Diabetic Macular Edema Grading based on Deep Neural Networks

Baidaa Al-Bander[1], Waleed Al-Nuaimy[1], Majid A. Al-Taee[1], Bryan M. Williams[2], Yalin Zheng[2]

[1]Department of Electrical and Electronic Engineering, University of Liverpool, UK
[2]Department of Eye and Vision Science, University of Liverpool, UK

{hsbalban, wax, altaeem, bryan, yalin.zheng}@liverpool.ac.uk

Abstract. Diabetic Macular Edema (DME) is a major cause of vision loss in diabetes. Its early detection and treatment is therefore a vital task in management of diabetic retinopathy. In this paper, we propose a new feature-learning approach for grading the severity of DME using color retinal fundus images. An automated DME diagnosis system based on the proposed feature-learning approach is developed to help early diagnosis of the disease and thus averts (or delays) its progression. It utilizes the convolutional neural networks (CNNs) to identify and extract features of DME automatically without any kind of user intervention. The developed prototype was trained and assessed by using an existing MESSIDOR dataset of 1200 images. The obtained preliminary results showed accuracy of (88.8 %), sensitivity (74.7%) and specificity (96.5 %). These results compare favorably to state-of-the-art findings with the added benefit of an automatic feature-learning approach rather than a time-consuming handcrafted approach.

1 Introduction

Diabetic Macular Edema (DME) or Diabetic Maculopathy (DM) is a condition characterized by the appearance of exudates on close to the macula. Consequently, the central vision of the patient is affected. DME usually develops at any time during the progression of Diabetic Retinopathy (DR). DR is associated with high blood glucose levels that cause damage to the vessels supplying blood to the retina. According to the Early Treatment Diabetic Retinopathy Study (ETDRS), DME is characterized by the thickening of the macula, hard exudate (HE) and blot Hemorrhage (HA) [1]. Clinically, the severity of DME is mainly divided into two classes: (i) non-clinically significant macular edema (non-CSME) and (ii) clinically significant macular edema (CSME). Non-CSME is a mild class of maculopathy in which the distance between the lesions and the center of the macula is greater than one optic disc diameter. The retinal thickening and hard exudate are considered the main clinical features for Non-CSME. The CSME represents the severe form of maculopathy in which lesions (blot hemorrhage and exudate) occur within a distance of less than one optic disc diameter from the center of macula [2], [3].

It is believed that the early detection and treatment of DME may improve the visual acuity. Different imaging techniques have been used for the diagnosis of DME

X. Chen, M. K. Garvin, J. Liu, E. Trucco, Y. Xu (Eds.): OMIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, Iowa Research Online, pp. 121–128, 2016. Available from: http://ir.uiowa.edu/omia/2016_Proceedings/2016/

such as retinal thickness analyzer (RTA), color fundus photographs, fluorescein angiography (FA) and optical coherence tomography (OCT). Recently, many automated and computerized systems for DME grading have been introduced along with associated image processing techniques for exudate, fovea detection and segmentation using retinal fundus images [4] – [10]. In order to detect and grade the severity of DME, existing methods in the literature have relied on either detection of the location and segmentation of exudate and the macula [4] – [7] or the extraction of texture and image based features [8] – [10].

In [6], Tariq *et al.* proposed a method based on extracting morphological features and the location of exudate after segmenting the exudate using a Gabor filter bank and mathematical morphology. Finally, the distance between the exudate and the center of the macula was calculated in order to grade the severity of DME in each image. Furthermore, Zaidi *et al.* [7] developed a grading method using Gabor filtering, mathematical morphology and Otsu thresholding with a Bayesian classifier to detect the location of exudate and positional constraints to grade the severity of DME. Moreover, Giancardo *et al.* [5] proposed an automated grading system based on exudate probability map and wavelet decomposition. The Kirsch edge operator and a region-growing algorithm were used to locate the hard exudate and the fovea region. In [8], the authors developed a method based on motion pattern analysis and the Radon transform to extract the features necessary to detect the presence of DME in the images. Baby *et al.* [9] used Gaussian data description (GDD) to extract features from the wavelet sub-bands that are obtained by a dual tree complex wavelet transform (DT-CWT). Another method based on higher order spectra features was also proposed by Mookiah *et al.* [10] to grade the severity of DME.

Performance of the aforementioned grading systems, however, relies on the exudate segmentation, anatomical structure localization and feature extraction strategies. The detection of anatomical structures (i.e. fovea and macula) and exudate segmentation in these methods are challenging and time-consuming. Furthermore, the features extraction is highly dependent on the dataset under study to evaluate the proposed methodology. Moreover, finding and engineering a feature set that is appropriate for different datasets is still a challenge since features that are representative of or descriptive for one dataset are often not representative of or descriptive for other datasets.

Automated feature learning algorithms depending on deep learning have recently emerged as a feasible approach and have proven to be effective in some computer vision applications such as ImageNet classification [11] and face recognition [12]. More recently, a primary step towards automatic deep learning method for location detection of important retinal landmarks, the fovea and optic disc (OD) in digital fundus retinal images was reported by Baidaa *et al.* [13]. However, its effectiveness in DME grading is not yet thoroughly explored in the literature. In this paper, we introduce a deep learning-based convolutional neural network (CNN) approach to address the problem of relying on handcrafted features as well as the time consumed in the segmentation of retinal landmarks such as the fovea and optic disc and lesions.

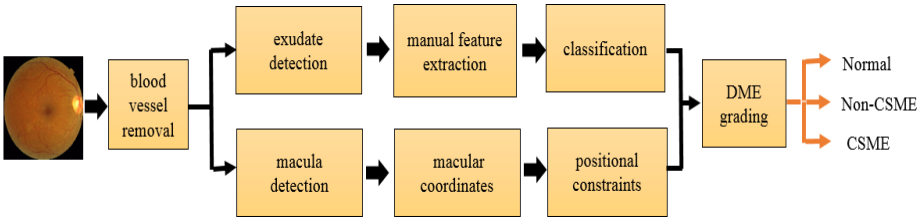
The remainder of this paper is organized as follows. Section 2 provides an overview of the proposed system along with a comparison with traditional systems, Section 3 presents the methodology and describes the deep learning algorithm used in

this work, Section 4 presents and discusses the obtained results. Finally; the work is concluded in Section 5.

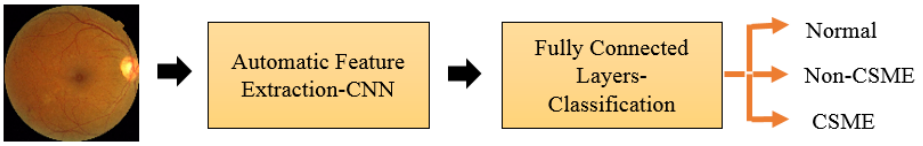
2 System Overview

In traditional automated grading systems [5] – [10], in order to grade DME, the contrast of images is enhanced as a pre-processing stage and then the blood vessels are removed using matched filtering or mathematical morphology. Further to this, the location of the macula is detected and exudate segmentation is applied. After that, different texture, morphological, and image-based features are extracted from the segmented exudates. In the last stage, the DME grading is calculated depending either on the distance of exudate from the macula or using machine-learning algorithms. Figure 1(a) shows a simplified diagram for traditional DME grading systems [3].

As the performance of the automated grading system is highly dependent on the extracted features, the performance of the existing methods may vary from one dataset to another because the extracted features are not always representative of different datasets. Also, these features are handcrafted which incurs a time cost and considerable effort. So, the need to adopt a generalized and automatic feature extraction method is the best solution to counter these issues. Furthermore, traditional approaches have relied on the position of exudates around and near the macula to grade the severity of DME. The extraction of exudates mainly depends on the efficiency of the segmentation algorithm while detection of the location of the macula and fovea highly depends on the accuracy of data mining and texture imaging techniques. Therefore, implementing a method of grading DME that is independent of the segmentation algorithms is a crucial task.



(a) Traditional DME grading systems



(b) The proposed DME grading system

Fig. 1. Comparison between traditional and proposed DME-grading systems - the proposed system does not depend on any kind of segmentation or handcrafted features and sums up many stages of traditional system in only two automatic stages.

Figure 1(b) shows a diagram of the proposed DME grading system. In the feature extraction stage, the features are automatically learnt by convolutional neural network (CNN) and fed into a classifier for classification. The CNN consists of convolutional layers, max-pooling layers, dropout layers, fully connected layers and activation functions. The convolutional layer is made up of a sequence of filters which are used to perform a two dimensional (2D) convolution with the input image. The output of this layer is called the feature map [14]. The max-pooling layer is a subsampling layer where the feature map is down-sampled [15]. The inclusion of a dropout layer is a regularization technique that is essential for reducing overfitting [16]. The fully connected layer is the final layer in the CNN where each neuron is completely connected to the other neurons. The proposed system is therefore represents a promising solution to address the above-mentioned concerns of traditional systems since it does not depend on any kind of segmentation or handcrafted features.

3 Materials and Methods

The proposed feature learning and grading approach comprises three main stages: (i) pre-processing, (ii) network design and training, and (iii) an evaluation stage. These stages are explained as follows. An existing dataset (called MESSIDOR) [17] is used in this study for training and evaluation purposes because it is a fairly large dataset and it is labelled. It comprises 1200 images classified as either normal (no DME), Non-CSME or CSME. These images were captured by using a color 3CCD camera on a Topcon TRC NW6 with 45-degree field of view (FOV) with resolutions of either (1440×960), (2240×1488) or (2304×1536) pixels.

In order to implement the system, a NVIDIA GTX TITAN X 12GB GPU card with 3072 CUDA parallel-processing cores is used. This GPU has more than the estimated 4GB required to load all training images and process them at one time. This GPU supports the CUDA Deep Neural Network library (cuDNN) for GPU learning. The Lasagne Python deep learning [18] and Theano [19] libraries were used to implement and train the networks.

3.1 Pre-processing

In this stage, the smallest rectangular region containing the entire FOV is automatically determined and used to crop each image. After that, the cropped images are resized into three different sizes 128×128, 256×256 and 512×512 pixels to obtain acceleration while keeping the images sufficiently large to identify features such as exudates. Finally, the red, green and blue (RGB) channels in the image are scaled to have zero mean and unit variance.

3.2 Network Design and Training

The structure of the proposed CNN is shown in Figure 2. Three CNN architectures are implemented and trained with one of three different image sizes 128×128, 256×256 and 512×512 pixels. The main objective of training three different networks

is to obtain fast computation time. We use the weights of the networks trained on the smaller images to initialise the networks trained on the larger images. This helps to speed up the process of training without resorting to resizing images below a level where key features may not be detectable for the final classification. The structure of the first network includes the layers in the first block along with the fully connected layers (shown in Figure 2). This network is trained from scratch using images with size of 128x128 pixel images. The architecture of the second network comprises the first and second blocks along with the fully connected layers. The weights are initialized from the first network and trained by using images with 256x256 pixels. Finally, the third network comprises the first, second and third block along with the fully connected layers and are trained on images with 512x512 pixels with the weights being initialized from the second network. The final network architecture comprises 13 convolutional layers with filter of sizes 5×5 and 3×3. Each convolutional layer is followed by a Leaky (0.01) rectified linear unit (ReLU) step. Four max-pooling layers with window size of 3×3, 2 fully connected layers with 1024 neurons each and two dropout layers between the fully connected layers are used.

The dataset is randomly divided into 70% for training and validation (10% of this data is used for validation), and the remaining 30% for testing. To increase the size of data artificially in order to decrease possible overfitting problem, the data is augmented. In every epoch (single pass of all training data through the network) during training, each image is randomly augmented with: random rotation between 0-360 degrees, random horizontal and vertical flipping, random translations of between -40 and 40 pixels, random zooming and random shearing. Furthermore, to counter the impact of unevenly distributed data, oversampling is applied on the imbalanced training set in order to get more uniform distribution of classes and increase detection performance on the rare classes.

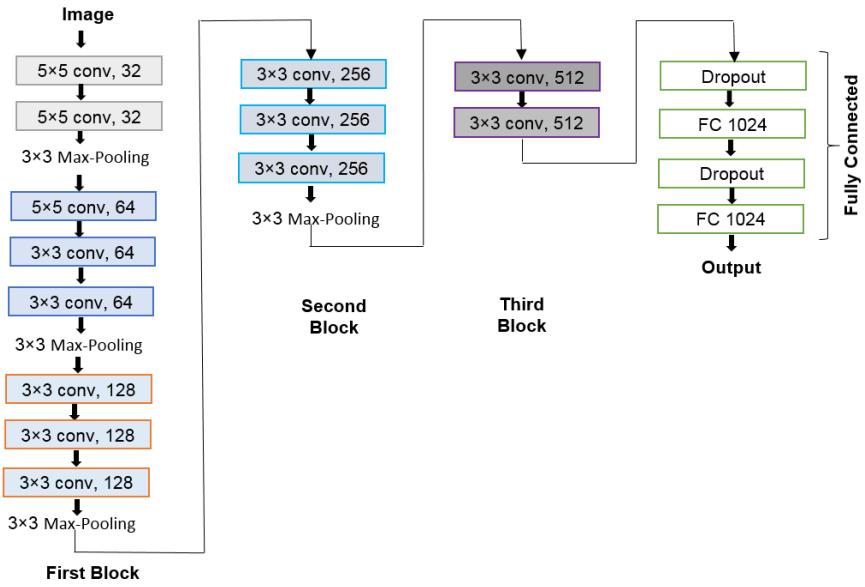


Fig. 2. Architecture of the proposed convolutional neural network

To train the networks, stochastic gradient descent SGD with the Nesterov momentum optimization algorithm is used with adaptive learning rate (start=0.003, stop=0.00003) and momentum parameter 0.9. The orthogonal weight initialization method proposed in [20] is considered to initialize the weights of filters in the first implemented network. The first and second networks are trained with 200 epochs while the third network is trained with 250 epochs. The loss function used for optimization is the mean squared error (MSE) with a thresholding value in order to predict and obtain the three classes (Normal (0), non-CSME (1), CSME (2)). Moreover, L2 regularization with weight decay factor 0.005 is used in the convolutional layers and a dropout rate of 0.5 is used between fully connected layers. These are implemented as regularization approaches to decrease overfitting in the network during training.

3.3 Evaluation

Once the network is trained, the test images are used to evaluate the performance of the implemented network by predicting the classification of previously unseen data. The performance of the implemented system are evaluated using three measurements; sensitivity, specificity and accuracy. In this study, sensitivity is defined as the percentage of images that are correctly classified as having DME out of the true total number of images with DME. Specificity is defined as the percentage of images that are correctly classified as not having DME out of the true total number of images without DME. Accuracy is the percentage of images that classified correctly.

4 Results and Discussion

The proposed system was evaluated on 30% of 1200 images (357 images) where the trained CNN achieved 88.8%, 74.7% and 96.5% in terms of accuracy, sensitivity and specificity, respectively. From the confusion matrix in Figure 3 that gives the prediction ratio per class, we notice that the prediction ratios for each of the classes are very encouraging. An important consideration is the imbalanced distribution of the images in the dataset, 154 images (13%) are grade 2, 75 images (6%) are grade 1 while 971 images (81%) are grade 0. Consequently, the non-CSME and CSME classes constitute only 19% of the data while the majority of the images are class 0 (normal). This imbalanced distribution may cause the CNN to overfit to the majority classes and result in the differences observed in prediction ratios between classes.

We have used an oversampling strategy in order to balance the distribution in the case of rare or less common classes. This has led to improved results for prediction but may be improved further. Our proposed system demonstrates comparable results with other methods in the literature; we obtain higher accuracy than [21] who achieved 85.2% and similar accuracy to machine learning methods such as [7] and [10] which had 94.1% and 95.56% respectively. Although our accuracy is lower than these two models, we achieve this without the need for prior feature extraction, exudate and macula segmentation or the removal of retinal blood vessels.

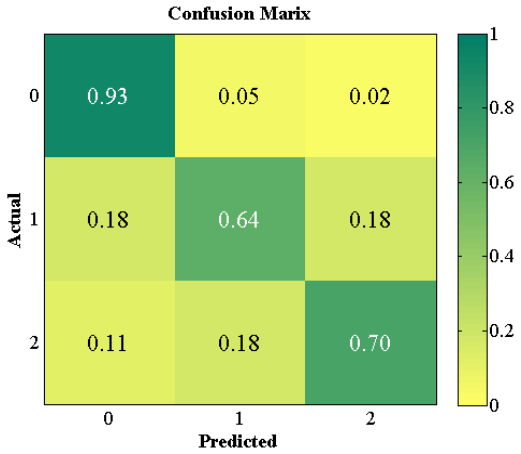


Fig. 3. Prediction ratio per class in confusion matrix, (0) refers to normal images, (1) non-CSME and (2) CSME. It shows good prediction results despite the rarity of classes 1 and 2.

5 Conclusions

An automated method for grading the severity of DME has been presented. We have shown that the proposed system based on CNN has convincing ability for automated feature extraction from retinal fundus images and grading of DME. This technique will be valuable for future automated DR grading system. This promising solution addresses the main concerns of traditional systems where the segmentation and handcrafted features are considered key requirements.

In future, we plan to improve the performance of the proposed CNN approach by considering better techniques for the training network parameters, implementing different loss functions, and considering further enhancements of the network architecture. We also attempt to improve the performance by addressing the data limitations in various ways, such as optimizing the method to allow for the training of higher-resolution images and increasing the quality of images using enhancement algorithms. We also aim to develop an improved solution for addressing data imbalance issues over that provided by our current oversampling strategy.

References

1. Early Treatment Diabetic Retinopathy Study Research Group. (1987). Treatment techniques and clinical guidelines for photocoagulation of diabetic macular edema: Early Treatment Diabetic Retinopathy Study report number 2. *Ophthalmology*, *94*(7), 761-774.
2. Bhagat, N., Grigorian, R. A., Tutela, A., & Zarbin, M. A. (2009). Diabetic macular edema: pathogenesis and treatment. *Survey of Ophthalmology*, *54*(1), 1-32.
3. Mookiah, M. R. K., Acharya, U. R., Fujita, H., Tan, J. H., Chua, C. K., Bhandary, S. V., & Tong, L. (2015). Application of different imaging modalities for diagnosis of Diabetic Macular Edema: A review. *Computers in Biology and Medicine*, *66*, 295-315.

4. Nayak, J., Bhat, P. S., & Acharya, U. R. (2009). Automatic identification of diabetic maculopathy stages using fundus images. *Journal of Medical Engineering & Technology*, *33*(2), 119-129.
5. Giancardo, L., Meriaudeau, F., Karnowski, T. P., Li, Y., Garg, S., Tobin, K. W., & Chaum, E. (2012). Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. *Medical Image Analysis*, *16*(1), 216-226.
6. Tariq, A., Akram, M. U., Shaukat, A., & Khan, S. A. (2013). Automated detection and grading of diabetic maculopathy in digital retinal images. *Journal of Digital Imaging*, *26*(4), 803-812.
7. Zaidi, Z. Y., Akram, M. U., & Tariq, A. (2013, December). Retinal image analysis for diagnosis of macular edema using digital fundus images. In *Applied Electrical Engineering and Computing Technologies (AEECT), 2013 IEEE Jordan Conference on* (pp. 1-5). IEEE.
8. Deepak, K. S., & Sivaswamy, J. (2012). Automatic assessment of macular edema from color retinal images. *IEEE Transactions on Medical Imaging*, *31*(3), 766-776.
9. Baby, C. G., & Chandy, D. A. (2013, February). Content-based retinal image retrieval using dual-tree complex wavelet transform. In *Signal Processing Image Processing & Pattern Recognition (ICSIPR), 2013 International Conference on* (pp. 195-199). IEEE.
10. Mookiah, M. R. K., Acharya, U. R., Chandran, V., Martis, R. J., Tan, J. H., Koh, J. E., & Laude, A. (2015). Application of higher-order spectra for automated grading of diabetic maculopathy. *Medical & Biological Engineering & Computing*, *53*(12), 1319-1331.
11. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097-1105).
12. Sun, Y., Liang, D., Wang, X., & Tang, X. (2015). Deepid3: Face recognition with very deep neural networks. *arXiv preprint arXiv:1502.00873*.
13. B. Al-Bander, W. Al-Nuaimy, M. A. Al-Taee, A. Al-Ataby& Y. Zheng (2016). Automatic Features Learning Method for Detection of Retinal Landmarks. *Proc. 9th International Conference on Developments in eSystems Engineering (DeSE '2016),* Liverpool & Leeds, England, 31st August – 2nd September 2016.
14. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.
15. Ranzato, M. A., Huang, F. J., Boureau, Y. L., & LeCun, Y. (2007). Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on* (pp. 1-8). IEEE.
16. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929-1958.
17. Decenciere, E., Zhang, X., Cazuguel, G., Laÿ, B., Cochener, B., Trone, C., & Charton, B. (2014). Feedback on a publicly distributed image database: The Messidor database. *Image Analysis and Stereology*, 33(3), 231-234.
18. The Lasagne, available online: https://github.com/Lasagne/Lasagne, (last accessed on 03 August 2016).
19. The Theano, available online: https://github.com/Theano/Theano, (last accessed on 31 July 2016).
20. Saxe, A. M., McClelland, J. L., & Ganguli, S. (2013). Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. *arXiv preprint arXiv:1312.6120*.
21. Lim, S. T., Zaki, W. M. D. W., Hussain, A., Lim, S. L., & Kusalavan, S. (2011, December). Automatic classification of diabetic macular edema in digital fundus images. In *Humanities, Science and Engineering (CHUSER), 2011 IEEE Colloquium on* (pp. 265-269). IEEE.