IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 6, 2021, accepted August 21, 2021, date of publication August 31, 2021, date of current version September 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3109240

Multitasking Deep Learning Model for Detection of Five Stages of Diabetic Retinopathy

SHARMIN MAJUMDER, (Graduate Student Member, IEEE), AND NASSER KEHTARNAVAZ, (Fellow, IEEE)
Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX 75080, USA
Corresponding author: Sharmin Majumder (sxm180109@utdallas.edu)

ABSTRACT Early diagnosis and treatment of diabetic retinopathy (DR) can reduce the risk of vision loss. There are five stages of DR consisting of no DR, mild DR, moderate DR, severe DR, and proliferate DR. This paper presents a multitask deep learning model to detect all the five stages of DR more accurately than existing methods. The developed multitask model consists of one classification model and one regression model, each with its own loss function. After training the regression model and the classification model separately, the features extracted by these two models are concatenated and inputted to a multilayer perceptron network to classify the five stages of DR. A modified Squeeze Excitation Densely Connected deep neural network is also developed as part of this multitasking approach. The developed multitask model is applied to the two large Kaggle datasets of APTOS and EyePACS. The results obtained indicate that the developed multitask model achieved a weighted Kappa score of 0.90 and 0.88 for the APTOS and EyePACS datasets, respectively. In addition, the micro and macro average area under the receiver operating characteristic (ROC) curve was found to be 0.96, and 0.93, respectively, which are higher than existing methods for detecting the five stages of DR.

INDEX TERMS Diabetic retinopathy (DR), eye fundus images, five stages of diabetic retinopathy, multitasking deep neural network, squeeze excitation densely connected network.

I. INTRODUCTION

International Diabetes Federation (IDF) states that there are more than 460 million adults (20-79 years) in the world living with diabetes. The number of adults with diabetes has more than tripled over the past 20 years and 1 in 2 people (about 230 million) with diabetes are undiagnosed [1]. A complication of diabetes is Diabetic Retinopathy (DR), an eye retinal disease that can lead to visual impairments and even blindness. DR is caused by microvascular complications of diabetes appearing as morphological changes in the eye fundus. Diagnosis and treatment of DR at its initial stages reduces the risk of vision loss to a great extent. With the development of color fundus photography, DR can be detected non-invasively at its early stages.

Along with the detection of DR, the severity level of DR also needs to be determined for treatment purposes. There are two major types or classes of DR: non-proliferative DR (NPDR) and proliferative DR (PDR) [2]. NPDR is further categorized into the following three stages: (i) mild NPDR, which is the earliest stage of DR, (ii) moderate NPDR, and (iii) severe NPDR. PDR denotes the advanced stage of DR. The severity level of DR is thus generally graded as these five stages: no DR, mild DR, moderate DR, severe DR, and proliferate DR.

Lesions in fundus images that appear as small circular red dots at the end of blood vessels indicate the earliest sign of DR. The tiny bulges in the blood vessels in the retina are called Microaneurysms. At least one Microaneurysm is present in the mild stage of DR [3], [4]. Microaneurysms, Hemorrhages and/or Exudates are signs of moderate DR. In the PDR stage, new blood vessels are formed along with the above abnormalities [5]. FIGURE 1 shows sample color fundus images of normal retina and different severity levels of DR. A major issue with DR detection involves the difficulty of identifying symptoms at its early stages due to visual similarities between no DR, mild DR, and sometimes moderate DR. If DR proceeds to the advanced stage, vision loss can occur.

Many computer-based methods have been developed in the literature for the detection of DR. In these previous methods, to mimic human experts, much attention has been paid to

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

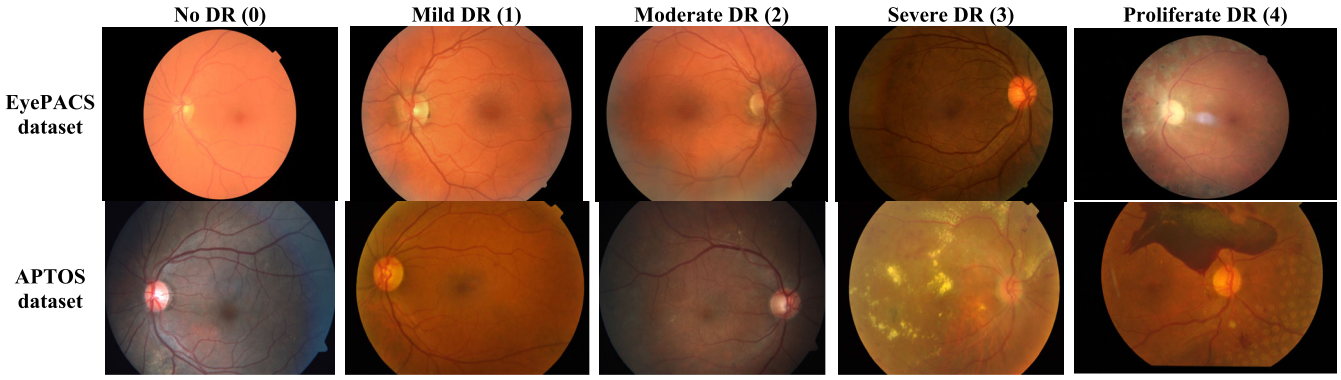

FIGURE 1. Sample fundus images from EyePACS and APTOS datasets.

the automatic detection of lesions for DR screening and grading. A representative set of methods already developed are stated here. Detection and segmentation of blood vessels in retinal images was discussed in [6] and [7]. In case of Microaneurysm detection, automated image processing approaches were developed in [8]–[11]. Several methods for detection of Exudates in color fundus images were introduced in [12]–[15]. In [16], the detection of neovascularisation and lesions was performed for the Messidor dataset.

Lately, due to the success of deep learning models in many image processing tasks, researchers have utilized them for DR detection. In [17], a region-based fully convolutional network (R-FCN) for lesion detection and DR grading into four stages was developed. In [18], an instance learning was used to detect lesions in fundus images for the Messidor dataset.

There have also been some works reported on detecting DR stages, that is conducting detection as well as classification of DR at the same time. These types of image classification tasks can be grouped into conventional image processing techniques where handcrafted features were considered [19]–[26], and more recent deep learning techniques [27]–[29]. In [23], an algorithm based on random forest was applied to handcrafted features to detect the presence of DR and assess its risk. In [24], a DR classification was performed by using BossaNova and Fisher Vector midlevel features which extended the classical Bags of Visual Words features. In [25], a two-step method based on handcrafted features was covered: one step for detecting the presence of DR and one step for detecting its severity level. A bag of features approach was developed for detection of DR stages by using the histogram of orientated gradients in [26]. In [30], both binary and multiclass classification of DR was achieved by using Haralick and multiresolution features.

Deep learning techniques, especially convolutional neural networks (CNNs), have generated much success in image classification due to their end-to-end learning capabilities or not requiring to devise handcrafted features [31]–[33]. In [29], three CNN models were utilized for the binary classification of DR, that is DR/no DR. Deep learning-based classification approaches were also discussed in [28] and [34]–[37]. References [28], [36], [37] focused on binary classification of DR as referable and non-referable. In [37], the EfficientNet-B5 model was used for this classification task. In [38], binocular fundus images from both eyes (left and right eye) were taken as the inputs to a transfer learning-based CNN model. In [39], a DCNN (Deep Convolution Neural Network) for detecting two stages of DR (normal and NPDR) was discussed. In [40], the right and left eye images were inputted to two separate CNN models with each model performing a binary classification of DR. In [27], a CNN based smartphone app was developed for binary classification of DR in real-time.

Classification of the severity stages of DR were presented in [41]–[43]. In [44], a deep neural network for four-degree classification of DR was covered. A hyperparameter tuning was done in the Inception-v4 model to obtain four classes of DR in [41]. A CNN model was developed to classify the five stages of DR in [42]. In [45], three deep learning models (Feed Forward Neural Network (FNN), Deep Neural Network (DNN), and Convolutional Neural Network (CNN) were applied to the EyePACS dataset for DR classification whereas the performance of the EyePACS dataset was examined for different CNN models in [45]. In [46], several deep learning models (AleXNet, VggNet, GoogleNet, ResNet) were compared for DR classification using the Kaggle EyePACS dataset with VggNet achieving the best accuracy. A transfer learning-based smartphone app using a pretrained Xception model was previously developed for the classification of the five stages of DR in [43] by our research group. This app runs in real-time on smartphone platforms based on fundus images that are captured via commercially available lenses that snap onto smartphones in front of their cameras.

A few recent papers have utilized ensembles of two or more deep learning models for DR classification. In [47], the integration of deep learning models was used to detect no DR, referable DR (rDR), vision threatening DR, and macular edema. In [48], an ensemble of five pretrained CNN models consisting of Resnet50, Inceptionv3, Xception, Dense121, and Dense169, were used for DR classification into five stages. All these papers considered ensemble of two or more classification models but did not employ any regression task.

Moreover, the above mentioned or existing five-stage DR classification papers have reported not high accuracy when considering all the stages. A few papers in the literature also proposed multitasking network for fundus image analysis mainly focusing on lesion segmentation task [49]–[51]. In [49], a weakly supervised multitask architecture is proposed for retinal lesions segmentation whereas simultaneous segmentation of bright and red lesions in fundus images is performed using a multitasking architecture in [50]. In [51], a region-specific multitask recognition model was proposed to classify 36 different retinal disease without examining the classification of different DR stages.

In this paper, a multitasking deep learning architecture is introduced for classifying fundus images into the five stages of DR (no DR, mild DR, moderate DR, severe DR, and proliferate DR). For example, severe DR stage comes after moderate DR stage, moderate DR stage comes after mild DR stage, etc. Thus, there is a dependency among different stages. This dependency between the stages can be learnt by a regression model towards the classification task. Keeping this in mind, we have introduced a multitasking model consisting of a regression and a classification model to classify or detect the five stages of DR (no DR, mild DR, moderate DR, severe DR, and proliferate DR). Normally, the classification task is done based on the difference or distinction between the classes using a single loss function. Our model uses two loss functions, one for the classification task and the other for the regression task. This approach to addressing classification of the DR stages is indeed novel and it is the first time such an approach is taken. In this work, a densely connected network modified with squeeze excitation (SE) layers is developed to implement the multitasking approach due to the capability of SE layers to learn channel interdependencies. A MultiLayer Perceptron (MLP) model is used at the end to classify the five stages of DR based on the features extracted from the two networks (classification and regression). Since deep neural networks require a large amount of training data, the introduced multitasking approach is examined based on an Xception transfer learning model. The two large public domain Kaggle datasets, namely APTOS [52] and EyePACS [53], are examined in this work to evaluate the performance of the introduced approach.

The rest of the paper is organized as follows: a description of the datasets considered together with a description on the proposed multitasking approach are provided in Section II. The developed multitasking squeeze excitation densely connected model and multitasking Xception transfer learning model are then presented in Section III. Section IV describes the experimentations carried out together with their results followed by the conclusion in Section V.

II. MATERIALS AND METHODS

A. DATASETS

Choice of dataset is important as it needs to contain a rich collection of images. In this work, two publicly available Kaggle datasets that incorporate a large number of images of all the five stages of DR are considered. These datasets are EyePACS (Eye Picture Archive Communication System) [53] and APTOS 2019 (Asia Pacific Tele-Ophthalmology Society) Blindness Detection Dataset [52].

1) APTOS 2019 BLINDNESS DETECTION DATASET

This dataset includes fundus images for the five stages of DR labeled by the severity level 0 to 4, where label 0 indicates no DR, label 1 mild DR, label 2 moderate DR, label 3 severe DR, and label 4 proliferate DR. It contains a total of 3,662 retinal images where 1,805 images belong to no DR, 370 images to mild DR, 999 to moderate DR, 193 images to severe DR, and 295 images to proliferate DR. The resolution of the images is 3216×2136.

2) EyePACS DATASET

Similar to the APTOS dataset, the EyePACS dataset also contains fundus images belonging to the five stages of DR. This dataset contains 35,126 retina images of size 3888×2951 for both the left and right eyes, with 25,810 images labeled as 0 DR (no DR), 2,443 mild DR, 5,292 moderate DR, 873 severe DR, and 708 proliferate DR images. Here, 10,000 images are randomly selected from the no DR stage and our model is trained on a total 19,316 images (10,000 no DR, 2,443 mild DR, 5,292 moderate DR, 873 severe DR, and 708 proliferate DR).

It is to be noted that the above datasets are highly imbalanced thus introducing bias. To have a balanced dataset, a class weighting method is applied to weigh classes inversely proportional to their frequency according to (1).

$$w_j = \frac{n}{kn_j} \tag{1}$$

where w_j denotes the weight of class j, n the total number of samples, n_j the number of samples in class j, and k the total number of classes. This class weighing method provides different class weights to the cost function to reduce errors for the minority class.

B. DATA PREPROCESSING AND AUGMENTATION

The datasets considered in this paper contain images of high resolution. The images are resized to 299×299 to feed into the networks. It should be noted that although the aspect ratio is altered as a result of this resizing, the DR related lesions including Microaneurysms, Hemorrhages, Exudates still remain preserved. After resizing, input image intensity values are normalized between 0 and 1. Deep learning is data hungry. The amount of data in the above two datasets are not sufficient for training a deep neural network from scratch. Therefore, data augmentation techniques such as rotation, horizontal flip, width shift, height shift, zooming, and shearing are applied to the original data.

C. PROPOSED MULTITASKING METHOD

DR progresses to higher severity levels after the lower severity levels leading to a dependency among different stages.

Algorithm 1 Algorithm of Multitasking Approach (lr=Learning Rate, β_1, β_2=Exponential Decay Rate in Adam Optimization for the First Moment and Second Moment Estimates, Respectively)

Require: Fundus Images and Labels *(X, Y)*, where $Y = \{y/y \in \{0, 1, 2, 3, 4\}$ [0: No DR, 1: Mild DR, 2: Moderate DR, 3: Severe DR, 4: Proliferate DR]
Input: fundus images $x \in X$
Output: Trained model predicts probability scores corresponding to$\forall y$ for an input x Perform Preprocessing:

- Resize the image to $299 \times 299 \times 3$
- Perform Data Augmentation techniques: rotation, horizontal flip, width shift, height shift, zooming, and shearing.

Design a Classification Model and a Regression Model $H = \{ClassificationModel, RegressionModel\}$
For $\forall h \in H$ (**do**
  lr=0.001, momentum=0.7
  for epochs=1 to 250 **do**
    for each minibatch $(X_{\text{mini}}, Y_{\text{mini}}) \in (X, Y)$ **do**
      if $h = Classification\ Model$ **then**
        Update the parameters of the Classification Model using Stochastic Gradient Decent optimization (SGD).
        if epochs>150**then**
          lr=0.0001
        end
        if epochs>200 **then**
          lr=0.00001, momentum=0.5
        end
      end
      if $h = Regression\ Model$ **then**
        if epochs<50 **then**
          Update the parameters of the Regression Model using Adaptive Moment Estimation (Adam).
        end
    end
  end
end
Concatenate the features extracted from $\forall h \in H$ and fed to a MLP classifier to generate the Multitasking Model.
for Multitasking Model **do**
  lr=0.001, β_1=0.9, β_2=0.999
    for epochs =1 to 50 **do**
      for each minibatch $(X_{\text{mini}}, Y_{\text{mini}}) \in (X, Y)$ **do**
        Update the parameters of the Multitasking Model using Adaptive Moment Estimation.
        if the validation error is not improving for four epochs **then**
          $lr = lr \times 0.01$
        end
      end
    end
  end
for $x \in X_{test}$ **do**
Trained Multitasking Model predicts probability scores for $\forall y$
end

A regression model can learn the dependency or progressive characteristics between the stages. Here, a multitasking model consisting of a classification model and a regression model is considered in order to classify or detect the five stages of DR.

The classification model learns the distinguishing characteristics between the five stages whereas the regression model learns the inter-dependency characteristics among the stages. Classification task usually works based on one loss function. More than one loss function may improve the classification performance [54]. In this paper, two loss functions are considered, *cross entropy* loss function as given in (2) is used for classification task and *mean square error* loss function as defined by (3) is used for regression task. The regression model and the classification model are trained separately on the fundus images using a linear

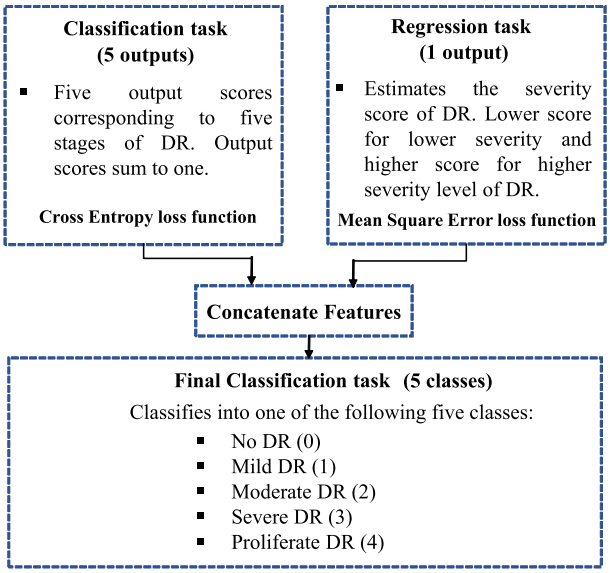

FIGURE 2. Developed multitasking approach for detecting the five stages of DR.

activation function and a softmax activation function, defined by (4), in the last layer of the regression model and the classification model, respectively.

$$Cross\ entropy = -\sum_{i}^{M} y_i \log(\widehat{y_i}) \tag{2}$$

$$MSE = \frac{1}{M}\sum_{i=1}^{M}(y_i - \widehat{y_i})^2 \tag{3}$$

where, $\widehat{y_i}$ is the predicted value and y is the true value. M is the number of classes.

$$Soft\ \max(i) = \frac{e^{\delta(i)}}{\sum_{j=1}^{K} e^{\delta(j)}}, \quad j = 1, ..i, ..K \tag{4}$$

where K denotes the total number of classes, and δ denotes the output of the last fully connected layer. The output probabilities of each class lie between 0 and 1 with all the values adding up to 1.

The classification model outputs five probability scores (sum to one) corresponding to the five classes or stages of DR. The regression model has one output providing the severity level of DR. The regression model is trained with the output labels 0 for no DR, 0.2 for mild DR, 0.4 for moderate DR, 0.6 for severe DR, and 0.8 for proliferate DR. In other words, the regression model outputs one score corresponding to the five stages of DR. Features learnt by the classification model and the regression model are concatenated and fed into a multilayer perceptron network (MLP) classifier for the final classification of the five stages of DR. The introduced approach is presented as an algorithm in Algorithm 1 and the steps to implement it are shown in FIGURE 2.

III. IMPLEMENTATION OF MULTITASKING DEEP LEARNING MODELS

A. SQUEEZE EXCITATION DENSELY CONNECTED MULTITASKING NETWORK (MSEDenseNet)

A modified densely connected network (DenseNet) is developed here to implement the multitasking approach. A basic DenseNet with compression is combined with a squeeze-excitation (SE) network. SE network introduces a building block that improves channel interdependencies to improve the performance of the model.

The Multitasking Squeeze Excitation Densely Connected Network (MSEDenseNet) consists of a SEDenseNet classification model, a SEDenseNet regression model, and a MLP classifier. The architecture of the developed MSEDenseNet is shown in FIGURE 3.

1) MODEL ARCHITECTURE

As shown in FIGURE 3, SEDenseNet consists of five dense blocks and four transition blocks each in between two dense blocks. In each dense block, a SE-dense module has been repeated for 16 times. A SE-dense module consists of a batch normalization layer, ReLU activation, a 3×3 convolution layer, and a SE block. A SE block comprises a squeeze layer which is a global average pooling layer, and an excitation layer with two 1×1 convolution layers. The first convolution layer is followed by a ReLU activation and the second convolution layer is followed by a sigmoid activation. In SE block, each channel is squeezed to a single numeric value using average pooling. The ratio to reduce the channel complexity is set to 16. Finally, each channel of the input to the SE block is scaled by the respective weight obtained from the SE block. Down-sampling is achieved by the transition layer between two dense blocks. A transition layer is made of batch normalization, ReLU, 1×1 convolution, and average pooling. A SE block is also added to the transition layer. The last fully connected layer of the original network is replaced with 2×2 convolution layer to reduce the number of parameters.

The depth and growth rates of the developed SEDenseNet network is set to 164 and 18, respectively. Therefore, the number of dense modules in one dense block is 16. The number of filters for the first convolution layer is $2\times$ growth rate whereas for the convolution layers in the dense block and in the transition block are $2 \times$ growth rate $\times$ compression ratio. The compression ratio for the network is set to 0.5.

In the SEDenseNet multitasking model, a SEDenseNet classification model, and a SEDenseNet regression model are combined to enrich the learned features. Fig. 3 illustrates the concatenation of the regression model and the classification model. Outputs from the last average pooling layer of the trained classification model and the regression model are fused together to feed into the MLP. The MLP comprises a batch normalization layer, a fully connected layer of 512 units with ReLU activation, and another fully connected layer with softmax activation function as the output layer.

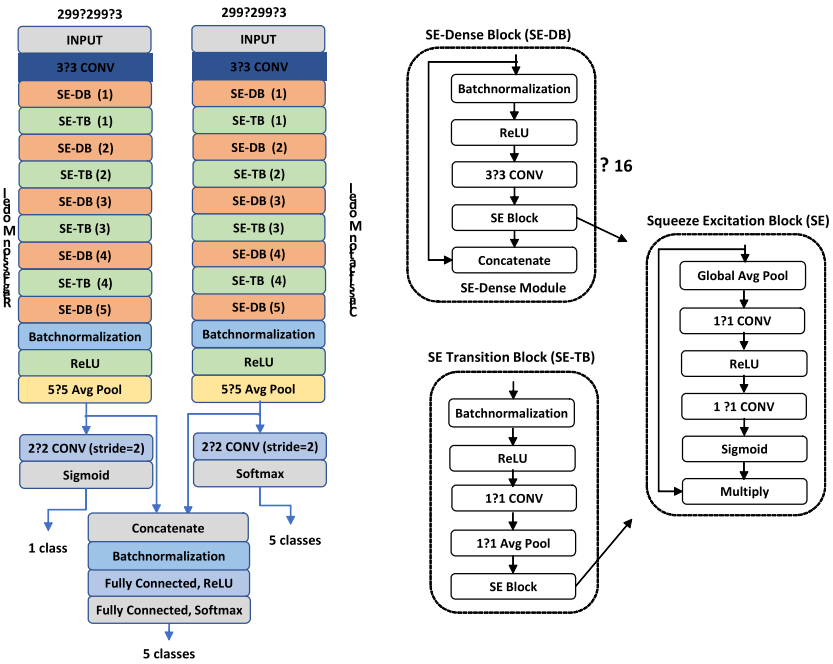

FIGURE 3. Architecture of the developed Multitasking Squeeze Excitation Densely Connected Deep Neural Network (MSEDenseNet) - CONV: Convolution.

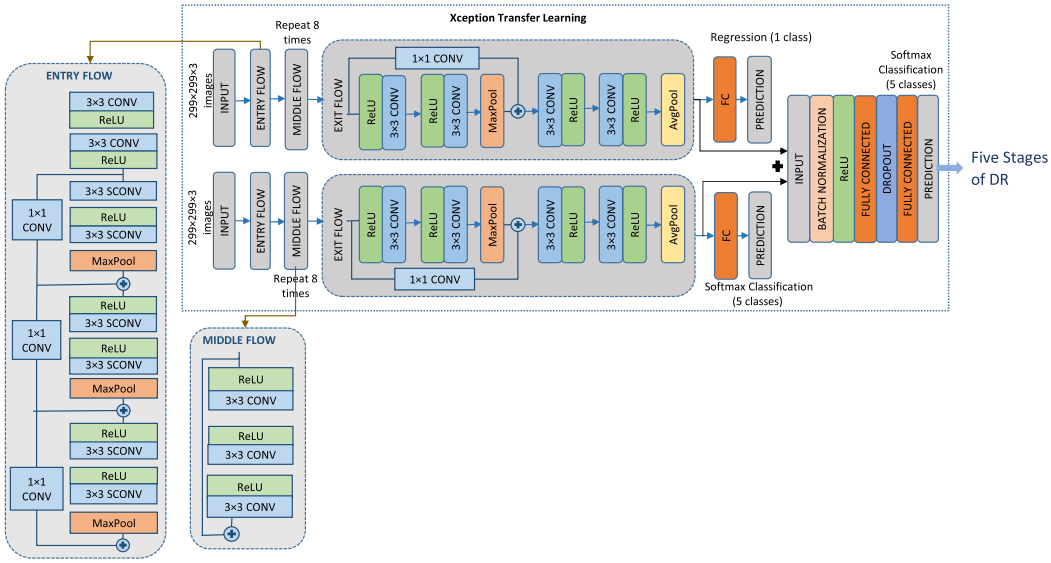

FIGURE 4. Architecture of the developed multitasking Xception transfer learning model (MXception) - CONV: Convolution, FC: Fully Connected Layer, SCONV: Separable Convolution.

2) MODEL TRAINING

The developed SEDenseNet classification model is trained with 250 epochs using Stochastic gradient descent (SGD) optimization algorithm and Categorical Cross-entropy (CCE) loss function. For the first 150 epoch, the learning rate, batch size, and momentum are set to 0.001, 2, and 0.7, respectively. For the next 50 epochs, the learning rate is reduced to 0.0001. For the last 50 epochs, the learning rate and momentum are changed to 0.00001 and 0.5, respectively. The Validation accuracy is checked in every epoch and the model with the highest validation accuracy is saved using the model checkpoint feature of the Keras callback. The output layer of the classification model is a convolution layer with softmax activation function to generate five probability scores corresponding to the five classes of DR.

The developed SEDenseNet regression model is trained with 50 epochs. Adaptive moment estimate (Adam) optimization algorithm with a learning rate of 0.001 and mean square

error (MSE) loss function are used to train the model. The mini batch size is kept 2. The output layer of the regression model is a convolution layer with linear activation function to score one output corresponding to the severity level of DR.

After concatenation of the features from the previously trained classification and regression model, a batch normalization layer with momentum 0.9 is added for the purpose of normalizing data across a batch. A batch normalization layer is usually used to speed up the training and to reduce the sensitivity to initialization. The CCE loss function and the Adam optimization algorithm with a learning rate of 0.001 are considered during training for 50 epochs. The learning rate is reduced by a factor of 0.1 if the validation loss is not reduced for four consecutive epochs. The model which has the highest validation accuracy is then saved.

The so-called 'He normal' initialization and 'l2 kernel'' regularization are considered for the units of the convolution and fully connected layers of MSEDenseNet. The regularization protects the model against overfitting.

B. MULTITASKING XCEPTION TRANSFER LEARNING MODEL (MXception)

Due to the scarcity of sufficient training data to train a deep neural network from scratch, a widely used model with transfer learning is also considered. A pretrained Xception ImageNet model is fine-tuned to implement the multitasking approach to classify the five stages of DR.

1) MODEL ARCHITECTURE

The architecture of the Xception model [55] is based on depthwise separable convolution layers and consists of three major sections: entry flow, middle flow, and exit flow. FIGURE 4 shows the architecture of the Xception model. Image data first goes through the entry flow, then through the middle flow which is repeated eight times, and finally through the exit flow. Note that all the convolution and separable convolution layers are followed by batch normalization. This model is composed of 36 convolutional layers forming the feature extraction base of the network. The Xception model was previously trained with 299×299 ImageNet images for 1000 classes with the top-1 accuracy of 79%.

2) MODEL FINE TUNING

A pretrained Xception ImageNet model is fine-tuned as a regression model with one class in the output. The architecture of the regression Xception model is shown in Fig. 4. The last fully connected layer of the Xception model is chopped and then an average pooling layer is added. A dense layer consisting of one neuron is also added as the output layer with linear activation function. The adaptive moment estimate (Adam) optimization algorithm with a learning rate of 0.001 and MSE loss function are used to train the model for 25 epochs. During training, the image dataset is split into mini batches of size 16.

Another pretrained Xception model is fine-tuned with retinal images to classify the five stages of DR. The last fully connected layer is replaced with an average pooling layer, and a dense layer with softmax activation function. A dropout layer is also added before the output layer with 0.8 keep probability to regularize the model. The CCE loss function and Adam optimization with a learning rate of 0.001 (0.9 exponential decay rate for the first-moment estimates, $\beta 1$ and 0.999 exponential decay rate for the second-moment estimates, $\beta 2$) are used for training. Model with the highest validation accuracy is saved. The model is fine tuned for 25 epochs with minibatch size of 16. If the validation loss is not reduced for four consecutive epochs, the learning rate is reduced by a factor of 0.1.

Features generated from the last average pooling layer of the fine-tuned Xception classification model and regression model are concatenated and inputted to an MLP classifier. Similar to the SEDenseNet, MLP in the Multitasking Xception network also contains two fully connected layers with softmax activation function at the last layer to generate five score for the five classes or stages of DR. The training parameters to train the MLP classifier in the Xception multitasking network are similar to the parameters used in the MLP classifier of MSEDenseNet. The model which has the highest validation accuracy is then saved.

IV. EXPERIMENTATIONS AND RESULTS

A. IMPLEMENTATION FRAMEWORK

The experimentations reported here were carried out on a computer equipped with an NVIDIA Quadro P5000 GPU. The computer had an Intel® Core™ i9 processor with twenty 3.3GHZ cores and 32GBs of RAM. The software packages used for implementation of the models included Python 3.7 together with the deep learning libraries of Keras with Tensorflow, H5PY, OpenCV, and Scikit-Learn.

B. PERFORMANCE MEASURES

Performance was assessed based on the five widely used performance measures of *Precision, Recall, F1 Score, Accuracy, Receiver operating curve* [48] and *Weighted Kappa Score (WKS)* as stated in (5). *Precision* and *Recall* were computed for the five classes separately and then a macro average was taken for the multiclass classification. For *Kappa Score,* quadratic weight was considered as follows:

$$WKS = 1 - \frac{\sum_{i=1}^{N}\sum_{j}^{N} W_{ij}O_{ij}}{\sum_{i=1}^{N}\sum_{j}^{N} W_{ij}E_{ij}} \tag{5}$$

where i and j denote the indices associated with the true class and the classified class, respectively, O is actual observation counts, E is expected counts, N is the total number of classes, and W_{ij} is given by (6):

$$W_{ij} = \frac{(i-j)^2}{(N-1)^2} \tag{6}$$

C. CLASSIFICATION OUTCOME

1) RESULTS FOR APTOS DATASET

For APTOS dataset, three different experiment sets are considered where each experiment set contains randomly selected 90% of the data for training and the remaining 10% for validation. Confusion matrices of the validation datasets for experiment 1 using the SEDenseNet Classification and SEDenseNet multitasking model are shown in Table 1, and Table 2, respectively. The performance measures of *Precision, Recall*, *F1 Score*, *Accuracy*, and *Quadratic weighted Kappa Score* for experiment 1 are shown in Table 3. As can be seen from this table, the multitasking model improved the classification performance by nearly 4%.

TABLE 1. Confusion matrix on the APTOS dataset for SEDenseNet classification model: Exp 1.

		Classified				
	Class Name	Not DR	Mild DR	Moderate DR	Severe DR	Proliferate DR
Actual	Not DR	183	3	2	0	0
	Mild DR	4	15	12	0	2
	Moderate DR	3	5	82	2	8
	Severe DR	0	0	10	4	1
	Proliferate DR	0	2	15	1	13

TABLE 2. Confusion matrix on the APTOS dataset for SEDenseNet multitask model: Exp 1.

		Classified				
	Class Name	Not DR	Mild DR	Moderate DR	Severe DR	Proliferate DR
Actual	Not DR	183	4	1	0	0
	Mild DR	3	24	5	0	1
	Moderate DR	2	7	83	4	4
	Severe DR	0	0	7	8	0
	Proliferate DR	0	2	13	1	15

TABLE 3. Performance measures of the developed SEDenseNet models for Exp 1: APTOS dataset.

Models	Precision	Recall	F1-score	Accuracy	WKS
SEDenseNet Classification	0.67	0.59	0.61	0.81	0.84
SEDenseNet Multitask	0.75	0.70	0.72	0.85	0.88

The confusion matrices for the experiment 2 using the two models are shown in Table 4 and Table 5, respectively, whereas the performance measures are shown in Table 6. This table also shows an improvement in the classification performance for multitasking model.

The confusion matrices for experiment 3 in case of the classification and multitasking SEDenseNet are presented in Table 7 and Table 8, respectively. The performance measures for this experiment are shown in Table 9 indicating improved classification performance in the multitasking case by 2%.

As the Receiver Operating Curve (ROC) shows how the model distinguishes among classes, we also generated ROC curve for our proposed multitasking SEDenseNet model. The ROC curve of the proposed model is shown in FIGURE 5. The micro-average ROC sums up the individual true positive,

TABLE 4. Confusion matrix on the APTOS dataset for SEDenseNet classification model: Exp 2.

		Classified				
	Class Name	Not DR	Mild DR	Moderate DR	Severe DR	Proliferate DR
Actual	Not DR	176	5	0	0	0
	Mild DR	4	11	21	0	1
	Moderate DR	2	3	94	1	4
	Severe DR	0	0	11	2	5
	Proliferate DR	1	1	9	0	16

TABLE 5. Confusion matrix on the APTOS dataset for SEDenseNet multitask model: Exp 2.

		Classified				
	Class Name	Not DR	Mild DR	Moderate DR	Severe DR	Proliferate DR
Actual	Not DR	174	5	2	0	0
	Mild DR	3	25	9	0	0
	Moderate DR	0	11	88	4	1
	Severe DR	0	1	6	6	5
	Proliferate DR	1	3	8	0	15

TABLE 6. Performance measures of the developed SEDenseNet models for Exp 2: APTOS dataset.

Models	Precision	Recall	F1-score	Accuracy	WKS
SEDenseNet Classification	0.69	0.58	0.63	0.81	0.87
SEDenseNet Multitask	0.725	0.67	0.70	0.84	0.88

TABLE 7. Confusion matrix on the APTOS dataset for SEDenseNet classification model: Exp 3.

		Classified				
	Class Name	No DR	Mild DR	Moderate DR	Severe DR	Proliferate DR
Actual	Not DR	170	4	3	0	1
	Mild DR	4	20	15	0	3
	Moderate DR	2	4	91	2	1
	Severe DR	0	0	10	8	2
	Proliferate DR	0	2	11	2	12

TABLE 8. Confusion matrix on the APTOS dataset for SEDenseNet multitask model: Exp 3.

		Classified				
	Class Name	Not DR	Mild DR	Moderate DR	Severe DR	Proliferate DR
Actual	Not DR	171	4	3	0	0
	Mild DR	2	31	7	0	2
	Moderate DR	3	6	88	2	1
	Severe DR	0	1	9	8	2
	Proliferate DR	0	2	10	0	15

TABLE 9. Performance measures of the developed SEDenseNet models for Exp 3: APTOS dataset.

Models	Precision	Recall	F1-score	Accuracy	WKS
SEDenseNet Classification	0.73	0.64	0.68	0.82	0.85
SEDenseNet Multitask	0.79	0.71	0.74	0.85	0.87

TABLE 10. Performance measures of the developed models: APTOS dataset.

Models	Precision	Recall	F1-score	Accuracy	WKS
SEDenseNet Classification	0.70	0.60	0.64	0.81	0.85
SEDenseNet Multitask	0.76	0.69	0.72	0.85	0.88
Xception Classification	0.74	0.68	0.70	0.83	0.87
Xception Multitask	0.77	0.70	0.73	0.86	0.90

false positive and false negative and then map a value on a graph, where macro-average takes the average of precision and recall and map a value on a graph. This figure shows that the micro-average and the macro-average area under the curve (AUC) are 0.96, and 0.93, respectively. As the micro and macro average shows the overall performance of the model, this result indicates a good model.

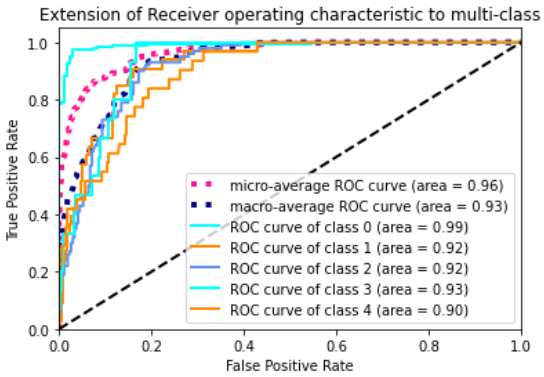

FIGURE 5. ROC of the proposed multitasking SEDenseNET model for APTOS dataset.

The highest AUC of 0.99 is achieved by the class 0 which shows our model predicts the Not DR class very well. The lowest AUC is 0.90, for the Proliferate DR class. By applying the class weighting method, we are able to generate balanced dataset which in turn improves the accuracy of the minority classes by reducing the error for minority classes. From the ROC curve of our model, AUC for all the classes are more than 0.90 which reflects a good performance of the model for all classes.

In addition, the Xception transfer learning was examined to show the effectiveness of the multitasking method. 90% of the data were randomly selected for fine tuning the model and the remaining 10% were used for validation. The performance of Xception Classification and Xception Multitask models along with the performance of the SEDenseNet Classification and SEDenseNet Multitask model averaged over the three experiments are shown in Table 10. As can be seen from this table, the multitasking approach improved the classification performance by nearly 3% when using the SEDenseNet and the Xception models.

2) RESULTS FOR EyePACS DATASET

The multitasking method was also applied to the EyePACS dataset. For this dataset, only the Xception multitasking transfer learning model was considered. The model was fine-tuned using a total of 19,316 images from the EyePACS dataset. 80% of the images were randomly selected for training and the remaining 20% were used for validation. For the APTOS dataset, a 90%-10% division of training and testing sets were considered due to the smaller number of training images in this dataset compared with the EyePACS dataset. The performance measures of *Recall*, *Precision*, *F1 Score*, *Accuracy*, and *Quadratic weighted Kappa Score* for the Xception multitask model are presented in Table 11.

TABLE 11. Performance measures of Xception multitask model: EyePACS dataset.

Models	Recall	Precision	F1-score	Accuracy	WKS
Xception Multitask	0.64	0.69	0.66	0.82	0.88

D. COMPARATIVE STUDY

The results of the developed MSEDenseNet model for the APTOS dataset was compared with four recent works where the same APTOS dataset was used. The comparison of the performance measures is shown in Table 12. As can be seen from this table, the developed multitasking model generated the highest performance measures for the detection of the five stages of DR. Table 13 shows the comparison of the performance measures of the multitasking model for the EyePACS dataset with two recent works by Pratt *et al.* [42] and Qummar *et al.* [48] where the EyePACS dataset was used. This table also shows the highest performance measures were obtained by the developed multitasking model.

TABLE 12. Comparison with recent works: APTOS dataset.

Model	Year	Specificity	Accuracy (%)	*WKS*, *KS*	No. of Class
Bodapati et al. [56]	2020	-	80.96	*KS*: 0.71	5
Kassani et al. [57]	2019	0.87	83.09	-	5
Shaban et al. [58]	2020	0.94-0.95	88-89	*WKS*: 0.91-0.92	3
MSEDenseNet	2021	-	85	*WKS*: 0.88	5
MXception	2021	-	86	*WKS*: 0.90	5

TABLE 13. Comparison with recent works: EyePACS dataset.

Model	Year	Recall (%)	Precision (%)	Specificity (%)	F1 Score (%)	Accuracy (%)	Class No.
Pratt et al. [42]	2016	30	51.06	95	41.6	75	5
Qummar et al. [48]	2019	51.5	63.8	86.7	53.7	80.8	5
MXception Model	2021	64	69	-	66	82	5

V. CONCLUSION

In this paper, a multitasking deep neural network is developed to classify all the five stages of diabetic retinopathy from eye fundus images based on the DenseNET architecture. The largest publicly available datasets of eye fundus images (EyePACS and APTOS datasets) were used to train and evaluate the developed model. The results show that the developed multitasking model generated the highest performance measures compared to the existing five-stage diabetic retinopathy classification methods. A limitation of the developed approach which is commonly encountered in deep learning models is the comprehensiveness of the datasets used and the training time associated with using a very large number of images. However, once the model is trained, it classifies a test or unknown image in a short time (<0.5 s). A possible future extension of this work includes the real-time implementation of this model as a smartphone app so that it can easily be deployed in clinical environments for diabetic retinopathy eye examination.

REFERENCES

[1] *IDF Diabetes Atlas*, 9th ed. International Diabetes Federation, Brussels, Belgium, 2019.

[2] R. Williams, M. Airey, H. Baxter, J. Forrester, T. Kennedy-Martin, and A. Girach, "Epidemiology of diabetic retinopathy and macular oedema: A systematic review," *Eye*, vol. 18, pp. 963–983, Jul. 2004, doi: 10.1038/sj.eye.6701476.

[3] *Medscape*. Accessed: Jun. 16, 2021. [Online]. Available: https://www.medscape.com/answers/1225122-100724/what-are-the-characteristics-of-nonproliferative-diabetic-retinopathy

[4] *WebMD, Stages of Diabetic Retinopathy*. Accessed: Jun. 16, 2021. [Online]. Available: https://www.webmd.com/diabetes/diabetic-retinopathy-stages

[5] M. Mateen, J. Wen, M. Hassan, N. Nasrullah, S. Sun, and S. Hayat, "Automatic detection of diabetic retinopathy: A review on datasets, methods and evaluation metrics," *IEEE Access*, vol. 8, pp. 48784–48811, 2020, doi: 10.1109/ACCESS.2020.2980055.

[6] D. Marín, A. Aquino, M. E. Gegundez-Arias, and J. M. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features," *IEEE Trans. Med. Imag.*, vol. 30, no. 1, pp. 146–158, Jan. 2011, doi: 10.1109/TMI.2010.2064333.

[7] K. A. Goatman, A. D. Fleming, S. Philip, G. J. Williams, J. A. Olson, and P. F. Sharp, "Detection of new vessels on the optic disc using retinal photographs," *IEEE Trans. Med. Imag.*, vol. 30, no. 4, pp. 972–979, Apr. 2011, doi: 10.1109/TMI.2010.2099236.

[8] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp, "Automated microaneurysm detection using local contrast normalization and local vessel detection," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1223–1232, Sep. 2006, doi: 10.1109/TMI.2006.879953.

[9] W. Cao, N. Czarnek, J. Shan, and L. Li, "Microaneurysm detection using principal component analysis and machine learning methods," *IEEE Trans. Nanobiosci.*, vol. 17, no. 3, pp. 191–198, Jul. 2018, doi: 10.1109/TNB.2018.2840084.

[10] B. Antal and A. Hajdu, "An ensemble-based system for microaneurysm detection and diabetic retinopathy grading," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 6, pp. 1720–1726, Jun. 2012, doi: 10.1109/TBME.2012.2193126.

[11] R. Pires, H. F. Jelinek, J. Wainer, S. Goldenstein, E. Valle, and A. Rocha, "Assessing the need for referral in automatic diabetic retinopathy detection," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3391–3398, Dec. 2013, doi: 10.1109/TBME.2013.2278845.

[12] A. Sopharak, B. Uyyanonvara, S. Barman, and T. H. Williamson, "Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods," *Computerized Med. Imag. Graph.*, vol. 32, no. 8, pp. 720–727, Dec. 2008, doi: 10.1016/j.compmedimag.2008.08.009.

[13] K. Wisaeng and W. Sa-Ngiamvibool, "Exudates detection using morphology mean shift algorithm in retinal images," *IEEE Access*, vol. 7, pp. 11946–11958, 2019, doi: 10.1109/ACCESS.2018.2890426.

[14] N. G. Ranamuka and R. G. N. Meegama, "Detection of hard exudates from diabetic retinopathy images using fuzzy logic," *IET Image Process.*, vol. 7, no. 2, pp. 121–130, Mar. 2013, doi: 10.1049/iet-ipr.2012.0134.

[15] H. F. Jaafar, A. K. Nandi, and W. Al-Nuaimy, "Automated detection and grading of hard exudates from retinal fundus images," in *Proc. 19th Eur. Signal Process. Conf.*, Aug./Sep. 2011, pp. 66–70.

[16] S. Sil Kar and S. P. Maity, "Gradation of diabetic retinopathy on reconstructed image using compressed sensing," *IET Image Process.*, vol. 12, no. 11, pp. 1956–1963, Nov. 2018, doi: 10.1049/iet-ipr.2017.1013.

[17] J. Wang, J. Luo, B. Liu, R. Feng, L. Lu, and H. Zou, "Automated diabetic retinopathy grading and lesion detection based on the modified R-FCN object-detection algorithm," *IET Comput. Vis.*, vol. 14, no. 1, pp. 1–8, 2020, doi: 10.1049/iet-cvi.2018.5508.

[18] P. Costa, A. Galdran, A. Smailagic, and A. Campilho, "A weakly-supervised framework for interpretable diabetic retinopathy detection on retinal images," *IEEE Access*, vol. 6, pp. 18747–18758, 2018, doi: 10.1109/ACCESS.2018.2816003.

[19] S. Majumder and M. A. Ullah, "Feature extraction from dermoscopy images for melanoma diagnosis," *Social Netw. Appl. Sci.*, vol. 1, no. 7, p. 753, Jul. 2019, doi: 10.1007/s42452-019-0786-8.

[20] S. Majumder and M. A. Ullah, "A computational approach to pertinent feature extraction for diagnosis of melanoma skin lesion," *Pattern Recognit. Image Anal.*, vol. 29, no. 3, pp. 503–514, Jul. 2019, doi: 10.1134/S1054661819030131.

[21] S. Majumder and M. A. Ullah, "Feature extraction from dermoscopy images for an effective diagnosis of melanoma skin cancer," in *Proc. 10th Int. Conf. Electr. Comput. Eng. (ICECE)*, Dec. 2018, pp. 185–188, doi: 10.1109/ICECE.2018.8636712.

[22] S. Majumder, M. A. Ullah, and J. P. Dhar, "Melanoma diagnosis from dermoscopy images using artificial neural network," in *Proc. 5th Int. Conf. Adv. Electr. Eng. (ICAEE)*, Sep. 2019, pp. 855–859, doi: 10.1109/ICAEE48663.2019.8975434.

[23] R. Casanova, S. Saldana, E. Y. Chew, R. P. Danis, C. M. Greven, and W. T. Ambrosius, "Application of random forests methods to diabetic retinopathy classification analyses," *PLoS ONE*, vol. 9, no. 6, 2014, Art. no. e98587, doi: 10.1371/journal.pone.0098587.

[24] R. Pires, S. Avila, H. F. Jelinek, J. Wainer, E. Valle, and A. Rocha, "Beyond lesion-based diabetic retinopathy: A direct approach for referral," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 193–200, Jan. 2017, doi: 10.1109/JBHI.2015.2498104.

[25] J. Wang, Y. Bai, and B. Xia, "Feasibility of diagnosing both severity and features of diabetic retinopathy in fundus photography," *IEEE Access*, vol. 7, pp. 102589–102597, 2019, doi: 10.1109/access.2019.2930941.

[26] M. Leeza and H. Farooq, "Detection of severity level of diabetic retinopathy using bag of features model," *IET Comput. Vis.*, vol. 13, no. 5, pp. 523–530, Aug. 2019, doi: 10.1049/iet-cvi.2018.5263.

[27] H. Wei, A. Sehgal, and N. Kehtarnavaz, "A deep learning-based smartphone app for real-time detection of retinal abnormalities in fundus images," *Proc. SPIE*, vol. 10996, May 2019, Art. no. 1099602, doi: 10.1117/12.2516665.

[28] W. M. Gondal, J. M. Kohler, R. Grzeszick, G. A. Fink, and M. Hirsch, "Weakly-supervised localization of diabetic retinopathy lesions in retinal fundus images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2018, pp. 2069–2073, doi: 10.1109/ICIP.2017.8296646.

[29] G. Quellec, M. Lamard, A. Erginay, A. Chabouis, P. Massin, B. Cochener, and G. Cazuguel, "Automatic detection of referral patients due to retinal pathologies through data mining," *Med. Image Anal.*, vol. 29, pp. 47–64, Apr. 2016.

[30] S. Gayathri, A. K. Krishna, V. P. Gopi, and P. Palanisamy, "Automated binary and multiclass classification of diabetic retinopathy using Haralick and multiresolution features," *IEEE Access*, vol. 8, pp. 57497–57504, 2020, doi: 10.1109/ACCESS.2020.2979753.

[31] B. Richey, S. Majumder, M. V. Shirvaikar, and N. Kehtarnavaz, "Real-time detection of maize crop disease via a deep learning-based smartphone app," , vol. 11401, Apr. 2020, Art. no. 114010A, doi: 10.1117/12.2557317.

[32] S. Majumder and N. Kehtarnavaz, "Vision and inertial sensing fusion for human action recognition: A review," *IEEE Sensors J.*, vol. 21, no. 3, pp. 2454–2467, Feb. 2021, doi: 10.1109/JSEN.2020.3022326.

[33] S. Majumder and N. Kehtarnavaz, "A review of real-time human action recognition involving vision sensing," in *Proc. SPIE*, vol. 11736, Apr. 2021, Art. no. 117360A.

[34] T. Chandrakumar and R. Kathirvel, "Classifying diabetic retinopathy using deep learning architecture," *Int. J. Eng. Res.*, vol. V5, no. 6, pp. 19–24, May 2016, doi: 10.17577/ijertv5is060055.
[35] D. W. Raja Memon, D. B. Lal, and D. A. Aziz Sahto, "Diabetic retinopathy; frequency at level of HbA1C greater than 6.5%," *Prof. Med. J.*, vol. 24, no. 2, pp. 234–238, Feb. 2017, doi: 10.17957/tpmj/17.3616.
[36] Z. Wang, Y. Yin, J. Shi, W. Fang, H. Li, and X. Wang, "Zoom-in-Net: Deep mining lesions for diabetic retinopathy detection," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes Artificial Intelligent Lecture Notes Bioinformatics)* (Lecture Notes in Computer Science), vol. 10435. 2017, pp. 267–275, doi: 10.1007/978-3-319-66179-7_31.
[37] A. Momeni Pour, H. Seyedarabi, S. H. Abbasi Jahromi, and A. Javadzadeh, "Automatic detection and monitoring of diabetic retinopathy using efficient convolutional neural networks and contrast limited adaptive histogram equalization," *IEEE Access*, vol. 8, pp. 136668–136673, 2020, doi: 10.1109/access.2020.3005044.
[38] X. Zeng, H. Chen, Y. Luo, and W. Ye, "Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network," *IEEE Access*, vol. 7, pp. 30744–30753, 2019, doi: 10.1109/ACCESS.2019.2903171.
[39] Y. Yang, T. Li, W. Li, H. Wu, W. Fan, and W. Zhang, "Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes Artificial Intelligent Lecture Notes Bioinformatics)* (Lecture Notes in Computer Science), vol. 10435. 2017, pp. 533–540, doi: 10.1007/978-3-319-66179-7_61.
[40] G. García, J. Gallardo, A. Mauricio, J. López, and C. Del Carpio, "Detection of diabetic retinopathy based on a convolutional neural network using retinal fundus images," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Lecture Notes in Computer Science), vol. 10614. 2017, pp. 635–642, doi: 10.1007/978-3-319-68612-7_72.
[41] K. Shankar, Y. Zhang, Y. Liu, L. Wu, and C.-H. Chen, "Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification," *IEEE Access*, vol. 8, pp. 118164–118173, 2020, doi: 10.1109/access.2020.3005152.
[42] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," *Procedia Comput. Sci.*, vol. 90, pp. 200–205, 2016, doi: 10.1016/j.procs.2016.07.014.
[43] S. Majumder, Y. Elloumi, M. Akil, R. Kachouri, and N. Kehtarnavaz, "A deep learning-based smartphone app for real-time detection of five stages of diabetic retinopathy," , vol. 11401, Apr. 2020, Art. no. 1140106, doi: 10.1117/12.2557554.
[44] Z. Gao, J. Li, J. Guo, Y. Chen, Z. Yi, and J. Zhong, "Diagnosis of diabetic retinopathy using deep neural networks," *IEEE Access*, vol. 7, pp. 3360–3370, 2019, doi: 10.1109/ACCESS.2018.2888639.
[45] S. Dutta, B. C. Manideep, S. M. Basha, R. D. Caytiles, and N. C. S. N. Iyengar, "Classification of diabetic retinopathy images by using deep learning models," *Int. J. Grid Distrib. Comput.*, vol. 11, no. 1, pp. 89–106, Jan. 2018, doi: 10.14257/ijgdc.2018.11.1.09.
[46] S. Wan, Y. Liang, and Y. Zhang, "Deep convolutional neural networks for diabetic retinopathy detection by image classification," *Comput. Electr. Eng.*, vol. 72, pp. 274–282, Nov. 2018, doi: 10.1016/j.compeleceng.2018.07.042.
[47] M. D. Abràmoff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. C. Folk, and M. Niemeijer, "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning," *Investigative Opthalmology Vis. Sci.*, vol. 57, no. 13, p. 5200, Oct. 2016, doi: 10.1167/iovs.16-19964.
[48] S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, Z. U. Rehman, I. Ahmed Khan, and W. Jadoon, "A deep learning ensemble approach for diabetic retinopathy detection," *IEEE Access*, vol. 7, pp. 150530–150539, 2019, doi: 10.1109/ACCESS.2019.2947484.
[49] C. Playout, R. Duval, and F. Cheriet, "A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2434–2444, Oct. 2019, doi: 10.1109/TMI.2019.2906319.
[50] C. Playout, R. Duval, and F. Cheriet, "A multitask learning architecture for simultaneous segmentation of bright and red lesions in fundus images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2018, pp. 101–108, doi: 10.1007/978-3-030-00934-2_12.
[51] X. Wang, L. Ju, X. Zhao, and Z. Ge, "Retinal abnormalities recognition using regional multitask learning," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2019, pp. 30–38, doi: 10.1007/978-3-030-32239-7_4.
[52] APTOS 2019 Blindness Detection Dataset. *Kaggle*. Accessed: Dec. 11, 2019. https://www.kaggle.com/c/aptos2019-blindness-detection
[53] EyePACS Dataset. *Kaggle*. Accessed: Dec. 11, 2019. [Online]. Available: http://www.eyepacs.com
[54] H. Wei and N. Kehtarnavaz, "Determining number of speakers from single microphone speech signals by multi-label convolutional neural network," in *Proc. 44th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2018, pp. 2706–2710, doi: 10.1109/IECON.2018.8592773.
[55] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807, doi: 10.1109/CVPR.2017.195.
[56] 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, doi: 10.3390/electronics9060914.
[57] 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, doi: 10.1109/ISSPIT47144.2019.9001846.
[58] M. Shaban, Z. Ogur, A. Mahmoud, A. Switala, and A. Shalaby, "A convolutional neural network for the screening and staging of diabetic retinopathy," *PLoS ONE*, vol. 15, no. 6, 2020, Art. no. e0233514, doi: 10.1371/journal.pone.0233514.

SHARMIN MAJUMDER (Graduate Student Member, IEEE) received the B.S. and M.S. degrees in electrical and electronic engineering from Chittagong University of Engineering and Technology (CUET), Chittagong, Bangladesh, in 2013 and 2019, respectively. She is currently pursuing the Ph.D. degree in electrical engineering with The University of Texas at Dallas, Richardson, TX, USA. In 2015, she joined the Department of EEE, CUET, as a full-time Faculty Member. Her research interests include signal and image processing, ultrasound elastography imaging, computer vision, pattern recognition, and machine learning.

NASSER KEHTARNAVAZ (Fellow, IEEE) is currently an Erik Jonsson Distinguished Professor with the Department of Electrical and Computer Engineering and the Director of the Embedded Machine Learning Laboratory, The University of Texas at Dallas, Richardson, TX, USA. His research interests include signal and image processing, machine learning and deep learning, and real-time implementation on embedded processors. He has authored or coauthored 11 books and more than 400 journal articles, conference papers, patents, manuals, and editorials in these areas. He is a fellow of SPIE, a licensed Professional Engineer, and the Editor-in-Chief of the *Journal of Real-Time Image Processing*.