DeepGF: Glaucoma Forecast Using the Sequential Fundus Images

Liu Li[1⋆[0000−0003−2376−8162]], Xiaofei Wang[1⋆[0000−0001−9000−9022]], Mai Xu[1,2(✉)[0000−0002−0277−3301]], Ximeng Chen[3], and Hanruo Liu[4]

[1] School of Electronic and Information Engineering, Beihang University, Beijing, China
correspondence author(✉): maixu@buaa.edu.cn
[2] Hangzhou Innovation Institute, Beihang University
[3] Air Force General Hospital, Beijing, China
[4] Beijing Tongren Hospital, Capital Medical University, Beijing, China

Abstract. Disease forecast is an effective solution to early treatment and prevention for some irreversible diseases, e.g., glaucoma. Different from existing disease detection methods that predict the current status of a patient, disease forecast aims to predict the future state for early treatment. This paper is a first attempt to address the glaucoma forecast task utilizing the sequential fundus images of a patient. Specifically, we establish a database of sequential fundus images for glaucoma forecast (SIGF), which includes an average of 9 images per eye, corresponding to 3,671 fundus images in total. Besides, a novel deep learning method for glaucoma forecast (DeepGF) is proposed based on our SIGF database, consisting of an attention-polar convolution neural network (AP-CNN) and a variable time interval long short-term memory (VTI-LSTM) network to learn the spatio-temporal transition at different time intervals across sequential medical images of a person. In addition, a novel active convergence (AC) training strategy is proposed to solve the imbalanced sample distribution problem of glaucoma forecast. Finally, the experimental results show the effectiveness of our DeepGF method in glaucoma forecast.

Keywords: Glaucoma forecast · SIGF database · DeepGF method.

1 Introduction

The early prediction of disease deterioration plays an important role in supporting clinicians, as an approximated 11% of deaths in hospital follow a failure of promptly recognization and treatment [17]. Therefore, forecasting the potential deterioration ahead of time has become an urgent demand in medical diagnosis, especially for the incurable disease like glaucoma, which is the leading causes of irreversible blindness [2].

Different from the existing disease detection methods that predict the current status of the patient, disease forecast aims to predict the potential of contracting

⋆ Liu Li and Xiaofei Wang contribute equally to this paper.

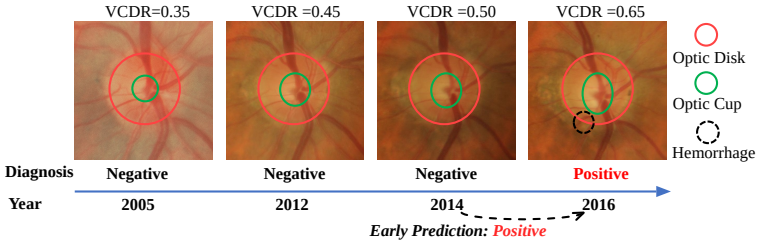

Fig. 1. Sequential fundus images of a patient transferring from negative to positive glaucoma between the year of 2005 and 2016.

a disease in the future, utilizing the variation tendency from the historical data, which is even more important for glaucoma treatment. An example is shown in Figure 1. As shown, the patient changes from negative to positive glaucoma in 2016 with a disc hemorrhage. Using glaucoma detection methods, this patient cannot be detected until 2016 with the hemorrhage. However, in 2014, glaucoma forecast with sequential fundus images can predict the deterioration in 2016, since the vertical cup-disc ratio (VCDR) increases from 0.35 to 0.5 during 2005-2014. As such, prevention and treatment can be conducted ahead of time. In this paper, we propose a deep learning method for glaucoma forecast (DeepGF), with an attention-polar convolution neural network (AP-CNN) and a variable time interval long-short term memory (VTI-LSTM) model to learn spatio-temporal transition for glaucoma forecast.

To our best knowledge, no previous works have focused on the glaucoma forecast task. While recently, a few works have been proposed for the early prediction of some other diseases, e.g., kidney injury [18] and chronic lung disease [5], based on the electronic health records (EHR). However, these EHR-based forecast methods faces the challenges by the trait of EHR: high-dimensionality, sparsity and irregularity [5, 10]. Different from the EHR-based methods, in this paper, we address glaucoma forecast by utilizing the sequential fundus images, which can capture the subtle changes of glaucoma pathologies [6]. Specifically, we first establish a large-scale sequential fundus image database for glaucoma forecast, called SIGF, which includes an average of 9 images per subject, corresponding to 3,671 fundus images in total. Besides, our DeepGF model is trained and validated with the sequential images.

Apart from leveraging the spatio-temporal features for glaucoma forecast, there are still two challenges due to the data limitation: (1) Data imbalance problem; (2) the irregular sampling intervals of the sequential data. Firstly, the data imbalance between positive and negative samples is a widely concerned problem in medicine [3, 18], e.g., the morbidity of glaucoma is only 3.5% among people over 45 years old [16]. General solution to this problem is data augmentation [12, 3], however, it may also lead to the overfitting problem [13]. In this paper, we solve such a data imbalance problem from both aspects of the network design and training strategy. Specifically, the attention mechanism is considered in the AP-CNN subnet highlighting the key features, so that less positive samples are

needed for training our DeepGF model. Then, a novel active convergence (AC) training strategy is proposed to actively choose the hard samples, which increases the sensitivity rate while simultaneously keeps the high specificity rate.

Secondly, it is hard to regularize patients to take medical examinations with a fixed time interval, which may vary from months to years, bringing the difficulties to predict the status transition across times. Different from the traditional LSTM-based methods for forecast that simplify the variant sampling intervals [1, 15], we propose a novel VTI-LSTM in the DeepGF, which can learn the temporal transition at different time intervals. Note that the input of the VTI-LSTM is the spatial features from AP-CNN.

To the best of our knowledge, our work is the first attempt for glaucoma forecast. The main contributions are concluded as follows. (1) We establish the SIGF database with sequential fundus images. (2) Based on the SIGF, we propose the DeepGF structure with AP-CNN and VTI-LSTM to predict the risk of glaucoma. (3) The attention mechanism and the AC training strategy are proposed in our DeepGF method to solve the imbalance sample distribution problem for glaucoma forecast.

2 Method

2.1 Framework

For glaucoma forecast, we develop a novel deep neural network (DNN) architecture that includes AP-CNN and VTI-LSTM. The architecture of our DeepGF is shown in Figure 2. As shown, the input is an RGB-channel fundus image $\mathbf{I}_t \in \mathbb{R}^{224\times224\times3}$ at time step t, and the output at the time step t is the binary label $\hat{l}_{t+1}$ ($\in \{0, 1\}$) of being glaucomatous or not at the next time step $t + 1$. Mathematically, the problem of glaucoma forecast can be formulated by

$$\hat{l}_{t+1} = f(\mathbf{I}_{t+1}|\mathbf{I}_1, \mathbf{I}_2, \cdots, \mathbf{I}_t), \tag{1}$$

where $f(\cdot)$ is the glaucoma forecast function to be learned.

In the DeepGF, the AP-CNN structure is composed of 2 subnets: the attention subnet and the polar subnet, inspired by the characteristics of glaucoma fundus images. Due to the characteristic that large redundancy exists in glaucoma detection [14], the attention subnet is proposed to remove the redundant regions on the fundus images by masking an attention map $\mathbf{A}_t \in \mathbb{R}^{224\times224\times1}$ on both the input fundus images and the extracted feature maps. As a result, the key features for glaucoma forecast can be highlighted. Besides, the polar subnet is introduced, because the positive glaucoma has larger VCDR [9, 7]. The fundus images mapped in polar coordinates, i.e., $\mathbf{I}_t^r(\rho, \theta) \in \mathbb{R}^{224\times224\times3}$ can enlarge the optic cup regions, which is effective in finding the larger VCDR. Note that ρ and θ denote the radius and angle in the polar coordinates, respectively. Finally, the features extracted by the attention and polar subnets are concatenated and further reduced to a vector $\mathbf{f}_t \in \mathbb{R}^{1\times320}$ as the final spatial feature.

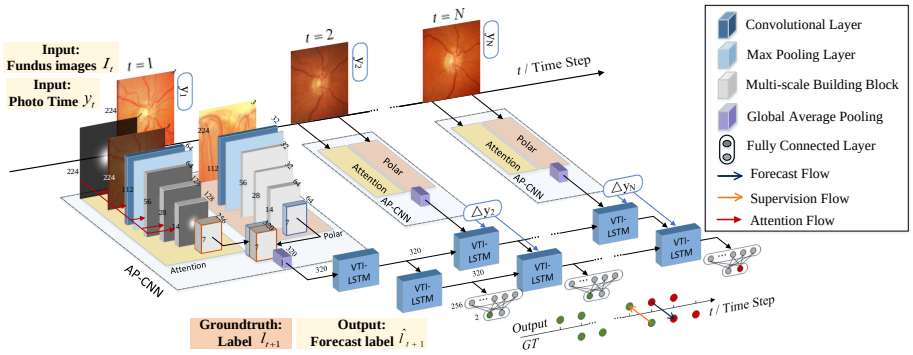

Fig. 2. Architecture of DeepGF.

Subsequently, the 2-layer VTI-LSTM is designed to learn the dynamic glaucoma transition, in which features $\{\mathbf{f}_t\}_1^T$ (T denotes the total number of time steps) from the AP-CNN serve as the input. In contrast to the traditional LSTM [8], the length of variant time interval $\triangle y_t$ is further encoded into the forget gate f_t of the VTI-LSTM. Finally, the output from the second VTI-LSTM layer $\mathbf{h}_t^2 \in \mathbb{R}^{1\times 320}$ at time step t is fed into the fully connected (FC) layers, outputting the probability of $\hat{l}_{t+1}$ at the next time step $t+1$. Note that the AC training strategy is adopted in our DeepGF to solve the imbalance data problem.

2.2 Attention and polar based CNNs

Attention subnet. In our AP-CNN, an attention subnet is designed for extracting the salient spatial features from the fundus images. As shown in Figure 2, the architecture of the AP-CNN is followed by one convolutional layer with a kernel size of 7×7, a max-pooling layer and four multi-scale building blocks [14]. Note that all convolutional layers in AP-CNN are followed by a batch normalization layer and an ReLU layer for increasing the nonlinearity of AG-CNN. Then, to remove the redundancy on the fundus images, the attention map $\mathbf{A}_t \in \mathbb{R}^{224\times 224\times 1}$ is weighted on each channel of its input fundus image $\mathbf{I}_t$ and the corresponding feature maps $\{\mathbf{F}_t^k\}_{k=1}^4$ at layer k, as shown in Figure 2. Note that, the attention map of each fundus image $\mathbf{A}_t$ is predicted by [14], as shown in Figure 3 left. For the detailed structure of AP-CNN, refer to Table 1 in Supplementary Materials.

Polar subnet. A polar subnet is proposed in the AP-CNN to extract features from the fundus images in polar coordinates (ρ, θ). The architecture of the polar subnet is similar to the attention subnet, and the input to the polar subnet is the ROIs of the fundus images mapped into polar coordinates, i.e., $\mathbf{I}_t^r(\rho,\theta) \in \mathbb{R}^{224\times 224\times 3}$. Consequently, the optic cup regions can be enlarged to effectively detect the larger VCDR, as illustrated in Figure 3. To be specific, the ROIs in the polar coordinates $\mathbf{I}_t^r(\rho,\theta)$ are generated by the ROIs in rectangular coordinates $\mathbf{I}_t^r(u,v)$. Note that the ROI $\mathbf{I}_t^r(u,v)$ is cropped with the center of

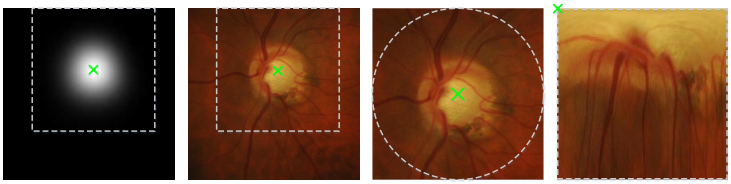

Fig. 3. An example of the input to the DeepGF. (Left): Attention map. (Middle-left): Original fundus images. (Middle-right):ROIs. (Right): ROIs in polar coordinates.

the optic cup center from original fundus image $\mathbf{I}_t(u, v)$. Mathematically, given $\mathbf{I}_t^r(u, v)$, ρ and θ of $\mathbf{I}_t^r(\rho, \theta)$ are calculated by

$$\begin{cases} \rho = \sqrt{(u - \frac{U_p}{2})^2 + (v - \frac{V_p}{2})^2}, \\ \theta = \tan^{-1}(\frac{v - v_0}{u - u_0}), \end{cases} \tag{2}$$

where U_p and V_p are the width and height of $\mathbf{I}_t^r(u, v)$, and (u_0, v_0) is the center coordinates of $\mathbf{I}_t^r(u, v)$.

Finally, the features $\mathbf{F}_t^a (\in \mathbb{R}^{7\times7\times256})$ extracted by attention subnet and $\mathbf{F}_t^p (\in \mathbb{R}^{7\times7\times64})$ extracted by the polar subnet are concatenated along with channels to obtain $\mathbf{F}_t (\in \mathbb{R}^{7\times7\times320})$, which is then down-sampled by global average pooling and reshaped to a vector $\mathbf{f}_t (\in \mathbb{R}^{1\times320})$. Note that the parameters ω_c of AP-CNN are shared across time steps.

2.3 VTI-LSTM

In addition to the AP-CNN, we propose a novel VTI-LSTM to learn transition across time steps. Mathematically, the VTI-LSTM can be formulated as:

$$\boldsymbol{i}_t = \sigma(\mathbf{W}_{x_i}\boldsymbol{x}_t + \mathbf{W}_{h_i}\boldsymbol{h}_{t-1} + \boldsymbol{b_i}), \tag{3}$$

$$\tilde{f}_t = \sigma(\mathrm{p} \triangle y_t + \mathrm{q}), \tag{4}$$

$$\boldsymbol{f}_t = \sigma(\tilde{f}_t(\mathbf{W}_{x_f}\boldsymbol{x}_t + \mathbf{W}_{h_f}\boldsymbol{h}_{t-1}) + \boldsymbol{b}_f), \tag{5}$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t \circ \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \circ \tanh(\mathbf{W}_{x_c}\boldsymbol{x}_t + \mathbf{W}_{h_c}\boldsymbol{h}_{t-1} + \boldsymbol{b}_c), \tag{6}$$

$$\boldsymbol{o}_t = \sigma(\mathbf{W}_{\boldsymbol{x}_o}\boldsymbol{x}_t + \mathbf{W}_{\boldsymbol{h}_o}\boldsymbol{h}_{t-1} + \boldsymbol{b}_o), \tag{7}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \circ \tanh(\boldsymbol{c}_t), \tag{8}$$

where $\sigma(\cdot)$ is the sigmoid function, and "$\circ$" denotes the Hadamard products. For time step t, we have the input $\boldsymbol{x}_t$, the cell state $\boldsymbol{c}_t$, the hidden state $\boldsymbol{h}_t$ and the gates of input $\boldsymbol{i}_t$, forget $\boldsymbol{f}_t$ and output $\boldsymbol{o}_t \in \mathbb{R}^{1\times320}$. In addition, $\mathbf{W}_{x_\sim}$ and $\mathbf{W}_{h_\sim} \in \mathbb{R}^{320\times320}$ are the corresponding parameters to be learned. In (4), $\triangle y_t (\in \mathbb{N})$ is the time length from time step $t-1$ and t, and $\tilde{f}_t$ $(0 < \tilde{f}_t < 1)$ denotes to the variable time interval (VTI) gate, which further controls the forgetting rate with the time interval. Besides, p and q$(\in \mathbb{R})$ are the parameters of the VTI gate. As such, different from the traditional LSTM, the VTI-LSTM capably encodes the variant length between time steps.

2.4 Imbalance sample training strategy: AC strategy

To train the DeepGF, we supervise the training process with the labels of the next time step l_{t+1} by the forecast loss function $\mathcal{L}_f$ as follows:

$$\mathcal{L}_f = -\frac{1}{T}\sum\nolimits_{t=1}^{T}\left(l_{t+1}\log p_{t+1} + (1-l_{t+1})\log(1-p_{t+1})\right), \tag{9}$$

where T is the maximum time step of the image sequence, and p_{t+1} is the forecast probability of positive glaucoma at time step $t+1$. Note that $p_{t+1} = \frac{1}{1+e^{-z_t}}$, where z_t is the output of the second FC layer at time step t. Given the trained DeepGF, the forecast binary label $\hat{l}_{t+1}$ ($\in \{0,1\}$) is decided by $\hat{l}_{t+1} = \mathbb{1}\left(p_{t+1} \geq \xi\right)$, where $\mathbb{1}(\cdot)$ is a conditional binary indicator and ξ is the probability threshold.

However, the morbidity of glaucoma is 3.5% among people over 45 years old [16], leading to the imbalance data distribution problem. To address this problem, we propose a novel AC training strategy to self-update the distribution of the training set actively and adaptively. The AC strategy is consisted of 2 steps, as summarized in Algorithm 1 in the Supplementary Materials. (1) $\mathcal{L}_f(\omega)$ is calculated and the parameters ω of the DeepGF are updated by decreasing with the gradients $\nabla_\omega \mathcal{L}(\omega)$. (2) Some training samples are dynamically removed from the training set $\{\mathbf{I}^s\}_{s=1}^{S}$ according to $\mathcal{L}_f(\omega)$, where S is the total number of the training sequences, and $\mathbf{I}^s = \{\mathbf{I}_t^s\}_{t=1}^{T}$ represents a fundus image sequence. To be specific, a number of βS sequences, the training loss of which (i.e., $\mathcal{L}_f(\omega)$) ranks the lowest, are discarded from the training set every δ epoches. Note that the discarding terminates until reaching the maximum abandon epoch N. As a result, the training set $\{\mathbf{I}^s\}_{s=1}^{S}$ can be actively updated, remaining the hardest samples to be further learned.

3 SIGF Database

For training and testing our DeepGF model, we establish a database of sequential fundus images for glaucoma forecast, called SIGF[5]. To our best knowledge, SIGF is the first database consisting of sequential fundus images. Our SIGF database contains 405 image sequences from different eyes, and each eye has at least 6 fundus images captured at different years (ranging from 1986 to 2018), with an average of 9 images per eye, corresponding to 3,671 images in total. An example of the SIGF database is shown in Figure 1 in the Supplementary Materials

In the SIGF database, all fundus images are annotated with binary labels of glaucoma, i.e., positive or negative glaucoma. The samples are labeled to positive glaucoma when they satisfy any of the three criteria, i.e., retinal nerve fibre layer defect, rim loss and optic disc hemorrhage. Consequently, all the fundus images are labeled and the sequences can be divided into 2 types: time-variant (37 sequences change from negative to positive glaucoma) and time-invariant (368 sequences keep negative glaucoma). Note that no sequence changes from positive

[5] The database is available at https://github.com/XiaofeiWang2018/DeepGF

to negative glaucoma, since glaucoma is irreversible. Our work is conducted according to the Declaration of Helsinki. As the fully anonymous usage of fundus images, we are exempted by the medical ethics committee to inform the patients.

4 Experiment

4.1 Setting

In the experiment, the 405 image sequences of our SIGF database are randomly divided into training (300), validation (35) and test (70) sets, and the proportion of the time-variant and time-invariant sequences keeps invariant in the training (27 and 273), validation (3 and 32) and test sets (7 and 63).

Specifically, we temporally segment the 405 sequences into 1,146 clips, with each clip containing $T(= 6)$ time steps. An overlap of 5 frames is allowed in cutting the sequence clips, for the purpose of data augmentation. Before feeding into the DeepGF, the fundus image are resized to 224×224. In training AP-CNN and VTI-LSTM, we learn the parameters using Adam optimizer[11], and the hyper-parameters are tuned over the validation set, as provided in Table 2 in the Supplementary.

4.2 Evaluation

Glaucoma forecast. In this section, we compare the glaucoma forecast performance with two other methods [14, 4]. To our best knowledge, there is no method for glaucoma forecast. Therefore, we convert [14, 4] to be glaucoma forecast methods by supervising their CNN models with the labels at the next time step. Table 1-left lists the results of accuracy, sensitivity and specificity. As shown in this table, our DeepGF achieves 80.7%, 85.7% and 80.6% in accuracy, sensitivity and specificity, respectively, which are considerably better than other two methods [14, 4]. Then, Figure 4 (b) plots the ROC curves of our and other methods, for visualizing the trade-off between sensitivity and specificity. As shown, our ROC curve is closer to the upper-left corner than other two methods [14, 4]. This indicates that given a referable specificity, our method has better performance in forecasting positive glaucoma than [14, 4]. Similar results can be found from the AUC metrics in Table 1. In a word, our DeepGF method is effective in forecasting glaucoma, significantly outperforming the state-of-the-art methods.

Network Interpretability. An additional visualization experiment is conducted by the CAM-based method [19] for glaucoma forecast. As shown in Figure 4 (a), the visualization result focuses on the potential pathological rim-loss regions and the optic cup regions (the biomarkers for glaucoma in clinic [6]), which verifies the interpretability of our DeepGF method for glaucoma forecast.

4.3 Results of ablation experiments

AP-CNN. In our ablation experiments, we first investigate the impact of the attention and polar subnets in the AP-CNN. To this end, we simply remove the

Table 1. Performance of glaucoma forecast over SIGF database.

Method	Glaucoma forecast, %			Ablation study (W/O), %				
	Ours.	Chen [4]	Li [14]	Attention	Polar	VTI gate	LSTM	AC
Accuracy	**80.7**	67.6	66.0	76.0	77.8	73.2	67.1	58.4
Sensitivity	**85.7**	62.2	68.9	73.8	73.8	73.8	69.0	66.7
Specificity	**80.6**	67.8	65.9	76.0	77.9	73.1	67.1	58.2
AUC	**87.0**	71.8	70.0	80.9	81.6	80.8	74.9	64.8

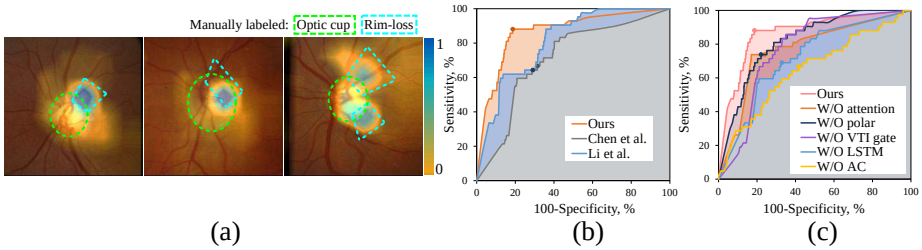

Fig. 4. (a): Visualization results of DeepGF for glaucoma forecast. (b): ROC curves of glaucoma forecast methods. (c): ROC curves of ablation experiments.

attention and polar subnets, respectively, and then compare the performance of glaucoma forecast. The quantitative results are shown in Table 1-right. As shown, the attention subnet can improve the accuracy, sensitivity, specificity and AUC by 4.7%, 11.9%, 4.6% and 6.1%, respectively, and the polar subnet can also increase 2.9%, 11.9%, 2.7% and 5.4%, respectively. Also, the ROC curves of removing the attention or polar subnet are shown in Figure 4-c. These results indicate the necessity of our AP-CNN model for glaucoma forecast.

VTI-LSTM. Additionally, we evaluate the impact of VTI-LSTM in the DeepGF by 2 stages. First, we replace the VTI-LSTM by a standard LSTM to show the impact of VTI gate encoding the variable time intervals. The result is shown in Table 1-right. We can see that the VTI gate improves the performance of glaucoma forecast in terms of accuracy by 7.5%. Similar results can be found for sensitivity, specificity and AUC. Second, we further remove the LSTM in our DeepGF, such that the temporal features are not considered. As shown in Table 1-right, all metrics decrease at least 12.1%. Such performance degradation is considerably more than that of the AP-CNN, also validated by the ROC curves in Figure 4 (c). Hence, the VTI-LSTM is vitally important for glaucoma forecast.

AC strategy. We further evaluate the effectiveness of our AC training strategy, by replacing the AC training strategy with a data augmentation strategy [3], which is the common solution to the imbalance problem. As shown in Table 1-right, the performance of training the DeepGF with [3], instead of our AC strategy, decreases in all metrics, , i.e., the reduction of 22.3%, 19.0%, 22.4% and 22.2%, for accuracy, sensitivity, specificity and AUC, respectively. Similar results can be found from the ROC curves on Figure 4 (c). These ablation results validate the effectiveness of the AC training strategy proposed in this paper.

5 Conclusion

Glaucoma forecast refers to predicting whether a person will suffer from glaucoma in the future. Since glaucoma is an irreversible blindness disease, glaucoma forecast is even more important than glaucoma detection that aims at diagnosing glaucoma at the present stage. To achieve glaucoma forecast, both the current and previous fundus images are useful, as the temporal tendency of glaucoma can be learned. Input with the sequence of the present and previous fundus images, DeepGF has been proposed in this paper as a novel deep learning method for glaucoma forecast. In our DeepGF method, we proposed two new DNN structures: AP-CNN and VTI-LSTM. The AP-CNN mainly focuses on extracting the spatial features of fundus images by taking into account the attention mechanism and polar projection. The VTI-LSTM was developed to learn the temporal features across fundus images at different time steps. Consequently, the spatio-temporal features can be obtained for glaucoma forecast. Finally, the experiment results show that our DeepGF is superior to other state-of-the-art methods in glaucoma forecast.

References

1. Amirkhan, R., Hoogendoorn, M., Numans, M.E., Moons, L.: Using recurrent neural networks to predict colorectal cancer among patients. In: SSCI. pp. 1–8. IEEE (2017)
2. Bourne, R.R.A., Stevens, G.A., White, R.A., Smith, J.L., Flaxman, S.R., Price, H., Jonas, J.B., Keeffe, J., Leasher, J., Naidoo, K.: Causes of vision loss worldwide, 1990-2010: a systematic analysis. Lancet GH **1**(6), e339–e349 (2013)
3. Chen, J., Lalor, J., Liu, W., Druhl, E., Granillo, E., Vimalananda, V.G., Yu, H.: Detecting hypoglycemia incidents reported in patients' secure messages: Using cost-sensitive learning and oversampling to reduce data imbalance. JMIR **21**(3), e11990 (2019)
4. Chen, X., Xu, Y., Wong, D.W.K., Wong, T.Y., Liu, J.: Glaucoma detection based on deep convolutional neural network. In: EMBC. p. 715 (2015)
5. Cheng, Y., Wang, F., Zhang, P., Hu, J.: Risk prediction with electronic health records: A deep learning approach. In: ICDM. pp. 432–440. SIAM (2016)
6. Fan, X., Wu, L.L., Ma, Z.Z., Xiao, G.G., Liu Jr, F.: Usefulness of frequency-doubling technology for perimetrically normal eyes of open-angle glaucoma patients with unilateral field loss. Ophthalmology **117**(8), 1530–1537 (2010)
7. Fu, H., Cheng, J., Xu, Y., Zhang, C., Wong, D.W.K., Liu, J., Cao, X.: Disc-aware ensemble network for glaucoma screening from fundus image. IEEE TMI (2018)
8. Gers, F.A., Schmidhuber, J., Cummins, F.: Learning to forget: Continual prediction with lstm (1999)
9. Harizman, N., Oliveira, C., Chiang, A., Tello, C., Marmor, M., Ritch, R., Liebmann, J.M.: The isnt rule and differentiation of normal from glaucomatous eyes. Archives of Ophthalmology **124**(11), 1579 (2006)
10. Jin, B., Che, C., Liu, Z., Zhang, S., Yin, X., Wei, X.: Predicting the risk of heart failure with ehr sequential data modeling. IEEE Access **6**, 9256–9261 (2018)
11. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
12. Krawczyk, B., Galar, M., Jeleń, Ł., Herrera, F.: Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy. ASC **38**, 714–726 (2016)
13. Le, T., Vo, M.T., Vo, B., Lee, M.Y., Baik, S.W.: A hybrid approach using oversampling technique and cost-sensitive learning for bankruptcy prediction. Complexity (2019)
14. Li, L., Xu, M., Wang, X., Jiang, L., Liu, H.: Attention based glaucoma detection: A large-scale database and cnn model. arXiv preprint arXiv:1903.10831 (2019)
15. Lipton, Z.C., Kale, D.C., Elkan, C., Wetzel, R.: Learning to diagnose with lstm recurrent neural networks. arXiv preprint arXiv:1511.03677 (2015)
16. Tham, Y.C., Li, X., Wong, T.Y., Quigley, H.A., Aung, T., Cheng, C.Y.: Global prevalence of glaucoma and projections of glaucoma burden through 2040 : A systematic review and meta-analysis. Ophthalmology **121**(11), 2081 (2014)
17. Thomson, R., Luettel, D., Healey, F., Scobie, S.: Safer care for the acutely ill patient: learning from serious incidents. London: National Patient Safety Agency (2007)
18. Tomašev, N., Glorot, X., Rae, J.W., Zielinski, M., Askham, H., Saraiva, A., Mottram, A., Meyer, C., Ravuri, S., Protsyuk, I., et al.: A clinically applicable approach to continuous prediction of future acute kidney injury. Nature **572**(7767), 116 (2019)
19. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: IEEE CVPR. pp. 2921–2929 (2016)