Enhancing Intelligent Anemia Detection via Unifying Global and Local Views of Conjunctiva Image with Two-Branch Neural Networks

Lijuan Zheng
South China University of Technology
Shaopeng Liu
Guangdong Polytechnic Normal University
Senping Tian
South China University of Technology
Jianhua Guo
Guangdong Polytechnic Normal University
Xinpeng Wang
Guangdong Polytechnic Normal University
Xiuxiu Liao
Guangdong Polytechnic Normal University
Jiaming Hong (hjm@gzucm.edu.cn)
Guangzhou University of Chinese Medicine

Research Article

Keywords: Intelligent Anemia Detection, Multi-View Learning, Two-Branch Neural Networks, Conjunctiva Image, Anemia

Posted Date: January 3rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1170958/v1

Enhancing Intelligent Anemia Detection via Unifying Global and Local Views of Conjunctiva Image with Two-Branch Neural Networks

Lijuan Zheng[1], Shaopeng Liu[2,*], Senping Tian[1], Jianhua Guo[2], Xinpeng Wang[2], Xiuxiu Liao[2] and Jiaming Hong[3,*]

[*]Corresponding author:
Shaopeng Liu
149265005@qq.com
Jiaming Hong,
hjm@gzucm.edu.cn
[2] Department of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China
[3] School of Medical Information Engineering, Guangzhou University of Chinese Medicine, Guangzhou, China
Full list of author information is available at the end of the article.

Abstract

Background: Anemia is one of the most widespread clinical symptoms all over the world, which could bring adverse effects on people's daily life and work. Considering the universality of anemia detection and the inconvenience of traditional blood testing methods, many deep learning detection methods based on image recognition have been developed in recent years, including the methods of anemia detection with individuals' images of conjunctiva. However, existing methods using one single conjunctiva image could not reach comparable accuracy in anemia detection in many real-world application scenarios.

Results: To enhance intelligent anemia detection using conjunctiva images, we proposed a new algorithmic framework which could make full use of the data information contained in the image. To be concrete, we proposed to fully explore the global and local information in the image, and adopted a two-branch neural network architecture to unify the information of these two aspects.

Conclusions: Compared with the existing methods, our method can fully explore the information contained in a single conjunctiva image and achieve more reliable anemia detection effect. Compared with other existing methods, the experimental results verified the effectiveness of the new algorithm.

Keywords: Intelligent Anemia Detection; Multi-View Learning; Two-Branch Neural Networks; Conjunctiva Image; Anemia

Background

Anemia is a typical clinical symptom that the capacity of red blood cells in human peripheral blood is lower than the normal range. As one of the most widespread clinical symptoms all over the world, anemia could bring adverse effects on people's daily life and work [1-2]. Therefore, timely diagnosis and treatment is particularly important for anemia. However, the traditional diagnosis of anemia requires blood sampling for test with professional equipment, which requires

extra cost and intervention of professionals, and brings obvious limitation to its wide application in daily life [2-4].

In recent years, in order to improve the convenience of anemia examination, many studies have been carried out to investigate the effective approaches of detecting anemia in a noninvasive manner [4-7]. One of the most promising directions was to deploy artificial intelligent algorithms to discover the correlation of anemia with images of human surface organs, including fingernails, retinas, conjunctivas, etc. In other words, instead of testing blood samples, artificial intelligent algorithms for image classification and recognition are applied in anemia diagnosis [8-9]. Due to their non-invasive manner and non-requirement of professional blood testing, such approaches are less expensive, and seem much attractive than the traditional ones.

In this paper, we focus on developing more effective algorithms for one of such approaches, that is, to detect anemia with photos of conjunctiva [10-12]. Comparing to other similar approaches like diagnosis with retina images, detecting with conjunctiva images requires no extra professional equipment, which brings much convenience to users. Therefore, it can be easily deployed in Web-based platforms like APPs, and people can take photos and upload them at any time to get a diagnosis result immediately.

However, there is some room for improvement in existing algorithms of detecting anemia with conjunctiva images. One of the main issues is that, as was demonstrated in previous reports and in our testing experiments, the existing algorithms could not reach comparable performance with human experts in many real-world applications and could be easily affected by image quality. Such drawbacks might limit its potential application prospects.

To further enhance the performance of anemia diagnosis with the photos of conjunctiva, we

proposed in this paper a new algorithmic framework to solve such problems. As was discussed in detail later in this paper, the main issue which prevents previous algorithms from achieving higher accuracy may lie in that, they cannot fully exploit the information of the image, especially the local information. Inspired by such analysis, we put forward a new perspective which could unify both global and local information of the image data, and proposed a simple yet effective deep convolutional neural network architecture to fuse them, and obtained a much robust and reliable deep learning model for anemia detection using only one single image of an individual's conjunctiva.

Results

GLUDA Model

In this article we design a new deep learning paradigm for intelligent anemia prediction using only one single conjunctiva image. Based on our previous results, we find that when observing the conjunctiva image for diagnosis, one of the main inspirations is that, two different views of information (i.e., the global view and the local view) are simultaneously extracted from one single conjunctiva image, and these two views of information are unified in a reasonable approach to help forming comprehensive diagnosis. Motivated by such observation, we propose that similar operations can be deployed into the deep learning architecture to enhance the effectiveness of the algorithm in anemia detection. Thus, we readjust the basic structure of the classic convolutional neural network model and add the structure of multi-view generation and multi-view data fusion to achieve this goal. The overall framework of the new model is depicted in Fig. 1. We refer the new model as GLUDA model (which is short for the full name: Global and Local view Unifying

for Detection of Anemia). We illustrated the overall workflow of GLUDA model in Fig.1.

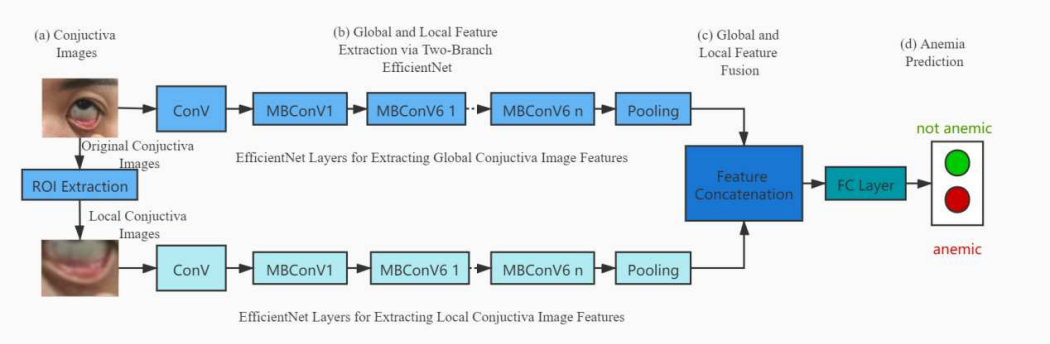

Fig.1 Overall Framework of GLUDA Model

Training framework of the GLUDA

The GLUDA model benefits from several key aspects, among which the two-branch architecture plays a central role in unifying both global and local information of a single conjunctiva image. We choose EfficientNet as the two sub-network structure for constructing the whole network.

Recently proposed by Google Brain Team, EffiecientNet is an efficient convolutional neural network, which can scale up CNN models by balancing network depth, width, and resolution in a systemic way. Formally, denote the input data image as an input tensor X, then the output of EfficientNet can be formally defined as:

$$\mathrm{Eff(X)} = F_K^d \odot F_{K-1}^d \ldots \odot F_2^d \odot F_1^d(X), \quad (1)$$

where each F_i^d is the ith layer of the network, for i=1,2,…,K, with K the total number of all the convolutional layers of the sub-network, and d the coefficients for scaling network depth. To get a more compact expression, we can further denote the ith input tensor as $X_{<r\cdot H_i, r\cdot W_i, w\cdot C_i>}$, where $< H_i, W_i, C_i >$ is predefined tensor shape of the corresponding layer , and $< d, r, w >$ is the parameter tuples for scaling network depth, width, and resolution. Therefore, for any input tensor

X (as a conjunctiva image), the two sub-network Net*i* (i=1,2) in GLUDA can be formally referred to as:

$$\mathrm{Net1(X, d, w, r)} = \bigodot_{i=1,2,\ldots,S} F_i^{d,L_i}\left(X_{<r\cdot H_i, r\cdot W_i, w\cdot C_i>}\right), \tag{2}$$

and

$$\mathrm{Net2}(local(\mathrm{X}), \mathrm{d}, \mathrm{w}, \mathrm{r}) = \bigodot_{i=1,2,\ldots,S} F_i^{d,L_i}\left(local(X)_{<r\cdot H_i, r\cdot W_i, w\cdot C_i>}\right). \tag{3}$$

Note that we would construct the two EfficientNet sub-branches with exactly the same structure, and therefore they would share the same predefined tensor shape $< H_i,\ W_i, C_i >$ in each layer. But the weight parameters would be different after training. Denote the concatenation operation as F_{conca}, and the fully connected layer as F_c, the output of the GLUDA network can be written as :

$$\mathrm{Net(X, d, w, r)} = F_c \odot \left(F_{conca}\left(Net1(X, d, w, r), Net2(local(X), d, w, r)\right)\right). \tag{4}$$

To bring all the formal discussions above, the overall training framework of the GLUDA model can be summarized as follows:

$$\begin{gathered}
\text{Max Accuracy}(\mathrm{Net(d, w, r)}) \\
\text{s. t.} \\
\mathrm{Net1(X, d, w, r)} = \bigodot_{i=1,2,\ldots,S} F_i^{d,L_i}\left(X_{<r\cdot H_i, r\cdot W_i, w\cdot C_i>}\right) \\
\mathrm{Net2(X, d, w, r)} = \bigodot_{i=1,2,\ldots,S} F_i^{d,L_i}\left(X_{<r\cdot H_i, r\cdot W_i, w\cdot C_i>}\right) \\
\mathrm{Net(X)} = F_c \odot \left(F_{conca}\left(Net1(X, d, w, r), Net2(local(X), d, w, r)\right)\right), \\
\text{Memory}(\mathrm{Net}) \leq \text{memory_constraints}, \\
FLOPS(\mathrm{Net}) \leq \text{flops_constraints},
\end{gathered} \tag{5}$$

where memory_constraints and flops_constraints are constraints of the computation resources. To obtain the optimal parameters for the whole network, compound scaling method and a small

grid search could be applied. Briefly speaking, the compound scaling method would set $d = \alpha^{\emptyset}, w = \beta^{\emptyset}, r = \gamma^{\emptyset}$, and $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$, with user-specific coefficient $\emptyset$ and search optimal $\alpha \geq 1, \beta \geq 1, \gamma \geq 1$ for appropriate depth, width and resolution settings for the network.

From the formal optimization target in Eq.(5), and considering the structure of conjunctiva images processed in our empirical experiments, below we show the experiment result of the proposed GLUDA model.

Experiment Results

To evaluate the proposed model, we collected conjunctiva image data from patients enrolled in Zhongshan Ophahalmic Center, Third affiliated Hospital at Sun Yat-sent University and People's Hospital of Guangxi Zhuang Autonomous Region, from June 2019 to November 2019. To make a more comprehensive comparison, we developed two series of model training and testing in our experiments, considering whether data augmentation operation was incorporated in the learning process.

We reported the experimental results of both series in Table 1, Fig. 2 and Fig.3. Row 2 to row 4 in Table 1 and Fig. 2 showed testing results of models without data augmentation. And row 6 to row 8 in Table 1 and Fig. 3 showed testing results of models using data augmentation, where random rotation, cropping and flipping and applied and the size of the augmented training set are extended to 1396 in total. In each series of the experiments, GLUDA models outperformed EfficientNet_Origin and EfficientNet_ROI, respectively. When data augmentation was not adopted, GLUDA model achieved the highest scores among all the three models, with $\mathrm{AUC} = 0.9087, \mathrm{Acc} = 0.7937$, and EfficientNet_ROI obtained the lowest score, with $\mathrm{AUC} = 0.9305, \mathrm{Acc} = 0.8313$. The sensitivity and specificity scores and their corresponding 95% CI

intervals are also the highest for the GLUDA model. When data augmentation was applied before training, we could observe that all three models could achieved higher scores comparing to their counterpart model without augmentation, and among the three models, GLUDA could still obtained the highest scores in AUC, Acc, Sen, and Spe, respectively. According to such comparison results, GLUDA model could fuse global and local views of conjunctiva images in an effective approach, therefore outperformed the other two models which only use global (EfficientNet_Origin) or local (EfficientNet_ROI) views of conjunctiva images.

Table 1. Performance of GLUDA vs. other models in the test set

Not Augmented	**Acc (95% CI)**	**Sen (95% CI)**	**Spe (95% CI)**
GLUDA	0.7937(0.8539,0.9442)	0.8542(0.7245,0.8492)	0.7679(0.7912,0.9005)
EfficientNet_Origin	0.7604(0.7801,0.9179)	0.8056(0.6661,0.8347)	0.7333(0.7153,0.8723)
EfficientNet_ROI	0.7125(0.7662,0.8818)	0.8043(0.638,0.777)	0.6754(0.736,0.8584)
Augmented	**Acc (95% CI)**	**Sen (95% CI)**	**Spe (95% CI)**
GLUDA	0.8313(0.8802,0.9606)	0.8723 (0.656, 0.8814)	0.8142(0.8118,0.9154)
EfficientNet_Origin	0.7812(0.6886,0.8522)	0.8649(0.7823,0.9193)	0.7419(0.6324,0.80779)
EfficientNet_ROI	0.75 (0.6776,0.8107)	0.7872(0.7175,0.8435)	0.7345(0.6612,0.7969)

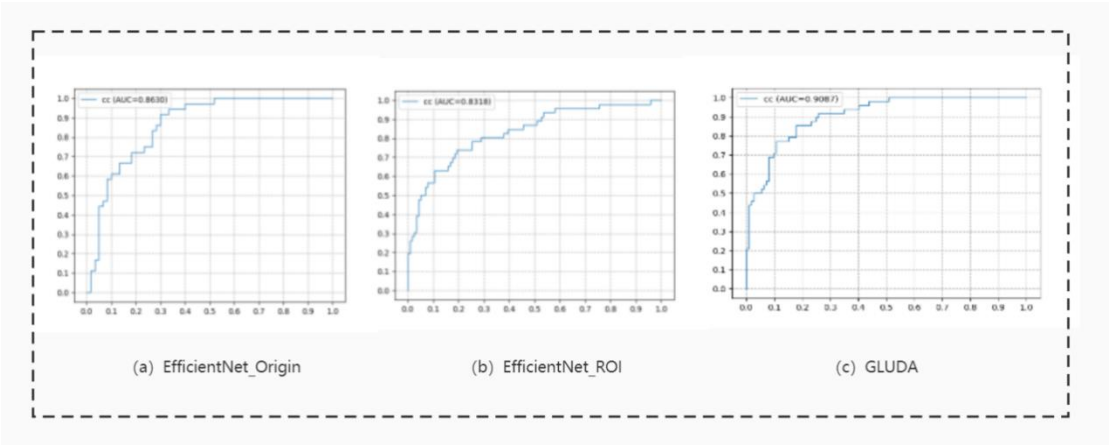

Fig 2. Performance of GLUDA vs. other models without data augmentation

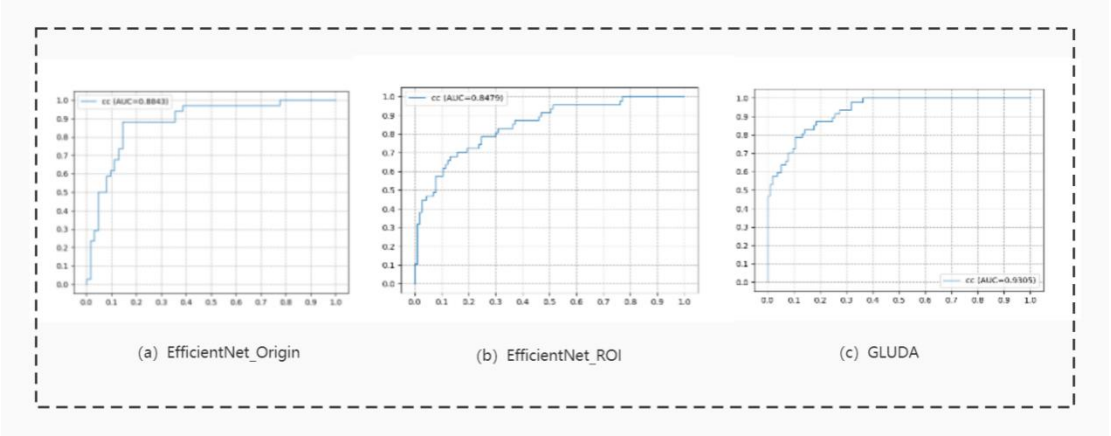

Fig 3. Performance of GLUDA vs. other models with data augmentation

Discussion

Model discussion

The heuristic of the new model was motivated by observing how professionals form their comprehensive diagnosis of anemia with the anemia images. Let's consider the following decision-making process: as a professional, when using a conjunctiva image for assisting anemia diagnosis, one would first make a preliminary judgment on the overall image of conjunctiva, considering the potential risk of whether the corresponding individual has anemia. Then, in order to further verify or reject this judgment, one will continue to carefully observe the regions of interest (ROI), and make comparative analysis with the whole image, and finally get a comprehensive judgment of the anemia diagnosis.

Several key steps in the GLUDA model can be summarized as follows:

Firstly, in addition to the original image data serving as the global view, the region of interest (ROI) from the original image, serving as the local view data. Formally, for any conjunctiva image data x_s, we perform ROI extraction algorithm to produce the local view data, denoted as:

$$x_s' = local(x_s), \tag{6}$$

where $local()$ denotes any appropriate ROI extraction algorithm. The obtained local view image

is then resized to the same size of x_s, which is also denoted as x_s' for convenience. x_s and x_s' is then treated as two equal parts to form a tuple as the input of the model, i.e.,

$$x_s \rightarrow (x_s, x_s'), \tag{7}$$

and the learning task is reframed to obtain a mapping function f which produce the following mapping for any (x_s, y_s) with high generalization ability:

$$f: (x_s, x_s') \rightarrow y_s. \tag{8}$$

Secondly, the overall framework comprises two branches of deep convolutional neural networks, which server as feature extractor for the original conjunctiva image and the ROI conjunctiva image, respectively. The network block architecture of each branch is identical to that of the EfficientNet model, which comprises convolutional layers, mobile inverted bottleneck MBConv layers and Pooling layers as its building blocks [25-27]. There are various versions of EfficientNet model, and we adopt EfficientNet-B0 in our study for convenience but note that other versions can substitute for EfficientNet-B0 here without substantial influence.

Thirdly, the extracted higher-level features of the two branches are fused together, and are passed through fully connected layers to learn a more comprehensive prediction. To be concrete, for each view of the image data, after passing through the layers of the branch, a set of higher-level features are obtained. To fuse these two sets of features, a simple concatenation operation was carried out, which would concatenate these two sources of features into a unified one. The fused features are then eventually passed through the FC layer to output a prediction of anemia. Note that alternative feature fusion operations other than concatenation might also be applicable here, but simple concatenation operation has well served the task here and works well in practice.

Experiment discussion

One of the aspects worth emphasizing is that, we could observe from the results that, even when training without data augmentation, the obtained GLUDA model could still outperformed EfficientNet_Origin and EfficientNet_ROI trained with data augmentation. We plot the comparing results in Fig.4. Such contrast is impressive, considering that in this case GLUDA only used 532 conjunctiva images for training, which is far less than the 1396 augmented images for training EfficientNet_Origin and EfficientNet_ROI. Given that data augmentation could provide extra information to boost model training in many cases, it might be reasonable to conclude that the way GLUDA fuse two views of conjunctiva images had successfully draw into valid discriminative information for training, which would be missing when only a single view of conjunctiva images was used. Such augmentation of information is parallel to that of data augmentation, since the performance of GLUDA could be further improved when trained with augmented set, which could be found in Figure 4.

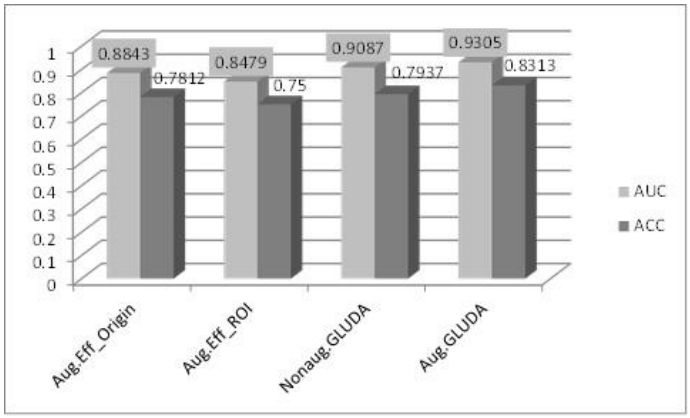

Fig.4 GLUDA obtained competitive results even without data augmentation

Conclusion

In this paper, we proposed a new algorithmic framework for the problem of anemia detection using individuals' images of conjunctiva. Compared with the traditional blood testing method, this method has great advantages in the convenience of application due to its inventiveness manner. However, existing work using one single conjunctiva image could not reach expected accuracy in many real-world application scenarios. Motivated by the observation that professionals can benefit from their abilities in comprehensively unifying both global and local information of a single conjunctiva image in enhancing the accuracy of anemia diagnosis, we proposed a two-branch neural network architecture to fully explore such two views of the conjunctiva image and achieved a much robust and reliable performance in intelligent diagnosis.

In our future work, we would continue to investigate other alternative approaches of unifying multiple views of conjunctiva image data other than the proposed two-branch neural network approach. Furthermore, considering that the paradigm proposed might benefit other clinical or medical application tasks involving image recognition, we will continue to explore more potential applications of this method.

Methods

Various approaches in detecting anemia

Here, we review some of the related work, including the various approaches in detecting anemia with images, and one of the main deep learning research in processing multi-modal data, i.e., multi-modal deep learning, whose idea was adopted in designing the proposed new algorithm.

Anemia detection methods without blood testing: Due to the universality of anemia detection and the inconvenience of traditional blood detection methods, several alternative non-invasive

approaches have been investigated in recent years. For instance, several previous works have been done to investigate the correlation of anemia and physical signs such as the pallor of conjunctivae, tongue, palms and nailbed [7, 12,13]. Since such approaches often need human intervention for estimation and are prone to be affected by subjective judgement of people, they have the disadvantages of instability and labor-intensive.

In order to enhance the objectivity and stability of diagnosis, more methods based on automatic image processing and recognition have been developed. For example, [9] proposed to use multi-linear regression to estimates hemoglobin levels with color data and metadata of fingernail bed smartphone photos, and use them as indicated index for detecting anemia. [10] proposed to use neural networks for building detecting models with conjunctiva images. [3] proposed a deep learning algorithm which was trained using retinal images and alternative metadata to output predicts of blood-haemoglobin levels for anemia diagnosis. [15] developed a smartphone App to estimate hemoglobin level for detecting anemia given a light source shining through a patient's finger.

The above approaches mainly rely on traditional image processing technology which is prone to be affected by the image defects in color and other aspects, and often needs experts' intervention to justify the image quality, making them instability and labor-intensive. Therefore, many researchers choose to switch to deep learning approach for more stable and intelligent detection. Such approach aims at training deep learning models with a set of annotated images, which would give automatic diagnosis without any intervention of professionals. [11] Developed a deep learning approach which uses fundus retinal images plus patients' metadata to perform automated anemia screening. [16] constructed neural network models for anemia detection using conjunctiva

images. [17] proposed a U-Net based conjunctiva segmentation model to detecting anemia. To summarize, previous work of this line have showed the huge potential of deep learning for anemia detection with conjunctiva images, but there is still more room for improvement to make such technology more feasible and robust.

Deep Learning for Multimodality Medical Data: Many practical medical or clinical applications involve multiple modalities rather than single modality data. Therefore, the analysis and processing of multimodality medical data with deep neural network has gradually become a research hotspot [18-19]. Hyperdensenet [20] is a 3D complete convolutional neural network, which extends dense connection to multimodal medical image segmentation task. Experiments show that hyperdensenet achieves satisfactory results in the brain tissue segmentation challenge. In [21], a network suitable for multimodal medical image scene was designed to fuse the complementary anatomical and functional information from different modal images of PET and CT, so as to realize the automatic detection of lung cancer. [22] uses multimodal (B-scan, color Doppler, elastography), multiple angle (cross section and vertical section) breast image data sets to construct an interpretable deep convolution neural network for multi-channel ultrasound image fusion. MMFNet [23] uses multiple encoders to extract multimodal MRI image features, and then fuses the features to complete the semantic segmentation task of nasopharyngeal carcinoma through the decoders. In [24], the novel coronavirus pneumonia severity was predicted with convolution neural network and regression analysis model, using both CT images and structured information from medical records. To the best of our knowledge, little work on multimodality deep learning for anemia detection has been done in previous works, and there is a lot of room for exploration in this research direction. And the approach proposed in this study might be a useful

exploration in applying multimodal algorithms in unifying both global and local information in this aspect.

Problem formulation

Formally, the learning task of this study is as follows: given a collection of conjunctiva images as training data instances, with each image labeled with Y/N annotated by professionals (Y meaning the corresponding individual is diagnosed as having anemia and N meaning otherwise), the target of the task is to construct a deep learning model, which could output an anemia prediction at professional level for any individual, after given his/her image of conjunctiva. To be concrete, we can formally describe the task as: given training data set $D = \{(X_s, y_s), s = 1,2, \dots, m\}$, where X_s denotes the conjunctiva image of the s-th individual, m denotes the sample size, and y_s denotes the diagnosis result of anemia status of the corresponding individual. We wish to construct a deep learning model for mapping X_s to y_s with high accuracy and generalization ability.

As was confirmed by previous clinical research and empirical studies, the rich information contained in conjunctiva images can really provide a sufficient basis for the diagnosis of anemia. Therefore, such learning task is empirically solvable, and all we need to do is to design more effective models based on the characteristics of the data and the task.

However, there is some room for improvement in existing algorithms of in this line of work. To be concrete, as was demonstrated by previous research and in our testing experiments, when using a single conjunctiva image with traditional CNN architecture for anemia detection, it might be difficult to achieve professional judgment accuracy in many cases, especially when there are problems such as unqualified shooting quality or color difference in the images.

Through careful analysis of this situation, we found that one of the main reasons for this situation is, the current deep learning model applied to anemia detection does not fully consider correlation of the global and local information of the image, which makes it easy to encounter problems when it encounters images with poor image quality. We believe that the ability to combine global and local information is an important reason why experts can obtain high accuracy diagnosis. To this end, we proceed to design a deep learning structure with this fusion ability.

Data preparation

All images obtained are anonymized before the experiment, and informed consent was obtained from all the patients. The whole data set contains a total of 852 conjunctiva images, with labels of anemic or non_anemic respectively, and each image's label was obtained according to the corresponding patient's laboratory Hb measurements within 24 hours of photography. All the images are resized to 224*224*3 before training. The whole data was randomly divided into three sub-sets, i.e., training set, verification set and test set. We randomly select around 60% of the whole anemic images and around 60% of the whole non_anemic images to form the training set, and randomly select about 20% of the whole anemic images and about 20% of the whole non_anemic images to form the verification set, and the rest of all the images are put together to form the test set. Note that we divided the whole data set in this manner to ensure that the proportion of positive and negative instances remains almost unchanged across the three sub-sets. The information of the data set with size of each class is showed in Table 2 .

Table 2 Conjunctiva Images Data

class label	non_anemic	anemic	total
whole data	575	277	852

training set	353	179	532
verification set	109	51	160
test set	113	47	160

Evaluation criteria

Considering that the sample sizes for different classes are unbalanced in our data set (the sample size of non_anemic conjunctiva images are more than two times of the size of anemic ones), we use focal loss [28] as loss functions for our GLUDA model and other comparing models. Formally, for every conjunctiva image, if the model output the probability of the ground truth label as p_t, then its focal loss would be:

$$\mathrm{FL}(p_t) = -(1-p_t)^{\gamma}\log(p_t), \tag{9}$$

where γ is a user-defined balanced parameter. As is universally acknowledged, focal loss function is specially designed for training problems with class imbalance. For all the models trained in this section, we simply set γ to be 0.5 across all experiments.

Several evaluation criteria were chosen to measure the performance of different models. The first one is the area under ROC curve (AUC), where ROC stands for the Receiver Operating Characteristic. The other three ones include Sensitivity (Sen), Specificity (Spe) and Accuracy (Acc). Denote TP and FP as the number of true positives and false positives, and TN and FN as the number of true negatives and false negatives, respectively, then the three criteria above can be written as follows:

$$\mathrm{Sen} = TP/(TP+FN) \tag{10}$$

$$\mathrm{Spe} = TN/(TN+FP) \tag{11}$$

$$\mathrm{Acc} = (TP + TN)/(TP + FP + TN + FN) \quad (12)$$

To evaluation the performance of GLUDA model, especially its effectiveness in unifying both global and local views of conjunctiva images, we choose the state-of-the-art EfficientNet as the base model, and use various forms of conjunctiva images as input to obtain different models for comparison. Since GLUDA model uses building blocks of EfficientNet as its main components, such choice of model is relatively reasonable. Specifically, to verify the effectiveness of GLUDA in unifying both global and local views of conjunctiva images, we benchmarked it with the following two models:

(1) EfficientNet_Origin: EfficientNet model trained using the original conjunctiva images;

(2) EfficientNet_ROI: EfficientNet model trained using the ROI of the original conjunctiva images.

The training details and parameter settings of GLUDA model and the comparing models are described as follows. First, as is pre-processed above, we set the input data format as 224*224*3, and use the default network structure of EfficientNet_B0 to form the framework of EfficientNet_Origin, EfficientNet_ROI, and the corresponding building blocks of GLUDA. We adopted transfer learning with the pre-trained EfficientNet_B0 to fine tune the models. We use Adaptive Moment Estimation (Adam) as the optimizer in training all the models, and set the learning rate as 0.003, the learning rate decay factor as 0.99 and the batch size as 16, respectively. Secondly, all the models are trained using the training set showed in Table. 1. And we set the total number of epochs to be 50, and use AUC on the verification set as reference value in performing model selection to output the final models. Thirdly, the AUC, Sen, Spe and Acc of the obtained

models (EfficientNet_Origin, EfficientNet_ROI and GLUDA, respectively) on test set are calculated and output for performance comparison.

Declarations

Ethics approval and consent to participate

This study procedure was approved by the Ethics Committee of Third affiliated Hospital at Sun Yat-sent University, and written informed consent was acquired from all participants.

Consent for publication

All participants gave written informed consent for the publication of findings.

Availability of data and materials

The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China, [Grant Number 62173151, Lijuan Zheng, Senping Tian], Natural Science Foundation of Guangdong Province [Grant Number 2019A1515012048, Shaopeng Liu, Jianhua Guo], The Opening Project of Guangdong Province Key Laboratory of Big Data Analysis and Processing at Sun Yat-sen University, [Grant Number 202001, Jiaming Hong], The Social Sciences Project of Guangzhou

University of Chinese Medicine [Grant Number 2021SKYB01, Jiaming Hong].

Authors' contributions

Conception and design: Lijuan Zheng, Shaopeng Liu, and Jiaming Hong. Administrative support: Senping Tian，Shaopeng Liu and Jiaming Hong ; Collection and assembly of data: Xinpeng Wang ,Xiuxiu Liao and Jianhua Guo, ; Data analysis and interpretation: Jiaming Hong, Shaopeng Liu, Xiuxiu Liao, Lijuan Zheng, and Senping Tian; Manuscript writing: All authors; Final approval of manuscript: All authors.

Acknowledgements

Not applicable.

Authors' information

[1] Department of Automation Science and Engineering, South China University of Technology, Guangzhou, China. [2] Department of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China. [3] School of Medical Information Engineering, Guangzhou University of Chinese Medicine, Guangzhou, China.

References

[1] Smith RE, JR. The clinical and economic burden of anemia. The American journal of managed care (2010)

[2] WHO. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity:

WHO; 2011

[3] Tham Y C , Cheng C Y , Wong T Y . Detection of anaemia from retinal images. Nature Biomedical Engineering, 2019, 4(1):1-2.

[4] Kim T , Choi S H , Lambert-Cheatham N , et al. Toward laboratory blood test-comparable photometric assessments for anemia in veterinary hematology[J]. Journal of Biomedical Optics, 2016, 21(10):107001.

[5] Shah N , Osea E A , Martinez G J . Accuracy of noninvasive hemoglobin and invasive point‐of‐care hemoglobin testing compared with a laboratory analyzer[J]. International Journal of Laboratory Hematology, 2014, 36(1)

[6] Wittenmeier E , Bellosevich S , Mauff S , et al. Comparison of the gold standard of hemoglobin measurement with the clinical standard (BGA) and noninvasive hemoglobin measurement (SpHb) in small children: a prospective diagnostic observational study[J]. Paediatr Anaesth, 2015, 25(10):1046-1053.

[7] Ashwini K , Mandar K , Rajnish J , et al. Accuracy and Reliability of Pallor for Detecting Anaemia: A Hospital-Based Diagnostic Accuracy Study[J]. Plos One, 2010, 5(1):e8545.

[8] Mitani A , Huang A , Venugopalan S , et al. Detection of anaemia from retinal fundus images via deep learning[J]. Nature Biomedical Engineering, 2020, 4(Suppl.):1-10.

[9] Mannino R G , Myers D R , Tyburski E A , et al. Smartphone app for non-invasive detection of anemia using only patient-sourced photos[J]. Nature Communications, 2018, 9(1).

[10] Collings S , Thompson O , Hirst E , et al. Non-Invasive Detection of Anaemia Using Digital Photographs of the Conjunctiva[J]. Plos One, 2016, 11(4):e0153286.

[11] Jain P , Bauskar S , Gyanchandani M . Neural network based non‐invasive method to detect

394 anemia from images of eye conjunctiva[J]. International Journal of Imaging Systems and
395 Technology, 2020.
396 [12] Sheth T N , Choudhry N K , Bowes M , et al. The Relation of Conjunctival Pallor to the
397 Presence of Anemia[J]. Journal of General Internal Medicine, 1997, 12(2):102-106.
398 [13] Chalco J P , Huicho L , Alamo C , et al. Accuracy of clinical pallor in the diagnosis of
399 anaemia in children: a meta-analysis[J]. BMC Pediatrics,5,1(2005-12-08), 2005, 5(1):46.
400 [14] Mitani A , Huang A , Venugopalan S , et al. Detection of anaemia from retinal fundus images
401 via deep learning[J]. Nature Biomedical Engineering, 2020, 4(Suppl.):1-10.
402 [15] Wang E J , Li W , Hawkins D , et al. HemaApp: noninvasive blood screening of hemoglobin
403 using smartphone cameras.[J]. GetMobile Mobile Computing and Communications, 2017,
404 21(2):26-30.
405 [16] Jain P , Bauskar S , Gyanchandani M . Neural network based non-invasive method to
406 detect anemia from images of eye conjunctiva[J]. International Journal of Imaging Systems
407 and Technology, 2020.
408 [17] Kasivisw An Ath An S , Vijayan T B , Simone L , et al. Semantic Segmentation of
409 Conjunctiva Region for Non-Invasive Anemia Detection Applications[J]. Electronics, 2020,
410 9(8):1309.
411 [18] Du J, et a. An Overview of Multi-Modal Medical Image Fusion. Neurocomputing, 2016, 215:
412 3–20.
413 [19] Zhou T, Ruan S, Canu S. A review: Deep learning for medical image segmentation using
414 multi-modality fusion. Array, 2019, 3-4: 100004.
415 [20] Dolz J, Ayed I B, Jing Y, et al. HyperDense-Net: A hyper-densely connected CNN for multi-

modal image segmentation. IEEE Transactions on Medical Imaging, 2019, 38(5): 1116-1126.

[21] Kumar A, Fulham M, Fen D, Kim J. Co-Learning Feature Fusion Maps From PET-CT Images of Lung Cancer. IEEE Transactions on Medical Imaging, 2020, 39, 204-217.

[22] Qian X, et al. Prospective assessment of breast cancer risk from multimodal multiview ultrasound images via clinically applicable deep learning. Nature biomedical engineering. 2021.

[23] Chen H, et al., MMFNet: A multi-modality MRI fusion network for segmentation of nasopharyngeal carcinoma, Neurocomputing., 2020(394):27-40.

[24] Lassau N, et al. Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID-19 patients. Nature Communications2021, 12:634.

[25] Tan M , Le Q V . EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 2019. Proceedings of the 36th International Conference on Machine Learning, 2019

[26] Sandler M , Howard A , Zhu M , et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.

[27] Tan M , Chen B , Pang R , et al. MnasNet: Platform-Aware Neural Architecture Search for Mobile[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019.

[28] Lin T, Goyal P , Girshick R , et al. Focal Loss for Dense Object Detection[C]// 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017:2999-3007.