Auto-Classification of Retinal Diseases in the Limit of Sparse Data Using a Two-Streams Machine Learning Model

C.-H. Huck Yang*[1,2], Fangyu Liu*[3], Jia-Hong Huang*[1], Meng Tian[4],
Hiromasa Morikawa[1], I-Hung Lin M.D.[5], Yi Chieh Liu[6], Hao-Hsiang Yang[6],
and Jesper Tegnèr[1,7]

[1] Biological and Environmental Sciences and Engineering Division, Computer,
Electrical and Mathematical Sciences and Engineering Division, King Abdullah
University of Science and Technology, Thuwal, Saudi Arabia {`chao-han.yang`,
`jiahong.huang`, `hiromasa.morikawa`, `jesper.tegner`}@kaust.edu.sa
[2] Georgia Institute of Technology, GA, USA
[3] University of Waterloo, Canada `fangyu.liu@uwaterloo.ca`
[4] Department of Ophthalmology, Bern University Hospital, Bern, Switzerland
[5] Department of Ophthalmology, Tri-Service General Hospital, Taiwan
[6] National Taiwan University, Taiwan
[7] Unit of Computational Medicine, Center for Molecular Medicine, Department of
Medicine, Karolinska Institutet, Sweden

Abstract. Automatic clinical diagnosis of retinal diseases has emerged
as a promising approach to facilitate discovery in areas with limited
access to specialists. Based on the fact that fundus structure and vascular
disorders are the main characteristics of retinal diseases, we propose a
novel visual-assisted diagnosis hybrid model mixing the support vector
machine (SVM) and deep neural networks (DNNs). Furthermore, we
present a new clinical retina labels collection sorted by the professional
ophthalmologist from the educational project Retina Image Bank, called
EyeNet, for ophthalmology incorporating 52 retina diseases classes. Using
EyeNet, our model achieves 90.40% diagnosis accuracy, and the model
performance is comparable to the professional ophthalmologists.

1 Introduction

Computational retinal disease methods [1, 2] has been investigated extensively
through different signal processing techniques. Retinal diseases are accessible to
machine learning techniques due to their visual nature in contrast to other common human diseases requiring invasive techniques for diagnosis or treatments.
Typically, the diagnosis accuracy of retinal diseases based on the clinical retinal images is highly dependent on the practical experience of a physician or
ophthalmologist. However, training highly-skilled ophthalmologists usually take
years and the number of them, especially in the less-developed area, is still far
from enough. Therefore, developing an automatic retinal diseases detection system is important, and it will broadly facilitate diagnostic accuracy of retinal

diseases. Moreover, for remote rural areas, where there are even no ophthalmologists locally to screen retinal disease, the automatic retinal diseases detection system also helps non-ophthalmologists find the patient of the retinal disease, and further, refer them to the medical center for further treatment.
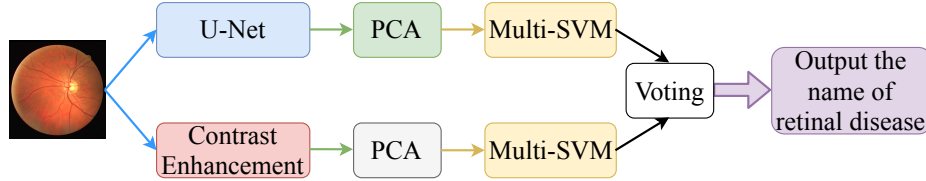
Fig. 1. This figure represents our proposed two-streams model. A raw retinal image as an input of DNNs, U-Net, and as the other input to a contrast enhancement algorithm. Then we pass the output of U-Net to two separated PCA processing. Finally, the output of these two PCA modules is sent as inputs to the retina disease classifier, SVM, which give the outcome of predicted retina disease.

The development of automatic diseases detection (ADD) [3] alleviates enormous pressure from social healthcare systems. Retinal symptom analysis [4] is one of the important ADD applications given that it offers a unique opportunity to improve eye care on a global stage. The World Health Organization estimates that age-related macular degeneration (AMD) and Diabetic Retinopathy, which are two typical retinal diseases, are expected to affect over 500 million people worldwide by 2020 [5].

Moreover, the increasing number of cases of diabetic retinopathy globally requires extending efforts in developing visual tools to assist in the analytic of the series of retinal disease. These decision support systems for retinal ADD, as [6] for non-proliferative diabetic retinopathy have been improved from recent machine learning success on the high dimensional images processing by featuring details on the blood vessel. [7] demonstrated an automated technique for the segmentation of the blood vessels by tracking the center of the vessels on Kalman Filter. However, these pattern recognition based classification still rely on hand-crafted features and only specify for evaluating single retinal symptom. Despite extensive efforts using wavelet signal processing, retinal ADD remains a viable target for improved machine learning techniques applicable for point-of-care (POC) medical diagnosis and treatment in the aging society [8].

To the best of our knowledge, the amount of clinical retinal images are less compared to other cell imaging data, such as blood cell and a cancer cell. However, a vanilla deep learning based diseases diagnosis system requires large amounts of data. Therefore, we propose a novel visual-assisted diagnosis algorithm which is based on an integration of the support vector machine and deep neural networks. The primary goal of this work is to automatically classify 52 specific retinal diseases for human beings with the reliable clinical-assisted ability

on the intelligent medicine approaches. To foster the long-term visual analytics research, we also present a visual clinical label collection, EyeNet, including several crucial symptoms as AMD, DR, uveitis, BRVO, BRAO.

Contributions.

- We design a novel two-streams-based algorithm on the support vector machine and deep neural networks to facilitate medical diagnosis of retinal diseases.
- We present a new clinical labels collection, EyeNet, for Ophthalmology with 52 retina diseases classes as a crucial aid to the ophthalmologist and medical informatics community.
- Finally, we visualize the learned features inside the DNNs model by heat maps. The visualization helps in understanding the medical comprehensibility inside our DNNs model.

2   Related Work

In this section, we review some works related to our proposed method. We divide the related works into three parts including medical dataset comparison, dimension reduction by feature extraction, and image segmentation by neural networks.

2.1 Medical Dataset Comparison

Large-scale datasets help the performance of deep learning algorithms comparable to human-level on the tasks of speech recognition [9], image classification and recognition [10], and question answering [11–15]. In the medical community, large scale medical datasets also help algorithms achieve expert-level performance on detection of skin cancer [16], diabetic retinopathy [17], heart arrhythmias [18], pneumonia [19], brain hemorrhage [20], lymph node metastases [21], and hip fractures [22].

Recently, the number of openly available medical datasets is growing. In Table 1, we try to provide a summary of the publicly available medical image datasets related to ours. According to Table 1, we notice that the recently released ChestXray14 [23] is the largest medical dataset containing 112,120 frontal-view chest radiographs with up to 14 thoracic pathology labels. Moreover, the smallest medical dataset is DRIVE [24] containing 40 retina images. Regarding the openly available musculoskeletal radiograph databases, the Stanford Program for Artificial Intelligence in Medicine and Imaging has a medical dataset containing pediatric hand radiographs annotated with skeletal age (AIMI). The Digital Hand Atlas [25] includes the left-hand radiographs which are from children of various ages labeled with radiologist readings of bone age. Then, our proposed EyeNet contains 52 classes of diseases and 1747 images.

2.2 Dimension Reduction by Feature Extraction

Feature extraction is a method to make the task of pattern classification or recognition easier. In image processing and pattern recognition, feature extraction is one of the special forms of dimensionality reduction [28] in some sense.

Name of Dataset	Study Type	Label	Number of Images
EyeNet	**Retina**	**Labels mining of Retinal Diseases**	**1747**
DRIVE [24]	Retina	Retinal Vessel Segmentation	40
MURA [26]	Musculoskeletal (Upper Extremity)	Abnormality	40,561
Digital Hand Atlas [25]	Musculoskeletal (Left Hand)	Bone Age	1,390
ChestX-ray14 [23]	Chest	Multiple Pathologies	112,120
DDSM [27]	Mammogram	Breast Cancer	10,239

Table 1. Overview of available different types of medical label collection and image datasets.

The purpose of feature extraction is to exploit the most relevant information based on the original data and describe the information in a space with lower dimensions. For example, typically the size of original medical image data, such as functional magnetic resonance imaging (fMRI) scans, is very large and it causes algorithms computationally inefficient. In this case, we will transform the original data into a reduced representation set of features. That is, we exploit a set of feature vectors to describe the original data and the process is called image feature extraction. In [29], the authors mention that the representation by extracted feature vectors should have a dimensionality that corresponds to the intrinsic dimensionality of the original data. Then, the intrinsic dimensionality of data is the minimum number of parameters required to account for the properties of the original data. Moreover, the authors of [30] claim that dimensionality reduction mitigates the curse of dimensionality and the other undesired properties of spaces with high dimensions. The dimensionality reduction by feature extraction method has been used in many different application fields such as document verification [31], character recognition [32], extracting information from sentences [14, 33, 13], machine translation [34, 35] and so on.

**2.3 Image Segmentation by Neural Networks**

Typically, researchers exploit the convolutional neural networks to do image classification tasks with a single class output label. However, in the biomedical image processing tasks, the output should contain localization. That is to say, a class label is assigned to each pixel. Furthermore, thousands of images in training set are typically beyond reach in the biomedical tasks. Therefore, the authors of [36] train a neural networks model, with sliding-window, to predict the output class label of each pixel by providing a sub-region, small patch, around that pixel as input.

In [36], we know that the neural network model can do localization and the number of training data, in the sense of patches, is much larger than the training images. Apparently, [36] has two drawbacks. First, there exists some trade-off between the use of context and localization accuracy. Then, since the model runs separately for each small patch and there is much redundancy due to overlapping patches, it is not efficient in the sense of computational speed. Recently, the authors of [37, 38] have proposed an approach which can do the good localization and use of context at the same time.

In the U-Net paper [39], the authors build upon an even more elegant neural network architecture, the so-called fully convolutional network [40]. The authors modify the architecture such that it works with very few training images and

produces even more accurate image segmentation. The main idea of [40] is to supplement a usual contracting neural network by the successive layers. Then, the authors exploit upsampling operators to replace pooling operators, so the resolution of output is enhanced by these layers. In order to do localization, the authors combine the upsampled output and high-resolution features from the contracting path. Furthermore, a successive convolutional layer learns to assemble a more accurate output based on this information. Due to the advantages of U-Net mentioned above, we modify and incorporate the U-Net to our proposed method.
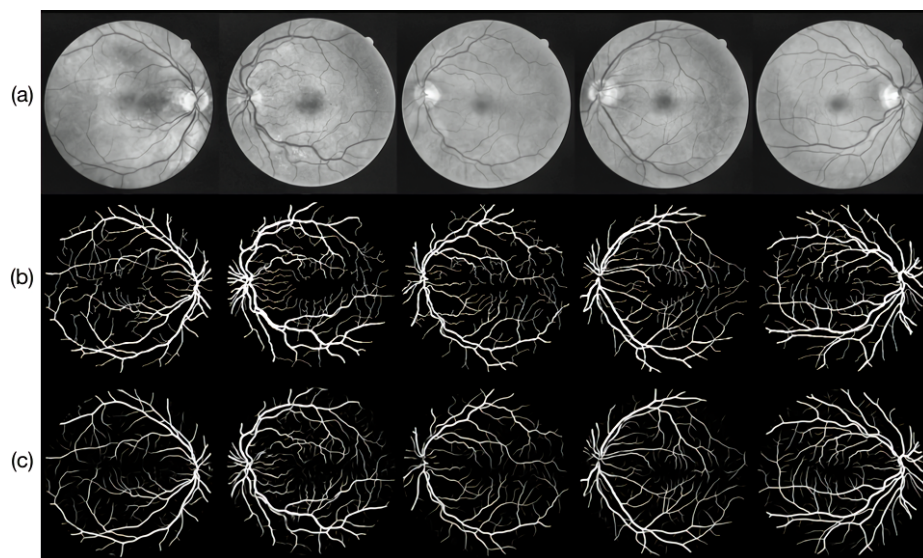
Fig. 2. The figure shows the result of U-Net effects on (a), unseen eyeball clinical images with different morphological shapes. (b) is the ground truth and (c) is the generated result of vessel-subtracted U-Net. Based on (b) and (c), we discover that the results are highly similar to the ground truth.

3 Methodology

In this section, we present the workflow of our proposed model, referring to Figure 1.

3.1 U-Net

DNNs has greatly boosted the performance of image classification due to its power of image feature learning [41]. The active retinal disease is characterized by exudates around retinal vessels resulting in cuffing of the affected vessels [42]. However, ophthalmology images from clinical microscopy are often overlayed

with white sheathing and minor features. Segmentation of retinal images has been investigated as a critical [43] visual-aid technique for ophthalmologists. U-Net [39] is a functional DNNs, especially for segmentation. Here, we proposed a modified version of U-Net by reducing the copy and crop processes with a factor of two. The adjustment could speed up the training process and have been verified as an adequate semantic effect on small size images. We use cross-entropy for evaluating the training processes as:

$$E = \sum_{x \in \Omega} w(x) log(p_l(x)) \qquad (1)$$

where p_l is the approximated maximum function, and the weight map is then computed as:
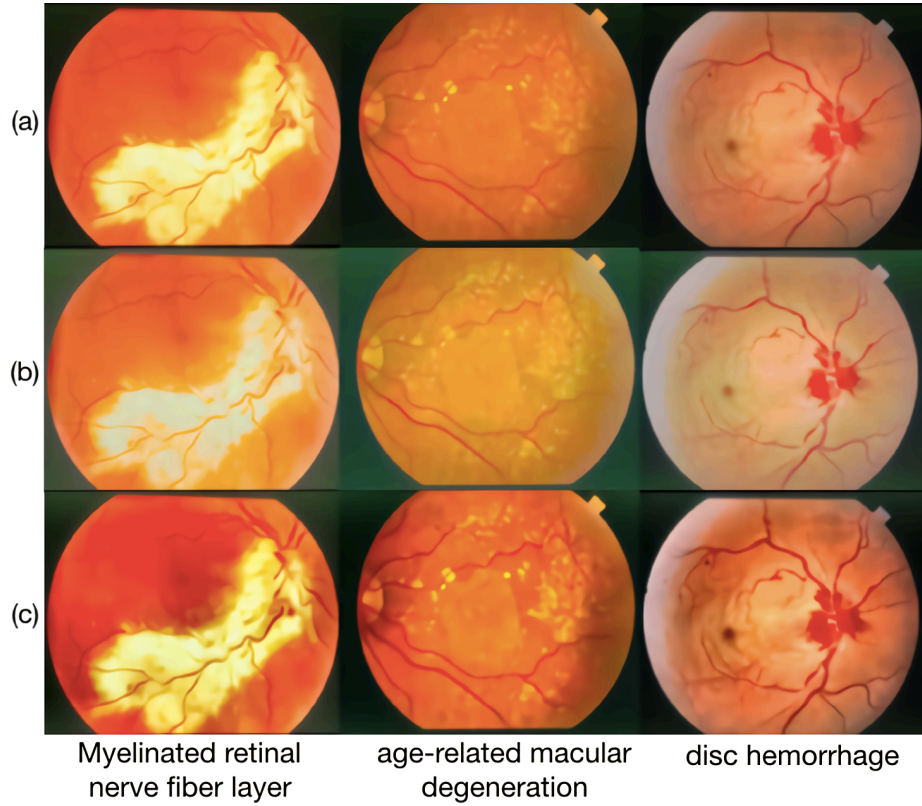
$$w(x) = w_c(x) + w_0 \cdot exp(\frac{-(d_{x1} + d_{x2})^2}{2\sigma^2}) \qquad (2)$$

d_{x1} designates the distance to the border of the nearest edges and d_{x2} designates the distance to the border of the second nearest edges. LB score is shown as [8]. We use the deep convolutional neural network (CNN) of two 3×3 convolutions. Each step followed by a rectified linear unit (ReLU) and a 2×2 max pooling operation with stride 2 for downsampling; a layer with an even x- and y-size is selected for each operation. For the U-Net model, we use existing DRIVE [24] dataset as the training segmentation mask. Then, we use Our proposed model converges at the 44th epoch when the error rate of the model is lower than 0.001. The Jaccard similarity of our U-Net model is 95.59% by validated on a 20% test set among EyeNet shown in Figure 2. This model is robust and feasible for different retinal symptoms as illustrated in Figure 3. The area under ROC curve is 0.979033 and the area under the Precision-Recall curve is 0.910335.

3.2 Principal Component Analysis as Eigenface in the limit of Sparse Data

$$\lambda_k = \frac{1}{M} \sum_{n=1}^{M} (u_k^T \Phi_n)^2 \qquad (3)$$

Eigenface [44] is classical and high-efficient image recognition technique derived from the covariance matrix of the probability distribution over the high-dimensional vector space of face images. Even with a single training image, previous works [45, 46] of eigenface already established robust automatic classification with confident accuracy (85.6%) by combined principal component analysis (PCA) and SVM classifiers. As a biological feature, retinal images share similar properties with the human face for a potential with eigenface recognition [47] included finite semantic layout between facial features and ophthalmological features [48]. The eigenface could be calculated [44] by maximizing the equivalent (3), where Φ_n represent the face differ, u_k is a chosen k_{th} vector, λ_k is the k_{th} eigenvalue, and M is a number of the dimension space. In our experiment, we

| Myelinated retinal nerve fiber layer | age-related macular degeneration | disc hemorrhage |

Fig. 3. This figure illustrates the qualitative results of contrast enhancement algorithm from the original clinical images to the (b) histogram equalization (c) contrast-limited adaptive histogram equalization.

select the $k_{UNet} = 40$ and $k_{RGB} = 61$ to generate a eigenface with highest accuracy for the U-Net-stream and RGB-stream separately.

3.3 Support Vector Machine

Support Vector Machine is a machine learning technique for classification, regression, and other learning tasks. Support vector classification (SVC) in SVM, map data from an input space to a high-dimensional feature space, in which an optimal separating hyperplane that maximizes the boundary margin between the two classes is established. The hinge loss function is shown as:

$$\frac{1}{n} \left[\sum_{i=1}^{n} max(0, 1 - y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} - b)) \right] + \lambda \|\boldsymbol{w}\|^2 \quad (4)$$

Where the parameter λ determines the trade off between increasing the margin-size and ensuring that the $\boldsymbol{x_i}$ lies on the right side of the margin. We use radial basis function (RBF) and polynomial kernel for SVC, which have been widely discussed [49] as a kernel-based fast SVC for images.

3.4 Contrast Enhancement

Contrast enhancement techniques play a vital role in image processing to bring out the information that exists within a less dynamic range of that image. As a major clinical feature, fundus [50, 51] structure is highly related [52] the image contrast [53]. Here, we use histogram equalization for the contrast enhancement in retinal images. Compare to the original images, images after histogram equalization show the light color detail as lesions in Figure 3(b). Images after contrast-limited adaptive histogram equalization (CLAHE) give further features as areas of retinopathy in Figure 3(c).

4   Efforts on Retinal Dataset

Retina Image Bank (RIB) [54] is an international clinical project launched by American Society of Retina Specialists in 2012, which allow retina specialists and ophthalmic photographers around the world to share the existing clinical cases online for medicine-educational proposes for patients and physicians in developing countries lack training resource. Any researcher could join as a contributor for dedicating the retinal images or as a visitor using the medical images and label for non-commercial propose. With the recent success on dataset collection, such as ImageNet [55], we believe that the effort of sorting and mining the clinical labels from RIB is valuable. With a more developer-friendly information pipeline, both Ophthalmology and Computer Vision community could go further on the analytical researches on medical informatics. Our proposed label collection, EyeNet is mainly based on the RIB and following the RIB's using guideline.

5   Experiments

In this section, we describe the implementation details and experiments we conducted to validate our proposed method.

5.1 Label Collection

For experiments, the EyeNet is randomly divided into three parts: 70% for training, 10% for validation and 20% for testing. All the training data have to go through the PCA before SVM. All classification experiments are trained and tested on the EyeNet.

5.2 Setup

The EyeNet has been processed to U-Net to generate a subset with a semantic feature of the blood vessel. For the DNNs and Transfer Learning models, we directly use the RGB images from the retinal dataset. EyeNet will be published online after getting accepted. For the CLAHE processing, we use *adapthisteq* function from the image toolbox in MATLAB.

5.3 Deep Convolutional Neural Networks

CNN has demonstrated extraordinary performance in visual recognition tasks [55], and the state of the art is in a great many vision-related benchmarks and challenges [56]. With little or no prior knowledge and human effort in feature
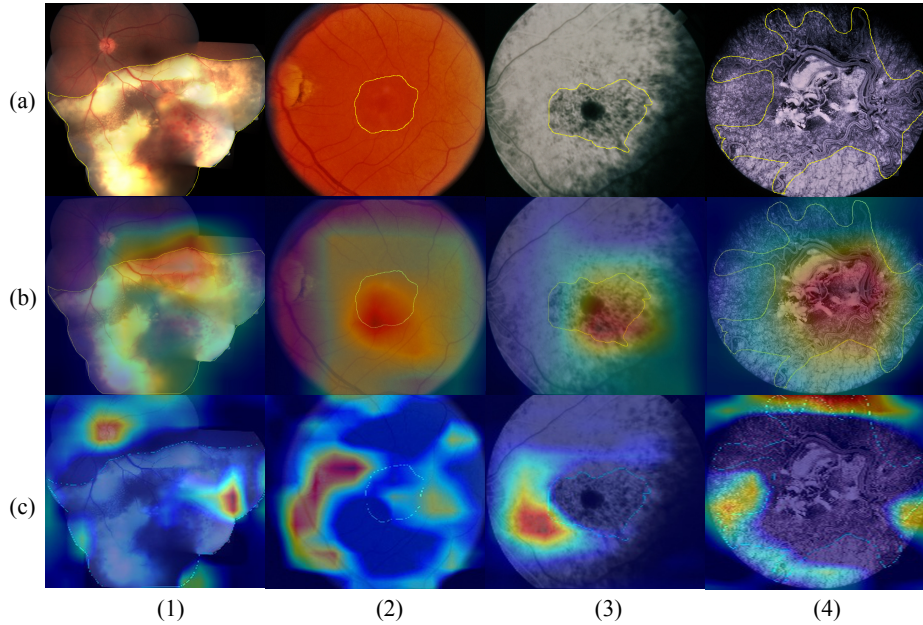
Fig. 4. We use figure (i, j), where $i = a, b, c$ and $j = 1, 2, 3, 4$, to demonstrate that our proposed method can capture the similar lesion areas as the ophthalmologist's manual annotations, i.e., the yellow sketches.

design, it yet provides a general and effective method solving variant vision tasks in variant domains. This new development in computer vision has also shown great potential for helping/replacing human judgment in vision problems like medical imaging [16], which is the topic we try to address in this paper. In this section, we introduce several baselines in multi-class image recognition and compare their results on the EyeNet.

Baseline1-AlexNet

AlexNet [55] brought up a succinct network architecture, with 3 fully connected layers, 5 convolutional layers, and the activation function is ReLU [57].

Baseline2-VGG

The authors of VGG [41] exploit the filters (3x3) repeatedly to replace the large filters (5x5,7x7) in traditional architectures. By increasing depths of the network, it achieved better results on ImageNet with fewer parameters.

Baseline3-ResNet

Residual Networks [58], one of the most popular neural networks today, utilize skip connections or short-cuts to jump over some layers. With skip connections, the network essentially collapses into a shallower network in the initial phase and this makes it easier to be trained, and then it expands its layers as it learns more of the feature space.

Baseline4-SqueezeNet

Hybrid-Ratio	RBF	Polyn.
0% : 100%	0.8159	0.8391
40% : 60%	0.8371	0.8381
50% : 50%	**0.9002**	0.8632
60% : 40%	0.8881	**0.9040**
100% : 0%	0.8324	0.8241

Table 2. Accuracy comparison of the two-streams model with Radial basis function (RBF) and polynomial kernel. We use the hybrid-ratio [61] of the mixed weighted voting between two multi-SVCs trained from images over U-Net and CLAHE.

Model	Pretrained	Random Init.
AlexNet	0.7912	0.4837
VGG11	0.8802	**0.7579**
VGG13	0.8721	0.7123
ResNet18	**0.8805**	0.7250
SqueezeNet	0.8239	0.5625

Table 3. Accuracy comparison of three DNNs baselines on EyeNet.

In real world, medical imaging tasks usually require a small and effective model to adapt to limited resources. As some deep neural networks cost several hundred megabytes to store, SqueezeNet [59] adopting model compression techniques has achieved the accuracy of AlexNet level with around 500 times smaller models.

5.4 Transfer Learning

We use a transfer learning framework from the normalized ImageNet [55] to the EyeNet for solving the small samples issue on the computational retinal visual analytics. With sufficient and utilizable training classified model, Transfer Learning resolves the challenge of Machine Learning in the limit of minimal amount of training labels and it drastically reduces the data requirements. The first few layers of DNNs learn features , similar to Gabor filters and color blobs, and these features appear not to be specific to any particular task or dataset and thus applicable to other datasets and tasks [60]. Our experiments show the significant improvement after we apply the pretrained parameters on our deep learning based models, referring to Table 3 and Table 4.

5.5 Two-Streams Results

All SVM has implemented in Matlab with libsvm [62] module. We separate both the original retinal dataset and the subset to three parts included 70% training set, 20% test set, and 10% validation set. By training two multiple-classes SVM models on both original EyeNet and the subset, we implement a weighted voting method to identify the candidate of retina symptom. We have testified different weight ratio as $Hybrid-Ratio$, SVM model with {Images over CLAHE: Image over U-Net }, with different accuracy at Table 2. We have verified the model without over-fitting by the validation set via normalization on the accuracy with $\tilde{2}.03\%$ difference.

5.6 Deep Neural Networks Results

All DNNs are implemented in PyTorch. We use identical hyperparameters for all models. The training lasts 400 epochs. The first 200 epochs take a learn-

Model	Pretrained	Random Init.
AlexNet	0.7952	0.4892
VGG11	0.8726	0.7583
VGG13	**0.8885**	**0.7588**
ResNet18	0.8834	0.6741
SqueezeNet	0.8349	0.5721

Table 4. Accuracy comparison of three DNNs baselines on EyeNet.

ing rate of 1e-4 and the second 200 take 1e-5. Besides, we apply random data augmentation during training. In every epoch, there is 70% probability for a training sample to be affinely transformed by one of the operations in {flip, rotate, transpose}×{random crop}. Though ImageNet and our Retinal Dataset are much different, using weights pretrained on ImageNet rather than random ones has boosted test accuracy of any models with 5 to 15 percentages, referring to Table 3. Besides, pretrained models tend to converge much faster than random initialized ones as suggested in Figure 4. The performance of DNNs on our retinal dataset can greatly benefit from the knowledge of other domains.

5.7 Neuron Visualization for Medical Images

Importantly, we verified the hypothesis that vessel-based segmentation and contrast enhancement are two coherent features to decide the type of retinal diseases. Using techniques of generating class activation maps introduced by [63], we visualized feature maps of the final convolution layer of ResNet18 (which is one of our deep learning model baselines). We notice that the features learned by deep learning models agree with our intuitions about developing the two-stream machine learning model. In fact, in the clinical diagnosis process, "vessel patterns" and "fundus structure" are also the two most crucial features to identify the symptom of different diseases. These two types of features actually cover more than 80% of retinal diseases [50, 51].

6   Conclusion and Future Work

In this work, we have designed a novel hybrid model for visual-assisted diagnosis based on the SVM and U-Net. The performance of this model shows the higher accuracy, 90.40%, over the other pre-trained DNNs models as an aid for ophthalmologists. Also, we propose the EyeNet to benefit the medical informatics research community. Finally, since our dataset not only contains images but also text information of the images, image captioning and Visual Question Answering [13–15] based on the retinal images are also the interesting future directions. Our work may also help the remote rural area, where there are no ophthalmologists locally, to screen retinal disease without the help of ophthalmologists in the future.

References

1. Tan, O., Chopra, V., Lu, A.T.H., Schuman, J.S., Ishikawa, H., Wollstein, G., Varma, R., Huang, D.: Detection of macular ganglion cell loss in glaucoma by

fourier-domain optical coherence tomography. Ophthalmology **116** (2009) 2305–2314

2. Lalezary, M., Medeiros, F.A., Weinreb, R.N., Bowd, C., Sample, P.A., Tavares, I.M., Tafreshi, A., Zangwill, L.M.: Baseline optical coherence tomography predicts the development of glaucomatous change in glaucoma suspects. American journal of ophthalmology **142** (2006) 576–582

3. Sharifi, M., Fathy, M., Mahmoudi, M.T.: A classified and comparative study of edge detection algorithms. In: Information Technology: Coding and Computing, 2002. Proceedings. International Conference on, IEEE (2002) 117–120

4. Abràmoff, M.D., Garvin, M.K., Sonka, M.: Retinal imaging and image analysis. Reviews in biomedical engineering **3** (2010) 169–208

5. Pizzarello, L., Abiose, A., Ffytche, T., Duerksen, R., Thulasiraj, R., Taylor, H., Faal, H., Rao, G., Kocur, I., Resnikoff, S.: Vision 2020: The right to sight: a global initiative to eliminate avoidable blindness. Archives of ophthalmology **122** (2004) 615–620

6. Bhattacharya, S.: Watermarking digital images using fuzzy matrix compositions and (α, β)-cut of fuzzy set. International Journal of Advanced Computing (2014)

7. Lin, C.Y., Wu, M., Bloom, J.A., Cox, I.J., Miller, M.L., Lui, Y.M.: Rotation-, scale-, and translation-resilient public watermarking for images. In: Security and watermarking of multimedia contents II. Volume 3971., International Society for Optics and Photonics (2000) 90–99

8. Cochocki, A., Unbehauen, R.: Neural networks for optimization and signal processing. John Wiley & Sons, Inc. (1993)

9. Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S., Sengupta, S., Coates, A., et al.: Deep speech: Scaling up end-to-end speech recognition. arXiv preprint arXiv:1412.5567 (2014)

10. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE (2009) 248–255

11. Rajpurkar, P., Zhang, J., Lopyrev, K., Liang, P.: Squad: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250 (2016)

12. Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Lawrence Zitnick, C., Parikh, D.: Vqa: Visual question answering. In: Proceedings of the ICCV. (2015) 2425–2433

13. Huang, J.H., Alfadly, M., Ghanem, B.: Vqabq: visual question answering by basic questions. arXiv:1703.06492 (2017)

14. Huang, J.H., Dao, C.D., Alfadly, M., Ghanem, B.: A novel framework for robustness analysis of visual qa models. arXiv:1711.06232 (2017)

15. Huang, J.H., Alfadly, M., Ghanem, B.: Robustness analysis of visual qa models by basic questions. arXiv:1709.04625 (2017)

16. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542** (2017) 115

17. Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. Jama **316** (2016) 2402–2410

18. Rajpurkar, P., Hannun, A.Y., Haghpanahi, M., Bourn, C., Ng, A.Y.: Cardiologist-level arrhythmia detection with convolutional neural networks. arXiv preprint arXiv:1707.01836 (2017)

19. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., et al.: Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225 (2017)

20. Grewal, M., Srivastava, M.M., Kumar, P., Varadarajan, S.: Radnet: Radiologist level accuracy using deep learning for hemorrhage detection in ct scans. In: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on, IEEE (2018) 281–284

21. Bejnordi, B.E., Veta, M., van Diest, P.J., van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J.A., Hermsen, M., Manson, Q.F., Balkenhol, M., et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. Jama **318** (2017) 2199–2210

22. Gale, W., Oakden-Rayner, L., Carneiro, G., Bradley, A.P., Palmer, L.J.: Detecting hip fractures with radiologist-level performance using deep neural networks. arXiv preprint arXiv:1711.06504 (2017)

23. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2017) 3462–3471

24. Staal, J., Abràmoff, M.D., Niemeijer, M., Viergever, M.A., Van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. TMI **23** (2004) 501–509

25. Gertych, A., Zhang, A., Sayre, J., Pospiech-Kurkowska, S., Huang, H.: Bone age assessment of children using a digital hand atlas. Computerized Medical Imaging and Graphics **31** (2007) 322–331

26. Rajpurkar, P., Irvin, J., Bagul, A., Ding, D., Duan, T., Mehta, H., Yang, B., Zhu, K., Laird, D., Ball, R.L., et al.: Mura dataset: Towards radiologist-level abnormality detection in musculoskeletal radiographs. arXiv preprint arXiv:1712.06957 (2017)

27. Heath, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, P.: The digital database for screening mammography. Digital mammography (2000) 431–434

28. Costa, J.A., Hero, A.: Classification constrained dimensionality reduction. In: Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on. Volume 5., IEEE (2005) v–1077

29. Fukunaga, K.: Introduction to statistical pattern recognition. Academic press (2013)

30. Jimenez, L.O., Landgrebe, D.A.: Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) **28** (1998) 39–54

31. Yang, Z., He, X., Gao, J., Deng, L., Smola, A.: Stacked attention networks for image question answering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 21–29

32. Trier, Ø.D., Jain, A.K., Taxt, T.: Feature extraction methods for character recognition-a survey. Pattern recognition **29** (1996) 641–662

33. Srihari, R., Li, W.: Information extraction supported question answering. Technical report, CYMFONY NET INC WILLIAMSVILLE NY (1999)

34. Somers, H.: Example-based machine translation. Machine Translation **14** (1999) 113–157

35. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)

36. Ciresan, D., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: Advances in neural information processing systems. (2012) 2843–2851

37. Hariharan, B., Arbeláez, P., Girshick, R., Malik, J.: Hypercolumns for object segmentation and fine-grained localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2015) 447–456

38. Seyedhosseini, M., Sajjadi, M., Tasdizen, T.: Image segmentation with cascaded hierarchical models and logistic disjunctive normal networks. In: Computer Vision (ICCV), 2013 IEEE International Conference on, IEEE (2013) 2168–2175

39. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on MICCAI, Springer (2015) 234–241

40. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2015) 3431–3440
41. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 (2014)
42. Khurana, A.: Comprehensive ophthalmology. New Age International Ltd (2007)
43. Rezaee, K., Haddadnia, J., Tashk, A.: Optimized clinical segmentation of retinal blood vessels by using combination of adaptive filtering, fuzzy entropy and skeletonization. Applied Soft Computing **52** (2017) 937–951
44. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on, IEEE (1991) 586–591
45. Lyons, M.J., Budynek, J., Akamatsu, S.: Automatic classification of single facial images. IEEE Transactions on pattern analysis and machine intelligence **21** (1999) 1357–1362
46. Wu, J., Zhou, Z.H.: Face recognition with one training image per person. Pattern Recognition Letters **23** (2002) 1711–1719
47. Moghaddam, B., Wahid, W., Pentland, A.: Beyond eigenfaces: Probabilistic matching for face recognition. In: Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on, IEEE (1998) 30–35
48. Akram, M.U., Tariq, A., Khan, S.A.: Retinal recognition: Personal identification using blood vessels. In: Internet Technology and Secured Transactions (ICITST), 2011 International Conference for, IEEE (2011) 180–184
49. Kuo, B.C., Ho, H.H., Li, C.H., Hung, C.C., Taur, J.S.: A kernel-based feature selection method for svm with rbf kernel for hyperspectral image classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **7** (2014) 317–326
50. Crick, R.P., Khaw, P.T.: A textbook of clinical ophthalmology: a practical guide to disorders of the eyes and their management. (World Scientific)
51. Akram, I., Rubinstein, A.: Common retinal signs. an overview. Optometry Today (2005)
52. Tang, S., Huang, L., Wang, Y., Wang, Y.: Contrast-enhanced ultrasonography diagnosis of fundal localized type of gallbladder adenomyomatosis. BMC gastroenterology **15** (2015) 99
53. Noyel, G., Thomas, R., Bhakta, G., Crowder, A., Owens, D., Boyle, P.: Superimposition of eye fundus images for longitudinal analysis from large public health databases. Biomedical Physics & Engineering Express **3** (2017) 045015
54. : Retina Image Bank: A project from the American Society of Retina Specialists. (http://imagebank.asrs.org/about) Accessed: 2018-06-30.
55. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in NIPS. (2012) 1097–1105
56. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: CVPR, IEEE (2017) 5987–5995
57. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10). (2010) 807–814
58. He, K., Zhang, X., Ren, S., Sun, J.: (Deep residual learning for image recognition)
59. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. arXiv:1602.07360 (2016)
60. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: NIPS. (2014) 3320–3328
61. Yang, C.H.H., Huang, J.H., Liu, F., Chiu, F.Y., Gao, M., Lyu, W., Tegner, J., et al.: A novel hybrid machine learning model for auto-classification of retinal diseases. arXiv preprint arXiv:1806.06423 (2018)

62. Chang, C.C., Lin, C.J.: Libsvm: a library for support vector machines. TIST **2** (2011) 27
63. Zhou, B., Khosla, A., A., L., Oliva, A., Torralba, A.: Learning Deep Features for Discriminative Localization. CVPR (2016)