knowledge gaps in fundus medical images. *Netw Model Anal Health Inform Bioinforma.* 2021;10:20.
11. Al-Fahdawi S, Al-Waisy AS, Zeebaree DQ, et al. Fundus-DeepNet: multi-label deep learning classification system for enhanced detection of multiple ocular diseases through data fusion of fundus images. *Inform Fusion.* 2024;102:102059.
12. Abdulsahib AA, Mahmoud MA, Aris H, et al. An automated image segmentation and useful feature extraction algorithm for retinal blood vessels in fundus images. *Electronics.* 2022;11:1295.
13. Xiong J, Li F, Song D, et al. Multimodal machine learning using visual fields and peripapillary circular OCT scans in detection of glaucomatous optic neuropathy. *Ophthalmology.* 2022;129:171–80.
14. Al-Waisy AS, Alruban A, Al-Fahdawi S, et al. CellsDeepNet: a novel deep learning-based web application for the automated morphometric analysis of corneal endothelial cells. *Mathematics.* 2022;10:320.
15. Li Z, Jiang J, Chen K, et al. Preventing corneal blindness caused by keratitis using artificial intelligence. *Nat Commun.* 2021;12:3738.
16. Tiwari M, Piech C, Baitemirova M, et al. Differentiation of active corneal infections from healed scars using deep learning. *Ophthalmology.* 2022;129:139–46.
17. Feng R, Xu Z, Zheng X, et al. KerNet: a novel deep learning approach for keratoconus and sub-clinical keratoconus detection based on raw data of the pentacam HR system. *IEEE J Biomed Health Inform.* 2021;25:3898–910.
18. Al-Timemy AH, Mosa ZM, Alyasseri Z, et al. A hybrid deep learning construct for detecting keratoconus from corneal maps. *Transl Vis Sci Technol.* 2021;10:16.
19. Hassani A, Walton S, Shah N, et al. Escaping the big data paradigm with compact transformers. 2022. <http://arxiv.org/abs/2104.05704>. Accessed 11 Dec 2022.

20. Li Y, Zhang K, Cao J, et al. LocalViT: Bringing locality to vision transformers. 2021. <http://arxiv.org/abs/2104.05707>. Accessed 23 May 2023.
21. Qu J-H, Qin X-R, Peng R-M, et al. A fully automated segmentation and morphometric parameter estimation system for assessing corneal endothelial cell images. *Am J Ophthalmol*. 2022;239:142–53.
22. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. 2017. <http://arxiv.org/abs/1706.03762>. Accessed 11 Dec 2022.
23. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale. 2021. <http://arxiv.org/abs/2010.11929>. Accessed 11 Dec 2022.
24. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines vinod nair. *Int Conf Int Conf Mach Learn*. 2010.
25. Shaw P, Uszkoreit J, Vaswani A. Self-attention with relative position representations. 2018. <http://arxiv.org/abs/1803.02155>. Accessed 23 May 2023.
26. Dai Z, Yang Z, Yang Y, et al. Transformer-XL: attentive language models beyond a fixed-length context. 2019. <http://arxiv.org/abs/1901.02860>. Accessed 23 May 2023.
27. Wu K, Peng H, Chen M, et al. Rethinking and improving relative position encoding for vision transformer. 2021. <http://arxiv.org/abs/2107.14222>. Accessed 23 May 2023.
28. Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell*. 2020;42:2011–23.
29. Zhong Z, Zheng L, Kang G, et al. Random erasing data augmentation. *Proc AAAI Conf Artificial Intell*. 2020;34:13001–8.
30. Yun S, Han D, Oh SJ, et al. CutMix: Regularization strategy to train strong classifiers with localizable features. 2019. <http://arxiv.org/abs/1905.04899>. Accessed 11 Dec 2022.
31. Cubuk ED, Zoph B, Shlens J, Le QV. RandAugment: Practical automated data augmentation with a reduced search space. 2019. <http://arxiv.org/abs/1909.13719>. Accessed 11 Dec 2022.
32. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vision*. 2015;115:211–52.
33. Loshchilov I, Hutter F. Decoupled weight decay regularization. 2019. <http://arxiv.org/abs/1711.05101>. Accessed 11 Dec 2022.
34. Loshchilov I, Hutter F. SGDR: Stochastic gradient descent with warm restarts. 2017. <http://arxiv.org/abs/1608.03983>. Accessed 11 Dec 2022.
35. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2015. <http://arxiv.org/abs/1512.03385>. Accessed 11 Dec 2022.
36. Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. 2020. <http://arxiv.org/abs/1905.11946>. Accessed 22 Jan 2023.
37. Touvron H, Cord M, Douze M, et al. Training data-efficient image transformers & distillation through attention. *arXiv*. 2021. <https://doi.org/10.48550/arXiv.2012.12877>.
38. Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows. *arXiv*. 2021. <https://doi.org/10.48550/arXiv.2103.14030>.
39. Abnar S, Zuidema W. Quantifying attention flow in transformers. *arXiv*. 2020. <https://doi.org/10.48550/arXiv.2005.00928>.
40. Aiello F, Gallo Afflitto G, Ceccarelli F, et al. Global prevalence of fuchs endothelial corneal dystrophy (FECD) in adult population: a systematic review and meta-analysis. *J Ophthalmol*. 2022;2022:3091695.
41. Zoega GM, Fujisawa A, Sasaki H, et al. Prevalence and risk factors for cornea guttata in the Reykjavik eye study. *Ophthalmology*. 2006;113:565–9.
42. Kitagawa K, Kojima M, Sasaki H, et al. Prevalence of primary cornea guttata and morphology of corneal endothelium in aging Japanese and Singaporean subjects. *Ophthalmic Res*. 2002;34:135–8.
43. Higa A, Sakai H, Sawaguchi S, et al. Prevalence of and risk factors for cornea guttata in a population-based study in a southwestern island of Japan: the Kumejima study. *Arch Ophthalmol*. 2011;129:332–6.
44. Gu SF, Peng RM, Xiao GG. Imaging features of posterior polymorphous corneal dystrophy observed by in vivo confocal microscopy. *Hong J Zhonghua Yan Ke Za Zhi*. 2022;58:103–11.
45. Bozkurt B, Irkek M, Mocan MC. In vivo confocal microscopic findings in posterior polymorphous corneal dystrophy. *Cornea*. 2013;32:1237–42.
46. Ding K, Nataneli N. Cytomegalovirus corneal endotheliitis In: *StatPearls*. Treasure Island: StatPearls Publishing; 2022.
47. Joye A, Gonzales JA. Ocular manifestations of cytomegalovirus in immunocompetent hosts. *Curr Opin Ophthalmol*. 2018;29:535–42.
48. Koizumi N, Inatomi T, Suzuki T, et al. Clinical features and management of cytomegalovirus corneal endotheliitis: analysis of 106 cases from the Japan corneal endotheliitis study. *Br J Ophthalmol*. 2015;99:54–8.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.