Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 219 (2023) 1097–1104

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2022

Diabetic Retinopathy Grading using Blended Deep Learning

Fernando C. Monteiro[a,b,*]

[a]Research Centre in Digitalization and Intelligent Robotics (CeDRI),
[b]Laboratório para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC),
Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal

Abstract

Diabetic retinopathy is a complication of diabetes that is mainly caused by the damage of the blood vessels located in the retina. Retinal screening contributes to early detection and treatment of diabetic retinopathy. DR has five stages, namely healthy, mild, moderate, severe and proliferative diabetic retinopathy. Computer-aided diagnosis approaches are needed to allow an early detection and treatment. Several automated deep learning (DL) based approaches have been proven to be a powerful tool for DR grading. However, these approaches are usually based on one DL architecture only which could produce over-fitted results. Another identified problem is the use of imbalanced datasets. In this paper, we proposed a blended deep learning approach obtained by training several individual DL models, using a 5-fold cross-validation technique and combining their predictions in a final score. This blended model highlights each individual model where it performs best and discredits where it performs poorly, increasing the robustness of the results. The experiments were conducted on a balanced DDR dataset containing 33310 retina fundus images equally distributed for the DR grades. An explainability algorithm was also used to show the efficiency of the proposed approach in detecting DR signs.

Keywords: Diabetic retinopathy grading; blended deep learning; retina fundus images; retinopathy levels

1. Introduction

Diabetic Retinopathy (DR) is a serious and common health condition caused by diabetes mellitus that affects the human retina causing damage to blood vessels which become leaky or blocked generating microaneurysms, haemorrhages , soft and hard exhudates [1] (see Fig 1). Vision loss most commonly occurs due to swelling in the central part

* Fernando C. Monteiro. Tel.: +351-273-303-012
E-mail address: monteiro@ipb.pt

10.1016/j.procs.2023.01.389

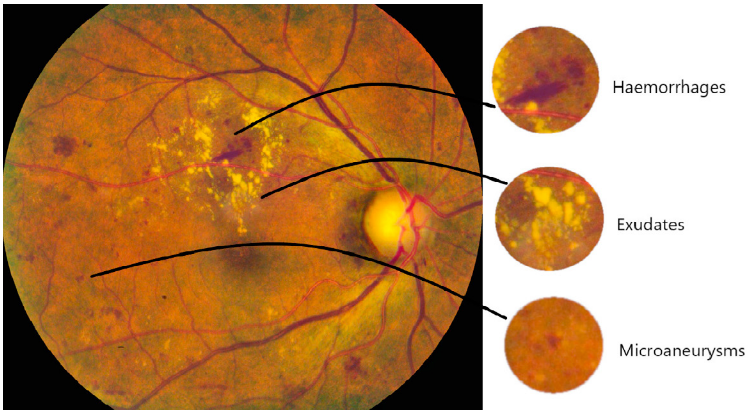

Figure 1. DR lesions in a retina fundus image.

of the retina which can lead to vision impairment. Abnormal blood vessels can also grow from the retina, which can bleed or cause scarring of the retina and blindness.

While the exact number is unknown, it is estimated that almost 12 million people have moderate or severe vision impairment or blindness due to glaucoma, diabetic retinopathy and other eye diseases that could have been prevented [2]. There is widespread knowledge that screening, early detection and prompt treatment of DR allow prevention of visual impairment. With the rising incidence and prevalence of DR [3], public health systems in both developing and developed countries should implement and maintain of a DR screening program for people with diabetes.

Accordingly with International Clinical Diabetic Retinopathy Disease Severity Scale [4] (ICDRDSS), there are five severity levels of DR depending on the presence of retina lesions, namely, *normal* (grade 0) no abnormalities; *mild DR* (grade 1) tiny bulges develop in the blood vessels, which may bleed slightly but do not usually affect vision; *moderate DR* (grade 2) swelling and distortion of blood vessels; *severe DR* (grade 3) widespread changes affect the blood vessels blocking them, including more significant bleeding into the eye; prominent microvascular abnormalities; *proliferative DR* (grade 4) neovascularization or vitreous/preretinal hemorrhage; these patients are at a high risk of disease progression and permanent vision loss, and are most likely experiencing neuropathy elsewhere. This grading is important to determine the exact stage of DR to begin a suitable treatment process and to prevent the retina deterioration and the possible blindness.

A worldwide problem associated with DR is the reduced number of specialists to follow the patients, e. g. in China there is a ratio of one ophthalmologist for 3000 diabetes patients [5]. To overcome this problem, several computer aided diagnosis techniques, mainly based on deep learning schemes, have been proposed to automatically detect DR and its different stages from retina fundus images [6, 7].

Deep learning models are inspired by the organization of the human visual cortex and are designed to automatically create and learn feature hierarchies through back-propagation by using multiple layer blocks, such as convolution layers, pooling layers, and fully connected layers from low to high level patterns [8]. This technology is especially suited for image processing, as it makes use of hidden layers to convolve the features with the input data. The automatic extraction of the most discriminant features from a set of training images, suppressing the need for preliminary feature extraction, became the main strength of deep learning approaches.

There still remain some concerns associated with the automated DR classification approaches proposed in the literature that are preventing its use in clinical applications. An important factor is the model reliability based in the fact that the majority of published approaches used imbalanced datasets in their studies. Using a dataset where more than 70% of the samples are healthy retina images and only 3% for severe DR grade will biased the results.

Deep learning network (DLN) automatically extracts the most discriminative features from training images, however, which features are extracted to make predictions remains unknown. Since ophthalmologists generally evaluate the severity of DR based on the presence of lesions in the retina, we believe that while automated techniques do not demonstrate that their results are based mainly on the lesions they will not be accepted as a reliable substitute of human ophthalmologists. Gradient-weighted class activation mapping (Grad-CAM) is an explainable technique that can be used to help understand the predictions made by a deep neural network [9].

In this paper, we introduce a blended DL approach for the problem of classifying DR in fundus imagery. To avoid over-fitting produced by imbalanced datasets in individual DLN, our approach applies several of the best known DLN models that were trained and tested in a balanced dataset, combining their predictions in a stacking model. This approach takes the advantage of training DLN in different resolutions and increases the robustness by reducing possible over-fitting models to produce reliable predictive results than those obtained by using individual DL architectures.

The remainder of this paper is organized as follows: Previous related works are reviewed in Section 2. In Section 3, we describe the used materials and the proposed method. Section 4 presents the results and some discussion of the findings. Finally, some conclusions are drawn in Section 5.

2. Related work

Many recent studies have investigated the use of deep learning techniques for detecting and grading DR. Readers are encouraged to read the excellent reviews [6, 7] for comprehensive details of research in the field of DR detection.

Ensemble learning has been played an important role in developing robust DL frameworks for DR classification, by combining the advantages of several classifiers [10, 11].

Qummar et al. [10] ensembled five DLN models to encode the best features and improve the classification for different stages of DR. They used stacking to combine the results of all different models and generate a unified output. They obtained a overall accuracy of 0.648. To overcome the bias created by imbalanced EyePACS dataset they balanced it via data augmentation. However, this process used only the original images in each grade set. For example, in grade 4 set they obtained 4119 images from just 113 original images. The variability of features in such augmented images is, of course, very small.

Jiang et al. [11] presented an approach using Inception V3, Resnet152 and Inception-Resnet-V2 architectures. They combined the DL predictions using the Adaboost algorithm by using the Class Activate Maps technique for each individual model as well as for the ensemble model, in order to reduce the bias of each single model. The analysis is done only for binary classification of the presence or absence of DR. Experiments shown that the proposed method has stronger robustness and achieved higher performance than that of individual deep learning model.

Zhang et al. [12] developed an ensemble model for the grading of DR in five levels. The individual models were based on several pretrained DL networks which acted as the feature extraction part, and a custom standard dense neural network which acted as the classifier. The ensemble model outperformed the individual ones.

Bodapati et al. [13] blended features extracted from multiple pretrained DL models using a multi-modal fusion module. These final representations are used to train a DLN used for DR identification and severity level prediction. They sustain that as each layer extracts different features, fusing them using 1D pooling and cross pooling will lead to better representation than using features extracted from a single DLN. They obtained an improvement in the performance of the proposed model over individual models.

Kaushik et al. [14] developed a deep learning based computer-aided diagnostic system, using a stacked generalization of convolution neural networks (CNN). Three custom CNN model weights are fed on the top of a single meta-learner classifier, which combines the most optimum weights of the three sub-neural networks to obtain superior metrics of evaluation and robust prediction results.

3. Proposed Method

In this section, we introduce the details of our framework. First, we introduce the used DR dataset, then the preprocessing methods and, finally, the blended DL approach to classify the samples in five DR stages.

3.1. Diabetic retinopathy dataset

Most of the publicly available datasets contain less than 2000 images, like the Indian Diabetic Retinopathy Image dataset (IDRiD) [15], with 516 images, or Messidor 2 [16], with 1748 images. The Asia Pacific Tele-Ophthalmology Society (*Kaggle* APTOS) [17] dataset contains 5590 images (3662 samples for training and 1928 samples for testing) collected by the Aravind Eye Hospital in India; however only the ground truths of the training images are publicly available. *Kaggle* EyePACS [18] is the largest DR dataset with 88702 images (35126 samples for training and 53576

Table 1. Number of samples in each grade for training, validation and test sets of the BDDR dataset.

DR class	Training	Validation	Test	Total	Percentage
Grade 0	3133	1253	1880	6266	20.0%
Grade 1	3133	1253	1880	6266	20.0%
Grade 2	3133	1253	1880	6266	20.0%
Grade 3	3133	1253	1880	6266	20.0%
Grade 4	3133	1253	1880	6266	20.0%
Total	15665	6265	9400	31330	

for testing) classified into five DR levels. This dataset was graded by only one ophthalmologist, which can potentially lead to annotation bias. It consists of a large number of images which were obtained under a variety of imaging conditions by various devices at multiple primary care sites throughout California [18].

The DDR dataset [19] is the second largest dataset when considering the DR grading task, consisting of 13673 images divided in 6835 for training, 2733 to validation and 4105 for test. The retina fundus images were collected between 2016 and 2018 across 147 hospitals in China, and annotated by seven ophthalmologists according to the ICDRDSS scale, using a majority voting approach. The appearance of low quality images is inevitable in clinical practice and it is meaningless for such images to be graded. To improve the DR detection accuracy, in addition to the five levels of DR severity, these ungradable images were classified into one new class (Grade 5). However, the majority of works that have used DDR dataset have removed the class 5 set. In order to do a fair comparative study, we also removed this class.

The automation of DR recognition and grading depends on large image datasets with many samples categorized by ophthalmologists. The results depend on the number of samples and their acquisition conditions. A small number of images may result in poor learning models, that are not sufficient to adequately train the deep learning architecture. Another factor to consider in DL approaches is the imbalance of the datasets used for training and testing. Over-fitting is a major issue in DLN, as imbalanced datasets make the network to over-fit to the most proeminent class in the dataset[7]. For example, in the *Kaggle* EyePACS dataset, more than 73% of the samples are annotated as normal, whereas severe and proliferative DR samples only account for less than 3% each.

To reduce the classification bias, several authors [19, 10, 20] have augmented existing samples from the 1, 2, 3, and 4 classes, to produce new samples. Although augmenting the dataset does not make it totally balanced, it does help to lessen the biases. However, as this augmentation process, in some classes, is based only in a very small number of images, the variability of features learned by the DLN does not increase.

To overcome this problem, in our experiment we have balanced the DDR classes in all the sets, using random samples of the same class, from the EyePACS, APTOS and IDRiD datasets, thus obtaining the balanced DDR dataset (BDDR) with a total of 31330 samples. Table 1 shows the distribution of training, validation and test sets per class for the balanced DDR dataset. Only a small number of samples from grade 3 and 4 of training set have been obtained by rotation since even when joining all the five datasets we cannot obtain the number of samples we need.

Using several datasets, that were collected on different resolution, equipment, and geographies allows to incorporate this variability into our model, which increases the generalization by reducing sensitivity to the image acquisition system. Using only one imbalanced dataset oversimplifies the identification process making it impractical to be used for recognizing DR in images obtained worldwide in different conditions [7].

3.2. Image preprocessing

As the used DR datasets were obtained from different locations with different type of equipment, they may contain images that vary in terms of resolution and aspect ratio. The images could also contain uninformative black space areas that may induce some bias in the training and test processes.

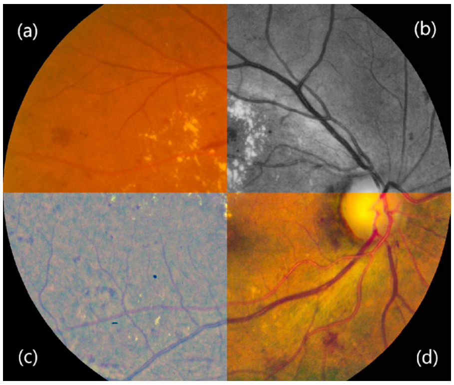

Figure 2. A mosaic of preprocessed fundus images: (a) original image, (b) CLAHE applied to green channel, (c) subtracting the local average colour, (d) applying CLAHE on each of the RGB channels.

Image preprocessing is thus a necessary step to enhance image features and to ensure the consistency of images. In order to reduce the black space areas, the images were cropped to a circle that limits retina. The image resolution was also adapted to each deep learning architecture, e.g. 224x224 for ResNet and 299x299 for Inception architectures.

In some works only the green channel of images was extracted due to its high contrast in blood vessels and red lesions [21]. However, these approaches loose the colour information. A technique often used is image normalization based on the min-pooling filtering [22, 23]. Contrast enhancement is a common preprocessing approach used to highlight the image features like blood vessels, or retina lesions. From this, in our experiments we adopted the Contrast Limited Adaptive Histogram Equalization (CLAHE), applying it to each colour channel. Figure 2 shows the outcome of each method.

3.3. Blended stacking approach

In our approach ten state-of-the-art DLNs were used: VGG 16/19 [24], ResNet 50/101 [25], Inception-V3 [26], Inception-ResNet [27], Xception [28], DenseNet201 [29], DarkNet53 [30] and EfficientNetB0 [31]. We proposed a blended deep learning model by combining those predictions in a final score.

To train each DL architecture, we used the fine-tuning strategy, as well as the stochastic gradient descent with momentum optimizer (SGDM) at their default values, dropout rate set at 0.5 and early-stopping to prevent over-fitting, and the learning rate at 0.0001. SGDM is used to accelerate gradients vectors in the correct directions, as we do not need to check all the training examples to have knowledge of the direction of decreasing slope, thus leading to faster converging. Additionally, to balance the memory consumption and the performance of the network, we used a batch size of 32 to update the network weights more often, and trained them in 30 epochs. To preprocess images further, all samples go through a heavy data augmentation process which include flipping, 180^o random rotation, rescaling factor between 70% and 130%, and horizontal and vertical translations between -20 and $+20$ pixels. The networks were trained using the ONNX package for Matlab® on a node of the CeDRI cluster equipped with a NVIDIA A100 GPU.

Different DL networks have different resolution and use the image features in a different way. This may produce over-fitting and some mistakes associated to an individual DL model. A blended method is a meta-algorithms that combine several deep learning techniques into one predictive model. Stacking is a model that combines the output of multiple predictive models and use a classifier to compute the final prediction [10]. The concept behind stacking is that certain models might correctly classify a test sample while others might fail to do so. It allows to use the strength of each individual estimator by using their output as input of a final estimator.

The stacking approach often performs better than individual models when base-models are configured to predict probabilities instead of crisp class labels, as the added uncertainty in the predictions provides more context for the meta-model when learning how to best combine the predictions. Stacking highlights each base model where it performs best and discredits each base model where it performs poorly. For this reason, we used stacking to improve the prediction of our model, which is evident from our results.

Table 2. Classification results (accuracy) of individual DLN classifiers on the BDDR test set and the blended results.

DLN	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4	OA
VGG16	0.991	0.693	0.359	0.619	0.429	0.618
VGG19	0.952	0.803	0.244	0.815	0.227	0.608
ResNet50	0.949	0.771	0.317	0.560	0.334	0.586
ResNet101	0.956	0.753	0.323	0.613	0.256	0.580
Inception-V3	0.960	0.732	0.292	0.703	0.289	0.595
Incep.-ResNet	0.970	0.750	0.262	0.673	0.359	0.603
Xception	0.952	0.697	0.261	0.668	0.323	0.580
DenseNet201	0.960	0.804	0.366	0.568	0.285	0.597
DarkNet53	0.940	0.805	0.452	0.539	0.301	0.607
EfficientNetB0	0.967	0.731	0.234	0.702	0.284	0.584
Blended	**0.986**	**0.822**	**0.326**	**0.694**	**0.397**	**0.645**

4. Results and analysis

For final scoring, we blended the results of ten state-of-the-art deep learning models for DR grading at different resolution. Since we have used 5-fold cross-validation, this gives a total of 50 models (10 architectures x 5 folds). The aim of this learning strategy is to reduce the generalization error of one DLN model and is a promising technique to fuse information from multiple models.

We experiment three ways of combining individual predictions: *majority voting* - for each test observation, the prediction is the most frequent class in all predictions; *averaging* - average over all the predictions (output of the softmax layer) from the different networks; *weighted average* - the weights are proportional to an individual model's performance. Averaging approach based on the individual DLN probability of belonging to each grade, achieved the best performance.

Table 2 summarizes the accuracy results for each class obtained from ten state-of-the-art DLNs, trained with 5-Fold cross-validation. These DLN were trained and tested over the balanced dataset described in Table 1 with 15665 samples for the training set, 6265 samples for the validation set and 9400 samples for the test dataset, equally distributed over all five DR grades.

As expected, the overall accuracy obtained with a balanced dataset is lower than the one obtained with an imbalanced dataset. From our current knowledge, this is the first study that used a dataset balanced with images from different datasets.

Although the model produced high accuracy for normal retina and mild DR, exhibits poor performance on samples with moderate and proliferative DR, suggesting that these models are weak in learning these two types of samples. Furthermore, swelling and distortion of blood vessels in moderate DR are difficult to identify, while samples of proliferative DR are easily misclassified as moderate or severe DR, specially when there is no severe hemorrhages. The blended approach increased the OA based on a better prediction of mild DR and severe DR. The high accuracy for grade 0 (normal retina) detection suggests that we can use these approaches to rapidly make a binary classification, in normal and DR retina.

To explain how the model learned to detect DR signs such as exhudates, hemorrhages and microaneurysms, we used an explainability algorithm based on the use of Gradient-weighted Class Activation Mapping (Grad-CAM) [9]. This technique was used to produce a visual explanation from the results of the proposed DL model. Grad-CAM maps are a common method where a heatmap is generated by projecting the gradients of the class specific weights of the output classification layer back to the feature maps of the last convolutional layer, thereby highlighting important regions in the image used for prediction. Figure 3 shows positive cases of DR accordingly with Grad-CAM map. The red coloured areas in the retinal images show the signs detected by our DL model for grading DR. We can see that the heatmap is accurately located around the signs of DR, such as exhudates and hemorrhages.

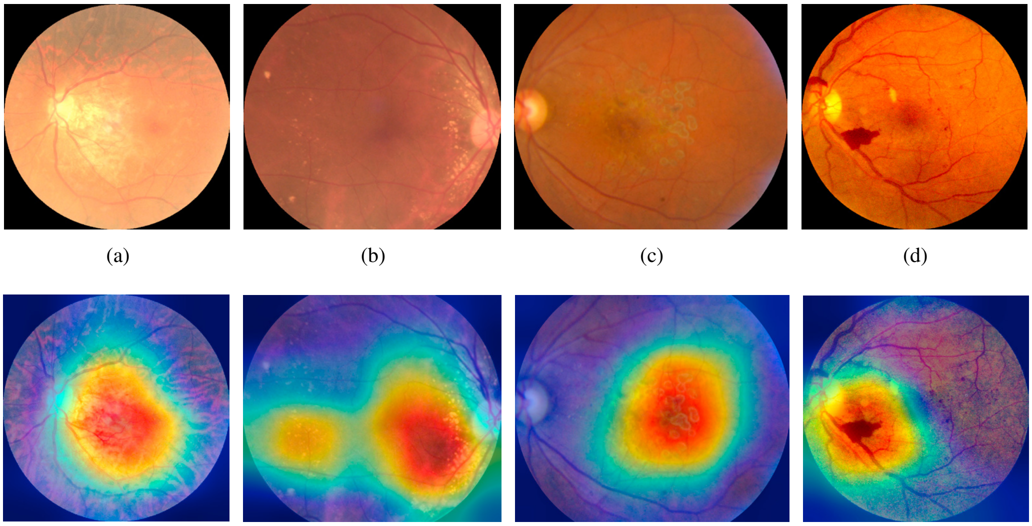

Figure 3. Grad-CAM explanation of positive cases of referable diabetic retinopathy: (a) mild DR with some microaneurysms between optic disc and macula, (b) moderate DR with exudates close to optic disc, (c) severe DR with hard exudates and hemorrhages located close to the macula region, (d) proliferative DR with preretinal hemorrhage.

5. Conclusion

In this work, we assessed the state-of-the-art deep learning models for DR grading in five levels. We proposed a blended grading predictor, by training ten individual DLN models, with 5-fold cross-validation, and combining their prediction in a final score. The aim of this learning strategy is to reduce the generalization error of one DLN model and is a promising technique to fuse information from multiple models. In our experiments we balanced the DDR dataset using images from from EyePACS, APTOS, Messidor-2 and IDRiD datasets such that all classes are represented equally. This new balanced dataset aims to reduce biased classification presented in approaches that used imbalanced datasets. The obtained results show that our model outperforms the individual score of each one of the available DL architectures. As future work, we aim to increase the detection of lesions of moderate and proliferative DR in order to achieve a better performance. Even more, we suggest to refine the datasets in order to remove ungradable samples and correct some misclassified notations that still remain.

Acknowledgements

The authors are grateful to the Foundation for Science and Technology (FCT, Portugal) for financial support through national funds FCT/MCTES (PIDDAC) to CeDRI (UIDB/05757/2020 and UIDP/05757/2020) and SusTEC (LA/P/0007/2021).

References

[1] R. Taylor, D. Batey, Handbook of retinal screening in diabetes: diagnosis and management, second ed., Wiley-Blackwell, 2012.

[2] World Health Organization, World report on vision - Licence: CC BY-NC-SA 3.0 IGO, accessed on 26 April 2022. (2019). URL https://www.who.int/publications/i/item/9789241516570

[3] Z. Teo, Y. Tham, M. Yu, M. Chee, T. Rim, N. Cheung, M. Bikbov, Y. Wang, Y. Tang, et al., Diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis, Ophthalmology 128 (11) (2021) 1580–1591.

[4] C. Wilkinson, F. Ferris, R. Klein, P. Lee, C. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, J. Verdaguer, Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales, Ophthalmology 110 (9) (2003) 1677–1682.

[5] L. Dai, L. Wu, H. Li, C. cai, Q. Wu, H. Kong, R. L. et al., A deep learning system for detecting diabetic retinopathy across the disease spectrum, Nature Communications 12 (1) (2021) 3242.
[6] W. L. Alyoubi, W. M. Shalash, M. F. Abulkhair, Diabetic retinopathy detection through deep learning techniques: a review, Informatics in Medicine Unlocked 20 (2020) 100377.
[7] N. Tsiknakis, D. Theodoropoulos, G. Manikis, E. Ktistakis, O. Boutsora, A. Berto, F. Scarpa, A. Scarpa, D. I. Fotiadis, K. Marias, Deep learning for diabetic retinopathy detection and classification based on fundus images: A review, Computers in Biology and Medicine 135 (2021) 104599.
[8] N. Baker, H. Lu, G. Erlikhman, P. J. Kellman, Local features and global shape information in object classification by deep convolutional neural networks, Vision Research 172 (2020) 46–61.
[9] K. Vinogradova, A. Dibrov, G. Myers, Towards interpretable semantic segmentation via gradient-weighted class activation mapping, Proceedings of the AAAI Conference on Artificial Intelligence 34 (10) (2020) 13943–13944.
[10] S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, Z. U. Rehman, I. A. Khan, W. Jadoon, A deep learning ensemble approach for diabetic retinopathy detection, IEEE Access 7 (2019) 150530–150539.
[11] H. Jiang, K. Yang, M. Gao, D. Zhang, H. Ma, W. Qian, An interpretable ensemble deep learning model for diabetic retinopathy disease classification, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2019, pp. 2045–2048.
[12] W. Zhang, J. Zhong, S. Yang, Z. Gao, J. Hu, Y. Chen, Z. Yi, Automated identification and grading system of diabetic retinopathy using deep neural networks, Knowledge-Based Systems 175 (2019) 12–25.
[13] J. D. Bodapati, V. Naralasetti, S. N. Shareef, S. Hakak, M. Bilal, P. K. R. Maddikunta, O. Jo, Blended multi-modal deep ConvNet features for diabetic retinopathy severity prediction, Electronics 9 (6) (2020) 914.
[14] H. Kaushik, D. Singh, M. Kaur, H. Alshazly, A. Zaguia, H. Hamam, Diabetic retinopathy diagnosis from fundus images using stacked generalization of deep models, IEEE Access 9 (2021) 108276–108292.
[15] P. Porwal, S. Pachade, M. Kokare, et al., IDRiD: Diabetic retinopathy - segmentation and grading challenge, Medical Image Analysis 59 (2020) 101561.
[16] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, J. Klein, Feedback on a publicly distributed image database: The Messidor database, Image Analysis & Stereology 33 (3) (2014).
[17] Asia Pacific Tele-Ophthalmology Society, Aptos 2019 blindness detection, accessed on 4 April 2022. (2019).
URL https://www.kaggle.com/competitions/aptos2019-blindness-detection
[18] EyePACS, Diabetic retinopathy detection, accessed on 4 April 2022. (2015).
URL https://www.kaggle.com/c/diabetic-retinopathy-detection
[19] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, H. Kang, Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening, Information Sciences 501 (2019) 511–522.
[20] D. A. Rocha, F. Ferreira, Z. Peixoto, Diabetic retinopathy classification using vgg16 neural network, Research on Biomedical Engineering 38 (2022) 761–772.
[21] J. Lu, Y. Xu, M. Chen, Y. Luo, A coarse-to-fine fully convolutional neural network for fundus vessel segmentation, Symmetry 10 (11) (2018) 607.
[22] R. Gargeya, T. Leng, Automated identification of diabetic retinopathy using deep learning, Ophthalmology 124 (7) (2017) 962–969.
[23] G. T. Zago, R. V. Andreão, B. Dorizzi, E. O. Teatini Salles, Diabetic retinopathy detection using red lesion localization and convolutional neural networks, Computers in Biology and Medicine 116 (2020) 103537.
[24] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Int. Conference on Learning Representations, 2015, pp. 1–14.
[25] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818–2826.
[27] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inceptionresnet and the impact of residual connections on learning, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, 2017, pp. 4278–4284.
[28] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1800–1807.
[29] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269.
[30] J. Redmon, A. Farhadi, YOLOv3: An incremental improvement, ArXiv (2018) 1804.02767.
[31] M. Tan, Q. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in: Proceedings of the 36th International Conference on Machine Learning, Vol. 97, 2019, pp. 6105–6114.