Journal of Ambient Intelligence and Humanized Computing (2021) 12:9825–9839
https://doi.org/10.1007/s12652-020-02727-z

ORIGINAL RESEARCH

Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification

Jyostna Devi Bodapati[1] · **Nagur Shareef Shaik**[1] · **Veeranjaneyulu Naralasetti**[2]

Received: 26 March 2020 / Accepted: 19 November 2020 / Published online: 2 January 2021

Abstract

Diabetic Retinopathy (DR) is a micro vascular complication caused by long-term diabetes mellitus. Unidentified diabetic retinopathy leads to permanent blindness. Early identification of this disease requires frequent complex diagnostic procedure which is expensive and time consuming. In this article, we propose a composite deep neural network architecture with gated-attention mechanism for automated diagnosis of diabetic retinopathy. The feature descriptors obtained from multiple pre-trained deep Convolutional Neural Networks (CNNs) are used to represent color fundus retinal images. Spatial pooling methods are introduced to get the reduced versions of these representations without loosing much information. The proposed composite DNN learns independently from each of these reduced representations through different channels and contributes to improving the model generalization. In addition, model also includes gated attention blocks which allows the model to emphasize more on lesion portions of the retinal images while reduced attention to the non-lesion regions. Our experiments on APTOS-2019 Kaggle blindness detection challenge reveal that, the proposed approach leads to improved performance when compared to the existing best models. Our empirical studies also reveal that, the proposed approach leads to more generalised predictions with multi-modal representations when compared to those of uni-modal representations. The proposed composite deep neural network model recorded an accuracy of 82.54% (↑ 2%), and a Kappa score of 79 (↑ 9 points) for diabetic retinopathy severity level prediction.

Keywords Diabetic retinopathy (DR) · Pre-trained convolutional neural network · Retinal fundus images · Spatial pooling · Composite deep Neural network · Multi-modal features · Gated-attention · Transfer learning

1 Introduction

Diabetic retinopathy (DR) is a serious disease that affects eyes of diabetic patients (Cheung et al. 2008). It is caused by damage to the blood vessels that nourish light-sensitive tissue called retina. It can lead to permanent blindness in humans if DR is not detected in the early stages. Conventional methods of identifying the presence of retinopathy includes manual observation of Optical Coherence Tomography (OCT) or color fundus images of retina, which requires expert ophthalmologists as it involves in using complex grading systems. Furthermore, it is error-prone and time consuming since it is done manually. According to recent statistics (Stitt et al. 2016), about 382 million people are suffering from DR, by 2013 and this figure could grow to 592 million by 2025. It is possible to prevent many people from going blind, if DR is detected in the early stages. Several scientific studies reveal that the presence of various lesions such as Micro Aneurysms (MA), Exudates (HE), Hemorrhages (HM), and Cotton Wool Spots (CWS) in the color fundus images of retina are symptoms of diabetic retinopathy (Bhandary and Rao 2018). Microaneurysms are bulges of retinal blood vessels which appear as sharp edged red-spots. Exudates are white or yellowish-white spots produced by protein leakage from thin retinal vessels located in the retinal region. Hemorrhages are deposits that are formed

✉ Jyostna Devi Bodapati
jyostna.bodapati82@gmail.com

Nagur Shareef Shaik
shaiknagurshareef6@gmail.com

Veeranjaneyulu Naralasetti
veeru2006n@gmail.com

[1] Department of Computer Science and Engineering, Vignan's Foundation for Science Technology and Research, Guntur 522213, India

[2] Department of Information Technology, Vignan's Foundation for Science Technology and Research, Guntur 522213, India

by thin and poor blood vessel leakage and appear as red dots with non-uniform margins. Figure 1 reflects multiple lesions detected in the retinal scan of the patient affected by diabetic retinopathy.

Diabetic Retinopathy can be categorized into Non-Proliferative (NPDR) and Proliferative (PDR) levels. NPDR can be further classified into Mild, Moderate and Severe phases depending on the formation of lesions (Gulshan et al. 2016). The early stage at which Micro Aneurysms are produced is Mild DR. As the disease progresses to a moderate level, blood vessel swelling is seen, resulting in blurred vision. Abnormal growth of blood vessels is observed during the severe stage. Proliferative DR is the advanced stage of DR during which large retinal breaks and retinal detachment takes place, leading to complete vision loss. (Williams et al. 2004). Figure 2 depicts color fundus images of retina affected with various severity levels of retinopathy.

In previous studies, various machine learning based solutions for automated DR diagnosis were proposed. Most of these approaches focused on detecting the presence of various lesions from color fundus retinal images. Habib et al. applied Gaussian matched filter on color fundus images to extract features and passed them to tree ensemble classifier to detect Micro Aneurysms on retinal surfaces (Habib et al. 2017). Otsu's threshold based method is applied on retinal fundus images to extract Hemorrhages features and fed to random forest classifier for detection of haemorrhages (Kaur et al. 2016). Shengchun Long et al. introduced a method to automate the Hard Exudate segmentation process using Support Vector Machine (SVM) and dynamic threshold based methods (Long et al. 2019). Haloi et al. used SVM-based classifiers to find retinal surface cotton wool spots that help the system to identify the severity level of the disease (Haloi et al. 2015).

Most of these traditional approaches show low generalisation, when compared to neural network models. Another set of techniques were developed using deep learning algorithms for recognizing the severity level of diabetic retinopathy. Noushin Eftekheri (Noushin et al. 2019) introduced a deep learning-based approach to spot MA in the given retinal scan images of diabetic patients. Another deep learning based approach is proposed by Van Grinsven et al. (Grinsven et al. 2016) for detecting hemorrhages. They introduced a sampling method in their study to speed up the CNN training with the aid of DR as an application. Srivastava et al. introduced a bounding box method to identify the region of interest in the retinal image (Srivastava et al. 2017). A deep neural based method is introduced in Haloi (2015), to find

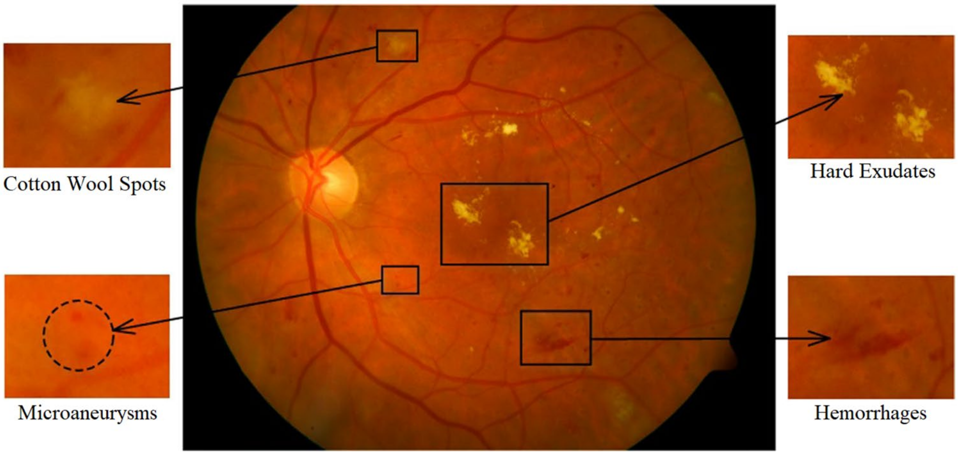


Fig. 1 Retina colour fundus image depicting various levels of lesions developed in retinopathy retina affected patients (Long et al. 2020)

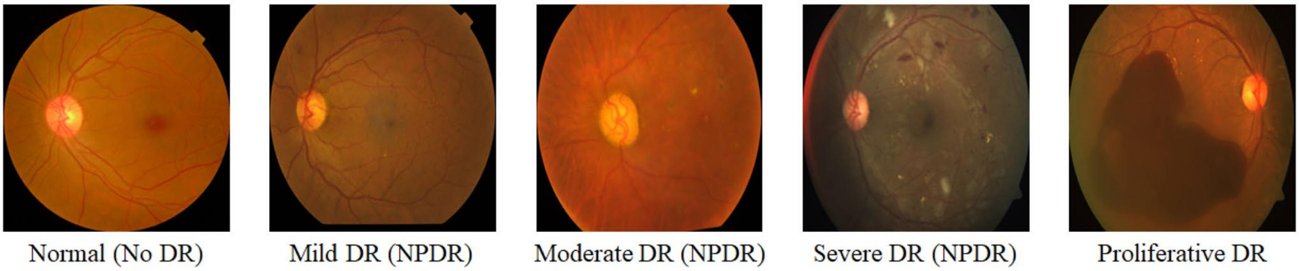


Fig. 2 Retinal color fundus images from APTOS 2019 Kaggle Challenge illustrating various severity levels of diabetic retinopathy ("APTOS 2019 Blindness Detection Challenge", n.d.)

MA, and a max-out activation method is implemented to boost the model's performance.

In this article, we propose a composite deep neural network architecture with a gated attention mechanism to classify the severity level of diabetic retinopathy. Initially, all color fundus retinal images were pre-processed and transferred to pre-trained deep convolutional networks such as VGG16, VGG19, Xception and Inception-ResNetV2 to extract initial deep feature representations of the fundus images. Processing these high-dimensional descriptors is computationally intensive and hence we propose a collection of spatial pooling strategies that generate compact representation. This spatial pooling eliminates noisy and redundant features from the data and produces a reduced feature representations that are useful for discriminating severity levels of DR. Our idea is to improve efficiency of the DR classification model by providing multiple such representations of the DR images. Accomplishing this, a composite deep neural network architecture is implemented, which embraces multiple feature representations of color fundus retinal images and can learn to predict more generalizable predictions from both sets of representations in parallel. The proposed composite neural network model can be trained in an end-to-end manner. Each feature descriptor of the input representation provided to the model may not contribute equally to the severity classification of the DR images. Providing greater emphasis on lesion-representing descriptors and less focus on the rest may lead to progress in the DR severity level predictions. Another major contribution of this work is, introducing a gated attention block in our proposed architecture, enables the model to stress more on lesion-representing descriptors when learning from the DR image representations. By giving adequate focus on the affected regions of the fundus retinal images, this gated attention mechanism contributes to producing more accurate predictions. Experimental studies on the benchmark APTOS 2019 challenge dataset show that, compared to the current models, the proposed composite deep neural network model produces superior in terms various performance measures. Furthermore, the proposed model trained on multi-modal deep features generates more generalised predictions compared to the models trained on uni-modal features. The efficacy of the proposed model is evaluated by adjusting the severity levels of the dataset to 2 levels and 5 levels, in order to report fairer comparison with existing models.

Major contributions of the proposed work includes:

- Using multi-modal feature representations of fundus retinal images, extracted from penultimate layers of different pre-trained deep convolutional networks
- Applying spatial Pooling to obtain reduced representations of the high-dimensional representations
- Building a composite deep neural network architecture that can learn in parallel from multiple feature representations
- Introducing gated attention block that enables the model to give the lesion regions more emphasis compared to the rest of the retinal image

2 Related work

The development of Computer Aided Diagnostic Systems (CAD) for the diagnosis of automatic diabetic retinopathy has been an active area of research in recent years. Several solutions based on machine learning and deep learning have been proposed in literature. In this section, we provide a brief overview of various approaches proposed for diabetic retinopathy detection and severity classification.

2.1 Conventional machine learning models for diabetic retinopathy detection

Most of the conventional approaches to DR severity prediction uses traditional machine learning models trained using the feature descriptors obtained from retinal images using algorithms that requires domain knowledge of the input data.

A hybrid ensemble classifier based on the Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) was used to assess the existence of lesions in the retina (Akram et al. 2013). A Similar approach has been used to enhance the model's classification accuracy by using shape augmented feature set along with the intensity features (Akram et al. 2014). As a step forward for these approaches, a model was implemented that uses genetic algorithm based feature representations of the retinal images (Welikala et al. 2015). In order to distinguish lesions from non-lesions in the given retinal images, various shallow classifiers such as Gaussian Mixture Model (GMM), K-Nearest Neighbors (KNN), Support Vector MAchine (SVM) and AdaBoost were implemented and their performances were analysed. (Roychowdhury et al. 2013). A hybrid feature extraction based approach has been proposed in which retinal images were processed for extraction of irregular signs, such as the area of hard exudates, the area of blood vessels, bifurcation points, texture and entropy (Mookiah et al. 2013). Efficacy of Various feature engineering techniques like Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP) and Gist was studied for fundus image classification (Dondeti et al. 2020).

Performance of the DR prediction techniques discussed in this section so far is sub-optimal since they use conventional models for classification that may not be sufficient for discriminating against complex real time data such as lesion and non-lesion images. Furthermore, techniques used in these methods for feature engineering require domain awareness of the input data.

2.2 Deep learning models for diabetic retinopathy severity detection

With recent developments (Bodapati and Veeranjaneyulu 2019) in artificial intelligence and pattern recognition in the field of medical image processing, deep learning models have become popular for providing new state-of-the-art results for various challenging tasks such as diabetic retinopathy severity classification (Jyostna Devi et al. 2020), brain tumor prediction (Deepika et al. 2019) and many other tasks. A novel convolutional neural network model was implemented with Siamese-like architecture that can accept two fundus images corresponding to the left eye and right eye at the input layers. Then the model transmits the input images to the Siamese-like block and learns from their correlation to detect diabetic retinopathy. In addition, the model adopts the same architecture as Inception V3 and uses two weight sharing blocks to extract high-dimensional characteristics of binocular fundus images. This model is computationally intensive and expects retinal images from both the eyes and does not fit well with the single retinal images for the prediction of DR (Zeng et al. 2019). Fuzzy C-Means clustering (FCM) is used along with deep CNN, which is trained on more than a million of retinal images for the task of DR severity identification (Dutta et al. 2018). This model shows that deep CNNs are capable of quantifying the features as blood vessels, fluid drip, exudates, hemorrhages and micro aneurysms into different classes. Statistical features derived from unprocessed high-dimensional RGB images, and edge based features are extracted from gray-level versions of the original images. A large deep CNN with 18 layers along with 3 fully connected layers was introduced to classify high-resolution color fundus images of retina into 3 one of the three severity levels of DR (Shaban et al. 2020). Their studies show that simple pre-processing and class-specific data augmentation, small kernel sizes, and dropout lead to improving the performance of the model trained with imbalanced data. Most of the recent DR diagnosis literature focused on the implementation and training of deep CNNs with large volumes of data. Although few of these models were trained on small sample sizes, data augmentation techniques such as rotation, scaling, shear etc. were widely used to enhance the original datasets. In order to achieve good efficiency, even with datasets that were originally small, these augmented datasets were used to train the deep models (Porter et al. 2019).

2.3 Transfer learning aproaches for diabetic retinopathy severity identification

Achieving reasonable performance even with small sample sizes become feasible with the introduction of transfer learning. This new paradigm makes it possible to transfer the knowledge gained by a model trained on large sample sizes to other models when performing tasks close to those of the original models.

Since 2013, various deep CNN architectures have been implemented and outperformed the earlier traditional machine learning models and occupied the leader board of Large Scale Visual Recognition Challenge (ILSVRC). Visual Geometric Group (VGG), GoogLeNet, ResNet, Xception, InceptionResNetV2 are, to name a few, are the winners of ILSRVC and are extremely complex and computationally intensive, and are trained on ImageNet dataset (Deng et al. 2009) containing millions of images of 1000 different classes. Significant work has been done to transfer the knowledge acquired by these pre-trained deep CNNs while implementing the DR severity prediction models. Muhammad Mateen et al. propsoed an approach that extracts deep features from VGG19 and applies Principal Component Analysis (PCA) and Singular Value Decomposition (SVD) on those high-dimensional features, before applying classification (Mateen et al. 2019). In similar lines, feature representations from Neural Architecture Search Network (NASNet) are used to represent fundus images and are ultimately projected to a non-linear space using t-distributed stochastic neighbour embedding (t-SNE) transformation. This study demonstrates that two features in the reduced non-linear space can effectively represent the retinal images compared to those in the deep feature space (Dondeti et al. 2020). Compared to those using dense layer descriptors, descriptors from the convolutionary layers of of Xception proved to be more effective for representing retinal images with diabetics retinopathy (Kassani et al. 2019). Their methodological studies emphasize that, aggregation of features derived from multiple layers of pre-trained models boost the performance of the multi-layer perceptron used for classification (Fig. 3).

Another significant contribution to retinopathy prediction tasks is the integration of feature representations from two different pre-trained networks, such as VGG16 and Xception (Bodapati et al. 2020). A set of approaches are proposed to blend the features derived from two distinct networks and a DNN is used for DR detection and severity prediction. The authors contend that such blended representations of fundus images boost the performance of classification models as each representation derives different descriptors and when used together, each set complements the other. Based on the recent works in the literature it is evident that the deep descriptors facilitate better representations of retinal images

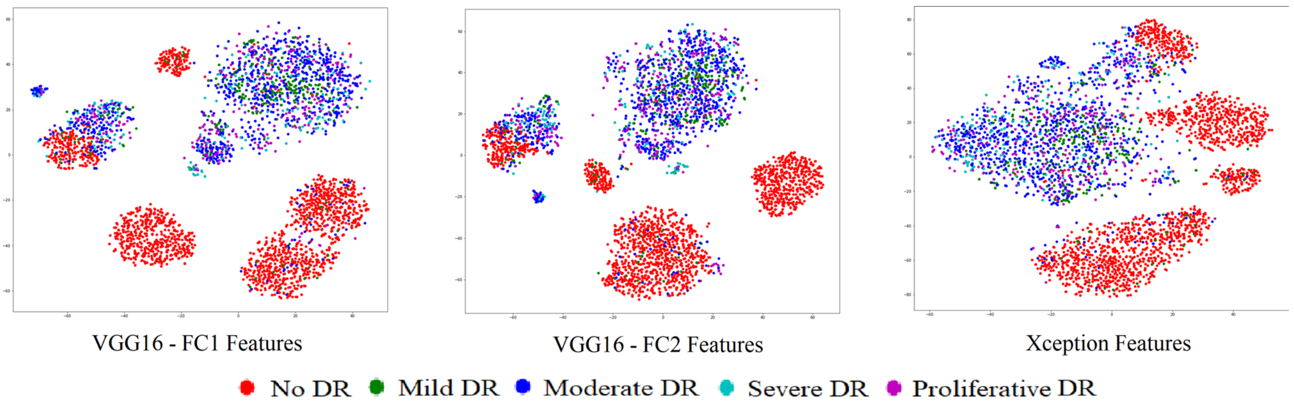


Fig. 3 t-SNE plots for features extracted from penultimate layers of VGG16 and Xception pre-trained networks

and contribute to strong generalization. There is plenty of space to further optimise the efficiency of the DR model by taking several deep descriptors into consideration and process them methodically using aggregation methods.

2.3.1 Pre-trained convNets for feature extraction

Brief overview of the widely used pre-trained deep CNNs:

- *Visual Geometric Group (VGG 16)* (Simonyan and Zisserman 2014): VGG16 is proposed by K. Simonyan and A. Zisserman and has been widely used for various visual recognition problems such as image recognition and object recognition. The architecture, uses 3×3 filters with same padding and stride 1 at each convolution layer and 2×2 non-overlapping filters with stride 2, and same padding. The simple homogeneous structure of VGG enables a deeper model and is crucial for its success. The network, with approximately 138 million parameters is designed to be fairly large with 16 layers, including the 2 fully connected soft-max layers each of size 4096 at the end. Consecutive convolution layers of the model allows the model to train more complex non-linearity from the input images.
- *Extreme version of Inception (Xception)* (Chollet 2017): Xception is another deep ConvNet introduced by Google in 2017 as an extended version of Inception. The model with 71 layers relies solely on depth-wise separable convolutions. Standard convolutions look simultaneously at both channel and spatial correlations. Unlike standard convolutions, depth-wise separable convolutions look at channel and spatial correlations independently in successive steps. The 36 convolutional layers of Xception architecture, organised into 14 modules, seem capable of learning richer representations with fewer parameters, forming the network's feature extraction base.

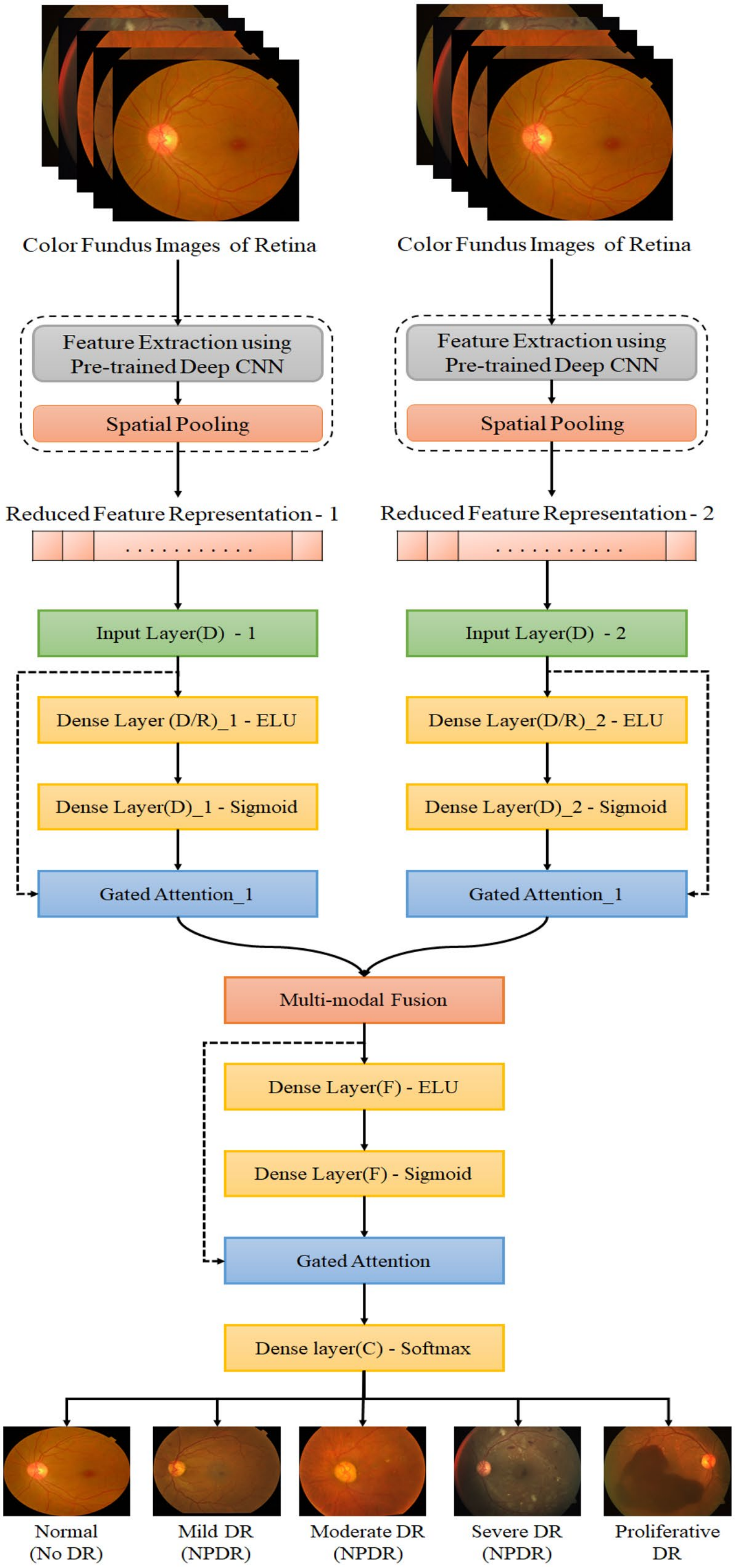
Different from these models, our proposed model derives feature representations from multiple deep CNNs to represent retinal images and propose aggregation methods.

3 Proposed methodology

The purpose of this work is to extract a universal representation of retinal images that ultimately leads to the model trained on them being improved. We propose to use the integration of multiple representations of the retinal images obtained from Xception and VGG16. The reduced representations of this integrated representation is eventually transferred as input to a composite deep neural network model that can accept multiple representations of the input data at separate receiving ends. The proposed composite network is designed with gated attention blocks that allows to emphasize the input descriptors representing retinal image lesion portions while providing less emphasis on others. It is anticipated that the proposed method would produce more generalized predictions for the task of severity classification of diabetic retinopathy even with limited datasets and minimum computing resources.

Another significant contribution to this work is, the novel composite DNN architecture that can be trained jointly in an end to end manner. This model is capable of learning from multiple input channels in parallel using gated convolution blocks. The layers close to the final classification layer integrates the outputs from multiple channels and learn discriminating information required for DR prediction/classification. The proposed approach is simple in terms of number of parameters and yet leads to improved performance and generalized predictions without any overheads involved for pre-processing the input images. Architectural details of the proposed composite deep neural network with gated attention mechanism is depicted in Fig. 4. This section includes

Fig. 4 Architectural details of the proposed Composite Deep Neural Network with Gated Attention Mechanism



more detailed descriptions of the proposed approach, along with the functionality of each module.

3.1 Multi-modal high-dimensional retinal image representations

During this phase, two different pre-trained models like VGG16 and Xception are used to extract most significant features of color fundus images. The original images are reshaped according to the fixed input dimensions, depending on the pre-trained model used.

Initially each input image X_i is passed through the Xception architecture G_1 and the VGG architecture G_2. Then, three different feature representations ($F_1(X_i)$, $F_2(X_i)$, and $F_3(X_i)$) of the input image are collected. $F_1(X_i) \in R^{d_1}$, $F_2(X_i) \in R^{d_2}$ and $F_3(X_i) \in R^{d_3}$ represent the feature descriptors obtained from Xception, first fully connected layer of VGG and second fully connected layer of VGG respectively. The high-dimensional descriptors $F_2(X_i)$ and $F_3(X_i)$ are processed using spatial pooling operations to obtain reduced representations of the retinal images. Figure 3 shows the t-distributed Stochastic Neighbor Embedding (t-SNE) based projections of $F_1(X_i)$, $F_2(X_i)$ and $F_3(X_i)$ for every i^{th} image X_i of the training dataset.

3.2 Spatial pooling for reduced retinal image representation

The raw features ($F_1(X_i)$, $F_2(X_i)$ and $F_3(X_i)$) obtained from dense layers of deep pre-trained models are high-dimensional in nature and may lead to model over fitting when used for model training without any processing. It also possible that these high-dimensional features might include noisy and redundant features. The proposed Spatial pooling is applied on descriptors $F_2(X_i)$ and $F_3(X_i)$ independently to get a reduced versions of them. Let each $F_l(X_i)$ is a vector of d_l descriptors where $F_l(X_i) = [F_l^1(X_i), F_l^2(X_i), ..., F_l^{d_l}(X_i)]$.

Spatial pooling block receives input F_l of dimension d_l and produces an output of dimension $d_{l/4}$. The proposed pooling operation requires two hyper parameters, their spatial extent t, and the stride value S. Let the output produced by the spatial pooling as $F_l(X_i)' = [F_l^1(X_i)', F_l^2(X_i)', ..., F_l^{d_l}(X_i)']$ and is computed as:

$$F_l^k(X_i)' = \frac{1}{t}\sum_j F_l^{k+j}(X_i^p) \tag{1}$$

In Eq. 1, when p=1 then the operation works similar to average pooling and when p value tends to infinity then that refers to max pooling operation. Three different representations ($F_1(X_i) \in R^{d_1}$, $F_2^{'}(X_i) \in R^{d_2}$ and $F_3^{'}(X_i) \in R^{d_3}$) of retinal image X_i are available by the end of this spatial pooling operation. Where the unprocessed descriptors obtained from Xception are $F_1(X_i)$, while $F_2^{'}(X_i)$ and $F_3^{'}(X_i)$ correspond to the reduced versions of $F_2(X_i)$ and $F_3(X_i)$ respectively after spatial pooling is applied. The reduced representations of $F_2^{'}(X_i)$ and $F_3^{'}(X_i)$ are concatenated together to get, ($F_{23}(X_i)$), a d_2 dimensional representation of the retinal images.

3.3 Composite deep neural network

The aim of this work is to devise an effective representation of retinal images for the task of diabetic retinopathy severity prediction. It has been found, based on the literature, that it is often beneficial to use multiple image representations as the information from each will complement the other. We adopt the same for our work and introduce a model that can handle multiple such representations.

We introduce a composite deep neural network model that can handle multiple feature representations that are produced by the earlier stages of the proposed approach. The composite DNN model is implemented with two-channels which can accept two different representations of the retinal representations, process them, effectively learn the lesion information present in the images. The model can be extended to multiple channels with out any loss of generality. The implemented DNN model can be trained jointly in an end-to-end manner. The model accepts multiple sets of representations of retinal images X as input. Let $X^1, X^2 ... X^k$ be the k representations of the input image. The model is trained such that it minimizes the following loss function:

$$\mathcal{L}(\subseteq) = *argmin_\theta \frac{1}{n}\sum_{i=1}^{n} L(Y_i, f(Y_i^1, Y_i^2 \dots Y_i^k|\theta)) \tag{2}$$

Here in Eq. 2, n refers to size of the training dataset $\mathcal{D}$, where θ refers to the set of parameters to be optimized by the model, $X_i^1, X_i^2 ... X_i^k$ be the k different feature descriptors corresponding to the ith image and y_i be the associated label. $L(p_i, q_i)$ is the loss suitable for multi-class data and is computed as:

$$L(p_i, q_i) = -\sum_{k=1}^{C} p_{ik} \; log \; q_{ik} \tag{3}$$

Here in Eq. 3, p_i represent the one hot vector vector representation of the ith image and q_i is the predicted scores produced at the output layer of the model when the ith image is presented to the model. Similarly, p_{ik} and q_{ik} represent the actual label and the predicted scores by the model respectively for the ith training sample w.r.t. the k^{th} class.

In addition to accepting input from multiple channels, the proposed DNN model is also designed to have gated attention blocks. Three layers of the network together are designed to act as a gated attention block. This attention block allows the model to emphasis on descriptors representing lesion portions while giving less emphasis on those that representing non-lesion portions of the retinal images. Multiple attention blocks, one for each channel and one block between the output layer and merge layer of the model, are used in the proposed model. The significance of the positioning of the attention blocks is: the block placed with each channel allows the model to learn in parallel from each of the retinal representations that are presented at each channel. Whereas the block placed after the merge layer allows the model to learn from the merged representation and allows to learn discrimination information from the merged representations.

After the final attention block, a soft-max layer is added at the end to enable the model to generate the scores for the severity level of diabetic retinopathy. Though the model is simple it generalizes well, as it includes attention blocks as well as it can learn from multiple input data representations through different channels in parallel. The proposed architecture of the model along with details of the attention block is depicted in the Fig. 4. Detailed working of each stage of the proposed model is described using pseudo code in Algorithm 1.

3.4 Gated attention mechanism

The proposed model is designed such that it pays high attention to specific features which in turn benefit the classification model in discriminating the retinal images. Unlike the conventional neural network models that treat each neuron output with equal importance, our proposed attention block allows the DNN model to learn better and contributes to optimizing the the model's efficiency as it treats each input descriptor with distinct emphasis. The model initially captures non-linearity from the input to accomplish this, then learns a non-mutually-exclusive relationship between the inputs by emphasizing specific features that represent lesion portions. A ReLU layers is used to achieve the former objective and the sigmoid layer is used to achieve the latter, which in turn serves as a gating mechanism in the model. The mathematical formulation of the proposed gating attention mechanism can be expressed as:

$$Z = \sigma(g(F, W_1) = \sigma(\delta(F, W_1), W_2)) \tag{4}$$

$$\hat{F} = Z \circ F \tag{5}$$

Algorithm 1: Composite Deep Neural Network with Gated Attention Mechanism

```
Input: Let the given fundus image dataset is partitioned into train, $D^{Tr}$, and test, $D^{Tst}$, splits where
        $D^{Tr} = \{(X_i, Y_i)_{i=1}^{N_{Tr}}\}$ and $D^{Tst} = \{(X_i)_{i=1}^{N_{Tst}}\}$. $X_i$ represents $i^{th}$ color fundus image in the dataset and $Y_i$
        refers to the severity level of DR associated with $X_i$. In case of DR identification task $Y_i \in \{0, 1\}$ whereas in
        the case of DR severity classification task $Y_i \in \{0, 1, 2, 3, 4\}$.
Output: Label $Y_i$ for each $X_i \in D^{Tst}$
For each sample $X_i \in D^{Tr}$, Step1, Step2 and Step3 are repeated
Step1: Generate different resized versions of image $X_i$ in the dataset so that they are compatible with various deep
  pre-trained models.
Step2: Feature Extraction
    For each resized image $X_i$ three different feature representations $(F_1(X_i), F_2(X_i),$ and $F_3(X_i))$ are extracted.
      $F_1(X_i) \in \mathcal{R}^{d_1} \leftarrow$ Feature representations from the penultimate layer of Xception network
      $F_2(X_i) \in \mathcal{R}^{d_2} \leftarrow$ Feature representations from the first fully connected layer (fc1) of VGG16
      $F_3(X_i) \in \mathcal{R}^{d_3} \leftarrow$ Feature representations from the second fully connected layer (fc2) of VGG16
Step3: Spatial Pooling
A spatial pooling operation is applied on the high-dimensional representations, $F_2(X_i)$ and $F_3(X_i)$ to get reduced
  versions.
Each $k^{th}$ descriptor $F_1^k(X_i)'$ corresponding to $k^{th}$ descriptor of $F_1^k(X_i)$ is computed as: $F_1^k(X_i)' = \frac{1}{t}\sum_j F_1^{k+j}(X_i)$
Each $k^{th}$ descriptor $F_2^k(X_i)'$ corresponding to $k^{th}$ descriptor of $F_2^k(X_i)$ is computed as: $F_2^k(X_i)' = \frac{1}{t}\sum_j F_2^{k+j}(X_i)$
Step4: Composite Model Training
   Let $X_i^1, X_i^2, ...X_i^k$ be the k different representations corresponding to the $i^{th}$ retinal image of the original fundus
  image dataset. Now let the dataset is formulated with multiple representations of $X_i$. Now
  $D^{Tr} = \{(\hat{X}_i^1, \hat{X}_i^2, ...\hat{X}_i^k Y_i)\}_{i=1}^{N_{Tr}}$
   Train the proposed composite DNN with gated attention model on $D^{Tr}$
Step5: Model Inference
   While inference stage, each retinal image is processed through step1 to step3 to get multiple representations. Now
  $D^{Tst} = \{(\hat{X}_i^1, \hat{X}_i^2, ...\hat{X}_i^k Y_i)\}_{i=1}^{N_{Tst}}$
   Evaluate the performance of $D^{Tst}$ using the composite DNN trained in Step4
```

In Eq. 5, $\hat{F}$ represents the reduced representation of the retinal image after applying attention. Similarly, in Eq. 4, δ and σ represents Rectified Linear Unit (ReLU) and sigmoid activation functions respectively, while Z refers to the attention scores computed by the gated attention block. The gating mechanism is parameterised by two dense layers with parameters $W_1 \in R^{\frac{d}{r}}$ and $W_2 \in R^d$, where d refers to the dimension of the vector F and r is reduction ratio. The first dense layer, bottleneck layer, of the attention block serves as a dimension-reduction layer and captures non-linearity from the input using ReLU activation function. Second dense layer serves as an up sampling layer and learns non-mutually-exclusive relationship among the inputs using sigmoid activation function. The output neuron values, (Z), of the sigmoid layer, serves as attention scores. Final attention feature vector ($\hat{F}$), is computed by applying point wise multiplication of input feature vector (F) and the attention scores (Z).

The attention block introduced in the model allows it to learn better and contributes to enhancing the precision of DR severity predictions as it treats the neuron outputs of the previous layers with varying significance. This makes our model far superior to the standard neural network models that treat all neuron outputs of the previous layers without any variation.

4 Experimental results

In this section, a detailed discussion of the experimental studies that we have conducted to prove the effectiveness of proposed composite neural network with gated attention approach for DR Severity level classification.

4.1 DR identification vs DR severity prediction

The proposed model can be used to detect the presence of DR in the given retinal images and, if present, it can be extended to reveal the level of severity of DR as well. Throughout this section, task1 refers to the detection of the presence of DR while task2 refers to the identification of severity level if the presence of the DR is observed.

Table 1 Distribution of fundus images to different severity levels of retinopathy; 2 Stage label for DR Identification; 5 Stage label for DR Severity Classification

Severity level	#Images	5 Stages	2 Stages
Normal stage (No DR)	1805	0	0
Mild stage (NPDR)	370	1	1
Moderate stage (NPDR)	999	2	
Sever stage (NPDR)	193	3	
Proliferative stage (PDR)	295	4	
Total	3662	–	–

4.2 Summary of the dataset

In this work, we use Asia Pacific Tele-Ophthalmology Society (APTOS) dataset that is released as part of Kaggle blindness detection challenge - 2019 ("APTOS 2019 Blindness Detection Challenge", n.d.). This dataset contains 3661 high resolution color fundus images captured by different types of clinical cameras under laboratory conditions. All these images are labelled on a scale of 0–4 (0—No DR, 1—Mild, 2—Moderate, 3—Severe, 4—Proliferative DR) representing the severity levels of diabetic retinopathy. Table 1 gives distribution of color fundus images over different levels of severity. Figure 2 shows sample images from APTOS-2019 Kaggle challenge representing potential severity levels of diabetic retinopathy.

The original dataset comes with five levels of severity and is appropriate for the task of severity classification. For our convenience, to make the dataset suitable for DR identification task, the images of Mild, Moderate, Severe and Proliferative (0—No DR, 1—Mild + Moderate + Severe + Proliferative) stages are united under a single category and are relabelled with the same label. As the dataset does not have any train/test splits, following the standard hold-out method for validation, we choose 80% of the data for training and the remaining 20% of the data is preserved for testing.

4.3 Performance evaluation measures

We use various standard evaluation metrics like accuracy, precision, recall, F1-Score and AUC to evaluate the effectiveness of the proposed model. Besides all these we also use another popular metric named kappa as it is appropriate for evaluating the performance of imbalanced data. Accuracy measure indicates the fraction of correctly predicted samples and is very popular for the classification tasks. As accuracy may mislead the model performance, other metrics such as precision, recall and F1-Score are used. Precision is the proportion of positive predictions from actual true positives where as recall is the proportion of actual positives, predicted as positives. Harmonic mean of precision and recall is referred as F1-score. Accuracy, precision,recall and f1 score can be computed by using the following expressions.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{6}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{7}$$

Table 2 Performance of machine learning models (task1) trained using the representations from the first fully connected layer of VGG16

Model	Accuracy	Precision	Recall	F1	Kappa
LR	96.18	96	96	96	92.36
k-NN	95.63	96	96	96	91.26
DT	90.45	90	90	90	80.91
SVM-linear	96.04	96	96	96	92.07
SVM-RBF	95.77	96	96	96	91.54
MLP	96.18	96	96	96	92.36
DNN	97.27	97	98	97	95.12

Table 3 Performance of machine learning models (task1) trained using the representations from the second fully connected layer of VGG16

Model	Accuracy	Precision	Recall	F1	Kappa
LR	97.13	97	97	97	94.27
k-NN	95.36	96	95	95	90.73
DT	91.27	91	91	91	82.52
SVM-linear	96.58	97	97	97	93.17
SVM-RBF	96.86	97	97	97	93.73
MLP	96.45	97	96	96	92.91
DNN	97.32	98	98	98	94.63

$$Recall = \frac{TP}{(TP + FN)} \tag{8}$$

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{9}$$

In the above expressions from Eq. 6–9 TP refers to the number of true positive cases while FP refers to the number of false positive. Similarly, TN refers to the number of true negatives, and FN represents false negatives. In addition, we also use Kappa Statistic to report efficiency of proposed method. Kappa statistic compares the observed accuracy against the expected accuracy. Kappa Statistic is calculated as:

$$\text{Kappa Score} = \frac{(\text{Observed accuracy} - \text{Expected acuracy})}{(1 - \text{Expected accuracy})} \tag{10}$$

4.4 Experimental studies on uni-modal features for DR detection and severity level prediction

Our initial experiments are intended to verify the efficiency of various traditional machine learning models when a single representation of the retinal images is taken into account for the task of DR identification (task1) and severity prediction (task2). The performance of these models provides a basis for the task under consideration and is further used to understand the level of improvement achieved by the proposed model when compared to these models. Various conventional machine learning models such as Logistic Regression (LR), K Nearest Neighborhood (k-NN), Decision Tree (DT), Support Vector Machine (SVM), Multi Layer Perceptron (MLP) and Deep Neural Network models (DNN) are used as baseline models.

4.4.1 Task1: diabetic retinopathy detection

As the first experiment, the features extracted from first fully connected layer of VGG16 are used to represent retinal images and these feature descriptors are used to train the machine learning models.

From Table 2, one can observe that Linear regression and MLP gives same performance, while DNN outperforms rest of the models with FC1 features of VGG16. Now in the next experiment, the features extracted from second fully connected layer of VGG16 are used to represent retinal images and these feature descriptors are used to train the machine learning models.

From Tables 2 and 3, we can notice that representations from FC2 layer lead to better performance compared to those from FC1 layer for task1. Though accuracy, precision, recall and F1-score are better we can notice that kappa score is low with FC2 features. Using FC2 features, SVM and MLP gives approximately same performance, while Linear regression and DNN outperforms rest of the models with FC2.

Now in this experiment, the features extracted from the penultimate layer of Xception are used to represent retinal images and these feature descriptors are used to train the machine learning models.

Based on these set of experiments using uni-modal features for the task of DR identification each set of features collected from different networks leads to different performances. This gives us the clue that the features learned by each of these models are different and hence leads variation in the performances.

4.4.2 Task2-diabetic retinopathy severity level prediction

This severity prediction task is much complex similar to that of DR identification task as the task is now no more binary, it is a multi-class classification with five classes. Another challenge here is the number of samples available in the higher level of severity are very low and this imbalance in the dataset raises a bigger challenge and hard to deal with.

Similar to the experiments carried out for DR detection task, the same set of experiments are followed for DR severity prediction task also. As a first experiment, the features

Table 4 Performance of machine learning models (task1) trained using the representations from the penultimate layer of Xception Network

Model	Accuracy	Precision	Recall	F1	Kappa
LR	96.45	96	96	96	93
k-NN	95.50	96	95	95	91
DT	87.59	88	88	88	75.17
SVM-linear	96.18	96	96	96	92.36
SVM - RBF	97.40	97	97	97	94.82
MLP	96.00	96	96	96	91.89
DNN	97.41	97	97	97	94.82

Table 5 Performance of machine learning models (task2) trained using the representations from the first fully connected layer of VGG16

Model	Accuracy	Precision	Recall	F1	Kappa
LR	77.9	77	78	77	66.47
k-NN	74.76	74	75	72	60.58
DT	65.35	65	65	65	48.13
SVM-linear	77.22	76	77	76	65.18
SVM-RBF	77.63	77	78	74	64.65
MLP	78.04	77	78	77	66.47
DNN	80.06	80	81	80	70.02

Table 6 Performance of machine learning models (task2) trained using the representations from the second fully connected layer of VGG16

Model	Accuracy	Precision	Recall	F1	Kappa
LR	77.49	77	77	77	65.77
k-NN	75.44	76	75	73	61.49
DT	68.75	69	69	69	52.85
SVM-linear	76.94	76	77	77	65.15
SVM-RBF	77.89	78	78	75	65.12
MLP	77.63	77	78	77	65.63
DNN	79.81	79	80	79	68.88

extracted from first fully connected layer of VGG16 are used to represent retinal images and these feature descriptors are used to train the machine learning models for the task of severity prediction.

From Table 5, one can observe that Linear regression and SVM gives approximately same performance, while MLP is better, and DNN outperforms rest of the models with FC1 features of VGG16. Now in the next experiment, the features extracted from second fully connected layer of VGG16 are used to represent retinal images and these feature descriptors are used to train the machine learning models.

Table 7 Performance of machine learning models (task2) trained using the representations from the the penultimate layer of Xception Network

Model	Accuracy	Precision	Recall	F1	Kappa
LR	76.81	76	77	76	64.95
k-NN	71.49	66	71	67	54.66
DT	60.03	60	60	60	40.33
SVM-linear	75.17	75	75	75	62.47
SVM-RBF	76.81	76	77	72	63.12
MLP	77.90	77	78	77	66.42
DNN	78.99	78	79	78	67.67

From Tables 5 and 6, we can notice that representations from FC2 layer lead to degradation in performance compared to those from FC1 layer for task2. The same trend could be observed with all the evaluations metrics including accuracy. Surprisingly Decision tree based classifier alone leads to an improvement using FC2 features. Linear regression, SVM and MLP produces identical performances without significant deviations. We can notice that DNN is consistently out performing rest of the models irrespective of the feature representations, model and task.

Now in the next experiment, the features extracted from the penultimate layer of Xception are used to represent retinal images and these feature descriptors are used to train the machine learning models.

From Tables 5, 6 and 7, we can notice that representations from FC1 layer representations from VGG16 network lead to better performance compared to those of others for task1. This could be because of the consecutive convolution layers provide more non-linearity and the size of the kernels used in VGG16 allows the model to learn fine details of lesions from the retinal images. DNN significantly better compared to the rest of the models in terms of various performance measures.

Based on these set of experiments using uni-modal features for the task of DR severity prediction each set of features collected from different networks leads to different performances. This gives us the clue that the features learned by each of these models are different and hence leads variation in the performances.

4.5 Experimental studies on proposed composite DNN with multi-modal features for DR detection and severity level prediction

Based on the experience from the experimental studies carried out using uni-modal representations, we conclude that though there is slight variation in the models using feature representations from different networks the variation is not much significant and the performance of these models can

Table 8 Class-wise performance of proposed composite DNN gated attention model trained using features from FC1 and FC2 of VGG16 along with Xception features for task1

Class	Accuracy	Precision	Recall	F1	AUC
DR negative	97.78	98	98	98	98
DR positive	97.84	98	98	98	98
Overall	97.82	98	98	98	98

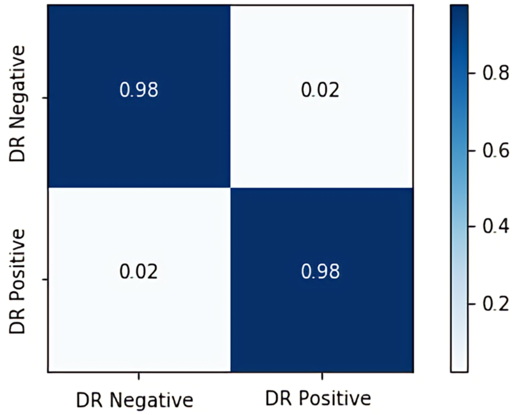


Fig. 5 Confusion matrix of the proposed composite DNN gated attention model for the DR identification task

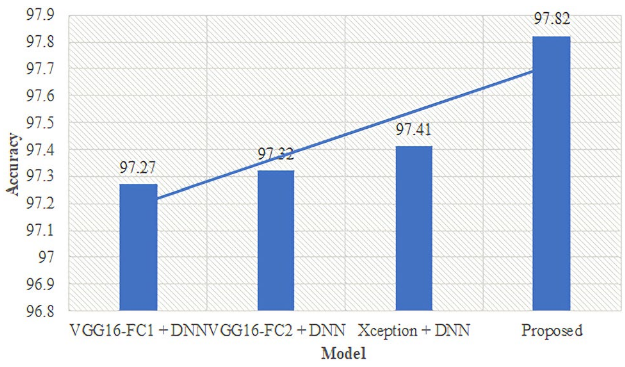


Fig. 6 Comparing the performance of uni-modal features with proposed multi-modal features for DR identification

Table 9 Class-specific performances of the proposed composite DNN Gated attention model trained using features from FC1 and FC2 of VGG16 for DR severity classification

Severity	Accuracy	Precision	Recall	F1	AUC
No DR	96.94	98	97	97	97
Mild	48.65	68	49	57	73
Moderate	90.05	70	91	79	88
Severe	43.59	63	44	52	71
Proliferative	37.29	59	37	46	68
Overall	82.54	82	83	82	79

not be accepted in the real-time environment for the DR severity prediction task.

4.5.1 Task1-diabetic retinopathy detection

The next set of experiments are carried out with an assumption that representations from multiple models leads to better model performance as each set of representations can complement each other and hence allow the model better learn better. Now multiple representations of the retinal images are passed as input to the models for DR prediction or for DR severity classification. The standard ML models used in case of the uni-modal representations can not handle multiple feature representations at the same time in parallel. Hence we use our proposed model which can handle multiple representations of the input data to carry out these set of experiments. For these set of experiments, we use feature representations from both FC1 and FC2 layers of VGG16 along with penultimate layer features from Xception network and are processed as described in proposed work.

In Table 8, class-specific measures obtained for the DR identification are reported. Compared to the results from Tables 2, 3 and 4 the performance of the proposed model is superior in terms of all the metrics used. The model is much stable compared to those trained using uni-modal features and has become robust with negligible mis-classifications. The proposed model is superior not only in terms of accuracy but also in terms of other scores including kappa. Though the proposed model leads to an improvement in accuracy of 0.5%, the model improves in terms of kappa score by more than 3 points when compared to those of the models trained using uni-modal features.

Figure 5 shows the confusion matrix of the proposed model for the task of DR identification. We can notice that only 0.02% of the retinal images are mis-classified, which is fairly acceptable for deploying the model in real-time.

The graphical representation provided in Fig. 6, clearly indicates that the performance of the proposed model contributed to more accurate identification of DR from the given retinal images compared to those trained using uni-modal feature representations.

Based on the series of experiments we have seen so far, one can infer that the proposed model generalizes well for the DR identification task. It is very interesting to verify whether the model is pursuing the similar trends in terms of performance for the DR severity prediction as well.

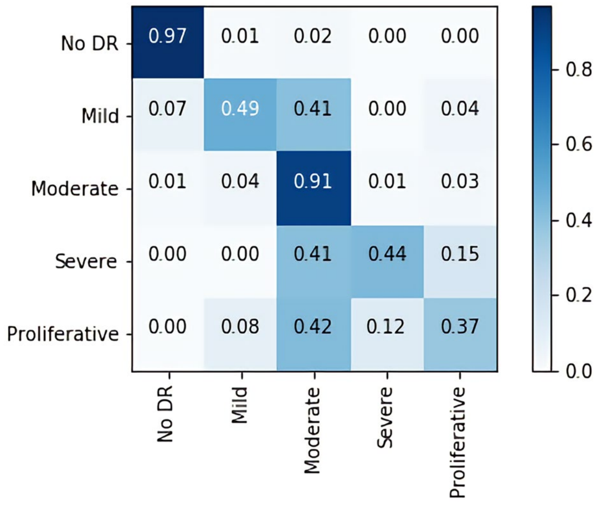


Fig. 7 Confusion Matrix for five-stage DR severity classification using proposed composite DNN with VGG16 and Xception features

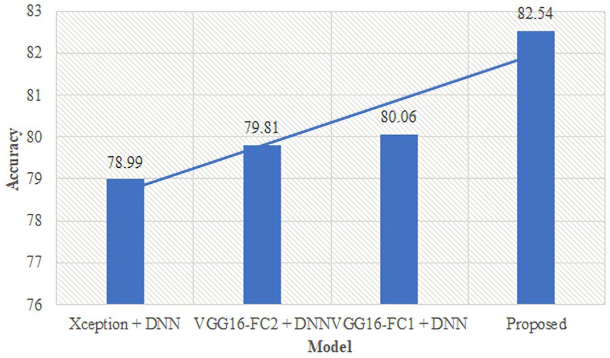


Fig. 8 Comparing the performance of Uni-modal features with proposed multi-modal features for DR severity classification

4.5.2 Task2-diabetic retinopathy severity level prediction

In the next few experiments, we check the potential of the proposed model in terms of generalization for the most challenging DR severity classification task.

Based on the results reported in Table 9, we can notice that the model generalizes well for predicting No-DR samples as well as moderate DR samples. Whereas the model observes significant mis-classifications for the other classes. The reason for this kind of behavior by the model could be attributed to the inadequate samples in those classes. Though there are mis-classifications, the proposed model exhibits superior performance compared to that of those trained using uni-modal features. When compared to the models trained using uni-modal representations the proposed model achieves close to 3% improvement in accuracy (from 80.06 to 82.54%), 1% improvement in terms of precision (from 81 to 82), 2% improvement in terms of recall (from 81 to 83), 2% improvement in terms of F1-score (from 80 to 82), 9% improvement in terms of kappa (from 70.02 to 79). From this we can notice that the proposed model is robust compared to those trained using uni-modal representations especially in terms of kappa scores.

Table 10 Performance comparison of the proposed approach with existing methods for DR severity level prediction

Approach	Accuracy (%)
Xception + MLP (Kassani et al. 2019)	79.59
InceptionV3 + MLP (Kassani et al. 2019)	78.72
MobileNet + MLP (Kassani et al. 2019)	79.01
ResNet50 + MLP (Kassani et al. 2019)	74.64
NASNet + t-SNE + SVM (Dondeti et al. 2020)	77.90
VGG+Xception (DNN) (Bodapati et al. 2020)	80.96
Composite Gated attention DNN (Proposed)	82.54

Figure 7, shows the confusion matrix of the proposed model for the task of DR severity prediction. We can notice that more mis-classifications for the mild, severe and proliferate DR cases. This could be due to over-fitting observed by the model due to the limited retinal images available in those severity levels.

The graphical representation provided in Fig. 8, clearly shows that the performance of the proposed model leads to accurate severity level prediction of DR compared to those trained using uni-modal feature representations.

4.6 Comparative study

Based on the results from our earlier experiments we prove that the proposed model is far superior in terms of various metrics when compared to the models trained using uni-modal representation. The same kind of trend is observed for DR identification as well as DR severity level prediction. It is very much essential to compare the performance of the proposed multi-channel DNN with Gated Attention mechanism with the recent and current best models in order to prove the efficiency of the proposed model.

From the above Table 10, we can observe that the proposed multi-channel DNN with gated attention mechanism is superior compared to those of the deep CNN architectures. Moreover the proposed model avoids pre-processing and accepts raw images without much pre-processing. This helps to save time during the model inference stage which demands real-time response. The model achieved approximately 2% of improvement in accuracy compared to the recent best model and sets the new state of the art scores for the APTOS challenge dataset. One reason for the improvement in performance could be due to the usage of multiple representations of the retinal images which are learned in parallel by the model through separate channels. Moreover the proposed gated attention mechanism allows the model to learn on which portions of the retinal images emphasis has to be given while giving less focus on the remaining portions. Based on the extensive empirical studies presented in this section, we claim that the proposed model is robust with negligible mis-classifications.

5 Conclusion and future scope

The major focus of this work is to devise an efficient model for DR identification and DR severity level prediction. To this end, we propose a multi-channel deep neural network model with gated attention mechanism that helps in improving the efficiency of DR predictions. The proposed model accepts multiple representations and learns from each one through different channels in parallel. In addition, the model comes with multiple attention blocks that allows the model to serve high emphasis on the lesion portions of the retinal images while giving less emphasis on other regions. Efficiency of the proposed model is compared with the recent best models and also multiple machine learning models that handle unique representations of the retinal images. Based on the results from the experimental studies on APTOS 2019 benchmark dataset, we claim that the proposed model helps in achieving better scores compared to the existing models in terms of various performance metrics. The model proposed is simple in terms of model complexity and does not use much pre-processing which in turn reduces inference time and makes the model much suitable for real-time application.

Limitation of the model is, at higher severity levels, it's performance is not acceptable for the task of severity level prediction. Inadequate samples relevant to the extreme severity levels may be one of the reasons for the poor performance in higher severity levels. As an extension of this work we would like to address this issue in future by using various augmentation strategies that help in generating images and promote model generalization by avoiding model over-fitting. In addition, our experiments are constrained to the model with two channels where this can be tested with multiple channels in future.

References

Akram MU, Khalid S, Khan SA (2013) Identification and classification of microaneurysms for early detection of diabetic retinopathy. Pattern Recogn 46(1):107–116

Akram MU, Khalid S, Tariq A, Khan SA, Azam F (2014) Detection and classification of retinal lesions for grading of diabetic retinopathy. Comput Biol Med 45:161–171

APTOS 2019 Blindness Detection Challenge. (n.d.). https://www.kaggle.com/c/aptos2019-blindness-detection. Accessed 30 Dec 2019

Bhandary SV, Rao KA (2018) Automated screening system for retinal health using bi-dimensional empirical mode decomposition and integrated index. Comput Biol Med 75:54–62

Bodapati JD, Veeranjaneyulu N (2019) Facial emotion recognition using deep cnn based features. Int J Innov Technol Explor Eng

Bodapati JD, Veeranjaneyulu N, Shareef SN, Hakak S, Bilal M, Maddikunta PKR, Jo O (2020) Blended multi-modal deep convnet features for diabetic retinopathy severity prediction. Electronics 9(6):914

Casanova R, Saldana S, Chew EY, Danis RP, Greven CM, Ambrosius WT (2014) Application of random forests methods to diabetic retinopathy classification analyses. PLOS One 9(6):e98587

Cheung N, Rogers SL, Donaghue KC, Jenkins AJ, Tikellis G, Wong TY (2008) Retinal arteriolar dilation predicts retinopathy in adolescents with type 1 diabetes. Diabetes Care 31(9):1842–1846

Chollet F (2017) Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1251–1258

Deepika K, Bodapati JD, Srihitha RK (2019) An efficient automatic brain tumor classification using lbp features and svm-based classifier. In: Proceedings of international conference on computational intelligence and data engineering. Springer, pp 163–170

Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE, pp 248–255

Dondeti V, Bodapati JD, Shareef SN, Naralasetti V (2020) Deep convolution features in non-linear embedding space for fundus image classification deep convolution features in non-linear embedding space for fundus image classification. Revue d'Intelligence Artificielle 34(3):307–313

Dutta S, Manideep B, Basha SM, Caytiles RD, Iyengar N (2018) Classification of diabetic retinopathy images by using deep learning models. Int J Grid Distrib Comput 11(1):89–106

Grinsven M, Ginneken B, Hoyng C, Theelen T, Sanchez C (2016) Fast convolution neural network training using selective data sampling. IEEE Trans Med Imaging 35(5):1273–1284

Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Cuadros J et al (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316(22):2402–2410

Habib M, Welikala R, Hoppe A, Owen C, Rudnicka A, Barman S (2017) Detection of microaneurysms in retinal images using an ensemble classifier. Inform Med Unlock 9:44–57

Haloi M (2015) Improved microaneurysm detection using deep neural networks. arXiv preprint arXiv : 1505.04424

Haloi M, Dandapat S, Sinha R (2015) A gaussian scale space approach for exudates detection, classification and severity prediction. arXiv preprint arXiv:1505.00737

Jyostna Devi B, Nagur Shareef S, Veeranjaneyulu N, Mundukur NB (2020) Joint training of two-channel deep neural network for brain tumor classification. Signal Image Video Process

Kassani SH, Kassani PH, Khazaeinezhad R, Wesolowski MJ, Schneider KA, Deters R (2019) Diabetic retinopathy classification using a modified xception architecture. In: 2019 IEEEE international symposium on signal processing and information technology (isspit). IEEE, pp 1-6

Kaur N, Chatterjee S, Acharyya M, Kaur J, Kapoor N, Gupta S (2016) A supervised approach for automated detection of hemorrhages in retinal fundus images. In: 2016 5th international conference on wireless networks and embedded systems (wecon). IEEE, pp 1-5

Long S, Chen J, Hu A, Liu H, Chen Z, Zheng D (2020) Microaneurysms detection in color fundus images using machine learning based on directional local contrast. BioMed Eng OnLine 19:1–23

Long S, Huang X, Chen Z, Pardhan S, Zheng D (2019) Automatic detection of hard exudates in color retinal images using dynamic threshold and svm classification: Algorithm development and evaluation. BioMed research international, 2019

Mateen M, Wen J, Song S, Huang Z et al (2019) Fundus image classification using vgg-19 architecture with pca and svd. Symmetry 11(1):1

Mookiah MRK, Acharya UR, Martis RJ, Chua CK, Lim CM, Ng E, Laude A (2013) Evolutionary algorithm based classifier parameter tuning for automatic diabetic retinopathy grading: a hybrid feature extraction approach. Knowl-based Syst 39:9–22

Noushin E, Pourreza M, Masoudi K, Ghiasi Shirazi E (2019) Microaneurysm detection in fundus images using a two step convolution neural network. Biomed Eng Online 18(1):67

Porter LF, Saptarshi N, Fang Y, Rathi S, Den Hollander AI, De Jong EK, Liloglou T et al (2019) Whole-genome methylation profiling of the retinal pigment epithelium of individuals with age-related macular degeneration reveals differential methylation of the ski, gtf2h4, and tnxb genes. Clin Epigenet 11(1):6

Rahim SS, Jayne C, Palade V, Shuttleworth J (2016) Automatic detection of microaneurysms in colour fundus images for diabetic retinopathy screening. Neural Comput Appl 27(5):1149–1164

Roychowdhury S, Koozekanani DD, Parhi KK (2013) Dream: diabetic retinopathy analysis using machine learning. IEEE J Biomed Health Inform 18(5):1717–1728

Shaban M, Ogur Z, Mahmoud A, Switala A, Shalaby A, Abu Khalifeh H, Sandhu H et al (2020) A convolutional neural network for the screening and staging of diabetic retinopathy. Plos One 15(6):e0233514

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556

Srivastava R, Duan L, Wong DW, Liu J, Wong TY (2017) Detecting retinal microaneurysms and hemorrhages with robustness to the presence of blood vessels. Comput Methods Programs Biomed 138:83–91

Stitt AW, Curtis TM, Chen M, Medina RJ, McKay GJ, Jenkins A, Simo R et al (2016) The progress in understanding and treatment of diabetic retinopathy. Progress Retinal Eye Res 51:156–186

Verma K, Deep P, Ramakrishnan A (2011) Detection and classification of diabetic retinopathy using retinal images. In: 2011 annual IEEE india conference. IEEE, pp 1–6

Welikala RA, Roshan A, Fraz MM, Dehmeshki J, Hoppe A, Tah V, Mann S, Barman SA (2015) Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy. Comput Med Imaging Graph 43:64–77

Welikala R, Dehmeshki J, Hoppe A, Tah V, Mann S, Williamson TH, Barman S (2014) Automated detection of proliferative diabetic retinopathy using a modified line operator and dual classification. Comput Methods Programs Biomed 114(3):247–261

Williams R, Airey M, Baxter H, Forrester J-M, Kennedy-Martin T, Girach A (2004) Epidemiology of diabetic retinopathy and macular oedema: a systematic review. Eye 18(10):963–983

Wu L, Fernandez-Loaiza P, Sauma J, Hernandez-Bogantes E, Masis M (2013) Classification of diabetic retinopathy and diabetic macular edema. World J Diabetes 4(6):290

Zeng X, Chen H, Luo Y, Ye W (2019) Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network. IEEE Access 7:30744–30753

Zoph B, Vasudevan V, Shlens J, Le QV (2018) Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 8697–8710

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.