Pattern Recognition Letters 135 (2020) 293–298

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset

Abhishek Samanta [a], Aheli Saha [a], Suresh Chandra Satapathy [a,*], Steven Lawrence Fernandes [b], Yo-Dong Zhang [c]

[a] School of Computer Engineering, Kalinga Institute of Industrial Technology (Deemed to Be University), Bhubaneswar, Odisha 751024, India
[b] Department of Electronics and Communication Engineering, Sahyadri College of Engineering and Management, Mangaluru 575007, India
[c] Department of Informatics, University of Leicester, Leicester LE1 7RH, UK

ARTICLE INFO

Article history:
Received 30 December 2019
Revised 13 April 2020
Accepted 17 April 2020
Available online 12 May 2020

Keywords:
Diabetic Retinopathy
CNN architecture
Colour fundus photography

ABSTRACT

Diabetic Retinopathy is a complication based on patients suffering from type-1 or type-2 diabetes. Early detection is essential as complication can lead to vision problems such as retinal detachment, vitreous hemorrhage and glaucoma. The principal stages of diabetic retinopathy are non-Proliferative diabetic retinopathy and Proliferative diabetic retinopathy. In this paper, we propose a transfer learning based CNN architecture on colour fundus photography that performs relatively well on a much smaller dataset of skewed classes of 3050 training images and 419 validation images in recognizing classes of Diabetic Retinopathy from hard exudates, blood vessels and texture. This model is extremely robust and lightweight, garnering a potential to work considerably well in small real time applications with limited computing power to speed up the screening process. The dataset was trained on Google Colab. We trained our model on 4 classes - I)No DR ii)Mild DR iii)Moderate DR iv)Proliferative DR, and achieved a Cohens Kappa score of 0.8836 on the validation set along with 0.9809 on the training set.

1. Introduction

Deep learning has enhanced the purpose of computer vision in identifying and classifying images and are a key tool used to automate tasks in our daily lives. Convolutional networks have been consistently developed for object detection, classification, segmentation. The use of convolutional neural networks (CNNs) on medical images has helped the medical sector immensely due to it's ability to learn representations of data [1].

Diabetic Retinopathy turns out to be a major cause of blindness in the western world, and regular screening of the patients reduces the risk of blindness. There are a number of features pertaining to the recognition of retinopathy in fundus photography, and computer vision based trained classifiers work pretty well in classification. Promising work has been displayed in the detection of retinopathy using k-NN classifiers and vector machines. CNNs have also been used for the classification of Diabetic Retinopathy [13,15], given a big dataset and considerable computing power [18]. They have been instrumental in detecting the features such as haemorrhage and hard exudes that identify retinopathy [8]. Deep architectures of CNN have been instrumental in providing the finesse and high performance to trained models by learning patterns from raw images [2]. Due to the availability of annotated data and evolution of GPUs, CNNs have been increasingly applicable in a number of cases. However, in case of medical datasets, huge amounts of annotated data are not readily available yet as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Thus, transfer learning has not been very useful for medical datasets as most networks have been trained well to recognize objects present in the ImageNet dataset as shown in Fig. Fig. 1 and Figs. 5–13.

A major problem faced in training the model on less data is underfitting. Moreover, the presence of skewed classes causes the model to overfit on the largest class, which is turn, decreases the the corresponding F1 scores and Cohen's Kappa. Large datasets can often be over-sampled on the lower class, but oversampling on a small dataset will not be of much help against overfitting.

In this paper, we propose a deep learning based CNN method to classify images from a small and skewed dataset of 3050 training images belonging to 4 classes and 419 validation images to achieve a considerably good result. The accuracy metric used by us is Cohen's Kappa.

* Corresponding author.
E-mail addresses: sureshsatapathy@ieee.org, suresh.satapathy@kiit.ac.in (S.C. Satapathy).

https://doi.org/10.1016/j.patrec.2020.04.026
0167-8655/

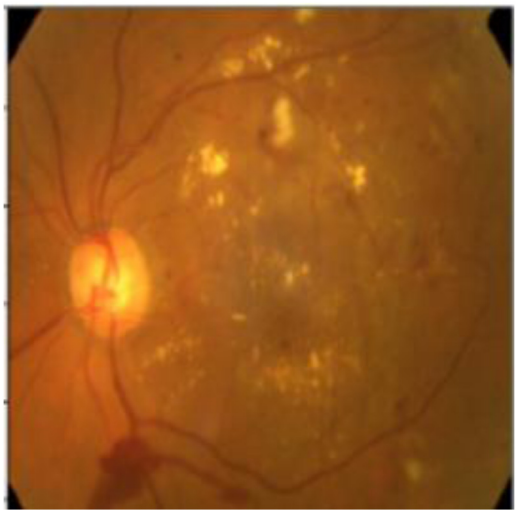

Fig. 1. An fundus image of a retina with Proliferative DR.

2. Related work

There has been considerable work regarding binary classification of Diabetic Retinopathy. Gardner et al used neural networks on 200 images by splitting an image into patches to achieve a sensitivity of 88.4% and specificity of 83.5% for binary classification. His work had been aided by a clinician in classifying the patches prior to using SVM.

Dr.Nayak et al also used neural networks to classify Diabetic Retinopathy Based on 3 classes by recognizing blood vessels and hard exudes from 140 images [3]. His model worked on classifying between normal, non-Proliferative and Proliferative retinopathy. His results were validated by clinical practitioners and showcased an accuracy of 93%, specificity of 100% and sensitivity of 90%.

Most research in the automated detection of Diabetic Retinopathy exceeding 3 classes have been done using SVMs. Acharya et al had worked on the detection of retinopathy using 5 classes [4]. Features extracted from images were used as input into the SVM classifier to capture the contours and variation of shapes. It resulted in an accuracy and specificity of 82% and 88% respectively. The areas of features such as hard exudates, hemorrhages, blood vessels and micro-aneurysms were calculated from the images and further used as input to the SVM classifier. Sensitivity and specificity were observed to be 82% and 86% respectively with an accuracy of 85.9% using this method.

Adarsh et al used image processing for automated Detection of Diabetic Retinopathy by detecting the features associated with retinopathy [5]. Texture features and the area of lesions were used for the construction of the feature vector for the SVM. Accuracies of 94.6% and 96% were obtained on the public image databases of DIARETDB1 and DIARETDB0 respectively. DIARETDB1 had 130 images and DIARETDB0 had 89.

Harry Pratt et al had proposed an approach using CNN to diagnose Diabetic Retinopathy from digital fundus images. 80,000 training images were used along with 5000 validation images belonging to 5 classes to achieve an accuracy and sensitivity of 75% and 95% respectively [6]. The training was hardware intensive with the requirement of NVIDIA K40c. Normalization was used in preprocessing the images prior to feeding them to their customized network of stacked convolution layers followed by fully connected layers.

Xiaogang Li et al had proposed a CNN based transfer learning approach on AlexNet, VGG16, VGG19. Experiments were performed on 1200 and 1014 fundus images from the MESSIDOR and DR1 datasets. Pre-trained models on the ImageNet were fine tuned on the DR datasets. Feature extraction was also experimented upon in this paper for transfer learning.

Carson Lam et al worked in classifying 4-ary data on a large dataset of 35,000 training images [9]. Their pre-processing involved systematic cropping using the Otsu's method followed by normalization. Contrast limited adaptive histogram equalization was further used for contrast adjustment. The model was trained on the 22 layered GoogleNet by transfer learning after removing the last dense layer. A Tesla K80 GPU hardware was used to aid training. Peak test accuracies of 74.5%, 68.8% and 57.2% were obtained on 2-ary, 3-ary and 4-ary classes respectively.

Maithra Raghu et al had explored transfer learning for the purpose of medical imaging [23]. Their dataset consisted of fundus photographs to diagnose a number of eye diseases including Diabetic Retinopathy graded into 5 classes. They paved an insight into the effects of transfer learning from an unrelated dataset to medical data. Evidence of feature reuse at the lowest layers and overparametrization of models leading to the deviations from transfer learning had been observed by them.

3. Method and structure

We decided upon our network after studying baseline literature [7,22] and testing the performance of other models [3–5,9,23]. It was observed that deeper layers cause overfitting as our dataset was comparatively smaller. In our network, we used CNN-based transfer learning on the DenseNet model pre-trained on ImageNet.

3.1. Dataset and hardware

The training and testing fundus images were obtained from Kaggle (https://www.kaggle.com). 3050 training images were used belonging to 4 classes. The number of images were 1805, 370, 999 and 295 for 1-ary (A), 2-ary (B), 3-ary (C) and 4-ary (D) classes respectively (Fig. 2).

We trained our model on the publicly available Google Colaboratory which is a free Jupyter notebook environment that runs on the cloud [10]. Keras was used as the deep learning package with Tensorflow at the backend.

3.2. Pre-processing

Different pre-processing methods were experimented with, to determine the one which outperforms the others on the task, when the images are fed through our model. In medical image analysis, it is essential to enhance contrast, meanwhile preserving the brightness, for effective classification. Further, different images have different lighting conditions, which needs to be addressed. Contrast enhancement was used to adjust the bright or dark pixels of an image in order to extract the hidden features. The contrast between the retinal background and the blood vessels observed in the fundus images are very low. Thereby, analysis and study of the tiny retinal vasculature and othersuch abnormalities are difficult. Due to this problem, enhancing the retinal area in the fundus photography images are important in order to provide better visualization of hard exudates, blood vessels and in turn the accuracy of detecting the abnormalities increases.

Fundus photography images are reddish in color consisting of a black background. Necessary details on fundus are not present in the background and can be removed to decrease noise. Equalization of the fundus images with a black background results the darkness to increase within the details of the image [20]. Considering this issue, we decided in eliminating the black background with pre-processing. The pixel values in the black background is 0, with the realistic brighter regions having other non 0 values. The

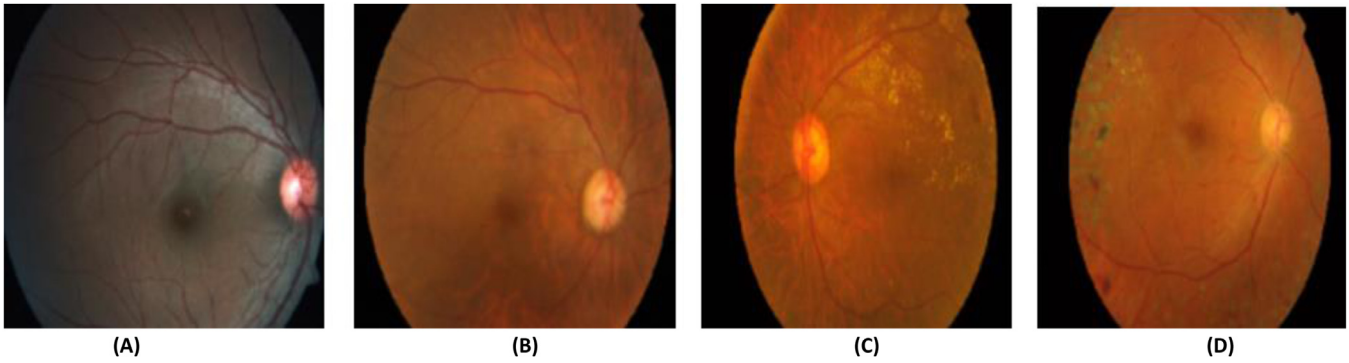

Fig. 2. Data used of 4 classes.

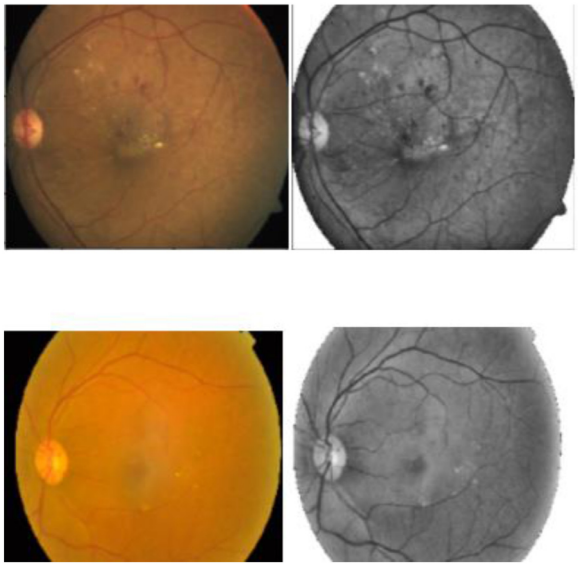

Fig. 3. Depiction of our attempted pre-processing (left-original, right-pre-processed).

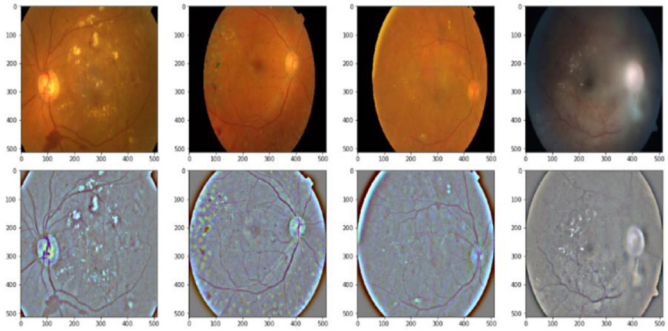

Fig. 4. Depiction of the finalized pre-processing results.

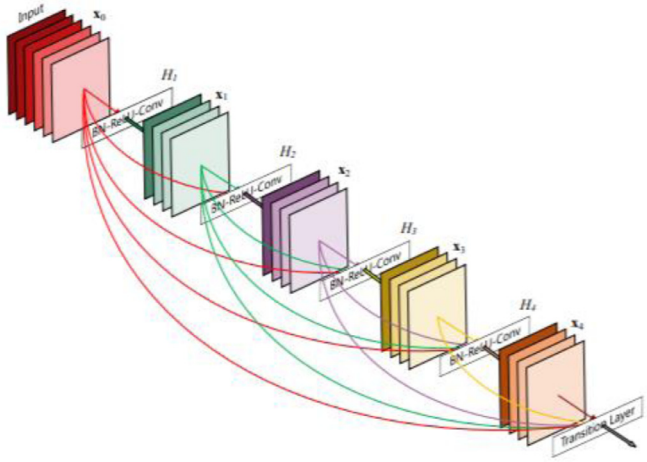

Fig. 5. An illustration of a 5-layer Dense block from Gao Huang et al [11].

black background was converted to 1 and other non black regions were converted to 0 by fixing pixel threshold. After thresholding was performed, the pixel regions containing 0 are substituted with gray scale of the input image, while the grayscale pixels are in turn substituted with the input fundus image. The green plane was extracted thereafter. Contrast Limited Adaptive Histogram Equalization (CLAHE) was then applied to enhance the small regions of relevance in the image. Clipping limit was set to 2.0, and tile grid size was taken as 8. This method, inspired from [21], though produced visually pleasing images and succeeding in bringing out certain subtleties, did not perform exceptionally well on the task (Fig 3).

To further enhance performance, we experimented with a pre-processing method inspired from Ben Graham, used in the Kaggle competition [14], and applied weighted Gaussian blur to the images. It is a 2D convolution operation which reduces noise. The values of sigmaX and sigmaY were taken to 10, as it produced comparatively better results and the kernel size was computed from the sigma values itself. This pre-processing method outperformed our previous experiments, and therefore was used in our final model (Fig 4).

In 2 dimensions: $G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2=y^2}{2\sigma^2}}$

3.3. Model

Stacking of convolution layers were a necessity for the classification of images. The first layers work on classifying the major distinctions and curves while the last layers are responsible for identifying the distinctive features of classification such as hard exudates and haemorrhage. Custom networks did not prove to be working well in case of a small dataset. We fine tuned and tested our model using several architectures such as Inceptionv1, Inceptionv2, Inceptionv3, Xception, VGG16, ResNet-50, DenseNet and AlexNet [16]. DenseNet121 was chosen as the final network as it was the most effective in classification. This network is the simplest among the other DenseNet architectures.

DenseNet works well in evading the vanishing gradient problem and enabling feature reuse and has achieved state of the art (SOTA) results on the ImageNet, CIFAR and SVHN datasets [11,17]. It is composed of dense blocks where every layer receives concatenated data from all the previous layers, thereby simplifying the residual network pattern. Moreover, it requires fewer parameters than the other convolutional networks. In contrast to ResNets, which also

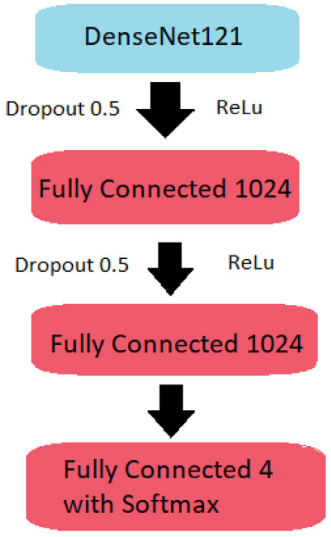

Fig. 6. Model Architecture.

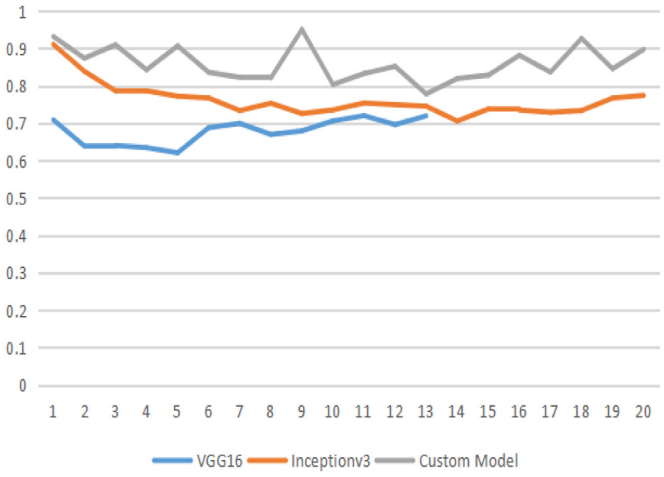

Fig. 7. Monitoring progress between VGG16, Inceptionv3 and a custom built model. X axis denotes epoch and Y axis denotes validation loss.

use skip connection, DenseNets concatenate the output and the input feature maps. Due to input concatenation, ease of access to learned maps for the corresponding layers are increased. Fine tuning on the pre-trained DenseNet helped in preserving the salient image features, along with learning the DR classification features.

The fully connected layers were chopped off and the output was flattened to one dimension. 2 fully connected layers of size 1024 were added. ReLu activation (y=max(0,x)) was used in the fully connected layers. To prevent overfitting, a dropout of 0.5 was used. The last layer used Softmax activation function for classification between 4 classes of Diabetic Retinopathy.

$$\text{Softmax Activation}: F(x_i) = \frac{e^{x_i}}{\sum_{j=0}^{k} e^{x_j}}$$

(i=0,1,2,..,k)

Gradient Descent was used as the optimizer. The loss function was computed using the popularly used categorical cross-entropy. The learning rate was varied with training, with the usage of Nesterov momentum.

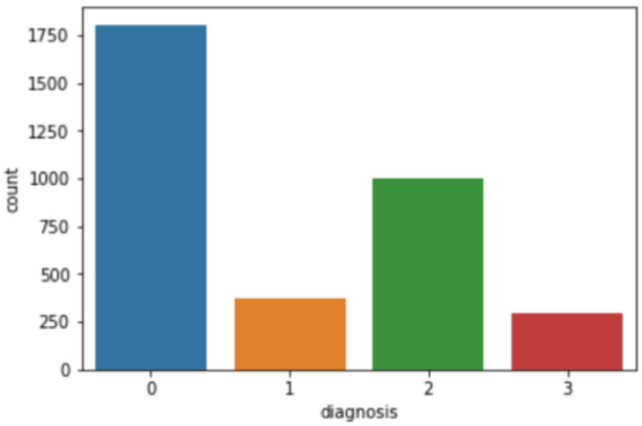

Fig. 8. Data distribution across the 4 classes depicting skewness.

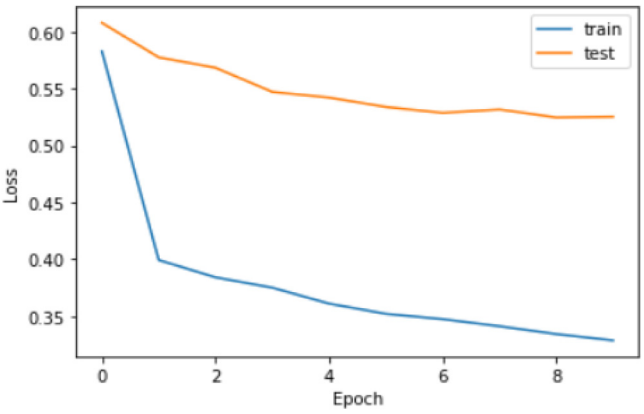

Fig. 9. Epoch till 10.

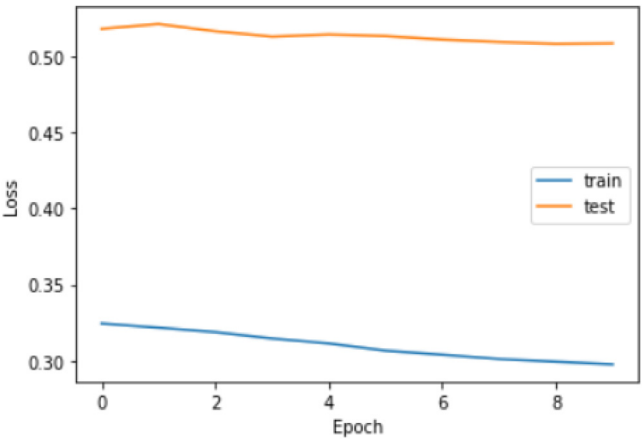

Fig. 10. Epoch till 20.

3.4. Training

Custom models deep enough were prone to overfitting, which performed well on the training set but suffered on the validation set, while shallow networks cause an underfit. Training time for the Inception networks were considerably high which restricted the number of training epoch. The usage of Xception took less time than the other counterparts. Experimentation among various networks by us has been given below.

As per the observation above, most custom models did not learn enough to find the right trade-off between high variance and bias. Training the model on VGG16 or Inceptionv3 did not prove to

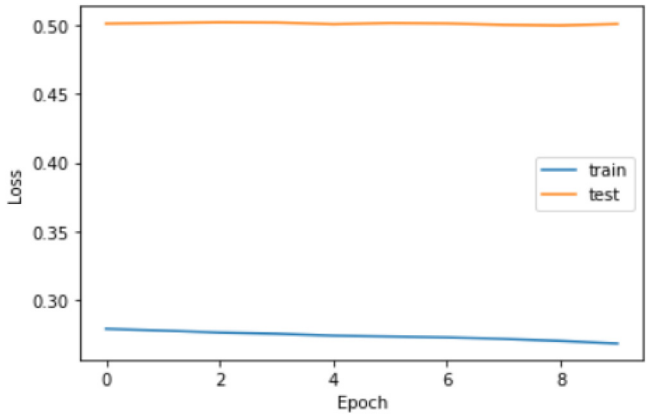

Fig. 11. Epoch till 30.

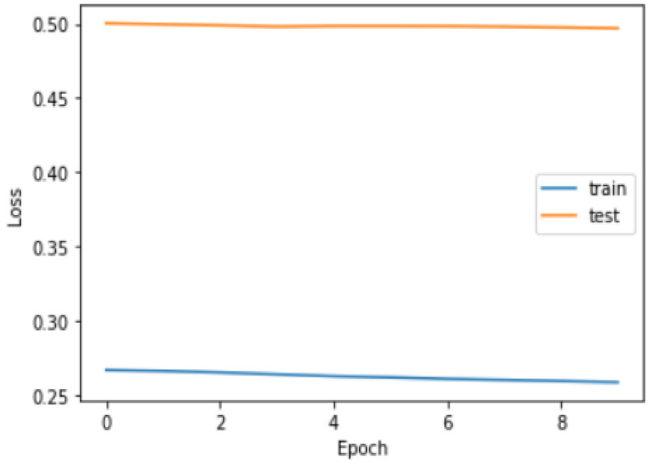

Fig. 12. Epoch till 40.

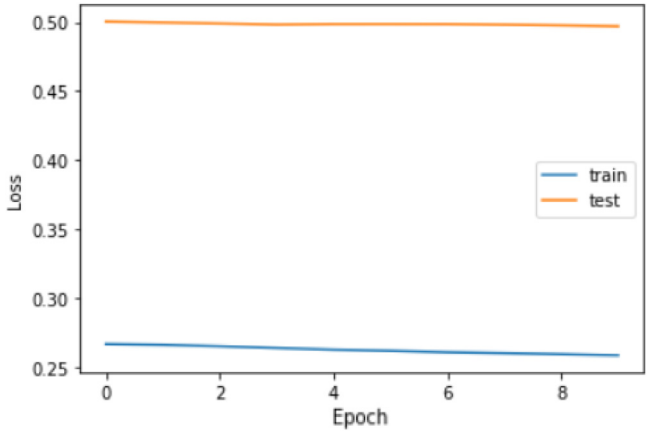

Fig. 13. Epoch till 50.

be of much use either as either of the models were unable to learn useful patterns from the limited amount of data. In Inceptionv3, a slight learning curve was observed, which plateaued for the validation loss while the training loss kept decreasing, indicating a clear situation of high variance. Several custom networks were also experimented by us, inspired from literature on other biomedical images [22]. However, neither were able to learn useful patterns from fewer data representations.

Our final model resorted to using DenseNet121 pre-trained on ImageNet as the base layer. The fully connected layers of the DenseNet121 were removed and 2 fully connected layers of 1024 were introduced along with a dropout of 0.5. The model was trained for a total of 50 epoch on the pre-processed RGB images, each of dimension size 360 × 360. Due to lack of continued training time, the model was saved at regular intervals of 10 epoch and trained. The last 2 layers of the DenseNet121 and the 2 fully connected layers were trained to fine-tune the model in learning the salient features owning to classification between the 4 classes. The implementation of transfer learning and freezing the other layers of the DenseNet helped in training by reducing the training time for the entire network along with preserving the image recognition features learned by the network from ImageNet.

Data skewness was a serious problem, and the small number of training images added to the problem. Class weights were used in ratio of the images present to counteract the problem. Their introduction was observed to boost the corresponding F1 scores of the classes and Cohen's Kappa. Oversampling did not prove to be beneficial in boosting the accuracy of the classes.

The model was trained using gradient descent. Adam and RMSprop were experimented upon but they did not turn out to be instrumental in better classification between the classes or decrease the validation loss. A learning rate of 0.003 was used for the first 30 epoch with a decay of 0.02. Nesterov momentum was used with a momentum of 0.01 in order to speed up training time by the accumulation of past gradients and move over local minima [12]. The usage of Nesterov momentum helped in going back the in case of a gradient offshoot due to the momentum. After 30 epoch, the learning rate was set to 0.001 and the decay was kept constant at 0.02. Training beyond 50 epoch did not cause the model to learn enough and in turn, resulted in overfitting on the training data. EarlyStopping was used for this purpose in order to monitor the validation loss and accuracy.

The plots above represent the losses over the training and test sets for 10 epoch each, the weights being saved and trained over regular intervals.

4. Results

The model was validated on 419 fundus images. The validation process was fast. We obtained a validation accuracy of 84.10 %. Accuracy was not used by us as the final metric due to the skewness of data. We chose Cohen's Kappa as our final metric than F1 score as Kappa provides a relativistic accuracy with respect to the other nearby classes, thus imparting a sense a reliability and originality in identification in case of a medical diagnosis [19].

$$\text{Cohen's Kappa } (\kappa) = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}$$

A literature study on past work highlighted low performance on the middle classes of fundus images. Our model provides F1 scores of 0.64 and 0.74 on 1-ary (Mild DR) and 2-ary (Moderate DR) classes respectively. The Kappa score obtained by us is 0.8836 on the validation data and 0.9809 on the training data.

5. Discussion and conclusion

Our model has approached the classification of 4-class problem in Diabetic Retinopathy on a small dataset using deep learning. Most earlier algorithms dedicated to the classification of DR fundus images on a small dataset evaded the use of deep learning. Our method has produced comparable results with previous literature given data and hardware constraints and the presence of skewed classes. Transfer learning and fine-tuning on the pre-trained DenseNet has proved to be extremely effective on this dataset and the training technique experimented by us paid off well in terms of achieving considerably good classification results.

Our model has no issues in detecting a healthy eye from a fundus photography. The F1 score of our model on a healthy eye is 0.97. The trained network can be used on an user interface and made available at public hospitals for initial screening, with each retinal photograph taking around 0.99 seconds to be graded with minimal hardware requirements, making it robust.

In future, we will be working on enhancing this work by grading the classes based on a semantic segmentation output highlighting the retinal arteries and vessels, which might provide detailed insights to the features that truly enable algorithms to recognize patterns in a fundus photograph responsible for retinopathy. Promising results are expected from the work.

Declaration of Competing Interest

None.

References

[1] Yann Lecun, Yoshua Bengio, Geoffrey Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[2] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. pages 1–9, 2014.
[3] J. Nayak, P.S. Bhat, R. Acharya, C. Lim, M. Kagathi, Automated identification of diabetic retinopathy stages using digital fundus images, J. Med. Syst. 32 (2) (2008) 107–115.
[4] R. Acharya, C.K. Chua, E. Ng, W. Yu, C. Chee, Application of higher order spectra for the identification of diabetes retinopathy stages, J. Med. Syst. 32 (6) (2008) 481–488.
[5] P. Adarsh, D. Jeyakumari, Multiclass svm-based automated diagnosis of diabetic retinopathy, in: Communications and Signal Processing (ICCSP), 2013 International Conference on, IEEE, 2013, pp. 206–210.
[6] Harry Pratt, Frans Coenen, Deborah M. Broadbent, Simon P. Harding, Yalin Zheng, Convolutional neural networks for diabetic retinopathy, 2016.
[7] Xiaogang Li, Tiantian Pang, Biao Xiong, Weixiang Liu, Ping Liang, Tianfu Wang, Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification, 2017.
[8] O. Faust, U.R. Acharya, E.Y. Ng, K.H. Ng, J.S. Suri, Algorithms for the automated detection of diabetic retinopathy using digital fundus images: a review, J. Med. Syst. 36 (1) (2012) 145–157.
[9] Carson Lam, Darvin Yi, Margaret Guo, Tony Lindsay, Automated detection of diabetic retinopathy using deep learning, 2018.
[10] colab.research.google.com/notebooks/welcome.ipynb#.
[11] Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger, Densely connected convolutional networks, 2016.
[12] Ilya Sutskever, James Martens, George Dahl, Geoffrey Hinton, On the importance of initialization and momentum in deep learning, 2013.
[13] R. Gargeya and T. Leng. Automated identifification of diabetic retinopathy using deep learning. Elsevier, 2017.
[14] B. Graham. Kaggle diabetic retinopathy detection competition report. 2015.
[15] V. Gulshan, L. Peng, M. Coram, M.C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, JAMA 316 (22) (2016) 2402–2410.
[16] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
[17] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet classification with deep convolutional neural networks, 2012.
[18] M.D. Abramoff, J.M. Reinhardt, S.R. Russell, J.C. Folk, V.B. Mahajan, M. Niemeijer, G. Quellec, Automated early detection 'of diabetic retinopathy, Ophthalmology 117 (6) (2010) 1147–1154.
[19] Anthony J. Viera, Joanne M. Garrett, Understanding interobserver agreement: the kappa statistic, 2005.
[20] K.G. Suma, V. Saravana Kumar, Chapter 5 A Quantitative Analysis of Histogram Equalization-Based Methods on Fundus Images for Diabetic Retinopathy Detection, Springer Science and Business Media LLC, 2019.
[21] K.G. Suma and V. Saravana Kumar,A quantitative analysis of histogram equalization-based methods on fundus images for diabetic retinopathy detection, 2019.
[22] Jongwon Chang; Jisang Yu; Taehwa Han; Hyuk-jae Chang; Eunjeong Park, A method for classifying medical images using transfer learning: a pilot study on histopathology of breast cancer, 2017.
[23] Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, Samy Bengio, Transfusion: understanding transfer learning for medical imaging, 2019.