IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 5, 2022, accepted May 20, 2022, date of publication May 23, 2022, date of current version May 27, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3177651

Texture Attention Network for Diabetic Retinopathy Classification

MOHAMMAD D. ALAHMADI
Department of Software Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah 23890, Saudi Arabia
e-mail: mdalahmadi@uj.edu.sa

ABSTRACT Diabetic Retinopathy (DR) is a disease caused by a high level of glucose in retina vessels. This malicious disease put millions of people around the world at risk for vision loss each year. Being a life-threatening disease, early diagnosis can be an effective step in the treatment and prevention of vision loss. To automate the early diagnosis process, computer-aided diagnosis methods are not only useful in detecting the diabetic signatures but also provide information regarding the diabetic grade for the optometrist to determine an appropriate treatment. Several deep classification models are proposed in the literature to solve the diabetic retinopathy classification task, however, these methods usually lack incorporate an attention mechanism to better encode the semantic dependency and highlight the most important region for boosting the model performance. To overcome these limitations, we propose to incorporate a style and content recalibration mechanism inside the deep neural network to adaptively scale the informative regions for diabetic retinopathy classification. In our proposed method, the input image passes through the encoder module to encode both high-level and semantic features. Next, by utilizing a content and style separation mechanism, we decompose the representational space into a style (e.g., texture features) and content (e.g., semantic and contextual features) representation. The texture attention module takes the style representation and applies a high-pass filter to highlight the texture information while the spatial normalization module uses a convolutional operation to determine the more informative region inside the retinopathy image to detect diabetic signs. Once the attention modules are applied to the representational features, the fusion module combines both features to form a normalized representation for the decoding path. The decoder module in our model performs both diabetic grading and healthy, non-healthy classification tasks. Our experiment on APTOS Kaggle dataset (accuracy 0.85) demonstrates a significant improvement compared to the literature work. This fact reveals the applicability of our method in a real-world scenario.

INDEX TERMS Diabetic retinopathy, deep learning, attention, classification.

I. INTRODUCTION

In the healthcare field, early diagnosis of diseases is a vital step since diseases are more treatable in their early stages. Annually, millions of people around the world suffer from diabetes. Diabetes is a disease that increases the amount of glucose in the blood due to a lack of control over the amount of insulin. According to the International Diabetes Federation [1], 425 million adults in the world are affected by diabetes. If not controlled well, diabetes could damage various parts of the human body including the heart, kidneys, feet, nerves, and eyes [2]. Meanwhile, eyes retina disease (Diabetic Retinopathy (DR)) is extremely sensitive and, if left untreated, can result in vision loss. Figure 1 shows the effects of the DR on the retina vessels.

The retina is a thin tissue of the eye which lines the surface of the back of the eye excluding the area of the optic nerve. The retina contains light-sensitive cells which receive and transfer the light through neural signals and coordinate with the brain to process visual information. Like other organs in the human body, the retina receives its nourishment through blood vessels. If the blood sugar (glucose) level is high, it will cause DR, which will block the tiny blood vessels that nourish the retina, cut off its blood supply, and eyes will try to grow new blood vessels, but they won't develop well and will start to weaken. In other words, DR causes the blood vessels of the retina to swell, leak fluid, or bleed, which often leads to vision impairment or blindness [4]. DR causes 2.6% of

The associate editor coordinating the review of this manuscript and approving it for publication was Charalambos Poullis.

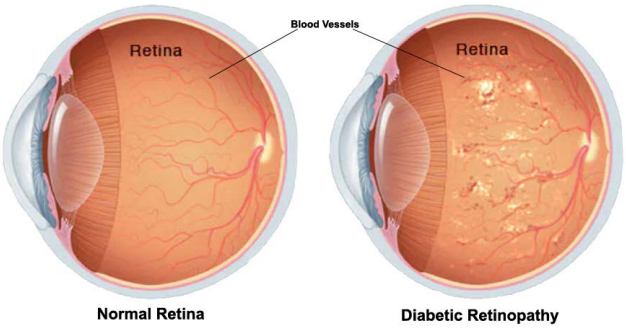

FIGURE 1. Effects of diabetic retinopathy on the retina vessels [3]. According to the Figure, in the early stages of the disease, the walls of the retina blood vessels are weakened. The action protrudes tiny bulges from the vessel walls, may leak or ooze fluid and blood into the retina and cause issues in the retina swell, producing white spots in the retina. As DR progresses, new blood vessels may grow and threaten human's vision.

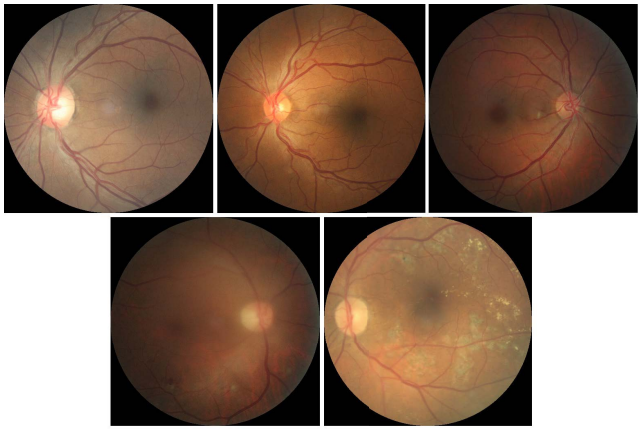

FIGURE 2. Sample of diabetic retinopathy images with stage zero (healthy) to four (proliferative diabetic retinopathy) starting from top left to bottom right. Samples from [6].

blindness worldwide [4] and it is the most prevalent microvascular complication among patients with diabetes mellitus [5]. Figure 2 shows the sample of diabetic retinopathy images for the different grade levels.

Diabetic retinopathy is a progressive eye disease classified into two types and four stages. The two types are non-proliferative (NPDR) and proliferative (PDR). NPDR refers to the early stages of the disease and is characterized by lesions such as microaneurysms (MAs) and exudates, whereas PDR is an advanced form of the disease, indicated by neovascularization of weak blood vessels. Bedsides, the four stages of diabetic retinopathy show the evolution cycle and the intensity of DR. These stages are:

- **Mild nonproliferative diabetic retinopathy:** It is the earliest stage of DR. Tiny areas of swelling in the blood vessels of the retina, microaneurysms, is the characteristic of this stage. In this stage, it is possible that a small portion of fluid leaks into the retina and triggers swelling of the macula.
- **Moderate nonproliferative diabetic retinopathy**: In this stage, nourishment cannot reach the retina due to the swelled blood vessels and blockage of the ways for blood to reach the retina. This process accumulates blood and other fluids in the macula.
- **Severe nonproliferative diabetic retinopathy:** At this stage, the number of blocked blood vessels increases and causes a considerable reduction in blood flow. At this point, new blood vessels start to grow in the retina.
- **Proliferative diabetic retinopathy:** This stage is considered the most dangerous because a large number of fragile blood vessels have formed in the retina. At this stage, there is a possibility of fluid leakage and visual disturbances such as blurriness, reduced field of vision, and even blindness at any time.

The risk of DR-induced vision loss is increasing every year. In this regard, early diagnosis of the disease in the early stages can be an effective step in the treatment and prevention of the disease. By investigating some types of retina lesions, such as microaneurysms (MA), hemorrhages (HM), soft and hard exudates (EX), it is possible to detect DR by medical experts. However, it is not always possible to do this due to a lack of expertise or the high cost of this process. Furthermore, in many cases, due to human errors and parameters such as fatigue, normal methods of examining medical images by human resources are not very accurate.

The use of automated DR identification methods not only gives many people the benefits of early diagnosis but also reduces costs, saves time, and increases the accuracy of the diagnosis. Numerous studies in recent years have shown that Computer-aided diagnosis (CAD) methods are extremely useful in medical image processing [7]. In this field, advanced machine learning-based methods have been proposed to automatically segment and classification of retina images. Segmentation and classification methods process images taken from the retina and identify areas of the disease and detect the stage of the disease. This process allows ophthalmologists to focus directly on the disease areas and apply appropriate treatments to fight the disease. Machine learning techniques try to extract and utilize retina features such as optic disk detection, vessel enhancement, and lesion segmentation from the original input image. Then, classification techniques are utilized to categorize the images such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naive Bayes [8]–[10]. These typical machine learning methods use specific handcrafted features to learn discriminative patterns from the image itself. As a result, these methods lack to learn the complex structure of the pattern and usually are deficient to model the underline structure of the abnormality. Consequently, the applicability of these engineering methods in a clinical domain is limited.

Recently, deep learning (DL) techniques, especially convolutional neural networks (CNN), has seen numerous advancements in medical image processing methods [11]–[18]. Numerous studies have been conducted in recent years for DR grade classification. K. Xu *et al*. [19] proposed a CNN-based

method for classifying the images to the normal and DR images. During the preprocessing step, they utilized data augmentation techniques, resizing images, and normalization. Their model includes eight CONV layers, four max-pooling layers, two FC layers, and a SoftMax function at the last layer to create a binary class. Li *et al*. [16] employed a Deep CNN (DCNN) for classifying DR images. More specifically, they utilized fractional max-pooling to extract more discriminative features from the input data and classified the obtained features using SVM. R. Pires *et al*. [20] proposed a 16 layers CNN model to classify DR images into the referable and non-referable classes and used drop-out and L2 regularization techniques to avoid overfitting during the training process. Sungheetha *et al*. [21] proposed a method to classify the diabetic condition of a retina image into five classes: No DR, Mild DR, Moderate, Severe, and Proliferative DR. The condition of the diabetic retinopathy were detected by analyzing the Hard Execute spotted in the blood vessel of an eye using a CNN model.

Although the proposed DL methods boosted the diabetic retinopathy grading performance, these methods still lack to model the local contextual dependency inside the representation space to recognize the diabetic retinopathy level. To address this limitation, we propose to incorporate the attention mechanism in both texture and semantic levels. In our design, the representation space is decomposed into style and content features in which we perform two parallel attention to attenuate more informative texture and spatial features. By fusion of both normalized features and following the decoding path, the model produces a grading label for each retina image. The contribution of this paper can be summarized as follows:

- Incorporating attention mechanisms to adaptively highlight texture information, which plays a significant role in recognizing diabetic retinopathy
- Style content decomposition module to separate texture and semantic representation
- State-of-the-art (SOTA) results on the public dataset for diabetic retinopathy classification

The rest of the paper is organized as follows. Section 2 reviews related work. The proposed network is presented in Section 3. The experimental results are described in Section 4. Finally, Section 5 concludes the paper.

II. RELATED WORK

Diabetic retinopathy classification plays an important role in the diagnosis of DR disease and prevents blindness by early detection of the disease. Similar to other research lines in the computer vision field, diabetic retinopathy classification approaches can be categorized into handcrafted (engineering features based) and DL-based approaches. The handcrafted methods focus on designing the specific feature to learn discriminative patterns from the the image itself, while DL methods can learn and make intelligent decisions on their own and discover hidden patterns inside the input data that explicitly do not exist. In the following, we will review some of the proposed methods for each of the handcrafted and DL methods.

A. HANDCRAFTED APPROACHES

Akram *et al*. [22] mixed the structure of the support vector machines and Gaussian Mixture Model (GMM) to identify and classify microaneurysms in the retina for early detection of diabetic retinopathy. Furthermore, they improved their method via enriching the feature set with shape, intensity, and statistics of the affected region in their follow-up study [23]. Roychowdhury *et al*. [24] proposed a two-step hierarchical classification approach that uses a computer-aided screening system (DREAM) for classifying the severity grade of DR. They investigated varying illumination and fields of view for generating the severity grade.

Zhang *et al*. [25] utilized a three-step method for DR screening. First, they apply a series of normalization and denoising steps to detect reflections and artifacts in the image. Then, the images are segmented using a mathematical morphology operation. Finally, the images classify into the lesion and non-lesion areas using a binary random forest classifier. The random forest classifier may fail in some cases where images do not have many distinctive features. Adal *et al*. [26] classify the DR severity grade by calculating the absolute difference between two-time points of the extreme's multiscale blobness responses of fundus images and applying SVM and K-nearest Neighbour (KNN) classification algorithms. The approaches based on the handcrafted features need a heuristic feature extraction stage and this made these methods challenging and the results less satisfying. In complex tasks, manual feature extraction may not work well.

Thus, alternative methods are needed in which they do not require a manual feature extraction process and be able to discover hidden patterns within the data itself. To this end, DL models were proposed with the capability to automatically extract features, reveal hidden patterns in the data, and handle large amounts of data, which outperformed typical handcrafted methods significantly.

B. DEEP LEARNING APPROACHES

DL methods, specially CNN, have made great strides in recent years and have been able to successfully perform several complex tasks. The hierarchical learning capability and extraction of high-level features of CNNs has made them much more powerful than methods that solely work with typical raw image features. Various architectures of CNN have been introduced in recent years, some state-of-the-art architectures include but not limit to: Fully Convolutional Neural Network (FCN) [27], U-Net [28], SegNet [29], hourglass [30], and DeepLab [31]. In the following, we review some recent methods in the field of diabetic retinopathy classification using CNNs.

Quellec *et al*. [32] proposed a method based on a heat map optimization scheme for identifying DR. They employed a back-propagation-based CNN for image-level

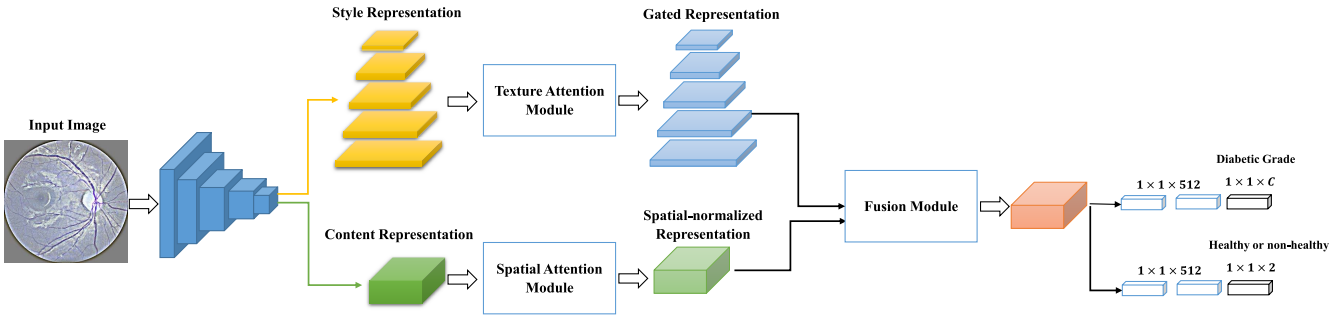

FIGURE 3. Structure of the proposed method for diabetic retinopathy classification. The proposed method decomposes the representational space into content and style features and then it applies texture attention and spatial attention modules to re-calibrate the feature sets. Finally, by fusion of both features, it recognizes the diabetic grade along with the healthy and non-healthy retina.

classification to automatically detect lesions in retinal images. Zang *et al.* [33] utilized a three-level classification method based on CNN to fulfill a DR classification. The first classifier determines whether the DR is referable or non-referable. Then, the second classifier classifies the eye as non-DR, non-proliferative DR (NPDR), or proliferative DR (PDR). Finally, the third classifier separates the case to no DR, mild and moderate NPDR, severe NPDR, and PDR. Kassani *et al.* [34] employed Xception model [35] concept and introduced a new model for classifying DR by inserting a deep layer aggregation that receives multilevel features from diverse convolutional layers. Then, a multi-layer perception (MLP) classifies these different features. Jain *et al.* [36] utilized different data augmentation techniques for balancing the input data during the preprocessing stage. Furthermore, they trained three different classifier networks including VGG16, VGG19, and InceptionV3 [37], [38]. These models are trained for both binary and 5-class DR classification. Based on the evaluation results, they have shown that the VGG19 network, which contains a large number of convolutional and pooling layers, was more efficient than the other models. Mateen *et al.* [39] fine-tuned a pre-trained VGG19 model bypassing input data through its layers to extract its features and classify DR. They utilize Principal Component analysis (PCA) and singular value decomposition (SVD) techniques to reduce the feature dimension and avoid overfitting during the model training stage.

Dai *et al.* [40] proposed a method for detecting microaneurysms from fundus images by integrating an image-to-text mapping scheme with a multi-sieving CNN framework. Using this approach, they handled one of the major challenges in retina image classification, which is that the percentage of relevant information (microaneurysms that are critical for ophthalmologists) in the retina images is lower than irrelevant information. The image-to-text mechanism is used as a clinical report. Alryalat *et al.* [10] presented a two-stage DL model for retina segmentation and predicting response to intravitreal anti-VEGF injections among Diabetic Macular Edema (DME) patients. They first utilized an attention-based U-Net for the segmentation task, then they passed the segmentation map through a classifier network to determine whether the patient would response to the anti-VEGF injection or not (a binary-classification task). Jaskari *et al.* [41] addressed the uncertainty challenges in the clinical application by developing a Bayesian-based classification method to model the underlying uncertainty in grading the diabetic retinopathy through the retina images. The evaluation results showed that using entropy uncertainty estimation improved the within-distribution uncertainty performance. Zia *et al.* [42] utilized an Inception-V3 with VGG network to distinguish the key precursors of Dimensionality Reduction. After extracting features from input data, they used an entropy concept to select the most discriminating features. Their model is capable of highlighting the veins, liquid dribble, exudates, hemorrhages, and miniaturized scale aneurysms in the input retina images.

One of the central limitations of the previous work on recognizing the diabetic retinopathy grade is the lack to model the hidden structure that exists in the texture of the retina images. To adaptively highlight these types of hidden information inside the representational space, we propose to incorporate the textual and spatial attentions mechanism on top of the network bottleneck. In the next section, we will present our method in more detail.

III. PROPOSED METHOD

Automatic diabetic retinopathy detection provides an early signal for designing a specific treatment by an expert doctor. Thus, it plays a critical role in the diagnosis and treatment process. To automate this process in an end-to-end manner several DL-based research works are proposed in the literature [5], [41], [42]. As described in the previous section, these methods lack the incorporation of an attention mechanism to highlight more informative regions and patterns that existed inside the retina images for detecting the grade of diabetes from the image itself. To address these challenges, we design a network to learn the intrinsic pattern that existed inside the retina images by utilizing a parallel attention mechanism. The general structure of the proposed method is visualized in Figure 3. In the next subsections, we will discuss each part in more detail.

A. PREPROCESSING

Nowadays with the rapid progress in the development of the imaging system, several clinical imaging devices are produced by different companies. Although these retinopathy

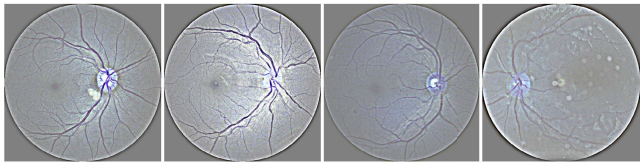

FIGURE 4. The normalized retina images are produced by the normalization step. The normalization process reduces the intensity shift that existed inside the dataset for better generalization performance.

imaging equipments follow the same pipleline to produce the retina images, due to the intrinsic characteristics of these devices the produced images may vary in terms of intensity, colour, and shape. If such variations are not addressed, it can affect the network training process, and consequently, the trained model will be biased towards a specific imaging standard. Furthermore, the input data need to be preprocessed in a way that would be suitable for a deep model. In this regard, similar to the work done in [34], we have added a series of preprocessing steps to prepare the dataset for neural network training. First, based on the input images aspect ratio we resized them to the size of 512*512 pixels using the bicubic interpolation technique. Next, the retina circle location in the resized input images is centred by cropping each image from the centre to a size of 320*320 pixels. Moreover, we utilized Graham [43] approach to enhance the clarity of blood vessels and lesion areas. For this purpose, all the black pixels have been removed from the input images and a min-pooling filtering technique used to normalize the images as [43] (1):

$$Ic = \alpha I + \beta G(\rho) * I + \gamma \tag{1}$$

where the convolution operation is denoted by $*$, the input image represented by I, and $G(\rho)$ marks the Gaussian filter with a standard deviation of ρ. Pre-defined parameters are also used as α, β, and γ. In addition, to achieve uniform distribution across the dataset and terminate feature bias, all the input images' cross channels' intensity values have been normalized to $[-1, 1]$. Figure 4 illustrates the result of these preprocessing stage effects on the input retina images.

B. INCEPTION ENCODER

DL architectures usually consist of two main parts, encoder, and decoder. Encoders are the first part of the network whose task is to encode input data into a format from which the network can extract numerous useful features to reveal existing patterns. In architectures related to the application of machine vision, the encoder section consists of a series of successive convolutional layers followed by the pooling and activation layers to represent the data in a high-level space. In our proposed method, the encoder is presented by the use of inception module. The concept of inception was first introduced by Szegedy *et al.* [44] and later several follow-up versions were made to improve its performance [39], [45], [46]. Unlike the regular CNN networks, the inception block consists of applying several parallel convolution operation to

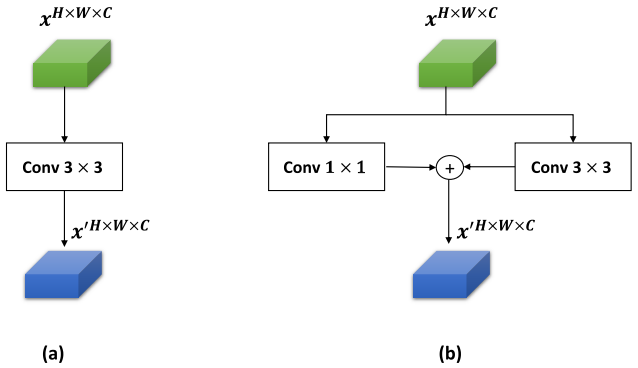

FIGURE 5. Structure of the (a): regular single convolution and (b) Inception module, which is designed to capture multi-scale representation.

encode the object of interest in various scales. Without using an inception block, the multi-scale representation feature map might not be obtainable using the regular convolution layer. Figure 5 shows the inception blocks architecture. According to Figure 5, this block is consisted of one convolutional path with two 3×3 convolutional layers and a short-cut path with a 1×1 convolutional layer which is designed to increase filter depth during encoding, or decrease filter depth during decoding, and ensure the pixel-wise summation by projecting the input feature map into the same space as output. Given the fact that there are multiple inception paths, with different scales, and the output feature maps of the different inception paths are concatenated together, additional parameters would be created for each inception block. Furthermore, in order to increase the performance of the model and make the model focus on specific areas of the image, we have used the inception network as our encoder module. Our encoder network E parametrized with θ which receives the normalized input image Ic and generates the encoded feature $x^{H\times W\times C}$:

$$x = E(\theta; Ic) \tag{2}$$

It is worthwhile to mention that the main idea behind choosing the inception module was its capability in learning a rich and generic representation compared to the counterpart baseline models. Although our proposed method does not rely on any particular baseline model, the reason for choosing the inception model was its better performance throughout our experiments.

C. STYLE AND CONTENT DECOMPOSITION

The deep encoding module usually performs a set of convolution operations followed by the pooling and activation layers to represent the object of interest in hierarchical representation space. This representation space can be divided into style and content features, where the style shows the common representation shared among layers such as colour, texture while the content representation contains more core features like structure, semantic and shape information [47]. In our strategy, on top of the encoder features, we apply the style content decomposition technique to separately perform attention in

each feature set. To perform that, we build a pyramid representation using the feature map derived from each block (the output of the first convolution in each block) of the encoding path. The resulted feature pyramid contains both deep and shallow features to represent the textural information, where we can perform a texture attention mechanism to adaptively recalibrate the most important regions. In the meanwhile, we create the content representation using the output of the last convolution operation on the encoding path. The content representation contains the semantic information, where we can apply the spatial normalization technique to determine important locations for diabetic retinopathy classification. Hence our content and style representation can be achieved by:

$$\begin{aligned} x_{content} &= E_l\left(\theta; Ic\right), \quad l = L \\ x_{style} &= concat(E_l\left(\theta; Ic\right), \quad l = 1, 2, \ldots L) \end{aligned} \tag{3}$$

where L is the number of convolutional block in the encoder module. In our model to ensure the style matching mechanism, we initially train the encoder network using the perceptual loss then the main training part uses the obtained weight to initiate the encoder parameters weight. To this end, a pair of retinopathy images are fed to the encoder module to generate both content and style representation, then similar to [48] by maximizing the correlation between the style of both images and keeping the content representation as same as possible we iteratively adjust the style matching mechanism. Both style and content losses are used to model the perceptual loss:

$$\begin{aligned} &\mathcal{L}_{\text{style}}^{texture}\left(x1_{style}, x2_{style}\right) \\ &\qquad = \sum_{l=0}^{L} w_l \frac{1}{4C_l^2 N_l^2} \sum_j \left(x1_{style}^{l,j} - x2_{style}^{l,j}\right)^2 \end{aligned} \tag{4}$$

where N determine the spatial dimension and C_l stands for the number of channels in the layer l. The content loss can be defined as:

$$\begin{aligned} &\mathcal{L}_{\text{content}}\left(x1_{content}, x2_{content}\right) \\ &\qquad = \frac{1}{2} \sum_j \left(x1_{content_j} - x2_{content_j}\right)^2 \end{aligned} \tag{5}$$

The main objective of the content loss is to keep the representation unchanged as much as possible.

D. ATTENTION MODULE

The idea of the attention mechanism is derived from the real world, where humans seek to focus on specific parts of their vision, such as particular food, road, text, etc., and think about why or how it happens. In the machine vision concept, attention is a technique by which the model can weigh features by the level of their importance, and use this weighting to help achieve the task. In our proposed method, we employ two different attention mechanism on top of the style and content modules in parallel.

1) TEXTURE ATTENTION MODULE

The texture representation in diabetic retinopathy images provides significant information regarding the abnormal regions. Hence, our texture attention mechanism aims to highlight these regions through the frequency domain. To this end, each level of the style pyramid passes through the Laplacian pyramid to modify the frequency information. To model the Laplacian operation, we use the difference of Gaussian operation applied on each level of the pyramid with varying variances. The Gaussian operation to generate different scales can be formulated as:

$$G_l(x_{style}) = x_{style} * \frac{1}{\sigma_l\sqrt{2\pi}} e^{-\frac{i^2+j^2}{2\sigma_l^2}} \tag{6}$$

where x_{style} represents the style feature pyramid, σ_l indicates the variance of the l^{th} Gaussian function, i and j show the spatial location. To highlight high-frequency information (relates to the texture), we simply use the difference of each pyramid level by increasing variance value:

$$LP_l = \begin{cases} G_l - G_{l+1}, & 1 \le l < L \\ G_L, & l = L \end{cases} \tag{7}$$

where LP_l is the l^{th} number of feature maps in the pyramid level, G_l indicates the output of the l^{th} Gaussian operation and L shows the total number of pyramid levels.

2) SPATIAL ATTENTION MODULE

Unlike texture attention which desires to realize 'what' is meaningful in the input image, spatial attention looks 'where' are informative parts of the input images. This spatial attention map generates by using the inter-spatial relationship of features and is complementary to texture attention. We calculate spatial attention by applying pooling operators (average-pooling) on the content feature map channel axis to generate a robust feature descriptor. In our proposed architecture, we generate a spatial attention map $\mathbf{M_s}(\mathbf{x_{content}}) \in \mathbf{R}^{H\times W}$ by utilizing a convolution layer on the features descriptor which emphasize or suppress special parts. The details of this process are described below.

We generate channel information features by $\mathbf{x_{content}}^{\mathbf{s}}_{\text{avg}} \in \mathbb{R}^{1\times H\times W}$, which indicates average-pooled features across the channel. Next, we acquire a 2D spatial attention map by applying a convolving operation on top of the resulted feature map. The following equation shows how attention is computed.

$$\begin{aligned} \mathbf{M_s}(\mathbf{x_{content}}) &= \sigma\left(f^{7\times7}(\text{Avg Pool}(\mathbf{x_{content}}))\right) \\ &= \sigma\left(f^{7\times7}\left(\left[\mathbf{x_{content}}^{\mathbf{s}}_{\mathbf{avg}}\right]\right)\right) \end{aligned} \tag{8}$$

where the sigmoid function is marked by σ, and a convolution operation with the filter size of 7×7 is denoted by $f^{7\times7}$ [49]. By applying spatial attention to the content module, we generate the spatial-normalized representation feature map to guide the network to emphasize more on the informative regions related to the structural information of the diabetic

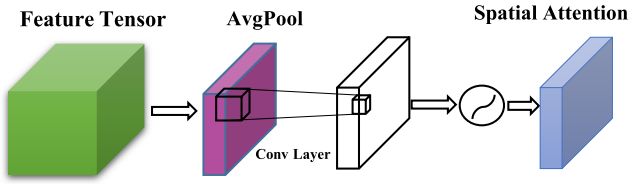

FIGURE 6. The spatial normalization module [49] is utilized in our model to highlight the important location inside the input image to tune the representational space accordingly.

retinopathy patterns. Figure 6 shows the spatial normalization process.

3) FUSION MODULE

The resulted feature maps (Gated representation and spatial-normalized representation) are then concatenated and followed by the convolution operation to form an aggregated feature map which is then fed to the decoder part to produce the classification label.

E. DECODER

For mapping the features vectors obtained from the encoder part to the desired output, we utilize a fully connected layers decoder block. The purpose of the proposed method is depicted in two main goals:

DR classification: classifying the retinopathy images into five classes, according to the DR grade, is the main objective of this study. In this regard, the classification model learns to predict the diabetic retinopathy classes. For calculating loss between the predicted class and the true class, we employed a cross-entropy loss function.

Healthy and non-healthy retina: Since retinal fundus images classification and grading is a complex and costly task, as well as the main purpose of retinopathy detection, is to help the ophthalmologist/hospitals reduce the monitoring burden, the healthy and non-healthy retina binary-classification is intended as an auxiliary task to assist ophthalmologists in recognizing retinopathy. Therefore, this binary annotation can be provided more comfortably. In our structure besides the main classification task, we include this auxiliary task to provide a second signal for the ophthalmologists in recognizing healthy and non-healthy cases. A binary cross-entropy loss function is used for training the auxiliary task model.

IV. EXPERIMENTAL RESULTS

Our proposed method has been evaluated on the APTOS kaggle dataset [6] for diabetic retinopathy classification. Besides the dataset description, this section provides detailed information regarding the training process, evaluation metrics, comparison results for both multiclass diabetic retinopathy classification and healthy or non-healthy retinopathy classification, and finally the ablation study to emphasize the contribution of each proposed module on the model generalization performance. In the next subsections, we will elaborate on each part in more detail.

A. APTOS DATASET

The APTOS dataset [6] is a large collection of retina fundus images prepared using various medical imaging techniques. Aravind Eye Hospital prepared this database for a classification task which is used for developing an automatic approach for detecting and classifying the severity of diabetic retinopathy on a scale from 0 to 4 where the numbers represent the extent of the disease. The dataset consists of a total of 3,662 retina images with a class label for each image that is rated by a clinician according to the severity of the diabetic retinopathy it contains, including No DR (Class0), Mild DR (Class 1), Moderate DR (Class 2), Severe DR (Class 3), Proliferative DR (Class 4). Some samples of the APTOS dataset are illustrated in Figure 2. Given the fact that the classes distribution of this dataset is highly imbalanced, i.e., 49%, 10%, 27%, 5%, and 8% of images belong to normal, mild, moderate, severe, proliferative DR, respectively, we take it into account in the training section with defining weighted loss for each class. Similar to a previous work [34], we use 10% of the labelled samples as a test set and the rest for the training.

B. TRAINING PROCESS

The proposed method is implemented in the Pytorch library and has been carried out on an NVIDIA RTX 3090GPU with a batch size of 8 without any data augmentation. We trained all the models with an initial learning rate $1e-3$ and the decay rate $1e-4$ for 200 epochs using the Adam optimization. In case the validation performance does not change in 10 consecutive epochs, we stop the training process. The baseline network utilized in our experiments has the same structure as the U-Net model without the proposed attention mechanism. It is worth mentioning that during the training process, we do not use transfer learning, instead we train each model using the random weights generated by the standard normal distribution.

C. EVALUATION METRICS

To evaluate our proposed method performance, we have used standard and well-known metrics including accuracy (AC), sensitivity (SE), specificity (SP), F1-Score, and Kappa coefficient. In the following, the terminologies are employed to explain how metrics are calculated.

True-Positive (TP) shows the predicted label that is correctly predicted as a retinopathy class.

False-Positive (FP) shows the predicted label that is falsely predicted as a retinopathy class.

True-Negative (TN) shows the predicted label that is truly labelled as a non-retinopathy pixel.

False-Negative (FN) shows the predicted label that is falsely labelled as a non-retinopathy pixel.

TABLE 1. Performance comparison of the proposed method vs well-know deep classification models on the APTOS dataset [6].

Methods	Accuracy	Sensitivity	Specificity	Kappa
MobileNet [50]	0.790	0.764	0.846	0.770
VGG [51]	0.800	0.853	0.866	0.787
ResNet50 [52]	0.746	0.565	0.857	0.786
Hybrid [53]	0.821	0.852	0.867	-
Modified Xception [34]	0.830	0.882	0.870	-
Multi-scale attention [5]	0.846	0.910	0.905	0.878
Blended VGG+ Xception + DNN [54]	0.809	-	-	-
Composite Gated Attention [55]	0.825	-	-	-
Baseline	0.787	0.636	0.853	0.792
Proposed Method	**0.851**	**0.903**	**0.920**	**0.881**

Accuracy indicates the percentage of correct prediction,

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{9}$$

Specificity indicates the proportion of FP that are correctly identified by model,

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{10}$$

Sensitivity indicates the proportion of predicted TP that are correctly identified by model,

$$\text{Sensitivity (Recall)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{11}$$

F1 score also known as balanced F-score or F-measure, is a weighted average of the precision and recall,

$$\text{F1 score} = \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}} \tag{12}$$

Kappa coefficient indicates the reliability between two raters who each classify N items into C mutually exclusive categories,

$$\text{Kappa} = 1 - \frac{\sum_{i,j} w_{i,j} O_{i,j}}{\sum_{i,j} w_{i,j} E_{i,j}} \tag{13}$$

D. DIABETIC RETINOPATHY CLASSIFICATION RESULTS

To evaluate the performance of the proposed method, we have used the publicly available APTOS dataset. To provide a fair evaluation, we followed the same setting as mentioned in [34] to divide our dataset into train and test sets. In our first evaluation strategy, we applied well-known classification models to classify diabetic retinopathy images. To this end, we slightly modified the classification layer (last fully-connected layer) of the MobileNet, VGG, and Resnet models to produce the classification label for diabetic classes. Both baseline models and our proposed method are trained for 200 epochs using the same training strategy we explained earlier in Section IV-B. We experienced that the results of the baseline models are almost the same as the results mentioned in [34]. Table 1 provides the comparison results.

The experimental results presented in Table 1 show that our approach that uses attention mechanism in texture and content feature outperformed every other model as it enhances the representation space and generate a more rich and generic feature set. Moreover, in comparison with the recent modified

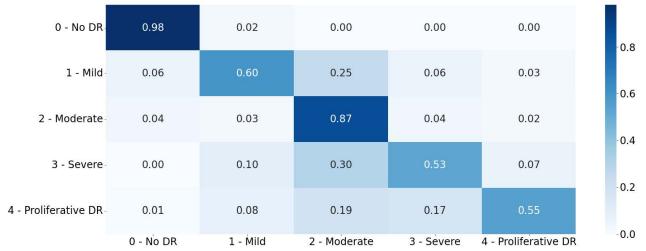

FIGURE 7. Confusion matrix achieved by applying the proposed method on the Kaggle APTOS dataset [6]. The confusion matrix reveals that the method is highly capable of separating healthy and non-healthy retinopathy images from each other, however, in determining the diabetic level there is some miss classification among high grade diabetic classes.

TABLE 2. Performance comparison of the proposed method against the competitive counterparts for healthy and non-healthy diabetic retinopathy classification on the APTOS dataset [6].

Method	F1	Accuracy	Sensitivity	Specificity
MobileNet [50]	0.969	0.963	0.974	0.969
VGG [51]	0.971	0.969	0.974	0.971
ResNet50 [52]	0.965	0.958	0.968	0.962
Multi-scale attention [5]	0.982	0.981	0.983	**0.982**
Baseline	0.968	0.965	0.972	0.969
Proposed Method	**0.984**	**0.985**	**0.985**	**0.981**

Xception [34], Hybrid [53], and the Multi-scale attention [5] methods, our attention-based strategy produces a better classification results. To better analyze the performance of the proposed method for the diabetic classes, we have provided the confusion matrix in Figure 7.

According to Figure 7, the proposed approach can effectively classify the healthy retinopathy images (with 98% confidence) from the diabetic classes. In other words, it provides remarkable classification confidence for separating non-healthy samples from the healthy class. However, the classification performance largely decreases in recognizing the Sever and Proliferative diabetic classes. This is mainly due to the high features similarity that exists among the different diabetic classes that are close to each other (e.g., classes 3 and 4), which makes it extremely cumbersome for the deep model to distinguish them.

E. HEALTHY AND NON-HEALTHY CLASSIFICATION RESULTS

From a clinical perspective, the classification of the healthy and non-healthy retinopathy images not only reduces the burden of the optometrists but also facilitates the screening process. Thus, we provide experimental results regarding the healthy and non-healthy diabetic retinopathy classification problem. First, we summarize and compare the classification results of the proposed method with both baseline and the literature work in Table 2.

According to Table 2, it is obvious that the proposed approach significantly classifies the healthy and non-healthy samples and outperforms the literature work in all metrics. Compared to the recent Multi-scale attention [5], our method slightly produces better classification

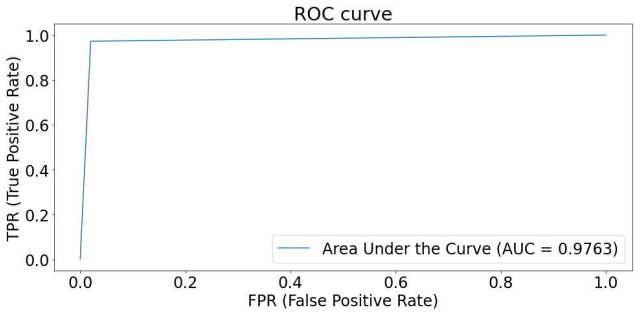

FIGURE 8. ROC curve achieved by applying the proposed method on the Kaggle APTOS dataset [6] for classifying healthy and non-healthy retinopathy images.

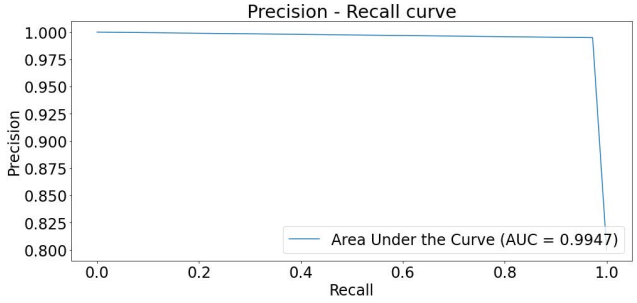

FIGURE 9. Precision-Recall curve achieved by applying the proposed method on the Kaggle APTOS dataset [6] for classifying healthy and non-healthy retinopathy images.

results. In Figure 8 and Figure 9, we provide the ROC and precision-recall curves to analyze the true positive detection vs the false positive rate. The ROC and precision-recall curves demonstrate the trade-off between the sensitivity/specificity and precision/recall metrics. As it can be seen from the curves that our model is highly effective in classifying healthy or non-healthy sample, which is useful in determining the list of patients that need to be check by a specialist.

F. ABLATION STUDY

In this section, the effect of decomposing the representation space into content and style features, as well as the impact of applying texture and spatial attention modules on the model performance are discussed. To investigate the effect of decomposing the representational space into content and style features, the proposed model was trained with and without decomposing the representational space. Instead of decomposition, we simply applied the spatial attention module on top of the bottleneck features and eliminated the parallel attention structure. The obtained results shown in Table 3 demonstrated a performance loss in diabetic retinopathy classification task. In another setting, we evaluated the model performance by simply dropping one attention module and only utilizing the other one to check the contribution of each attention mechanism separately to the model performance. The experimental results demonstrate that each module contributes to the model performance, and all together they provide a powerful features representation for classifying retinopathy, as shown in Table 3. It is worthwhile to mention that our suggested

TABLE 3. Contribution of each proposed module to the model performance. Results are presented using the multi-class diabetic retinopathy classification on the APTOS dataset [6].

Methods	Kappa Score
Baseline	0.792
Baseline+ without decomposition	0.832
Baseline+ only texture attention	0.850
Baseline+ only spatial attention	0.837
(proposed method (baseline+decomposition+both attention))	**0.881**

TABLE 4. Computational complexity of the proposed method.

Method	Inference (T)	Operation (M)	Parameters (M)
Ophthalmologists	80m	-	-
MSA Net [5]	2.5s	31.5	15.7
Proposed Method	**2.0s**	**30.2**	**15.5**

attention mechanism is light and does not bring a considerable number of parameters to increase model complexity, thus, we haven't observed any convergence issues throughout the training process. Moreover, The experiments showed that using separated content and style feature maps can effectively provide a regional-based feature recalibration process, which is critical for diabetic retinopathy classification. Eventually, experimental results indicate that applying the attention modules in the proposed model helps the model to focus on the more informative area and scale the representation space which increases the model performance in recognizing the DR. According to the experimental results, omitting the attention modules (e.g., baseline) from the model decreased the kappa coefficient score of the proposed method by 8.9% on the APTOS dataset [6], as shown in Table 3.

G. COMPUTATIONAL TIME

As we stated earlier in the introduction section, analyzing retinopathy images may take up to five minutes for the ophthalmologists to closely check the state of the diabetic retinopathy [56]. In addition, in rare cases such as the presence of macular degeneration, the screening process may even take more time. Hence, one major factor to determine the effectiveness of the machine learning algorithm is to evaluate the inference time of the algorithm. Besides that, we are also interested to analyze the complexity of the model in terms of the required arithmetic operation. Table 4 shows the obtained results. As depicted in Table 4, comparing to the ophthalmologists our machine learning algorithm predicts a batch of 16 images in two second, while this process can take up to 80 minutes for the ophthalmologists. Besides that, our proposed attention module only adds small number of parameters and effectively enhances the performance. Overall, our method has has 30.2 million and is able to run on a single GPU device with 8 GB memory, which is comparatively is more effective than the recent MSA Net [5]. Last but not least, our method comparatively contains less parameters due to the less overhead of the attention mechanism we included inside the model. In addition, the fast processing time and high performance of the proposed method make it a suitable solution for the clinical application.

V. CONCLUSION

In this paper, we proposed a deep neural network by combining a style and content recalibration mechanism that adaptively scales informative regions for the classification of diabetic retinopathy images. Our proposed model performs both diabetic grading and healthy, non-healthy classification tasks. To improve the representation power of the network, we utilized a separation mechanism that decomposes style and content representation. Furthermore, we employed an attention module along with a spatial normalization module. The texture attention module highlights the texture information by taking the style representation and applying a high-pass filter, and the spatial normalization module determines the more informative region inside the retinopathy image using a convolutional operation. Next, we applied a fusion module to combine both features to form a normalized representation. Our experiment on the APTOS Kaggle dataset shows an improvement over the work of the literature. In future work, a parametric approach to jointly model both local semantic and global contextual representation can provide a good solution for retinopathy signs detection.

REFERENCES

[1] *IDF Diabetes Atlas*, 7th ed., Diabetes Atlas, Int. Diabetes Fed., Brussels, Belgium, 2015.

[2] D. Mellitus, "Diagnosis and classification of diabetes mellitus," *Diabetes care*, vol. 28, no. S37, pp. S5–S10, 2005.

[3] *Diabetic Retinopathy*, Mayo Clinic, Rochester, MN, USA, 2021.

[4] R. R. A. Bourne, G. A. Stevens, R. A. White, J. L. Smith, S. R. Flaxman, H. Price, J. B. Jonas, J. Keeffe, J. Leasher, K. Naidoo, K. Pesudovs, S. Resnikoff, and H. R. Taylor, "Causes of vision loss worldwide, 1990–2010: A systematic analysis," *Lancet Global Health*, vol. 1, no. 6, pp. e339–e349, Dec. 2013.

[5] M. T. Al-Antary and Y. Arafa, "Multi-scale attention network for diabetic retinopathy classification," *IEEE Access*, vol. 9, pp. 54190–54200, 2021.

[6] *Aptos Dataset*, Kaggle, San Francisco, CA, USA, 2019.

[7] A. Maier, C. Syben, T. Lasser, and C. Riess, "A gentle introduction to deep learning in medical image processing," *Zeitschrift Medizinische Physik*, vol. 29, no. 2, pp. 86–101, May 2019.

[8] A. Pinz, S. Bernogger, P. Datlinger, and A. Kruger, "Mapping the human retina," *IEEE Trans. Med. Imag.*, vol. 17, no. 4, pp. 606–619, Aug. 1998.

[9] M. Al-Antary, M. Hassouna, Y. Arafa, and R. Khalifah, "Automated identification of diabetic retinopathy using pixel-based segmentation approach," in *Proc. 2nd Int. Conf. Watermarking Image Process.*, Sep. 2019, pp. 16–20.

[10] S. A. Alryalat, M. Al-Antary, Y. Arafa, B. Azad, C. Boldyreff, T. Ghnaimat, N. Al-Antary, S. Alfegi, M. Elfalah, and M. Abu-Ameerh, "Deep learning prediction of response to anti-VEGF among diabetic macular edema patients: Treatment response analyzer system (TRAS)," *Diagnostics*, vol. 12, no. 2, p. 312, Jan. 2022.

[11] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional ConvLSTM U-Net with Densley connected convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1–10.

[12] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Attention Deeplabv3+: Multi-level context attention mechanism for skin lesion segmentation," in *Proc. Eur. Conf. Comput. Vis.* Glasgow, U.K.: Springer, 2020, pp. 251–266.

[13] Z. Zeng, Y. Xulei, Y. Qiyun, Y. Meng, and Z. Le, "SeSe-Net: Self-supervised deep learning for segmentation," *Pattern Recognit. Lett.*, vol. 128, pp. 23–29, Dec. 2019.

[14] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: A review of literature," *Multimodal Technol. Interact.*, vol. 2, no. 3, p. 47, 2018.

[15] N. Eftekhari, H.-R. Pourreza, M. Masoudi, K. Ghiasi-Shirazi, and E. Saeedi, "Microaneurysm detection in fundus images using a two-step convolutional neural network," *Biomed. Eng. OnLine*, vol. 18, no. 1, pp. 1–16, Dec. 2019.

[16] Y.-H. Li, N.-N. Yeh, S.-J. Chen, and Y.-C. Chung, "Computer-assisted diagnosis for diabetic retinopathy based on fundus images using deep convolutional neural network," *Mobile Inf. Syst.*, vol. 2019, pp. 1–14, Jan. 2019.

[17] Y. Hatanaka, K. Ogohara, W. Sunayama, M. Miyashita, C. Muramatsu, and H. Fujita, "Automatic microaneurysms detection on retinal images using deep convolution neural network," in *Proc. Int. Workshop Adv. Image Technol. (IWAIT)*, Jan. 2018, pp. 1–2.

[18] G. S. Scotland, P. McNamee, A. D. Fleming, K. A. Goatman, S. Philip, G. J. Prescott, P. F. Sharp, G. J. Williams, W. Wykes, G. P. Leese, and J. A. Olson, "Costs and consequences of automated algorithms versus manual grading for the detection of referable diabetic retinopathy," *Brit. J. Ophthalmol.*, vol. 94, no. 6, pp. 712–719, Jun. 2010.

[19] K. Xu, D. Feng, and H. Mi, "Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image," *Molecules*, vol. 22, no. 12, p. 2054, Nov. 2017.

[20] R. Pires, S. Avila, J. Wainer, E. Valle, M. D. Abramoff, and A. Rocha, "A data-driven approach to referable diabetic retinopathy detection," *Artif. Intell. Med.*, vol. 96, pp. 93–106, May 2019.

[21] A. Sungheetha and R. Sharma, "Design an early detection and classification for diabetic retinopathy by deep feature extraction based convolution neural network," *J. Trends Comput. Sci. Smart Technol.*, vol. 3, no. 2, pp. 81–94, Jul. 2021.

[22] M. U. Akram, S. Khalid, and S. A. Khan, "Identification and classification of microaneurysms for early detection of diabetic retinopathy," *Pattern Recognit.*, vol. 46, no. 1, pp. 107–116, 2013.

[23] M. U. Akram, S. Khalid, A. Tariq, S. A. Khan, and F. Azam, "Detection and classification of retinal lesions for grading of diabetic retinopathy," *Comput. Biol. Med.*, vol. 45, pp. 161–171, Feb. 2014.

[24] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "DREAM: Diabetic retinopathy analysis using machine learning," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 5, pp. 1717–1728, Sep. 2014.

[25] X. Zhang, G. Thibault, E. Decencière, B. Marcotegui, B. Laÿ, R. Danno, G. Cazuguel, G. Quellec, M. Lamard, P. Massin, A. Chabouis, Z. Victor, and A. Erginay, "Exudate detection in color retinal images for mass screening of diabetic retinopathy," *Med. Image Anal.*, vol. 18, no. 7, pp. 1026–1043, Oct. 2014.

[26] K. M. Adal, P. G. Van Etten, J. P. Martinez, K. W. Rouwen, K. A. Vermeer, and L. J. van Vliet, "An automated system for the detection and classification of retinal changes due to red lesions in longitudinal fundus images," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 6, pp. 1382–1390, Jun. 2018.

[27] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2015, pp. 234–241.

[29] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[30] R. Azad, L. Rouhier, and J. Cohen-Adad, "Stacked hourglass network with a multi-level attention mechanism: Where to look for intervertebral disc labeling," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Springer, 2021, pp. 406–415.

[31] L. C. Chen, G. Papandreou, and I. Kokkinos, "DeepLab: Semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Jun. 2017.

[32] G. Quellec, K. Charrière, Y. Boudi, B. Cochener, and M. Lamard, "Deep image mining for diabetic retinopathy screening," *Med. Image Anal.*, vol. 39, pp. 178–193, Jul. 2017.

[33] P. Zang, L. Gao, T. T. Hormel, J. Wang, Q. You, T. S. Hwang, and Y. Jia, "DcardNet: Diabetic retinopathy classification at multiple levels based on structural and angiographic optical coherence tomography," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 6, pp. 1859–1870, Jun. 2021.

[34] S. H. Kassani, P. H. Kassani, R. Khazaeinezhad, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Diabetic retinopathy classification using a modified xception architecture," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2019, pp. 1–6.

[35] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.
[36] A. Jain, A. Jalui, J. Jasani, Y. Lahoti, and R. Karani, "Deep learning for detection and severity classification of diabetic retinopathy," in *Proc. 1st Int. Conf. Innov. Inf. Commun. Technol. (ICIICT)*, Apr. 2019, pp. 1–6.
[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[38] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
[39] M. Mateen, J. Wen, Nasrullah, S. Song, and Z. Huang, "Fundus image classification using VGG-19 architecture with PCA and SVD," *Symmetry*, vol. 11, no. 1, p. 1, Dec. 2018.
[40] L. Dai, R. Fang, H. Li, X. Hou, B. Sheng, Q. Wu, and W. Jia, "Clinical report guided retinal microaneurysm detection with multi-sieving deep learning," *IEEE Trans. Med. Imag.*, vol. 37, no. 5, pp. 1149–1161, May 2018.
[41] J. Jaskari, J. Sahlsten, T. Damoulas, J. Knoblauch, S. Särkkä, L. Kärkkäinen, K. Hietala, and K. Kaski, "Uncertainty-aware deep learning methods for robust diabetic retinopathy classification," 2022, *arXiv:2201.09042*.
[42] F. Zia, I. Irum, N. Nawaz Qadri, Y. Nam, K. Khurshid, M. Ali, I. Ashraf, and M. Attique Khan, "A multilevel deep feature selection framework for diabetic retinopathy image classification," *Comput., Mater. Continua*, vol. 70, no. 2, pp. 2261–2276, 2022.
[43] B. Graham, "Kaggle diabetic retinopathy detection competition report," Univ. Warwick, Coventry, U.K., Tech. Rep., 2015, pp. 24–26.
[44] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[45] R. Poplin, A. V. Varadarajan, K. Blumer, Y. Liu, M. V. McConnell, G. S. Corrado, L. Peng, and D. R. Webster, "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning," *Nature Biomed. Eng.*, vol. 2, no. 3, pp. 158–164, 2018.
[46] H. H. Vo and A. Verma, "New deep neural nets for fine-grained diabetic retinopathy recognition on hybrid color space," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2016, pp. 209–215.
[47] R. Azad, N. Khosravi, and D. Merhof, "SMU-Net: Style matching U-Net for brain tumor segmentation with missing modalities," 2022, *arXiv:2204.02961*.
[48] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," 2015, *arXiv:1508.06576*.
[49] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
[50] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
[51] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[53] A. K. Gangwar and V. Ravi, "Diabetic retinopathy detection using transfer learning and deep learning," in *Evolution in Computational Intelligence*. Springer, 2021, pp. 679–689.
[54] V. Naralasetti, S. N. Shareef, and S. Hakak, "Blended multi-modal deep convnet features for diabetic retinopathy severity prediction," *Electronics*, vol. 9, no. 6, p. 914, 2020.
[55] J. D. Bodapati, N. S. Shaik, and V. Naralasetti, "Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification," *J. Ambient Intell. Hum. Comput.*, vol. 12, no. 10, pp. 9825–9839, Oct. 2021.
[56] N. Aujla. (2022). *Retinal Imaging: How it Works & Why it's Important*. Accessed: May 2, 2022. [Online]. Available: https://visionaryeyecentre.com/retinal-imaging-how-it-works-why-its-important

MOHAMMAD D. ALAHMADI received the Ph.D. and M.Sc. degrees from Florida State University, in 2018 and 2020, respectively. He is currently an Assistant Professor with the Software Engineering Department, University of Jeddah. His research has been published in top journals and conferences in a wide variety of topics. His research interests include software engineering, computer vision, and machine learning.