Deep Learning with Class Imbalance for Detecting and Classifying Diabetic Retinopathy on Fundus Retina Images

Kamel Kamal (vrsc@azhar.edu.eg)  
Al-Azhar University

Rania. A. Mohamed  
Modern University for Technology & Information

Ashraf Darwish  
Helwan University

Aboul Ella Hassanien  
Cairo University

Research Article

Keywords: Diabetic retinopathy, deep learning, imbalanced data set, CNN architecture, and Convolutional neural network

Posted Date: August 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1935432/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Deep Learning with Class Imbalance for Detecting and Classifying Diabetic Retinopathy on Fundus Retina Images

Kamel K. Mohammed1 *, Rania. A. Mohamed2, Ashraf Darwish3 *, and Aboul Ella Hassanien4 *

1Center for Virus Research and Studies, Al-Azhar University, Cairo, Egypt.
2Modern University for Technology & Information, Faculty of Computers and Artificial Intelligence, Cairo, Egypt.
3Faculty of Science, Helwan University, Helwan, Egypt
4Faculty of Computer and Artificial Intelligence, Cairo University, Cairo, Egypt

*Scientific Research Group in Egypt (www.egyptscience.net) Abodicairo@cu.edu.eg

ABSTRACT Diabetes mellitus is a disorder that causes diabetic retinopathy and is the primary cause of blindness worldwide. Early detection and treatment are required to reduce or avoid vision degradation and loss. For that purpose, various artificial-intelligence-powered approaches for detecting and classifying diabetic retinopathy on fundus retina images have been proposed by the scientific community. This article explores solutions to diabetic retinopathy detection by using three recently developed deep neural networks that have proven effective and efficient. Densenet201, Resnet101, and EfficientNetb0 deep neural network families have been applied to detect and classify diabetic retinopathy on fundus retina images. The dataset was notably not equilibrium; the widespread majority had been normal images, while mild Diabetic retinopathy images made up a very minor percentage of the total dataset. To treat the skewed distribution and to keep away from biased classification results different scenarios have been used to balance the classes by utilizing (i) weight balancing with data augmentation; (ii) oversampling with data augmentation; (iii) focal loss with data augmentation, and (iv) a hybrid method of oversampling with a focal loss with data augmentation that improves the deep neural network performance of fundus retina images classification with the imbalanced dataset to build an expert system that can rapidly and adequately detect fundus images. The experimental results indicated that using Densenet201, Resnet101, and EfficientNetb0, with weight balancing on the dataset, substantially improves diabetic retinopathy prediction, by re-weighting each class in the loss function, a class that represents an under-classified class will receive a larger weight. The models yielded 94.74%, 94.74%, and 93.42%, respectively, on the test data set.

INDEX TERMS Diabetic retinopathy, deep learning, imbalanced data set, CNN architecture, and Convolutional neural network.

I. INTRODUCTION

It is considered diabetes mellitus a major public health concern that affects 463 million people worldwide and is estimated to reach 750 million by 2050 [1]. Not less than one-third of people with diabetes suffer from an eye disease related to their diabetes, the most prevalent of which is diabetic retinopathy (DR) [2]. Increasing vascular abnormalities brand diabetes-related retinopathy (DR) in the retina induced by continuous hyperglycemia can impact anybody with diabetes, regardless of how severe their condition [3]. So, it is the leading cause of blindness among adults worldwide, affecting an estimated 93 million people [4]. These numbers are anticipated to increase due to increased incidence as diabetes prevalence rises in developing Asian countries [5,6]. Even though the early stages of diabetic retinopathy are often asymptomatic, neuronal, retinal damage, and clinically undetected microvascular abnormalities occur [7]. Therefore, regular eye examinations are essential for diabetics, as early detection and treatment of the disease are important [8]. Early detection of DR is important because the only preventative medicine is to control hypertension, hyperlipidemias, and hyperglycemia [7]. Furthermore, if the eyes are treated early in the condition, modern therapies like laser photocoagulation can minimize the risk of blindness in proliferative retinopathy and diabetic macular disease by approximately 98% [9]. Early detection and treatment are critical for postponing or avoiding diabetic retinopathy-related blindness [10]. Diabetic retinopathy can be diagnosed clinically by examining the retinal fundus or imaging techniques like fundus photography or optical coherence tomography. The Early Treatment Diabetic Retinopathy Study [11] is one of several standard diabetic
retinopathy grading systems. Early Treatment Diabetic Retinopathy Study uses various layers to isolate fine detailed Diabetic Retinopathy features. This grade relates to the retinal fundus’ seven fields of vision. The Early Treatment Diabetic Retinopathy Study considers the standard method [12]. Still, due to implementation complexity and technical limitations [13], international clinical diabetic retinopathy [14] scales are used in clinical and computational diagnostics, etc. The alternative rating system setting for [15] is also used. There are five severity levels for Diabetic Retinopathy and four levels for Diabetic Macular Edema on the International Clinical Diabetic Retinopathy Scale, demanding fewer Fields of View. Table 2 discusses the International Clinical Diabetic Retinopathy levels, seen in Figure 1.

Diabetic retinopathy is detected in two stages, screening and diagnosis. Fine pathognomonic Diabetic retinopathy symptoms in the early stages are generally identified after dilating pupils (mydriasis). Slit-lamp biomicroscopy with a + 90.0 D lens and direct [17] indirect ophthalmoscopy [18] screen for diabetic retinopathy. Identifying diabetic retinopathy is the next step by locating diabetic retinopathy-associated lesions and comparing them to the standard grading system criteria. Manual evaluation of diabetic retinopathy is a time-consuming and subjective task and requires an ophthalmologist’s high degree of expertise. In addition, mild non-proliferative diabetic retinopathy can have a pathological pattern similar to moderate non-proliferative diabetic retinopathy and severe non-proliferative diabetic retinopathy. Therefore, it is difficult to distinguish mild non-proliferative diabetic retinopathy from moderate non-proliferative diabetic retinopathy and severe non-proliferative diabetic retinopathy. Even if all of these resources are available, there is always the chance of misdiagnosis [19]. Many healthcare systems, particularly in developing nations and isolated places, have a shortage of educated ophthalmologists. Furthermore, the number of diabetic retinopathy and vision-threatening diabetic retinopathy worldwide was expected to reach 132.12 million and 28.54 million in 2020. In 2045, that number is expected to grow to 160.5 million and 44.82 million, respectively [20]. As a result, diabetic retinopathy can induce weariness and put a load on medical professionals. These challenges need the creation of creative solutions that are both accessible and practical. AI-assisted tools have become an additional tool for doctors with the introduction of artificial intelligence (AI) and computer vision methods.

II. RELATED WORK

Detecting DR early is an exciting problem in the field of computer vision. With diagnostic transparency criteria, the purpose of detection is to classify diabetic retinopathy among grading systems used in the International Clinical Diabetic Retinopathy. So, the advancement of automated diabetic retinopathy (DR) pathology screening during the last few decades has been encouraging. Deep learning techniques have lately radically changed the area of computer vision. Many researchers are interested in using CNNs to accomplish image classification. A traditional technique based on the morphology of digital photographs extracts the number of such features from fundus images, according to (Hann et al., 2009) [21]. Dot hemorrhages (DH) and exudates are two of the most prevalent DR dysfunctions, and computer vision approaches have been developed to isolate and detect them. The algorithms employ specialized color channels and segmentation methods to distinguish these DR manifestations from physiological features in digital fundus images. The algorithms are tested on the first 100 images in a publicly available database. The positive and negative prediction values (PPV and NPV) are presented due to the diagnostic outcome. On a fundus image, the presence of exudates close to the macular is an essential diagnostic marker of diabetic macular edema. (Wang Shuangling, 2015) [22] Adopt a CNN (LeNet-5) model to extract image features for addressing blood vessel segmentation. It proposed a hierarchical retinal blood vessel segmentation method based on feature and ensemble learning. The proposed method has several unique characteristics. The experiment was conducted on two public retinal image databases (DRIVE and STARE). The numerous uses of deep CNNs in computer vision have formally mushroomed after Alex et al. [23,24] proposed AlexNet architecture for impressive performance gains at the 2012 ILSVRC competition. Deep transfer learning models for medical DR detection were examined (Khalifa et al., 2019) [25]. The numerical experiments were carried out on the 3600 images in the Asia Pacific Tele-Ophthalmology Society (APTOS) 2019 dataset. AlexNet, Res-Net18, SqueezeNet, GoogLeNet, VGG16, and VGG19 were the models used in this study. Data augmentation techniques were used to make the models more robust and solve the problem of overfitting. The AlexNet model was the most accurate (97.9 percent). Furthermore, this model only had a few layers, which reduced the training time and computational complexity. (Wanghu Chen, 2020). Proposes an approach to retinal image classification based on integrating multi-scale shallow CNNs with an accuracy of 0.85 [26]. It made use of 35,000 labeled images from the Kaggle platform. A clinician has graded each image for the presence of diabetic retinopathy on a scale of 0 to 4. As a result, an image’s label will represent the degree of DR determined by its characteristics, such as 0-No DR, 1-Mild, 2-Moderate, 3-Severe, and 4-Proliferative DR. (Mohamed Shaban,2020) [27] Proposed a low-complexity convolutional neural network (CNN) of 2D convolutional, Max pooling, fully connected, and Softmax layers trained to classify and stage DR subjects into normal, mild, moderate NPDR, severe NPDR, and PDR with 768 images of the fundus with an accuracy of 88.4%. After several excellent CNNs, architectures have been proposed, such as DenseNet201, ResNet101, and EfficientNetbo [28] are three new deep neural network families that have just been constructed. These
designs were chosen for their convenience of use and ability to perform a wide range of classification tasks in diabetic retinopathy conditions. [29-31]. Furthermore, compared to traditional machine learning approaches, these architectures have demonstrated greater accuracy and predictive efficiency [32]. The current research’ efficacy is hampered by imbalanced class distribution, which might result in biased classification results. To deal with the imbalanced data, we used five distinct approaches, all of which are alternatives to the strategy described by (Bridge et al., 2020) [33]. (Debasis Maji,2022) [34] proposed a system for grading retinal fundus images by the use of the Convolutional Neural Networks (CNNs) where the system can utilize the immense pattern learning ability of pre-trained CNNs like VGG16 and ResNet50 to grade the retinal fundus images they have used the two publicly available datasets on Kaggle and EIARG2 to train and validated this model with an accuracy 96%.

Table1: The previous literature papers.

<table>
<thead>
<tr>
<th>Paper name</th>
<th>Methods used</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Example of performance integration model</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Hann et al., 2009)</td>
<td>To distinguish these DR manifestation s from physiological features in digital fundus images, the algorithms employ specialized color channels and segmentation methods.</td>
<td>Transparency and simplicity are two of the approach's advantages. On a fundus image, the presence of exudates close to the macular is an essential diagnostic marker of diabetic macular edema.</td>
<td>The positive and negative prediction values (PPV and NPV) are presented due to the diagnostic outcome. The result didn't measure the severity of DR disease.</td>
<td>The accurate detection of diabetic retinopathy in retinal fundus images remains a serious difficulty because it includes both images in a single collecting environment, making it difficult to compare the performance of algorithms in the experiment.</td>
</tr>
<tr>
<td>(Wang Shuangling, 2015)</td>
<td>Using CNN performs as a trainable hierarchical feature extractor and ensemble RFs work as a classifier</td>
<td>Use ensemble classifier which can maintain advantages while get the best classifier</td>
<td>The dataset's features are manually and experimentally extracted; therefore, their correctness cannot be guaranteed. The data sets are typically small and of poor quality, consisting of only a few hundred or even hundreds of fundus</td>
<td>The proposed methodology performs well on small datasets in terms of both classification effect and efficiency.</td>
</tr>
<tr>
<td>(Khalifa et al., 2019)</td>
<td>The chosen deep transfer models were AlexNet, ResNet18, SqueezeNet, GoogleNet, VGG16, and VGG19.</td>
<td>The Performance integration model illustrates an advantage in accuracy.</td>
<td>To make the models more robust and solve the problem of over fitting, data augmentation techniques were used.</td>
<td>The image sample transformation and dataset are repeated, which effect performance of the integrated shallow CNN model.</td>
</tr>
<tr>
<td>(Wanghu Chen, 2020)</td>
<td>integration of multi-scale shallow CNNs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Mohamed Shaban,2020)</td>
<td>a low-complexity CNN of 2D convolutional</td>
<td></td>
<td>Transparency and the suggested method's simplicity make it appropriate for clinical adaption.</td>
<td></td>
</tr>
</tbody>
</table>
mild and moderate NPDR as one category, it can only classify the four phases of diabetic retinopathy. A limited dataset for training will prevent the model from extracting the necessary features to classify new data correctly.

Debasis Maji, 2022

CNNs like Vgg16, and ResNet50. Using the Focal Loss function has improved the model’s accuracy in computation. The suggested model has a complicated architecture that can be simplified to reach the same degree of accuracy.

<p>| Table 2: The International Clinical Illness Severity Scale (ICDSS) measures the severity of DR disease |</p>
<table>
<thead>
<tr>
<th>Level of seriousness</th>
<th>Observable findings during dilated ophthalmoscopy</th>
</tr>
</thead>
<tbody>
<tr>
<td>No DR</td>
<td>There are no anomalies.</td>
</tr>
<tr>
<td>Mild NPDR</td>
<td>It is considered the earliest phase of diabetic retinopathy, and it is distinguished by microaneurysms (MA), which are microscopic patches of enlargement in retinal blood vessels [16]. There is no substantial bleeding in the retinal nerves, and if DR is identified at this time, effective medical therapy can help save the patient's eyesight (Figure 1b).</td>
</tr>
<tr>
<td>Moderate NPDR</td>
<td>When blood leakage from the occluded retinal arteries occurs, mild NPDR escalates moderately. Hard Exudates (Ex) may also exist (Figure 1c). Furthermore, ophthalmoscopically visible Venous Beadings (VB) are caused by the retina’s dilatation and constriction of venules [16].</td>
</tr>
<tr>
<td>Severe and very severe NPDR</td>
<td>More retinal vessels are occluded at this stage, resulting in more than 20 intraretinal hemorrhages (IHE; Figure 1d &amp; e) or intraretinal microvascular abnormalities (IRMA) in all four fundus quadrants. This can be thought of as a thin, swollen blood vessel. IRMA appears as red dots with small, sharp edges in at least one quadrant. In addition, VB is in three or more quadrants [16].</td>
</tr>
<tr>
<td>PDR and advanced PDR</td>
<td>This is a more advanced condition stage that occurs when the problem is allowed to continue over an extended period. Neovascularization is the formation of new blood vessels in the retina (NV). These blood arteries are frequently weak, increasing the risk of fluid leakage and fibrous tissue growth [16]. Proliferative Diabetic Retinopathy causes various functional visual difficulties, including blurriness, a narrowed field of vision, and, in rare cases, total blindness (Figure 1f&amp;g).</td>
</tr>
</tbody>
</table>

In summary, the following are the important contributions of this work:

- To reduce the class bias induced by data imbalance, using four distinct methodologies which are straightforward, simple to execute and repeat, yet powerful, to handle skewed distribution of dataset or unbalanced dataset. These techniques, such as weight balancing, data augmentation, and focus loss methods, were studied to handle the unbalanced fundus image datasets.
- Displaying the suggestive regions, which are thought to be helpful for interpretability and explainability, that exceedingly impact CNNs’ expectation.

The following is how the paper is organized. Section 3 material and method that describes the data level approach in further detail and it also describes the algorithm-level approach. Section 4 shows clarify architecture and strategy of the CNNs models. Section 5 examines the key findings and comments. Finally, in Section 6, we describe conclusion of the study.
III. Material and Method

A. Description of diabetic retinopathy dataset

The diabetic retinopathy dataset is open to the public. Images of macula-centered retinal fundus were obtained from the Department of Ophthalmology, Ciencias Médicas Hospital, National University of Asuncion, Paraguay [35]. This imbalanced dataset contains 757 color fundus photos in this Imbalanced dataset. Fundus images have been classified into seven categories: As demonstrated in Figure 3, there were no DR indications (187 images), mild (or early) NPDR (4 images), moderate NPDR (80 images), severe NPDR (176 images), very severe NPDR (108 images), PDR (88 images), and advanced PDR (114 images).

B. Methodology

In this study, we propose deep-learning-based diabetic retinopathy predictions from fundus images. This study applied transfer learning of CNN models to classify fundus images. However, the dataset was highly unbalanced; therefore, we applied four different approaches: (i) weight balancing with data augmentation; (ii) oversampling with data augmentation; (iii) focal loss with data augmentation, and (iv) hybrid method of oversampling with the focal loss with data augmentation. The best-performing model from each approach was weight balancing on the dataset in substantially improves overall diabetic retinopathy prediction detected by deep learning for better interpretability. Figure 2 presents a block diagram of the anticipated study.

C. Unbalanced Data

In biomedical imaging, convolutional neural networks have shown significant performance. Its design, which is also influenced by class imbalance, consists of two or more convolutional layers that map input data into new representations or create predictions. Previous research [36] indicates that the majority class dominates the net gradient, which is responsible for updating the model's weight (or classes). This increases the minority class's (or classes') error in imbalanced settings. Deep learning approaches for class imbalance correction are divided into two types: data-level approaches and algorithm-level approaches.

D. Data level for Image Augmentation and oversampling on an imbalanced Dataset

To reduce the level of imbalance, data-level techniques employ data sampling methods. Overcome the effects of disproportionate histopathological images in deep learning-based breast cancer analysis by flipping the data up, down, left, and right, and rotating or augmenting the data [37]. Data augmentation was used to address the problem of class imbalance in deep learning-based brain cancer magnetic resonance imaging picture classification [38]. [39] Increased training samples based on class imbalances to overcome the problem of class imbalances for deep learning-based classification of histopathology breast cancer images. Under-sampling strategies are also used to handle the problem of unbalanced cancer image categorization. [40] Trained a lung nodule machine classifier using a support vector using a mixture of undersampling and oversampling. [41] addressed the problem of imbalances in breast cancer classification using a cluster-based under sampling strategy. The major disadvantage of random under-sampling is the loss of potentially useful data that may be important for learning.

E. Algorithm-Level Approach for Focal loss and Weight balancing

Algorithm-level solutions to imbalanced deep learning have been developed to reduce bias towards dominant groups, and class weights or penalties have been widely applied to address class imbalances. Lin et al. [42] proposed the focused loss function, which reshapes the cross-entropy (CE) losses and lessens the influence of readily identified data on the losses. This new loss function overcomes the problem of class imbalances; however, it also helps classify difficult-to-classify data. Authors in [43] used the focus loss function to build the deep learning model for optical detection of colorectal cancer. Authors in [44] combined focused loss and data augmentation to classify unbalanced lung nodules. Cost-sensitive learning, which reweights training data to provide further weight to minority classes, is also commonly employed in deep learning to overcome the class imbalance problem.
FIGURE 2. Block diagram of proposed system.
Furthermore, an ensemble approach has been proposed to address the imbalanced analysis of medical images that integrates multiple classifiers to achieve better performance than individual classifiers within the framework of deep learning [45-47]. This study uses strategies at both the data and algorithm levels to improve the neural network performance of fundus image classification with unbalanced multi-class datasets. Data imbalance is a typical difficulty in categorizing diabetic retinopathy images, and researchers have acknowledged and sought to fix it in prior studies [48]. We explored and evaluated the effectiveness of numerous techniques for unbalanced diabetic retinopathy image categorization. Our findings show that combining various data- and algorithm-level solutions can reduce class bias and use the performance measurements used in the study to analyze imbalance problems. As mentioned earlier, class bias is an imbalanced representation of a class in training data. It may result in a bias favoring the overrepresented class. Classifier bias, as opposed to class bias, is defined as the difference between the model's projected value and the anticipated correct value. The variability of a model prediction is represented by the variance error for a certain data point. A classifier's bias and variance must be balanced; excessive variance error leads to overfitting, while high bias error leads to underfitting. For the $n^{th}$ input to a neural network belonging to the $i^{th}$ class among $c$ total classes, Cross Entropy (CE) loss $L_{nCE}$ is a measure of the deviation between the predicted output $z_{in}$ and the expected output $y_{in}$, given by:

$$l_{nCE} = -\sum_i y_{in} \log z(i, n)$$  \hspace{1cm} (1)

To address imbalance dataset, a common practice is to use weighted cross entropy (WCE), $L_{nWCE}$ as the loss function. This is a variant of standard CE with an additional class weight parameter, $\omega_i$, inversely related to the number of classes in each class $i$, to balance the influence of each class. Thus, Weighted cross entropy loss is defined as:

$$l_{wnCE} = -\sum_i \omega_i y_{in} \log z(i, n)$$  \hspace{1cm} (2)

Focal Loss (FL) is a variant of WCE that formulates the weighting factor as a dynamic value by expressing it as a function of the error between:

$$z_{in} \text{ and } \{y_{in}; y_{in} = 1 \}, \text{ giving:}$$

$$l_{fl} = -\alpha \sum_i (1 - z_{in})^\gamma y_{in} \log(z_{in})$$  \hspace{1cm} (3)

Equation also shows that FL employs two additional scaling coefficients, $\alpha \in [0,1]$, $\gamma \geq 0$, common to all classes, to control the intensity of the loss.

**IV. Architecture and strategy of the CNN model**

Model architecture is an important aspect in enhancing the performance of many applications. From 1989 until the present, several changes have been made to the CNN architecture. Several CNN architectures have been introduced during the last ten years [49]. In this research, we used three different CNN architectures called Densenet201, Resnet101, and efficientnetb0.

1) **RESNET**

ResNet: He et al. [50] attempt to create an ultra-deep network devoid of the vanishing gradient problem, as opposed to earlier networks. Depending on the number of layers, many types of ResNet have been constructed, beginning with 34 layers and progressing to 1202 layers. ResNet's original innovation is using the bypass pathway concept; the next image shows Resnet101 and where it is employed in this study.

![Distribution of imbalanced fundus retina dataset](https://via.placeholder.com/150)

2) **DENSNET**

DenseNet: DenseNet was proposed to tackle the vanishing gradient problem in the same manner as ResNet and the Highway network [51]. One disadvantage of ResNet is that it conserves information by preserving individual changes, despite numerous levels providing a scarcity of information. Furthermore, ResNet contains many of them because each layer has its own set of weights. In an improved attempt to address this problem, DenseNet used cross-layer connections [52]. It employed a feed-forward strategy to connect each layer to the rest of the network's layers. As a result, the feature maps from the prior levels were used as input into all subsequent layers.

3) **EFFICIENTNETb0**

Two new families of deep neural networks known as EfficientNet have been developed [53]. Efficientnetb0: Efficientnetb0 is a convolutional neural network trained on over one million photos from the ImageNet collection.
EfficientNetb0 delivers an EfficientNet-b0 model network trained on the ImageNet dataset.

**V. Experimental Outcomes and Discussion**

The suggested algorithm was evaluated using images from the dataset and real photographs received from the hospital [35]. As depicted in **Figure 2**, we initially fragmented the dataset into training and testing datasets in a 90:10 ratio, respectively.

**A. Data collection and processing**

We experimented with our imbalanced fundus images datasets to see how well the methods above performed. We tried data-level approach for the training dataset before applying algorithm-level approach to the original imbalanced dataset. The result shows that it is difficult to classify fundus images with only data-level approach as shown in **Figure (4) “resnet101 with oversampling and data augmentation”**. The results also revealed that using only algorithm-level techniques to handle the imbalanced images collection is inadequate as shown in **Figure (5) “resnet101 with focal loss”**. After Applying data combination of data augmentation and weight cross-entropy (CE) loss the results improve the performance of the minority class classifier, as shown in **Figure (6) “resnet101 with data augmentation and weight cross-entropy.”**

**B. Detection-based Techniques**

In the diabetic retinopathy literature, detection and diagnosis are frequently used interchangeably. However, detection is described differently in computer vision since it is a process that involves locating an object. Depending on the image's shape, a bounding box such as a rectangle, square, or contour or interactive colors heat maps may be employed to accomplish localization. This section introduced two explainer modules, LIME and Grad-CAM that we have selected to compare to detect Diabetic Retinopathy that automatically explores infected fundus retina regions.

1) **LIME**

LIME (local interpretable model-agnostic explanations) [54] aims to balance interpretability and model fidelity by minimizing the following equation.

$$\varepsilon(x) = \min_{g \in G} L(f, g, \pi_x) + \Omega(g)$$

2) **GRAD-CAM**

Grad-CAM (gradient class activation mapping) [55] is a generalized version of CAM. In CAM, the method is designed for a specific type of CNN architecture where the global average pooling layer directly feeds into a softmax layer. In Grad-CAM, any convolution layer can be examined by first calculating the gradient using backpropagation then using global average pooling to assign weights to each feature map output in that layer.

$$a_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A^k_{ij}}$$

In the above equation, Z numbering of pixels, $$y^c$$ is class score finally $$A^k_{ij}$$ is the feature map activation of feature map k. The feature map outputs can now be weighted and summed before passing through a ReLU function. Only positive contributions to the class are presented when the ReLU function is used. The next **Figure 8** demonstration is generated using the open-source Grad-CAM implementation's default parameters.
FIGURE 5. Resnet101 with focal loss and data augmentation.

FIGURE 6. Resnet101 with data augmentation and weight cross-entropy.

FIGURE 7. LIME explainable AI to interpret the classification.

FIGURE 8. Grad-CAM explainable AI to interpret the classification
The fundus images were used as inputs to CNNs models to classify between seven categories considering the pathological conditions of diabetic retinopathy. The current popular optimizers, such as adaptive moment estimation (Adam) with a learning rate of 0.001 is used to train the network through time for 100 epochs and with the size of the minimum training batches set to 15. The Adam optimizer is an adaptive learning rate method based on stochastic gradient descent that iteratively optimize the gradient vector for each weight vector of the network during the training process. This gradient vector can deduce the error difference in case the weight vector has been adjusted by a particular value. This process is known as “backpropagation”. Regarding the computed gradients, the weights should be updated in the opposite direction, and the new error is calculated. Additionally, adam optimizer can provide a notable advancement on the computational time and performance of deep learning approaches developed for non-linear and complex problems. However, CNNs were trained on the normal cross-entropy loss to backpropagate the errors, minimizing the loss. Using a normal cross-entropy loss with our highly unbalanced dataset will incentivize the models to prioritize the majority class, because it contributes more to the loss. Simply put, the majority class will dominate the loss. Therefore, training the deep learning models with a uniformly balanced dataset is preferred so that the seven classes of each training class would have an equal contribution to the loss. Therefore, CNNs with experimental results that using those data-level and algorithm-level methods in the deep learning training process can result in good performance on imbalanced multi-class fundus retina images datasets that it was used to handle the highly unbalanced dataset. The first approach (Approach 1) was oversampling and data augmentation, oversampling which duplicated samples from the minority class to achieve equal distribution with the majority class. The normal class was the majority here; therefore, we oversampled mild (or early) NPDR to (168 images), moderate NPDR to (168 images), severe NPDR to (168 images), very severe NPDR to (168 images), PDR to (168 images), by randomly duplicating samples until they reached a quantity equal to the normal class, which helped to balance the distribution of different labels within each class. However, oversampling tends to be error-prone due to overfitting or added noise [56], therefore, we used data augmentation to effectively overcome class imbalance and help reduce over-fitting and could also improve the stability and classification accuracy at the same time and making the models more robust. Data augmentation is the technique that creates an artificial dataset by modifying the original dataset. It is known as the process of creating multiple copies of the original image with different scales, orientations, locations. We employed an image augmentation technique by randomly rotating the image between \(-30^\circ\) and \(30^\circ\), randomly flipping the image left, right, and randomly resizing the image to increase the number of samples since many authors believe that increasing data is fundamentally important for improving the performance of convolutional neural networks. The second approach (Approach 2) was using focal loss function with data augmentation.

Table 3: Classification results of diabetic retinopathy on the class of the whole image using Approach 1 (oversampling and data augmentation).

<table>
<thead>
<tr>
<th>Deep CNNs</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1-Score</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resnet10 1</td>
<td>85.53 NaN 97.58 NaN 74.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EfficientNetb0</td>
<td>94.74 92.70 98.99 92.7 4 92.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Densenet 201</td>
<td>90.79 90.40 98.15 90.1 4 90.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Classification results of diabetic retinopathy on the class of the whole image using Approach 2 (focal loss function with data augmentation).

<table>
<thead>
<tr>
<th>Deep CNNs</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1-Score</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resnet10 1</td>
<td>84.21 NaN 97.27 NaN 74.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EfficientNetb0</td>
<td>89.47 85.73 97.84 86.8 8 88.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Densenet 201</td>
<td>88.16 NaN 97.89 NaN 79.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Classification results of diabetic retinopathy on the class of the whole image using Approach 3 (focal loss function, data augmentation, and oversampling).

<table>
<thead>
<tr>
<th>Deep CNNs</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1-Score</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resnet10 1</td>
<td>80.26 76.11 95.93 79.4 7 76.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EfficientNetb0</td>
<td>89.47 89.53 97.97 88.3 7 88.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Densenet 201</td>
<td>85.53 81.33 97.06 83.5 0 80.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Classification results of diabetic retinopathy on the class of the whole image using Approach 4 with Weighted Loss.

<table>
<thead>
<tr>
<th>Deep CNNs</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1-Score</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resnet10 1</td>
<td>94.74 93.90 98.99 93.6 5 94.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EfficientNetb0</td>
<td>93.42 90.86 98.71 91.0 6 91.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Densenet 201</td>
<td>94.74 92.70 98.91 93.4 4 93.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
augmentation, there are two hyperparameters in focal loss function, namely, \( \alpha \) and \( \gamma \). In this work, we simply set \( \alpha = 0.05 \) and \( \gamma = 2 \) because it worked the best as in [57]. The third approach (Approach 3) was using the focal loss function with data augmentation and oversampling. The final approach (Approach 4) was the substitution of the pre-trained network’s loss with the weighted loss, which helped to balance the distribution of seven classes within each class. Each CNN was fine-tuned with different approaches to handle the unbalanced dataset. CNN models with each approach were evaluated on the testing dataset using the five-performance metrics: Accuracy, Sensitivity, Specificity, Precision and F1-Score. The classification results of Approach 1 are tabulated in Tables 3. The table exhibited that EfficientNetb0 yielded highest accuracy of 94.74% followed by Densenet201 and Resnet101 performed poorly, obtaining an accuracy of 85.53%. However, study of the literature shows that EfficientNet [58] models have achieved superior performance in natural and medical computer vision tasks, as compared to other models, in terms of accuracy, efficiency, and computational complexity. Hence, we decided to use the Approach 4 with resnet101. The classification results of with and without oversampling dataset with focal loss function and data augmentation are tabulated in Tables 4 and 5, respectively. The tables exhibited that with and without oversampling dataset produced relatively similar results. Additionally, most CNN models using the focal loss function and data augmentation generated slightly lower accuracy. However, this may be precise that focal loss function tends to produce a vanishing gradient during backpropagation. A limitation of focal loss is that its learning gradient becomes significantly smaller than that of the original cross-entropy function when the predicted output from the classification layer prematurely approaches the actual output. This introduces a “vanishing gradient” effect, that dramatically slows down the training of the network [59]. Table 6 presents the classification results of Approach 4, fine-tuning CNNs with the weighted loss. The empirical results show that using weighted loss marginally improved the overall accuracy of every CNN, with weight balancing on the dataset, substantially improves diabetic retinopathy prediction, by re-weighting each class in the loss function, a class that represents an under-represented class will receive a larger weight. From these empirical results, we found Approach 2 and Approach 3 produced lower accuracy. The approaches which involved increasing the number of fundus images performed better than other approaches 2 and 3. The Approach 4 classifier is the top four models achieved the highest accuracy, 94.74 percent, 94.74 percent, and 93.42 percent, respectively. To improve the understanding of the prediction made by CNN and visualize the features selected by it, we implemented Gradient-weighted Class Activation Maps or GradCAM. (Grad-CAM) method creates an activation map that highlights the crucial areas. In the Grad-CAM method, the gradients of the layers flowing into the final convolutional layer produce a rough localization map in which the important areas are highlighted. Grad-CAM uses the gradient information flowing into the last convolutional layer to assign significance values to each neuron which responds to class-specific information in the image.

VI. CONCLUSION

This study demonstrates how to eliminate the class bias caused by imbalanced fundus retina image datasets using both data- and algorithm-level strategies. Furthermore, it uses three newly developed classifiers to diagnose diabetic retinopathy, which proved effective and efficient. We used the Densenet, Resnet, and EfficientNet deep neural network families with balancing the class by utilizing (i) weight balancing with data augmentation; (ii) oversampling with data augmentation; (iii) focal loss with data augmentation and (iv) hybrid method of oversampling with the focal loss with data augmentation to build an expert system that can properly and rapidly detect fundus images by improving the deep neural network performance of fundus retina image classification using the imbalanced dataset. Our findings might be useful in understanding the impact of imbalanced datasets with deep learning classifiers for Diabetic Retinopathy and how to mitigate them. Deep learning enables more accurate diagnosis and treatment; ophthalmologists still need to improve performance, interpretability, and trustworthiness.

Declarations

Conflict of interest None.

Ethical considerations This study was performed by public dataset on reference [35].

Funding

No funding received.

Availability of data and materials

Data at this website

https://zenodo.org/record/4647952#.YGNjXVUzbIU

Author Contributions Statement

Aboul Ella Hassanien: Conceptualization, review & editing, Supervision.

Kamel K. Mohammed: Methodology, Software.

Rania. A. Mohamed: Visualization, Writing – original draft.

Ashraf Darwish: Validation, Data curation
REFERENCES


