Diabetic Blindness Detection System using Convolution Neural Network

[1]Akash Gajbhiye, [2]Parvez Iqbal Afsar Iqbal, [3]Dipali Hande, [4]Mayuri Gajbhiye

Information Technology Department
J D College of Engineering & Management, Nagpur

Abstract: Diabetic retinopathy could be a leading reason behind sightlessness among working-age adults. Early detection of this condition is crucial for a decent prognosis. During this paper, we tend to demonstrate the utilization of convolutional neural networks (CNNs) on color structure pictures for the popularity task of diabetic retinopathy staging. Our network models achieved check metric performance cherish baseline literature results, with a validation sensitivity of ninety fifth. Transfer learning on pre-trained GoogLeNet and AlexNet models from ImageNet improved peak check set accuracies to seventy four.5%, 68.8%, and 57.2% on 2-ary, 3-ary, and 4-ary classification models, severally.

Keywords: Blindness Detection, Diabetic Retinopathy (DR), Convolutional Neural Networks (CNNs), Fundoscopic images, Ophthalmology

1. Introduction

Approximately four hundred and twenty million folks worldwide are diagnosed with diabetes. The prevalence of this sickness has doubled within the past thirty years and is just expected to extend, notably in Asia. of these with polygenic disorder, just about tierce square measure expected to be diagnosed with diabetic retinopathy (DR), a chronic disease which will accomplish irreversible vision loss. Early detection, that is crucial permanently prognosis, depends on sure-handed readers and is each labor and time-intensive. This poses a challenge in areas that historically lack access to sure-handed clinical facilities. Moreover, the manual nature of DR screening ways promotes widespread inconsistency among readers. Finally, given a rise in prevalence of each polygenic disorder and associated retinal complications throughout the planet, manual ways of diagnosing is also unable to stay a pace with demand for screening services.

In this paper we tend to introduce associate automatic DR grading system capable of classifying pictures supported sickness pathologies from four severity levels. A convolutional neural network (CNN) convolves associate input image with an outlined weight matrix to extract specific image options while not losing placement data. we tend to ab initio valuate totally different architectures to see the most effective acting CNN for the binary classification task and aim to realize literature according performance levels. we tend to then look for to coach multi-class models that enhance sensitivities for the gentle or early stage categories, as well as numerous ways information|of knowledge|of information} preprocessing and data augmentation to each improve check accuracy moreover as increase our effective dataset sample size. we tend to address considerations of knowledge fidelity and quality by collating a group of eye doctor verified pictures. Finally, we tend to address the difficulty of lean sample size employing a deep bedded CNN with transfer learning on discriminant color area for the popularity task. we tend to then trained and tested 2 CNN architectures, AlexNet and GoogLeNet, as 2-ary, 3-ary and 4-ary classification models. they're tuned to perform optimally on a coaching dataset mistreatment many techniques as well as batch standardisation, L2 regularization, dropout, learning rate policies and gradient descent update rules3. Experimental studies were conducted mistreatment 2 primary information sources, the publically out there Kaggle dataset of thirty five,000 retinal pictures with 5-class labels (normal, mild, moderate, severe, finish stage) and a physician-verified Messidor-1 dataset of one,200 color complex body part pictures with 4-class labels. Throughout this study we tend to aim to elucidate a simpler means that of classifying early stage diabetic retinopathy for potential clinical advantages.

2. Background and Related Work

Diagnosis of pathological findings in fundoscopy, a medical technique to check the tissue layer, depends on a fancy vary of options and localizations among the image. The diagnosing is especially troublesome for patients with early stage diabetic retinopathy as this depends on discerning the presence of microaneursyms, little saccular outpouching of capillaries, retinal hemorrhages, damaged blood vessels—among different features—on the fundoscopic pictures. prototypic retinal sickness stages area unit shown in Fig. 1.
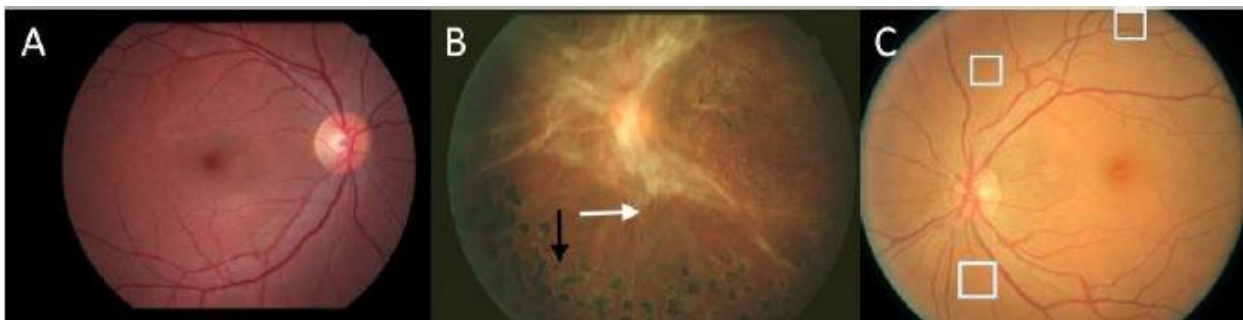
Figure 1.

Representative retinal images of DR at various stages of the disease, as labeled: A- normal, B- end stage, C- early stage. Arrows in B point to pathological indications. White boxes in C enclose very small lesions that the CNNs have difficulty discerning.

Computer-aided diagnosing of diabetic retinopathy has been explored within the past to scale back the burden on opthamologists and mitigate diagnostic inconsistencies between manual readers. machine-controlled ways to find microaneursyms and dependably grade fundoscopic pictures of diabetic retinopathy patients are active areas of analysis in laptop vision. the primary artificial neural networks explored the flexibility to classify patches of traditional membrane while not blood vessels, traditional retinas with blood vessels, pathologic retinas with exudates, and pathologic retinas with microaneurysms. The accuracy of having the ability to find microaneursyms compared to traditional patches of membrane was reported at seventy four.

Past studies victimisation numerous high bias, low variance digital image process techniques have performed well at distinguishing one specific feature utilized in the detection of refined malady like the employment of top-hat formula for microaneurysm detection. However, a spread of alternative options besides microaneurysms square measure efficacious for malady detection.

Additional ways of detective work microaneurysms and grading DR involving k-NN support vector machines, and ensemble-based ways have yielded sensitivities and specificities at intervals the ninetieth vary victimisation numerous feature extraction techniques and preprocessing algorithms.

3. Dataset

We used 2 fundo scope image datasets to coach an automatic classifier for this study. speedy prototyping was expedited by pretrained models obtained from the ImageNet visual beholding information. Diabetic retinopathy pictures were noninheritable from a Kaggle dataset of thirty five,000 pictures with 5-class labels (normal, mild, moderate, severe, finish stage) and Messidor-1 dataset of one,200 color structure pictures with 4-class labels (normal, mild, moderate, severe). each datasets incorporates color pictures that modify tall and dimension between the low tons of to low thousands. Compared to Messidor-1, the Kaggle dataset consists of a bigger proportion of uninterpretable pictures because of artefact preponderance, faulty labeling and poor quality. when coaching on the larger Kaggle datasets and distinguishing limitations of the standard approach to retinal image classification, we have a tendency to performed experiments on higher fidelity datasets with improved image quality and reliable labeling.

In the interest of economical model building, we have a tendency to progressed to a smaller however a lot of ideal dataset for learning tough options. The Revolutionary calendar month dataset was supplemented with a Kaggle partition (MildDR) consisting of 550 pictures that was verified for its effectivity by direct MD interpretation. The dataset contains pictures from a disparate patient population with extraordinarily varied levels of structure photography lighting and is tagged in an exceedingly consistent manner. The lighting affects element intensity values at intervals the photographs and creates variation unrelated to classification pathology. Our study solely uses the retinopathy grade as a reference, an outline of that is provided in Table one beside the amount of pictures for every class.

To assess the quality of transfer learning, we have a tendency to thought-about many frameworks to style, train and deploy a changed version of GoogLeNet as a baseline 2-ary, 3-ary and 4-ary classification model. Our final model was enforced in Tensorflow and was influenced by results from a deep learning GPU interactive coaching system (DIGITS) that enabled speedy neural network epitome coaching, performance observance and period of time visualizations.

Table 1.

Retinopathy grades in Messidor dataset

Grade	Description	Nb Images
R0	$(N_{MA} = 0)$ AND $(N_{HE} = 0)$	546
R1	$(0 < N_{MA} \leq 5)$ AND $(N_{HE} = 0)$	153
R2	$(5 < N_{MA} < 15)$ AND $(0 < NHE < 5)$ AND $(N_{NV} = 0)$	247
R3	$(N_{MA} \geq 15)$ OR $(N_{HE} \geq 5)$ OR $(N_{NV} > 0)$	254

N_{MA}, N_{HE}, N_{NV}: number of MAs, HEs and neovessels (NV), respectively

4. Methods

4.1. CNN Architectures

In order to assess the strengths and limitations of CNNs, many architectures were trained and tested with explicit concentrate on a twenty two layers deep model known as GoogLeNet. This terribly economical network achieves progressive accuracy employing a mixture of low-dimensional embeddings and heterogeneous sized spacial filters. Exaggerated convolution layers and improved utilization of internal network computing resources permit the network to find out deeper options. As an example, the primary layer may learn edges whereas the deepest layer learns to interpret exhausting exudate, a DR classification feature. The network contains convolution blocks with activation on the highest layer that defines complicated purposeful mappings between inputs and response variable, followed by batch normalisation when every convolution layer. Because the variety of feature maps increase, one batch normalisation per block is introduced in succession.

The liquid ecstasy pooling sample-based discretization method was performed with kernel size 3x3 and stride two. The network was then two-dimensional to at least one dimension when the ultimate convolutional block. Dropout of network layers was performed till reaching the dense 5 node output layer, that uses a softmax activation perform to cipher the likelihood of classification labels. Leaky corrected linear measure activation was conjointly applied with gradient worth zero.01 to mitigate dead vegetative cell bottlenecks throughout back-propagation. The network uses convolutional layer L2 regularization to cut back model overfitting, cross-entropy computed error loss, and also the missionary technique of initializing weights in order that vegetative cell activation functions begin enter unsaturated regions.

4.2. Preprocessing
All pictures were born-again to a graded format for preprocessing, information augmentation, and coaching. Preprocessing concerned many steps: pictures were cropped victimisation Otsu's methodology to isolate the circular coloured image of the tissue layer. pictures were normalized by subtracting the minimum constituent intensity from every channel and dividing by the mean constituent intensity to represent pixels within the vary zero to one. distinction adjustment was performed victimisation the distinction restricted adaptational bar chart equalisation (CLAHE) filtering algorithmic program.

4.3. Data Augmentation

We increased the quantity of pictures in time period to boost network localization capability and scale back overfitting. throughout every epoch, a random augmentation of pictures that preserve collinearity and distance ratios was performed. we have a tendency to enforced random artefact with zeros, zoom, rolling and rotation. These affine transformations area unit significantly effective once applied to illness category R1 that area unit the foremost tough to grade and fewest in variety.

4.4. Training and Testing Models

A Deep Learning GPU coaching System (DIGITS) with prebuilt convolutional neural networks for image classification expedited knowledge management, model prototyping and time period performance observance. DIGITS is Associate in Nursing interactive system and was 1st accustomed build a classification dataset by rending the Messidor and MildDR body structure folder into

coaching and validation subsets of 1077 and 269 pictures severally. the pictures were cropped to space size 256x256 and used as computer file by Image internet models antecedently trained for generic classification tasks. The take a look at set folder contained four hundred pictures from the Lariboisiere hospital Messidor partition and was disjoint from coaching knowledge. This coaching system, that offered in depth hyper parameter picks, was then accustomed build model prototypes over a hundred epochs requiring around twenty minutes every to complete.

4.5. Transfer Learning

Transfer learning primarily based approaches were dead with pretrained AlexNet and GoogLeNet architectures from ImageNet. The last totally connected layer was removed, then a transfer learning situation was followed by treating the remaining network elements as a hard and fast feature extractor for the new dataset. The transfer learning retains initial pretrained model weights and extracts image options via a final network layer.

5. Experiments

5.1. Digital image process improves sensitivity for gentle category detection

The dataset contained pictures from a disparate patient population with extraordinarily varied levels of lighting within the body structure photography. The lighting affects constituent intensity values inside the pictures and creates redundant variation unrelated to classification levels. A distinction restricted adjustive bar graph leveling filtering algorithmic program, mistreatment the OpenCV (http://opencv.org/) package was applied to deal with this object. Results from this preprocessing step area unit visually delineated in Fig. 2. we have a tendency to discovered that 3-ary classifier sensitivity for the gentle case exaggerated from zero to twenty nine.4%, whereas this live was around identical for the remaining 2 categories Fig. 3. Our digital image preprocessing technique enabled improved detection of pinpoint refined options and microaneurysms via convolutional filters, that were antecedently indiscernible by the CNN. we have a tendency to hypothesise this modification is due to the channel wise distinction enhancing result of bar graph leveling.
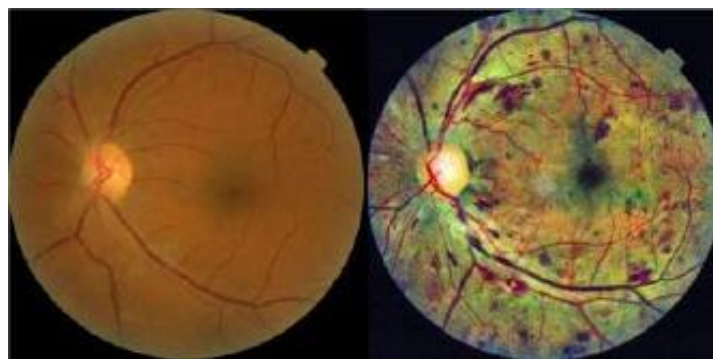

Figure 2.

Contrast Limited Adaptive Histogram Equalization enhances contrast and the detection of subtle features. Shown are fundoscopic illustrations before and after CLAHE application.
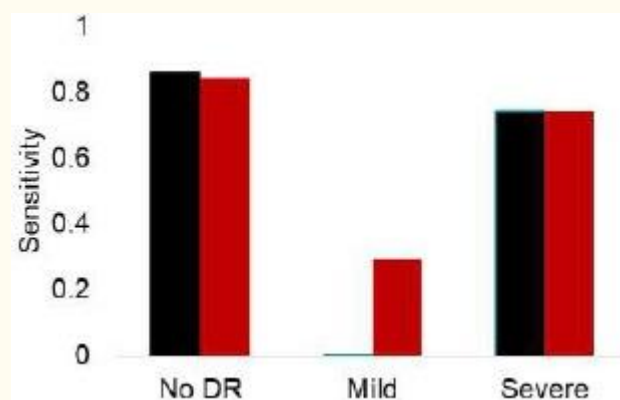

Figure 3.

Sensitivity of a 3-ary (no DR, mild, and severe classes) GoogLeNet classifier before (black) and after (red) CLAHE application on the Messidor dataset.

5.2. Binary model classification attains benchmark performance from literature

AlexNet, VGG16 and GoogLeNet models were trained on the binary-labeled (normal or gentle vs moderate to finish stage) Kaggle dataset to explore strengths and weaknesses of CNNs. This implementation was performed in Tensorflow, and every one model weights were allowed to be updated. The GoogLeNet model achieved the very best sensitivity of ninety fifth and specificity of ninety six mistreatment our real time information augmentation and preprocessing techniques see

 Fig. 4. Thus, we have a tendency to with success demonstrate progressive accuracy levels that have antecedently been revealed during this field.
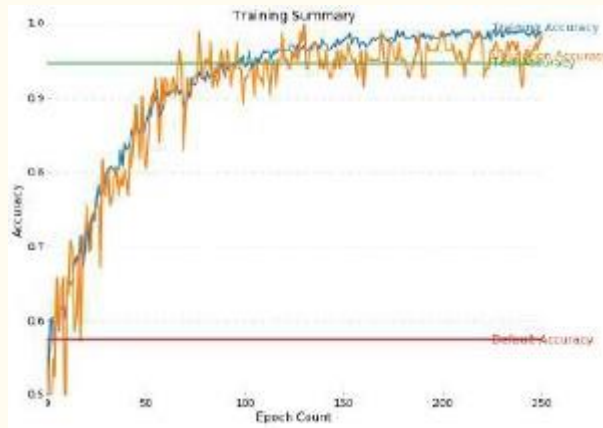

Figure 4.

Training Curve for model on the binary classified Kaggle data set of DR fundoscope images. Sensitivity of 95% and specificity of 96% was achieved.

5.3. Multi-class coaching sensitivities is very keen about dataset fidelity

However, once we trained 3-ary and 4-ary classifiers with a GoogLeNet model on the Kaggle dataset, we tend to were unable to realize vital sensitivity levels for the gentle category. As shown within the confusion matrix (Table 2), the sensitivity of the no DR and severe DR categories were ninety eight and ninety three respectively; but the sensitivity for the gentle category was solely seven-membered. Thus, we discover that our performance is restricted by the lack of CNNs to observe terribly delicate options. we tend to theorise that this will be explained by the sizeable noise and camera artifacts moreover as poor fidelity labeling.

Table 2.

Confusion matrix on test set of Kaggle dataset

-	Pred R0	Pred R1	Pred R2 or R3
True R0	149	1	2
True R1	21	2	7
True R2 or R3	1	15	202

We gained insight into ways in which which could improve our results by manufacturing visualizations that reveal at the image level our CNNs usage of high and low level options to enhance chance of sickness detection. mental image of pathological areas shown in a very heat map generated by window patch-wise occlusion, Fig. five demonstrate an outsized range of false positive and false negative zones, indicating that, whereas massive sickness options and people that have a unique color from the background area unit simply understood by the CNN, tiny and delicate options stay undetected .
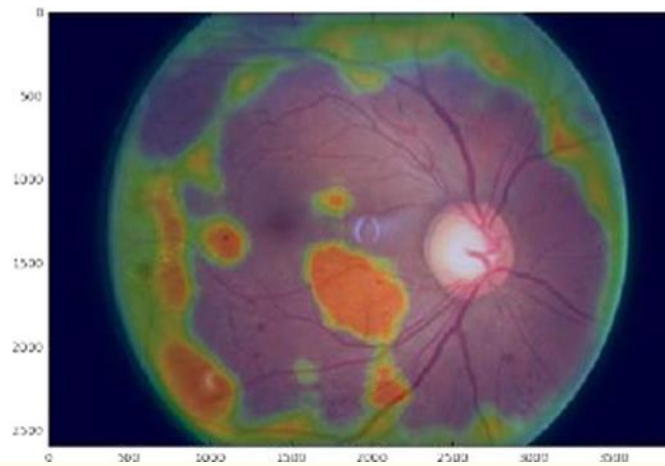
Figure 5.

Heat map on a representative DR image. Green: Regions that do not change the probability of an abnormal binary classification (neutral areas or unfamiliar areas); Orange: Regions that increase the probability of an abnormal binary classification (suspicious areas); Clear or light blue: Regions that decrease the probability of abnormal binary classification (normal areas).

Thus, we tend to take into account our Messidor assortment that addresses fidelity considerations of the Kaggle informationset (see Dataset section for a lot of details on data acquisition procedures).

After following our preprocessing steps, the 3-ary categoryifier achieves sensitivities for no DR and severe DR of eighty fifth and seventy fifth severally also as twenty ninth delicate class sensitivity. Given the dramatic activity increase for the delicate category, even with five-hitter the number of information, it's evident that information fidelity incorporates a robust impact on multi-class coaching model performance.

5.4. Transfer learning as a parallel means that of exploring best CNN models

Our previous models relied on a Tensorflow4 implementation while not fastened weights. The usefulness of transfer learning was investigated by employing a baseline model consisting of pretrained GoogLeNet model obtained from the ImageNet visual seeing information. The model was trained on the Messidor dataset for thirty epochs victimisation random gradient descent improvement with step decay learning rate initialized at zero.002. The classification model validation achieved sixty six.03% because the best accuracy.

Introducing loss (and accuracy layers) to intermediate representations of the deep network allowed for quicker propagation throughout coaching and avoided the vanishing gradients issue. we tend to found that coaching converged a lot of quicker victimisation the preprocessed images—in a matter of twenty five epochs for the transfer learning situation instead of ninety epochs once coaching on the information.

The accuracy measures indicate however well these intermediate origin layer representations have converged to supply the simplest classifier. we tend to ascertained spiking behavior within the coaching loss curves that steered a necessity for weight decay hyperparameter standardisation. to boot, standardisation hyper-parameters like L2 regularization, dropout and batch social control made a larger degree of accuracy layer convergence. Full coaching and take a look at results may be found within the Appendix Tables three and and44.

GoogLeNet Rapid Prototyping Results-Data Augmentation, Contrast Filtering & Regularization						
Model	Solver	Learning Rate	Policy	Drop Out[1]	Validation Accuracy%	Test Set Accuracy%
2-ary	Adam	1e-4	Step Down	(0.8,0.7,0.4)	88.35	74.50
2-ary	SGD	1e-3	Step Down	(0.6,0.6,0.4)	88.07	71.75
2-ary	NAG	1e-3	Step Down	(0.7,0.8,0.4)	87.50	70.00
2-ary	RMSProp	1e-4	Exponential Decay	(0.7,0.7,0.4)	85.80	69.50
2-ary	AdaGrad	1e-3	Exponential Decay	(0.7,0.7,0.5)	87.22	66.25
3-ary	AdaGrad	1e-3	Exponential Decay	(0.7,0.7,0.5)	63.28	68.75
3-ary	SGD	1e-3	Step Down	(0.6,0.6,0.4)	65.63	67.00
3-ary	NAG	1e-3	Step Down	(0.7,0.8,0.4)	67.71	64.75
3-ary	RMSProp	1e-4	Exponential Decay	(0.7,0.7,0.4)	60.68	64.00
3-ary	Adam	1e-4	Step Down	(0.8,0.7,0.4)	64.32	63.50
4-ary	SGD	1e-3	Step Down	(0.6,0.6,0.4)	60.00	51.25
4-ary	Adam	1e-4	Step Down	(0.8,0.7,0.4)	53.75	49.50
4-ary	NAG	1e-3	Step Down	(0.7,0.8,0.4)	55.00	47.75
4-ary	AdaGrad	1e-3	Exponential Decay	(0.7,0.7,0.5)	57.50	47.00
4-ary	RMSProp	1e-4	Exponential Decay	(0.7,0.7,0.4)	54.75	44.25

Table 3.

However, the ultimate check time 3-ary accuracy was similar for each, 67.2% coaching on the information and seventy one.25% for the transfer learning task. Grouping gentle and abnormal along the accuracy was seventy one.5% and 74.5% severally. the most effective check set accuracies will be found in Fig. 6. it's fascinating to notice that the accuracy of the 4-ary classifier decreases once preprocessing. we have a tendency to anticipate this can be because of the loss of necessary image options throughout downsampling. However, we have a tendency to note that there was a gain within the overall detection of refined options, microaneurysms, of Revolutionary Organization 17 November with transfer learning. This technique ought to be any explored to see if we will enhance sensitivity of those models to gentle category DR.

Figure 6.
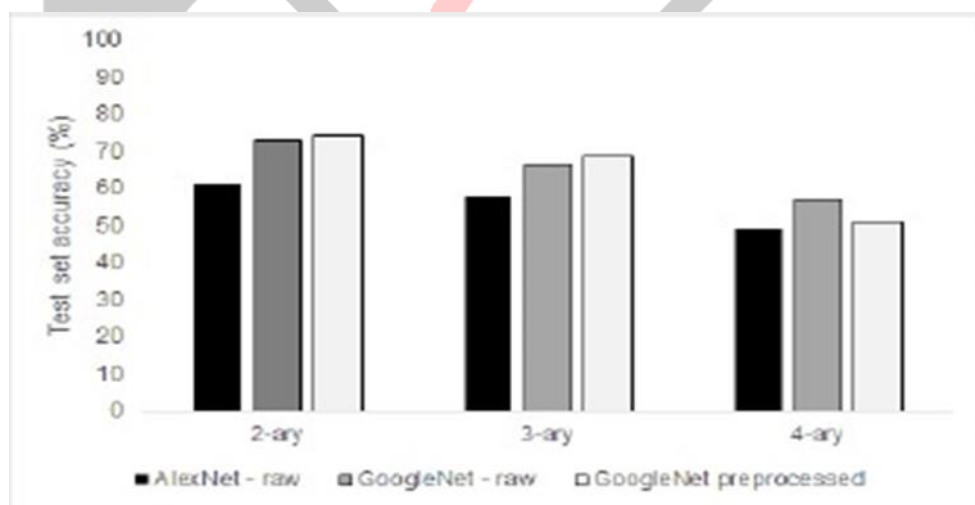
Test set accuracies for 2-ary, 3-ary, and 4-ary classifiers for transfer learning models based on AlexNet and GoogLeNet. Preprocessed images indicates the presence of real-time data augmentation and histogram equalization.

6. Conclusion

Automated detection and screening offers a novel chance to forestall a big proportion of vision loss in our population. In recent years, researchers have value-added CNNs into the set of algorithms wont to screen for diabetic sickness. CNNs promise to leverage the big amounts of pictures that are concentrated for doctor taken screening and learn from raw pixels. The high variance and low bias of those models might permit CNNs to diagnose a wider vary of nondiabetic diseases in addition.

However, whereas we have a tendency to win progressive performance with CNNs victimization binary classifiers, the model performance degrades with increasing range of categories. although it's tempting to surmise that a lot of information could also be higher, previous add the sphere has verified that CNN ability to tolerate scale variations is restricted et al. have prompt that within the case of retinal pictures, a lot of information cannot supplement for this inherent limitatio. Gulshan et al. rumored a 93-96% recall for binary classification of sickness however reports that recall isn't improved once coaching with sixty,000 samples vs one hundred twenty,000 samples of a personal dataset.

Medical pictures square measure fraught with delicate options that may be crucial for designation. as luck would have it, the foremost usually deployed architectures are optimized to acknowledge gross options like those gift within the ImageNet dataset. we have a tendency to could thus need a brand new paradigm for diagnosis diseases via CNN models. this might be a 2 stage lesion detection pipeline that involves feature localization followed by classification and additional preprocessing steps to phase out pathologies troublesome to make out by manual scrutiny, and eventually rebalancing network weights to account for sophistication imbalances seen in medical datasets. Overall, our future goals involve rising detection of gentle sickness and transitioning to tougher and helpful multi-grade sickness detection.

References

[1] Abràmoff M. D., Reinhardt J. M., Russell S. R., Folk J. C., Mahajan V. B., Niemeijer M., Quellec G. Automated early detection of diabetic retinopathy. Ophthalmology, 2010;117(6):1147–1154.

[2] Antal B., Hajdu A. An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. IEEE transactions on biomedical engineering, 2012;59(6):1720–1726.

[3] Chan J. C., Malik V., Jia W., Kadowaki T., Yajnik C. S., Yoon K.-H., Hu F. B. Diabetes in asia: epidemiology, risk factors, and pathophysiology. JAMA, 2009;301(20):2129–2140.

[4] Congdon N. G., Friedman D. S., Lietman T. Important causes of visual impairment in the world today. Jama, 2003;290(15):2057–2060.

[5] Decenciere E., Zhang X., Cazuguel G., Lay B., Cochener B., Trone C., Gain P., Ordonez R., Massin P., Erginay A., et al. Feedback on a publicly distributed image database: the messidor database. 2014;33:231–234.

[6] Gardner G., Keating D., Williamson T., Elliott A. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. British journal of Ophthalmology. 1996;80(11):940–944.

[7] Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. Elsevier. 2017

[8] Goh J. K. H, Cheung C. Y., Sim S. S., Tan P. C., Tan G. S. W, Wong T. Y. Retinal imaging techniques for diabetic retinopathy screening. Journal of diabetes science and technology, 2016;10(2):282–294.

[9] Graham B. Kaggle diabetic retinopathy detection competition report. 2015.

[10] 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, 2016;316(22):2402–2410.