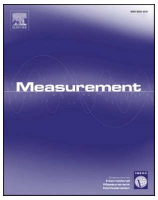
Measurement 200 (2022) 111485

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement



A novel few-shot classification framework for diabetic retinopathy detection and grading

M. Murugappan[a,b,*], N.B. Prakash[c], R. Jeya[d], A. Mohanarathinam[e], G.R. Hemalakshmi[f], Mufti Mahmud[g,h,i,*]

[a] Intelligent Signal Processing (ISP) Research Lab, Department of Electronics and Communication Engineering, Kuwait College of Science and Technology, Kuwait
[b] Department of ECE, School of Engineering, Vels Institute of Science, Technology, and Advanced Studies, Chennai, India
[c] Department of Electrical and Electronics Engineering, National Engineering College, Kovilpatti, India
[d] Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India
[e] Department of Biomedical Engineering, Karpagam Academy of Higher Education, Coimbatore, India
[f] Department of Computer Science and Engineering, National Engineering College, Kovilpatti, India
[g] Department of Computer Science, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, UK
[h] Computing and Informatics Research Centre, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, UK
[i] Medical Technologies Innovation Facility, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, UK

ARTICLE INFO

Keywords:
Diabetic Retinopathy
Detection
Grading
Aggregated transformations
Class activation

ABSTRACT

Diabetes Retinopathy (DR) is a major microvascular complication of diabetes. Computer-Aided Diagnosis (CAD) tools for DR management are primarily developed using Artificial Intelligence (AI) methods, such as machine and deep learning algorithms. DR diagnostic tools have been developed in recent years using deep learning models. Thus, these models require large amounts of data for training. Consequently, these huge amounts of data are not balanced due to fewer cases in the dataset. To solve the problems associated with training models with small datasets, such as overfitting and poor approximation, this paper proposes a paradigm called Few-Shot Learning (FSL) which uses a relatively small amount of training data to train the models effectively. This paper proposes a novel prototype network, a type of FSL classification network capable of grading and detecting DR based on attention. The DRNet framework uses episodic learning to train its model on few-shot classification tasks. We developed a DRNet based on the APTOS2019 dataset for diabetic detection and grading. In the proposed network, aggregated transformations and gradient activations of classes are leveraged to design the attention mechanism to capture image representations. As a result, the system achieves 99.73 % accuracy, 99.82 % sensitivity, 99.63 % specificity in DR detection, 98.18 % accuracy, 97.41% sensitivity, and 99.55% specificity in DR grading. An analysis of objective performance metrics and model interpretation shows that the proposed model can detect DR more efficiently and grade the severity more accurately when using unseen fundus images than existing state-of-the-art methods. Therefore, this tool could help provide a second opinion to an ophthalmologist about the severity level of DR.

1. Introduction

According to the recent statistics from the International Diabetes Federation, in 2019, there are 463 million people affected by Diabetic Mellitus (DM) in the age group of 20 – 79 years. It is expected to reach 700 million by 2045. Besides, 3 in 4 diabetic people live in low- and middle-income countries [1]. Management of DM is becoming a huge challenge for developing and developed countries globally. DR is the most common diabetic complication, a microvascular disorder of DM that causes visual impairment and blindness. According to international clinical standards, DR is categorized into four stages, namely, mild non-proliferative diabetic retinopathy (mnDR), moderate non-proliferative diabetic retinopathy (monDR), severe non-proliferative diabetic retinopathy (SnDR), and severe proliferative diabetic retinopathy (SPDR). At present, Early Treatment Diabetic Retinopathy Study (ETDRS) [2] scale is used for severity assessment of DR, which classifies the DR stage based on the presence of Diabetic Macular Edema (DME), neovascularization, hemorrhages, and exudates.

[*] Corresponding author.
E-mail addresses: m.murugappan@gmail.com (M. Murugappan), muftimahmud@gmail.com, mufti.mahmud@ntu.ac.uk (M. Mahmud).

https://doi.org/10.1016/j.measurement.2022.111485
Received 12 January 2022; Received in revised form 12 May 2022; Accepted 11 June 2022
Available online 9 July 2022

Early detection of DR and its severity level is essential to initiate clinical interventions to avoid adverse outcomes such as blurred vision, eye floaters, and vision loss. Identification of DR and grading through manual examination is a highly time-consuming error-prone process that requires highly skilled ophthalmologists to accurately evaluate the effects of DR. Pioneering works on automated [3] DR detection is based on fundus photography, as evident from the works reported in two decades ago [4–6]. However, fundus photography is an invasive technique that requires pupil dilation and is therefore impractical in elderly patients or those with poor manual dexterity. Furthermore, repeated dilation can cause ocular discomfort and lead to long-term complications, such as mydriasis, cataract, and retinal detachment.

Several machine learning models for DR diagnosis have been proposed in the early years of automated detection. In [7], discrete classifier models, including k-Nearest Neighbor (kNN), Gaussian Mixture Model (GMM), and Support Vector Machine (SVM), trained with handcrafted features extracted with AdaBoost algorithm are employed in the classification of DR and non-DR lesions. Further, a three-stage model called DREAM [8], comprising image segmentation, lesion classification from extracted features, and severity grading demonstrates superior sensitivity compared to baseline models. In this line, an ensemble [9] model with five base classifiers viz, AdaBoost, Decision Tree (DT), kNN, Logistic Regression (LR), and Random Forest (RF) classifiers is demonstrated to surpass the performances of the individual classifier models. Though the computational requirements of machine learning models are competitive, these models based on handcrafted features are not capable of learning and incorporating novel features from the training data, and they are subject to overfitting.

Recently, the evolution of ocular imaging technologies such as optical coherence tomography (OCT), confocal scanning laser ophthalmoscopy, etc., and the emergence of deep learning algorithms have simplified the process of DR screening through CAD systems based on retinal images [10]. A two-stage hybrid model proposed in [11] combines Convolutional Neural Network (CNN) for feature extraction and conventional machine learning approaches for classification. Experimental results with SVM classifiers trained on features extracted from retinal images with four different CNNs show that the Inception-v3 [12] can capture the most discriminating features. Most recent Deep Learning-based DR screening systems have attempted to detect microaneurysms by analysing the image content [13]. While these methods show promise, they also suffer from several limitations, including an imbalance in the number of healthy and unhealthy images. In addition, they have an insufficient number of images to train the networks and diverse morphologies and anatomical locations of the features. Therefore, it cannot be realistic to expect the same number of images from the same patient for each class. This results in the deep network learning only the typical pattern for each class, ignoring the individual differences of the patients.

It would be extremely difficult to collect a large number of samples of clinical information for any given application in a real-life scenario. In general, to develop a more robust CAD system for medical applications, most contemporary artificial intelligence methods, such as Deep Neural Networks (DNN), require a large amount of data. This problem could be solved with the FSL approach [14], where a single model is trained with the training images for a specific class and then tested with the unseen images of the class. Unlike traditional deep learning models, FSL models can learn the class-specific patterns from a few sample images. Attention is a powerful mechanism employed in deep learning to address the problem of long-term dependencies. In image processing problems, it is used to focus on a specific region of an image to capture significant features. This research employs this mechanism to capture intricate features for DR detection.

This paper presents a novel FSL-based framework called DRNet for DR detection and grading using an attention-based *meta*-learning mechanism. This framework is a prototypical network that constructs a *meta*-classifier from several base classifiers, trained on smaller subsets of the training data.

The major contributions of this research are:

1. We have designed and developed a novel prototypical network called DRNet with an inbuilt attention mechanism for DR detection and grading. The proposed model achieved a higher DR detection rate and grading accuracy than the state-of-the-art methods reported in the literature.
2. We have also proposed a mechanism for constructing the image embeddings with Gradient Class Activation Maps (GCAMs) and aggregated transformations for FSL.
3. The proposed DRNet has been trained and fine-tuned with the hyperparameters in multiple episodes on the open-source Asia Pacific Tele-Ophthalmology Society (APTOS2019) [15] dataset, and it exhibits superior performances compared to conventional deep learning models.

The system achieves an accuracy of 99.73%, a sensitivity of 99.82%, and a specificity of 99.63% in DR detection, and 98.18% accuracy, 97.41% sensitivity, and 99.55% specificity in DR grading. These results indicate the ability of the model to discern different types of DR unambiguously.

The paper is structured as below. In section 2, existing works on DR detection, grading, and FSL mechanisms are reviewed. Section 3 presents the dataset and the underlying methods employed in this research. The proposed prototypical network architecture and the training process are discussed in section 4. Experimental results with interpretations, comparative and explainable analyses, and advantages and limitations of the model are presented in section 5, and the paper is concluded in Section 6.

2. Related works

This section presents a comprehensive review of deep learning-based DR detection and grading models and briefly accounts for the FSL approach. Convolutional neural networks (CNN) are primarily used for image recognition and classification. Many researchers have demonstrated that deep CNN models can effectively solve complex real-life problems of diverse nature [16,17]. For feature extraction and classification of fish species into four categories, a VGG16-based CNN is used [18]. Based on the training samples, the model extracts hierarchical features that identify each type of fish and achieves a mean classification accuracy of 100%. The study found that the model's performance improves as the CNN depth increases.

Further, deep CNN models have been used in building intelligent systems in renewable energy [19]. Further, LSTM-based networks have been used in solving classification, prediction, and detection problems with time-series data, including model developments for thermal processes [20]. Due to their ability to extract and learn image features, flexible and scalable architectures, and portability, CNNs are essential in medical imaging-guided clinical interventions.

A detailed survey on deep learning-based DR detection models is presented in [21]. The pioneering work in deep learning-based DR detection was proposed in [22], which used a CNN with several convolutional blocks in the initial layers and fully connected layers as the final classification layers. This model achieves 75% accuracy and 95% sensitivity in DR grading with five classes (Non-Diabetic Retinopathy (NDR), mnDR, monDR, SnDR, and SPDR). An investigation by Gulshan et al. [23] employing the Inception-v3 for DR detection reportedly achieves 97.5% sensitivity and 93.4% specificity for the EyePACS [24] dataset and 96.1% sensitivity and 93.9% for Messidor-2 [25] datasets. A DR detection framework proposed in [26] employs a customized residual deep learning network for feature extraction, and a DT is used as a classifier. Here, the DT classifier is trained with meta image data appended with these features for binary classification (DM or Normal) of fundus images, and a maximum mean sensitivity of 93% and specificity of 87% are reported on the Messidor-2 dataset.

Similar to the model proposed in [23], an Inception-v3-based architecture is proposed in [27] for DR and macular edema grading, achieving an Area Under Curve (AUC) of 0.987, sensitivity and specificity of 89.6%, and 97.4% in referable DR detection, respectively. The CANet [28] performs joint grading of DR and DME with separate disease-specific and disease-independent attention networks. The image features maps are initially extracted with ResNet [29] and fed simultaneously to the subnetwork branches. The disease-specific attention modules leverage the relationship between features across channels and spatial locations to capture the disease characteristics. The disease-independent attention modules aggregate the channel features from the branches. This model achieves a joint accuracy of 85.1% in DR and DME grading. The Weighted Path Convolutional Neural Network (WP-CNN) [30] employs a weighted-path strategy to capture discriminative features, eliminating redundancies in referable DR detection. This approach enhances the network's attention to vital features, achieving an accuracy, sensitivity, and specificity of 94.23%, 90.94%, and 95.74%, respectively. A Region-based Fully Convolutional Neural Network (R-FCN) [31] is used for DR grading and lesion detection, and it is realized by modifying the ResNet-101 network for feature extraction and a Region Proposal Network (RPN) for DR detection. The R-FCN achieves a sensitivity of 92.59% and specificity of 96.20% on the Messidor dataset. However, the ability of the model to detect smaller lesions such as microaneurysms and hemorrhages is found to be less due to the lack of annotation in the training samples. A patch-based DR model employs a customized CNN as a selection model to process image patches to localize red lesions and generate the lesion probability map [32]. DR detection is performed by deriving a probabilistic value from this map. The CF-DRNet [33] for five-stage DR classification employs two subnetworks, a coarse network for DR detection and a fine network for grading the images that are classified as DR positive by the coarse network. Both the networks are modeled on the ResNet18, and an attention module is employed in the coarse network to capture the discriminative features for DR detection. However, this framework achieves a detection accuracy of only 56.19% on the IDRiD [34] dataset.

Synergic Deep Learning (SDL) is employed in a DR severity grading model which utilizes the histogram-based segmentation of the Region of Interest (RoI) in the medical images, followed by classification with a synergic network [35]. This framework employs two deep CNNs to process training images in parallel and the synergic labels generated by these networks are fed to the synergic network for DR grading. This model reportedly achieves a maximum mean classification accuracy of 99.28%, a sensitivity of 98.54%, and specificity of 99.38% on the Messidor dataset. However, this paper does not explicitly specify which pre-trained models are used in the first stage of classification. The two-stage hybrid architecture performs DR detection in the first stage, followed by four-class grading of images in the second stage on the images classified as DR in the first stage [11]. DR detection network is realized by transfer learning, by fine-tuning the pre-trained networks. DR detection with seven such pre-trained networks shows that the Inception-v3 achieves the best AUC of 0.993 and accuracy of 98.4% on the APTOS2019 dataset. DR grading is performed with AlexNet, VGG16, ResNet, and Inception-v3 as feature extractors, and Principal Component Analysis (PCA) for dimensionality reduction and training an SVM classifier with these features.

The most recent work in this context, DeepDR [36] consists of a base network and three subnetworks for image quality assessment, lesion segmentation, and DR grading. The DR base network is constructed from the pre-trained ResNet, and the weights of this network are shared with the subnetworks. A test image is given as input to each of the subnetworks to perform the specific tasks. The features extracted by the lesion-aware subnetwork are concatenated with those extracted by the DR grading network for classification. This network performs six stages of DR grading including NDR, Mild NPDR, moderate NPDR, severe NPDR, proliferative DR, and referable DR. This model achieves the best performance metrics for PDR detection such as 0.961 of AUC, 93.2 %of sensitivity, and 86.2% of specificity. However, DeepDR is relatively complex compared to the rest of the networks proposed in the literature, due to the constituent subnetworks.

In recent years, FSL based classifiers are slowly replacing conventional learning models owing to their ability to learn from few training samples.

There are several *meta*-learning frameworks reported in the literature to construct a classifier model, aggregating the base classifiers trained on data subsets [37]. A prototypical [38] network is a kind of *meta*-learning framework with the capability of learning from multiple instances of a task and can be used for classification, detection, segmentation, etc., with the goal of generalization. An attentive prototype learning network based on capsule networks for image embedding is demonstrated to achieve superior classification performances compared to that of baseline classifiers [39]. However, the application of FSL-based learning models in medical imaging-based diagnostics is extremely limited and there have been comparatively few investigations reported in the recent past. An FSL framework for DR detection and other common pathologies is based on a two-stage network consisting of a multi-task detector for detecting common pathologies and a probabilistic detector for detecting rare diseases [40]. Experimental results show that the best AUC value of 0.966 is achieved with the Inception networks for the classification of frequent conditions. Further, the learning and inference pipeline includes PCA projection and K Nearest Neighbour (KNN) regression. The t-distributed Stochastic Neighbour Embedding (t-SNE) method employed for dimensionality reduction in this model is also computationally expensive as the data must be log-transformed. FEDI [41], a recent FSL model based on a deep residual network and Earth Mover's Distance (EMD) algorithm performs classifications on 39 categories in a 1000 sample fundus image dataset. The residual network is used as feature extractor and the EMD algorithm is used to match the image features. Experimental results show classification accuracy of 95.87% is achieved in 3-way 10-shot classification. However, the authors do not present explicit results for the classification of 39 classes in their work.

Considering the literature reviewed above, it is evident that attention-based neural networks, which focus on significant image features, yield the best results for the detection and grading of DR. Additionally, existing FSL models for DR detection require more computational power, are inherently complex in design, incorporate other machine learning approaches like SVM and statistical exploratory techniques like PCA, and incorporate other machine learning approaches such as SVM. Therefore, the need for developing a less complex FSL framework for the detection and grading of DRs has become highly apparent.

3. Materials and methods

This section describes the dataset and the methods used in building the proposed model.

3.1. Database description

In this present work, we have used the open-source DM database for developing our model for DM severity detection using an FSL approach. The APTOS2019 [15] dataset comprises 5590 images and is grouped into training (3662 images) and testing (1928 images) datasets. Here, only the training datasets are labeled with different stages of DR, and testing images are not labeled. In this work, we have selected 3662 training images to design and develop a DRNet for DR detection and grading. The training dataset is organized into five classes namely NDR, mild DR (mnDR), Moderate DR (monDR), Severe DR (SnDR), and Severe ProliferativeDR (SPDR) with 1805, 370, 999, 295, and 193 images, respectively. The total number of training images is split into the training (2564 images) and testing (1098 images) set based on a

70:30ratio. In this work, we have performed binary classification (DR vs NDR) and multi-class (mmDR, monDR, SPDR, SnDR, and NDR). For binary classification, the images in four classes other than NDR are grouped under the DR class. The distribution of images used for developing our model is given in Table 1. All the images in the dataset are having a resolution of 3216 × 2136 in*.png* format and the images are resized to a lesser resolution of 256 × 256 to reduce the computational complexity of our proposed system.

3.2. Few-Shot learning

Conventional machine learning models are trained with large volumes of labeled data for classification. However, the availability of large amounts of labeled data is generally not feasible in many applications which are related to real-world problems or scenarios. This has led to the emergence of FSL [42], which aims to rectify this gap by leveraging *meta*-learning approaches to train a classification model with a limited number of samples. FSL methods can learn a representation of classes that can be used to generalize to new classes. In image classification [43] problems, FSL systems are trained to learn a function that transforms an input image into a class label, starting from an input feature representation, i.e., a high-level representation of images. FSL framework has been described as *meta*-learning mainly because it aims to develop a function that can be used to generalize the prediction of new classes without having direct access to the output layer of the model.

The *meta*-learning frameworks are divided into metric learning techniques [44] that minimize a distance function to learn the mapping between feature space and output layer and embedding techniques [45] that use feature embedding to learn the mapping between feature space and output layer. A metric learning approach is highly preferred in FSL since the mapping is generally learned unconstrainedly without the assumption of a pre-defined distance function between the classes. They capture high-level image representations of classes that are invariant to specific data distributions, providing robust predictions across domains and tasks.

The FSL problem is based on metric learning, and it is formulated to learn a mapping function from input data to class targets such that, the output distance is smaller for input–output pairs of similar classes and larger for input–output pairs of dissimilar classes. For a feature space X, this problem can be expressed as in Eq. (1).

$$\min_{f \in F} \sum_{(x,y) \in X \times Y} d(f(x), y)^2 \tag{1}$$

where,$d(\cdot, \cdot)$ is a distance metric on the input space, x is an instance of X, Y is a finite set of labels, y is an instance of Y and $f(x)$ is the transforming function.

Table 1
Dataset Distribution.

Class	DR Detection		DR Grading	
	No. of Training Images	**No. of Testing Images**	**No. of Training Images**	**No. of Testing Images**
NDR	1264	541	1264	541
DR	1300	557	–	–
Mild DR (mnDR)	–	–	259	111
Moderate DR (monDR)	–	–	699	300
Proliferative DR (SPDR)	–	–	207	88
Severe DR (SnDR)	–	–	135	58
Total	**2564**	**1098**	**2564**	**1098**

3.3. Prototype networks

Prototypical networks are based on the premise that there exists an embedded function that maps data points to a compact space, such that the similarity between any two data points can be computed by the distance between their corresponding embeddings. For a given support set of N labeled examples$S = \{(x_i, y_i)\}_{i=1}^{N}$, a prototypical network is a function $f_\theta(x)$ that maps data points to a compact space such that the similarity ($\mathscr{S}_\theta$) between any two data points can be computed by the distance between their corresponding embeddings: $\mathscr{S}_\theta(x_i, x_j) = \|f_\theta(x_i) - f_\theta(x_j)\|^2$. For each input class c_i, prototypical networks compute an embedding vector $e_{c_i} \in \mathbb{R}^d$ for that class. The prototype of a class c_i is the average of all the embeddings of the examples in the support set of the class c_i as given in Eq. (2).

$$e_{c_i} = \frac{1}{|S|} \sum_{x_i \in S} c_i \tag{2}$$

Given an input x belonging to a class c_i, the similarity between x and the prototypes of all the classes c_i is computed as $S_\theta(x, c_i) = \|f_\theta(x) - e_{c_i}\|^2 \forall i = 1, 2, \cdots C$, where c_i is class instance and C is the number of target classes. Finally, the network prediction is defined as the average of the similarities to all the prototypes of the classes in the support set of the input, as given in Eq. (3).

$$f_\theta(x) = \frac{1}{|S|} \sum_{i=1}^{|S|} S_\theta(x, c_i) e_{c_i} \tag{3}$$

During inference, given an input class c that has never been seen before, the network predicts the probability of the input being of class c with the Eq.(4).

$$P(c|x) = \frac{\exp(S_\theta(x, c))}{\sum_{j=1}^{|S|} \exp(S_\theta(x, c_j))} \tag{4}$$

For each class c_i, a prototypical network is trained with a set of labeled examples $S_{c_i} = \{(x_i, y_i)\}_{i=1}^{N'}$ and their prototypes $e_{c_i} = \frac{1}{|S_{c_i}|} \sum_{x_i \in S_{c_i}} e_i$, where N' is the number of samples in each class. The network parameter θ is optimized to minimize the cross-entropy between the predicted probability and the ground truth and the network is used for inference. The loss function is defined as in Eq. (5).

$$Loss(\theta) = \left(\sum_{i=1}^{|S_*|} -P(c_i|x_i) \exp(S_\theta(x_i, c_i) e_{c_i}) \right)^2 \tag{5}$$

where,$S_* = \{(x_i, c_i)\}_{i=1}^{N'}$.

The number of examples in the support set can differ for each input class depending on the training set size.

3.4. ResNeXt architecture

ResNeXt [46] is an extension of the conventional ResNet architecture, inspired by Inception networks. It is a highly modularized residual network architecture based on an aggregation of a set of similar modules, that can be used to model different types of relationships between the image regions. ResNeXt follows the split-transform-merge approach, where the network splits the input into several lower-dimensional embeddings, transforms the embeddings with a stack of convolutional layers, and concatenates the transformed embeddings. The final output is obtained by concatenating the input with the merged embeddings. The schematic of a ResNeXt block is shown in Fig. 1. It can be used to model different types of relationships between image regions, such as the relationship between the whole image and the parts of the image, the relationship between the parts of the image, and relationships between different parts of the image at multiple scales. This architecture

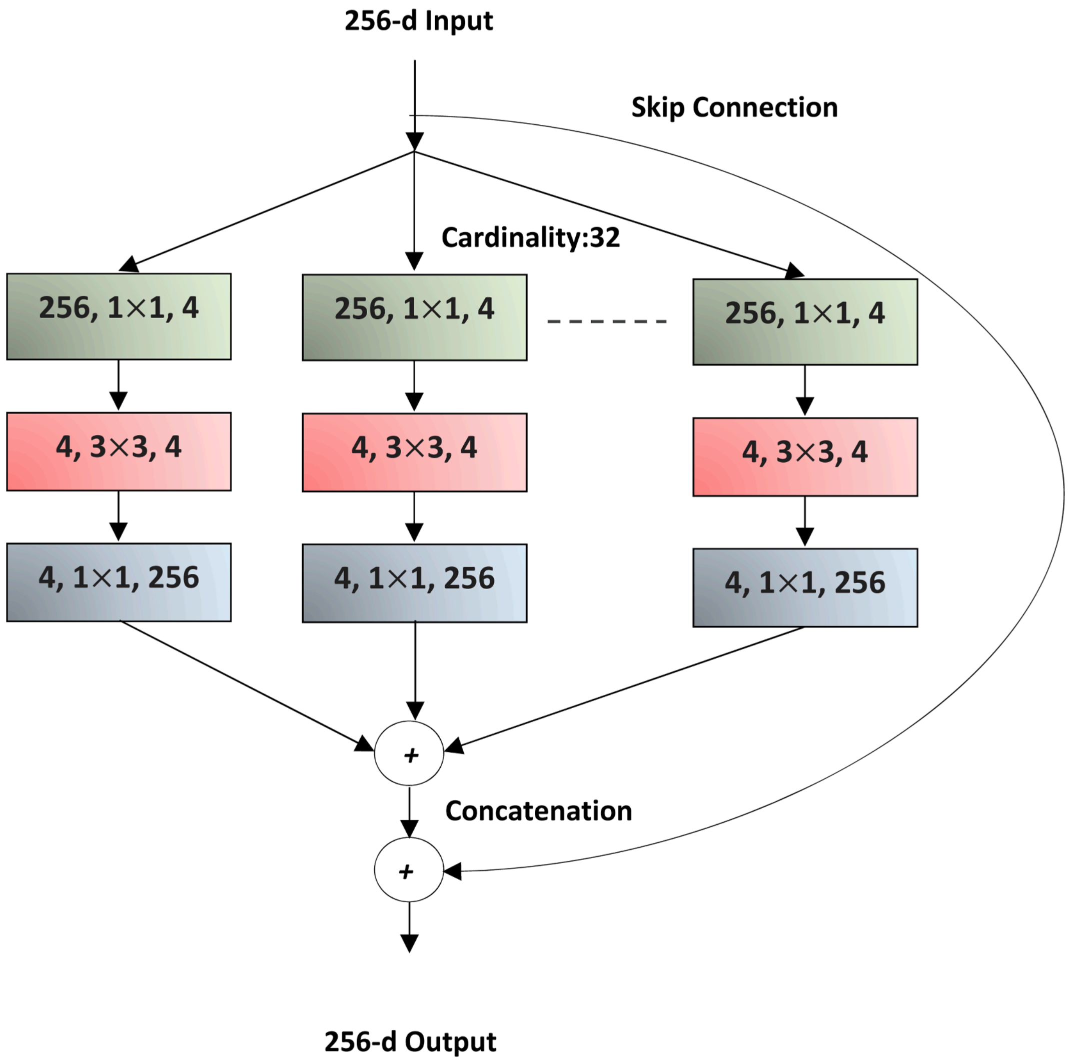


Fig. 1. ResNext Block. (Each layer is shown with # Input Channels, Filter Size, # of Output Channels).

introduces a new parameter called cardinality, the size of the set of transformations to be concatenated. Exercising a ResNeXt with different cardinalities shows that, the value of cardinality affects the quality of the predictions, with a higher value of cardinalities resulting in better predictions. Fig. 1 shows a 256-d input split into 32 paths (cardinalities), each with a stack of three convolutional layers for transforming the input and creating a final embedding by concatenating the outputs of these stacks with the original input.

4. Proposed system

This section presents the mathematical description of the DR detection and grading problems and describes the architecture of the proposed DRNet with schematic diagrams.

4.1. Problem definition

The DR detection and grading are formulated as binary and multi-class classification problems respectively. In binary classification, a binary classifier is trained to detect the presence or absence of DR in a test image. This problem is formulated as follows: Given a set of training images, each image has been annotated as DR or NDR. The binary classifier is trained to predict whether an arbitrary test image is DR or NDR.

Given a training set S of input-label pairs, where $S = \{(x_i, y_i)|x_i \in \mathbb{R}^d; y_i \in \mathbb{R}\}$, the task of the binary classifier is to learn a decision function $f_b(x)$ that maps a new test image x to a binary label. Here, the decision function is defined as in Eq. (6), where $\delta(x)$ is the predicted DR probability for x by the classifier.

$$f_b(x) = 1[\delta(x) > 0] \tag{6}$$

The multi-class classification problem is formulated as follows: A multi-class classifier is trained for five-class classification to discriminate NDR, mild, moderate, proliferative, and severe DR. Given a set of training images, where each image has been annotated with its DR severity, the multi-class classifier is trained to predict the severity of DR in the test image.

Given a training set S of input-label pairs, where $S = \{(x_i, y_i)|x_i \in \mathbb{R}^d; y_i \in \mathbb{R}\}$, is a training set of input images and labels, the task of the multi-class classifier is to learn a decision function $f_m(x)$ that maps a new test image x to a label as in Eq. (7),

$$f_m(x) = \arg \underset{k=1,\cdots,5}{max} [\delta(x) = k] \tag{7}$$

where $\delta(x)$ is the predicted DR severity grade for x and k is the number of classes.

4.2. Prototypical DR detection network

The proposed DR detection system is realized as a prototypical network. Initially, the training set (2564 images) is divided into support sets and query sets. The support set is further divided into two parts such as episodic learning, and embeddings of the support. This model is realized as a two-class two-shot and five-class two-shot classification network for binary and multi-class classifications, respectively. This model comprises a base layer, *meta*-layer, and classification layer, and embeddings of the support and query sets are constructed with an embedding module. The schematics of the prototypical architecture for five-class two-shot learning and the embedding module are shown in Figure 2.

Attention of the network to significant regions of the fundus images is realized with the embedding module. This module is designed to construct a representation of the input fundus images with the convolutional layers of the ResNext-50 network. This network is used for creating embeddings and classification as shown in Fig. 2b. The proposed DNN has five convolutional layers, in which each layer is stacked with different numbers of convolutional blocks and each block is configured with a diverse set of convolutional filters. The embedding of a fundus image is constructed from the gradients of the activation map of the final convolutional layer using the Grad-CAM [47] approach. This representation captures the significant regions of the input image which influence the classifier decision. As seen in Fig. 2b, the ResNext architecture consists of a Global Average Pooling (GAP) layer and a Fully Convolutional (FC) layer following the five convolutional layers. The base classifier is realized with the GAP and FC layers. As shown in Fig. 2a, the base classifiers are trained with the image embeddings of the training sets for classification, producing prototypes for each class. The prototypical network is constructed from these prototypes and is fine-tuned with the training set images for the given classes. In this manner, the prototypes are the parameters of the network and are trained with the support and query images of the training set. For example, the prototype for DR is the mean of the images with this label in the training set. For classification of each test image, the prototype is passed to an FC layer and the class with the highest activation score is chosen.

According to the schematic diagrams, it is evident that the DRNet is

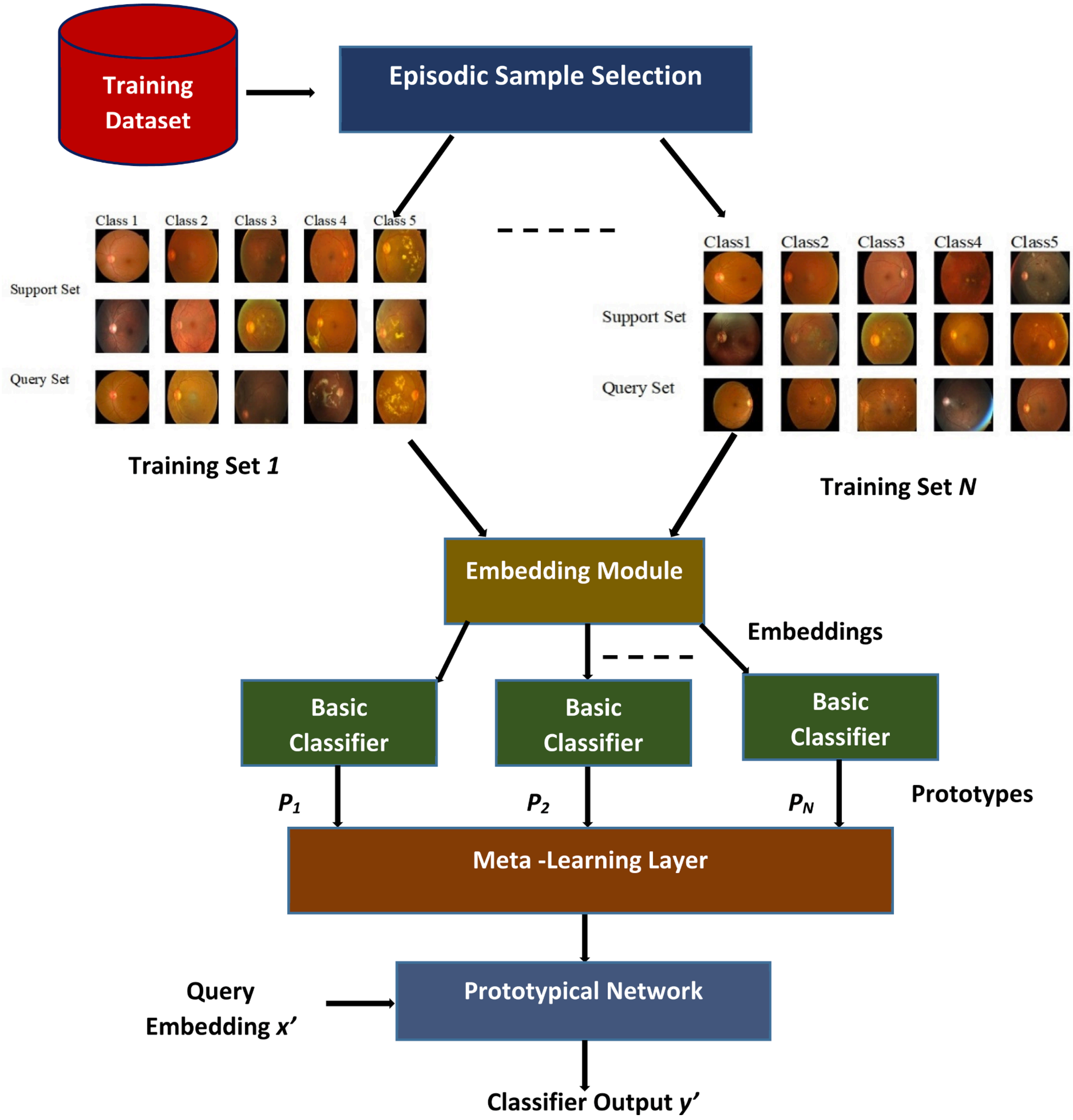


Fig. 2a. Prototypical DRNet.

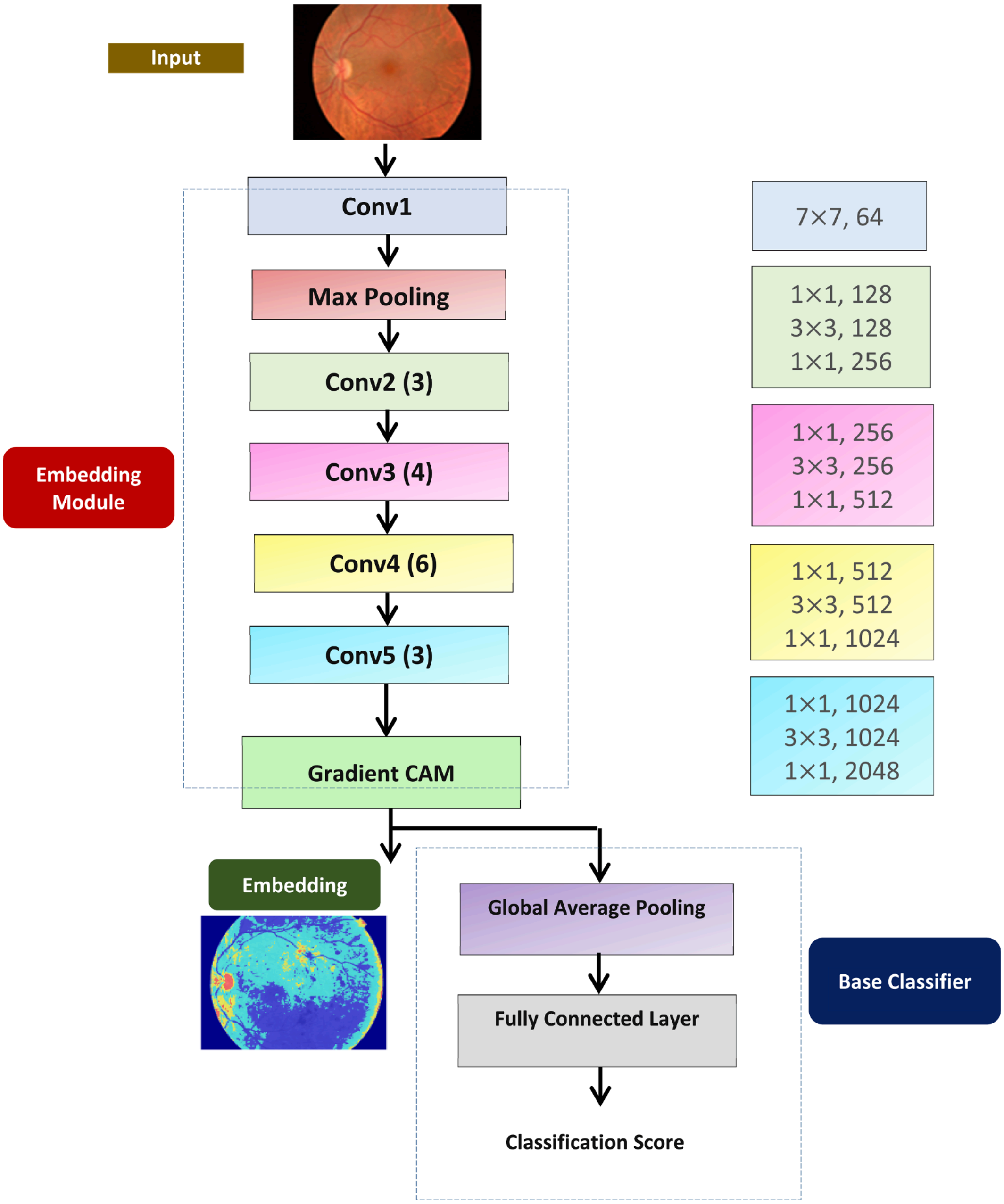


Fig. 2b. Embedding module.

not implemented with any additional explicit attention mechanism. However, it is attentive in virtue of the components it contains. Both the prototypical network and the ResNext are intrinsically attentive. The prototypical network learns from a compact image embedding carrying significant image features, and ResNext is capable of capturing the relationships between image components at multiple scales. With the prototypical network and the ResNext embedding module, the ability of the DRNet to focus on significant image features has been improved.

4.3. Prototypical DRNet training

Episodic learning is a standard training strategy in prototypical networks to learn a function f by minimizing the loss function with the overall samples as given in Eq. (8).

$$L_E = \frac{1}{\sum_i s_i} \sum_{(x,y)_i \in E} \frac{1}{s_i} ||y - f(x)||^2 \tag{8}$$

Given a training set S of input-label pairs, where $S = \{(x_i, y_i) | x_i \in \mathbb{R}^d; y_i \in \mathbb{R}\}$, training episodes are formed by randomly selecting an example x and its label y. The function f is defined to be the mean of all s -sized subsets of the input x and is given by: $f(x) = \frac{1}{s}\sum_{S\subseteq x} s|S|$. It is possible to learn the function f using the loss function for episodic learning. In this work, we have used the loss function to form episodes as above and then use those episodes to update f as given in Eq. (9).

$$f(x_n) = \frac{1}{\sum_i s_i} \sum_{S\subseteq x_n} s_n |S| \tag{9}$$

If an episode is formed by selecting the set x_n, where s_n is the number of samples in x_n, the update rule gives f and that is the average of all subsets of x_n.

The episodic learning procedure is as below.

1. Generate random samples $E = \{(x_i, y_i)\}_{i=1}^{m}$

2. Find s -sized subsets of E
3. Learn the function f by minimizing the episodic loss L_E
4. Generate a new sample, x_n and find s_n-sized subsets of x_n
5. Learn the function f by minimizing the update rule: $f(x_n) = \frac{1}{\sum_i s_i}\sum_{S\subseteq x_n} s_n|S|$.

It is worth noting that the learning algorithm can be extended to learning a mixture of prototypes. using a mixture of prototype functions as given in Eq. (10), where each p_j is a prototype.

$$f(x) = \sum_{j=1}^{k} p_j f_j(x) \tag{10}$$

Initially, the episodes are formed by selecting x_n and the s_n-sized subsets of x_n and then the prototype functions f_j is learned using the episodic learning procedure. For this, the prototype functions are trained one at a time, with each prototype function learning using the update rule. The learning update rule for a prototype function is given by Eq. (11).

$$f_j(x_n) = \frac{1}{\sum_i s_i}\sum_{S\subseteq x_n} s_n|S| \tag{11}$$

Further, a mixture of prototype learning can be done by learning all prototype functions, f_j, simultaneously using the update rule given in (12).

$$f(x_n) = \sum_{j=1}^{k} p_j f_j(x_n) \tag{12}$$

The DRNet is trained by episodic learning as described above. The hyperparameters for training the network are given in Table 2.

These parameters are optimized by grid search, with a grid G of hyperparameters, constructed by sampling the parameter space. The hyperparameter space is divided into N equal sized intervals, and for each interval, K samples are taken at random. The hyperparameter grid G is the union of all the sampled hyperparameter values. A typical training episode is given in Algorithm 1.

Algorithm 1. *Episodic Training Algorithm for Parameter Optimization.*

Input: Episodic input samples X_e,

Output: Episodic output samples Y_e, Episodic history of values for each parameter combination H_e, Episodic reward S_e, Model parameters.θ

```
i. Select $K$ samples from $X_e$
ii. Repeat until convergence:
iii. for each episode $e$:
iv. Predict the label $y_e$ for instance of the input $x_e$
    $y_e = f(x_e)$
v. Compute the loss for each set of samples
    $\mathscr{L} = \frac{1}{K}\sum_{n=1}^{K}[(1 - y_e^n)\log y_e^n + (y_e^n)\log(1 - y_e^n)]$
vi. Compute the Episodic reward
    $S_e = \mathscr{L} + \alpha_s \mathscr{S}(y_e, H_e)$
vii. Record history for each parameter combination
    $H_e^t = H_e^{t-1}, S_e^t = S_e^{t-1}$
viii. Set learning parameters to the model parameters from the sampled parameter grid. Store reward and learning parameters
    1. $S_e^{t+1} = S_e^t + \gamma S_e^{t+1}$
    2. $\theta^{t+1} = \theta^t - \mu\frac{\partial\mathscr{L}}{\partial\theta}$
ix. End
x. Go to step 2
```

Table 2
Training Hyperparameters.

Parameter	Values
Maximum Epochs	100
No. of Episodes	100,75,50
Momentum	0.9000
Learning Rate	0.001
Optimization	SGDM
L2 Regularization Parameter	0.001
Cardinality (ResNext)	32

Compared to the conventional randomized search approaches for model optimization, the grid search optimization employed in this research performs an exhaustive search and optimizes the hyperparameters, ensuring that the model converges for each episode.

5. Experimental results and discussions

This section presents the experimental setup for deploying DRNet, DR detection and grading results, comparative analyses, Explainable Artificial Intelligence (XAI) analysis of the proposed prototypical network, merits, and drawbacks of the model. The proposed model is implemented with Matlab 2021b software employing the deep learning and image processing toolboxes. It is trained and tested with a 64-bit *i7*-7700 K processor, with 4.5 GHz CPU speed, equipped with 32 GB RAM, and NVIDIA GeForce GTX 1080 GPU. Initially, the model is fine-tuned with a subset of the training and testing datasets under episodic learning until the loss function reaches the minimum.

5.1. DR detection and DR grading

The classification performances are evaluated with the objective metrics such as accuracy, sensitivity, specificity, precision, F1 score, and Matthews Correlation Coefficient (MCC). These metrics are based on the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values of the classifiers on the test dataset.

- **Accuracy**, a measure of the overall accuracy of a model, is defined as the ratio of the number of correctly classified instances to the total number of instances as given in Eq. (13).

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{13}$$

- **Sensitivity** is the measure of the performance of a model on predicting positive instances. It is expressed as the number of correctly detected positive samples out of the total number of positive samples, computed with the equation given in Eq.(14).

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{14}$$

- **Specificity** is a measure of the performance of a model on predicting negative instances. It is the ratio of the correctly predicted negative instances out of the total number of negative samples as in Eq.(15).

$$Specificity = \frac{TN}{(TN + FP)} \tag{15}$$

- **Precision** is the number of samples correctly identified out of the total number of positive samples predicted as in Eq.(16).

$$Precision = \frac{TP}{(TP + FP)} \tag{16}$$

The above metrics range from 0 to 1, signifying the worst and best performances of the model, respectively.

- **F1 score and MCC metrics** have been shown to better discriminate classifiers models under imbalanced dataset scenarios. F1 score gives equal weight to both precision and sensitivity metrics to provide a balance between the two. The F1 score can be computed using the equation given in Eq. (17) and it ranges from 0 to 1. The smallest value (0) refers to the worst performance of the classifier and the highest value (1) refers to the best performance of the classifier.

$$F1 = 2^{*}\frac{Precision \times Recall}{(Precision + Recall)} \tag{17}$$

- The MCC is defined as the harmonic mean of precision and sensitivity, measuring the correlation between the predicted output of the model with actual classes. By using the value of TP, TN, FP, and FN, the MCC can be computed using the equation as given in Eq. (18). It is a generalization of the well-known Pearson correlation coefficient (PCC), the value of MCC usually ranges from −1 to 1. The value of MCC closer to 1 indicates that the model classifies the positive and negative samples with equal accuracy otherwise with different accuracy.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{18}$$

The mean values of classification performance metrics of the proposed model are given in Table 3. It is seen that the performance of the binary classifier is better compared to the multi-class classifier, which shows that the proposed model is good in DR detection compared to severity grading.

The performances of the model are depicted with the confusion matrices in Figure 3. It is seen that, in DR detection, only a few misclassifications are evidenced with DR and NDR classes (Fig. 3a). Further, in DR severity grading, misclassifications are evidenced with all classes and the errors are highly pronounced with 6.7% for severe DR, followed by 4.5% for proliferative DR, 4.4% for mild DR, 1.7% for moderate DR, and 0.4% for NDR (Fig. 3b).

Due to the lower number of training samples (135 less than other classes as can be seen in Table 1), the high error rate for severe DR is attributable to the low number of samples over the class. Often, in multi-class classification problems, when the number of training samples is relatively small, a classifier may have difficulty distinguishing the target class from the other classes. It is very likely that the classifier will not be able to learn the distinguishing features of these classes from a few training samples. This is because it cannot generalize well to unknown data. Conventional deep learning models can resolve this issue through data augmentation. Nevertheless, the class imbalance issue in FSL is complex as it manifests at the *meta*-dataset or task level. Several FSL rebalancing techniques have been developed over the years, including random sampling, random shot *meta*-learning, and loss function rebalancing. However, the implementation and evaluation of the effectiveness of these approaches are extremely challenging because of the computational overhead.

Fig. 4 shows the Receiver Operator Characteristics (ROC) curves of the binary and multi-class classifiers. As indicated by the curve, the number of true positives can be plotted against the number of false positives in a continuous-valued feature space. AUC quantifies the accuracy of the classifier as it pertains to false positives. The optimal operating point (OOP), the point where the ROC curve reaches the upper-left corner, represents the balance between false positives and false negatives that is optimal. At this point, the number of true positives is equal to the number of false positives, and the classification process is optimal. AUC ranges from 0.5 (random chance) when the classifier cannot reliably differentiate the two classes, to 1.0 for accurate classification. The AUC value for binary and multi-class classifications in this research is 0.9999 and 0.9879, respectively. These values are consistent with the results presented in Table 3.

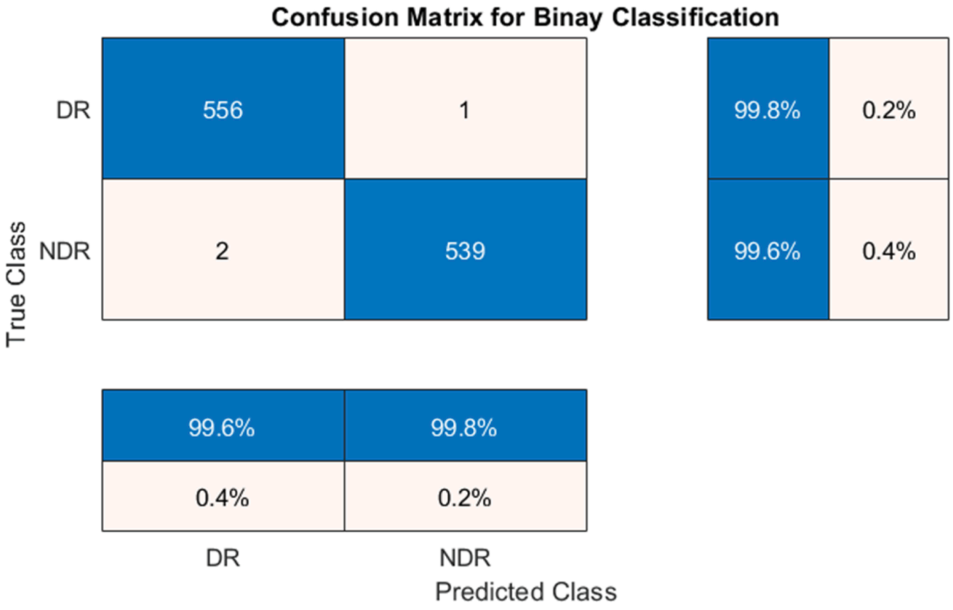
In Table 4, we have compared the performance of the proposed model for DR detection with the earlier work [11] which utilizes the same dataset with different deep learning models. In [11], DR detection is performed with seven classifiers constructed by fine-tuning the pre-trained network such as AlexNet, VGG16, ResNet, Inception-v3, NASNet, DenseNet, and GoogLeNet networks. The top two best values in Table 4 are highlighted in black and blue. It is seen that the best results are achieved by the proposed model without any preprocessing (color constancy, histogram equalization, and others) of the test images. Further, Inception-v3 exhibits the best accuracy and AUC values for the images preprocessed by the color consistency approach, rendering them invariant to the color of the source of illumination in [11].

In Table 5, the severity grading performance of the proposed model is compared with that of [11], in which the researchers have used four pre-trained classifiers as feature extractors, and SVM is used as a classifier. The two top values are shown in red and blue fonts. These results show that the best DR grading results achieved with the proposed model are far superior compared to other networks. Further, the SVM classifier achieves the best results with the features extracted with the Inception-v3 network from the fundus images preprocessed by the color constancy approach.

In addition to the above, a generic comparison with the state-of-the-art models is presented in Table 6, highlighting the significance of the proposed model. From the above analysis, it is seen that the proposed framework is better than the state-of-the-art methods concerning design and performance metrics. The size of the proposed model is smaller by a factor of 2 compared to the hybrid architecture [11] and the number of trainable parameters is higher by 2 million. While the number of parameters of the proposed DRNet is smaller by 0.6 million compared to DeepDR, the size is smaller by a factor of 4. Though the size of CF-DRNet is 4 times smaller than DRNet, the number of parameters is larger by 8 million and the accuracy of this model is very low compared to the other models. This analysis reveals that a larger number of trainable parameters do not always increase the accuracy and the underlying mechanisms of the model pipeline play a crucial role in classification and staging problems.

Table 3
Classification Performance Metrics.

Metrics	Classification Type	
	Binary (DR Detection)	Multi-class (Five classes) (DR Grading)
Accuracy	0.9973	0.9818
Sensitivity	0.9982	0.9741
Specificity	0.9963	0.9955
Precision	0.9964	0.9647
F1	0. 9973	0.9693
MCC	0.9945	0.9646
Per-class Accuracy	0.9982 (DR) 0.9963 (NDR)	0.9729 (Mild) 0.9733 (Moderate) 0.9926 (NDR) 0.9659 (Proliferative) 0.9655 (Severe)



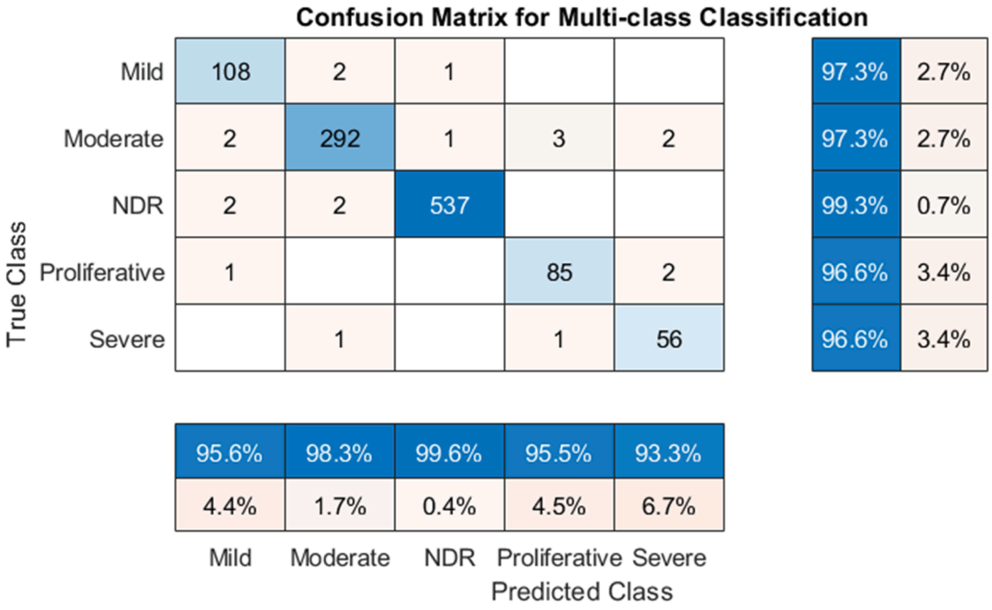
Confusion Matrix for Binay Classification

True Class

	DR	NDR		
DR	556	1	99.8%	0.2%
NDR	2	539	99.6%	0.4%
	99.6%	99.8%		
	0.4%	0.2%		

Predicted Class

Fig. 3a. Confusion Matrix for Binary Classification.



Confusion Matrix for Multi-class Classification

True Class

	Mild	Moderate	NDR	Proliferative	Severe		
Mild	108	2	1			97.3%	2.7%
Moderate	2	292	1	3	2	97.3%	2.7%
NDR	2	2	537			99.3%	0.7%
Proliferative	1			85	2	96.6%	3.4%
Severe		1		1	56	96.6%	3.4%
	95.6%	98.3%	99.6%	95.5%	93.3%		
	4.4%	1.7%	0.4%	4.5%	6.7%		

Predicted Class

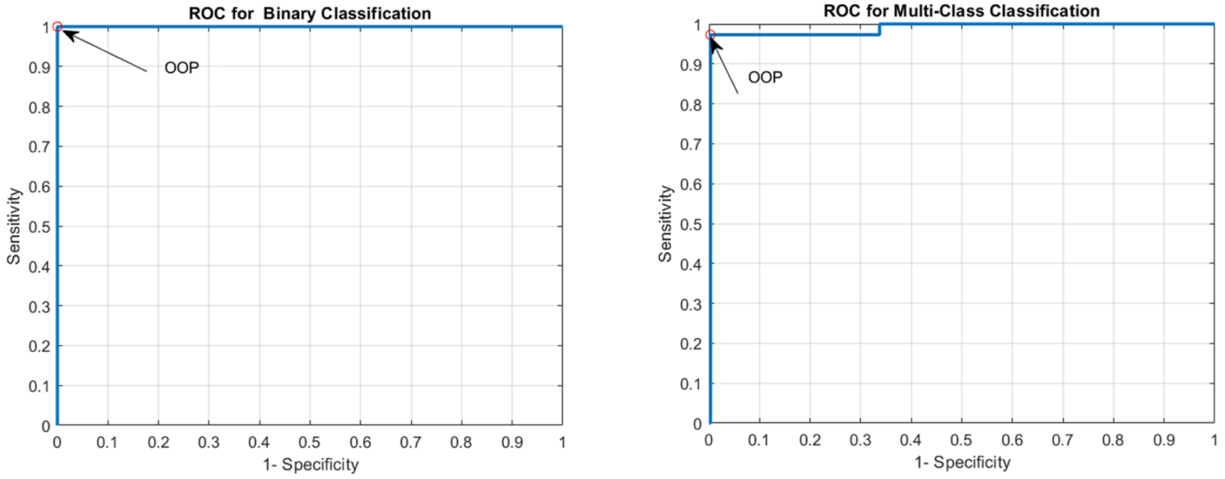
Fig. 3b. Confusion Matrix for multi-class (five class) Classification.

5.2. Explainable artificial intelligence analysis

XAI, also known as interpretable artificial intelligence, facilitates the interpretation of the behavior of machine learning models. In this work, the classification results in DR grading are analyzed by XAI analysis. Generally, this is performed by analyzing the Class Activation Map (CAM) of the final learnable layer output of a classifier. As GCAMs are used as embeddings of the support sets for the base classifiers in the proposed framework, XAI analysis does not require an explicit CAM construction and analysis.

Gradient Class Activation maps (GCAMS) and aggregated transformations are used in creating image embeddings in the proposed framework. The GCAMs and corresponding classification scores are presented in Fig. 5 for visual interpretation of the classifier activations. The GCAMs are heat maps in which the image components driving the classifier decision appear as bright red regions and rest have a dark blue hue. It is seen that the classification scores for correct classifications are closer to 1 as seen in the first two columns. Further, classification scores for misclassifications are greater than 0.5 for the last three images. These values signify ambiguity in the classification process which may be due to fine variations in features between target and misclassified classes. Further investigations on the GCAMs and classification scores can provide generalized definitions of the morphologies of OD under different severity levels and minimize the risks in treating the misclassified cases.

There are several CAM variants that can perform XAI analysis, including Grad-CAM++, Score-CAM, Ablation-CAM, and XGrad-CAM. GCAMs calculate the coefficients of the activation maps by averaging the gradients of the activated neurons that reflect the behavior of the model. It is imperative to note that Grad-CAM++ ignores subtle details which may be significant to clinical decisions, focusing only on higher-order derivatives and positive influences of neurons. The score-CAM and the ablation-CAM employ heuristic methods in the prediction of the



(a) (b)

Fig. 4. ROC Curves a) Binary Classification b) Multi-class Classification.

Table 4
Performance Comparison -DR Detection on APTOS2019 Dataset.

Method	Accuracy			AUC		
	Raw Images	Color Constancy	Histogram Equalization	Raw Images	Color Constancy	Histogram Equalization
DRNet (Proposed)	99.73 ± 0.105	NA	NA	0.9946 ± 0.003	NA	NA
AlexNet [11]	96.15 ± 1.7	96.80 ± 1.2	96.20 ± 1.8	0.981 ± 0.05	0.988 ± 0.03	0.982 ± 0.04
Inception-v3 [11]	96.60 ± 1.7	98.00 ± 1.3	97.20 ± 1.5	0.988 ± 0.03	0.993 ± 0.03	0.989 ± 0.03
ResNet [11]	96.70 ± 1.7	97.60 ± 1.4	96.80 ± 1.7	0.984 ± 0.02	0.990 ± 0.02	0.984 ± 0.02
VGG16 [11]	96.23 ± 1.6	97.00 ± 1.3	96.90 ± 1.9	0.982 ± 0.04	0.989 ± 0.03	0.988 ± 0.03
NASNet [11]	95.90 ± 2.0	96.70 ± 1.8	96.20 ± 1.9	–	–	–
DenseNet [11]	96.00 ± 1.4	96.30 ± 1.3	96.20 ± 1.9	–	–	–
GoogLeNet [11]	96.20 ± 1.6	96.70 ± 1.0	96.10 ± 1.5	–	–	–

NA- Not Applicable.
- Not Reported in their work.

Table 5
Performance Comparison -DR Grading on APTOS2019 Dataset.

Method	Accuracy in (%)		
	Raw Images	Color Constancy	Histogram Equalization
DRNet (Proposed)	98.18 ± 0.15	NA	NA
AlexNet[11]	75.7 ± 6.8	81.6 ± 5.4	80.5 ± 6.2
Inception-v3 [11]	79.8 ± 6.4	85.7 ± 5.4	83.7 ± 6.0
ResNet [11]	78.9 ± 7.1	84.9 ± 8.4	83.7 ± 7.7
VGG16 [11]	76.8 ± 5.3	83.5 ± 6.1	82.3 ± 5.7

NA- Not Applicable.

coefficients, and these methods are quite lengthy. As a result of the simplicity of GCAM implementation and its characteristic of considering all neurons, this research utilizes GCAMs for creating image embeddings. Additionally, these maps are captured from the ResNext model which performs aggregated transformations to improve the attention ability of the model.

5.3. Ablation study

An ablation study is performed in this work by reducing the number of training episodes. While the maximum number of episodes of the model is 100, 75 and 50 episodes are considered for ablation study. These assumptions are based on the experimental results produced by the model under 100 episodes. It is reasonable to evaluate the performance of the model with 75% and 50% of the maximum number of episodes. Though the support and query sets are constructed by selecting the samples randomly, it is ensured that the samples are not repeated across episodes. The performance of the model is evaluated with the same test dataset, training the model with 75 and 50 episodes. These results are presented in Table 7 for DR detection and grading. It is seen that there is a degradation in performance by 5% for every 25% of the episodes reduced.

5.4. Advantages and limitations

The DRNet proposed in this paper offers some advantages that might be of interest to the international DR research community.

1. Generalized Classification model

A DRNet trained on a smaller dataset facilitates DR classification and detection on arbitrary images with high accuracy [48,49]. In the light of the promising results produced by this model, DRNet can be extended to other pathologies such as Age-Related Macular Degeneration (AMD), glaucoma, diabetes, mineral deficiency, and any disorder with limited clinical presentations.

2. Adaptive embedding module

Embeddings are learned in *meta*-learning from a task-agnostic perspective. This is different from traditional methods that employ pre-defined feature extractors. This paper uses an embedding module to encode a set of inputs into a fixed-dimensional vector space in order to determine the intrinsic features of fundus images. It is possible to use this module discretely to create image embeddings for any pathology, modality, and task.

Table 6
Comparative analysis with State-of-the-art methods in DR detection and grading.

Models	Datasets	Detection/ Grading	Accuracy	AUC	Sensitivity	Specificity	Model Size/No. of Parameters	Model Characteristics
DRNet (Proposed)	APTOS2019	Detection Grading	0.9973 0.9818	0.999 0.9879	0.9982 0.9741	0.9963 0.9955	96 MB/25 × 10^6	• The same architecture is employed in DR detection and grading No exclusive classifier is employed
Hybrid Architecture Inception-v3 [11] Narayanan et al. (2020)	APTOS2019	Detection Grading	0.9800 0.857	0.993 –	– –	– –	2 × 97 MB/23.8 × 10^6	• Different Architectures are employed for DR detection and grading PCA used for dimensionality reduction SVM is used in DR grading
DeepDR [36] Dai et al. (2021)	EyePACS	Grading	–	0.961	0.932	0.862	4 × 98 MB/ 25.6 × 10^6	• Employs a base network and three subnetworks Overall Detection performance is not given
FEDI [41] Pan et al. (2021)	Kaggle Fundus Image	Grading (3 shot 10-way)	0.9587	–	–	–	–	• Results are not available for individual classes
CF-DRNet [33] Wu et al. (2020)	IDRiD EyePACS	Grading Grading	0.5619	0.8310	0.89	0.5399 0.9122	2 × 11.4 MB/ 33.3 × 10^6	• Two separate networks are used for DR detection and grading
Multi-task Detector [40] Quellec et al. (2020)	OPHDIAT	Grading (11 frequent disorders)		0.966	–	–	–	• Involves expensive computations
Customized CNN [23] Gulshan et al. (2016)	EyePACS Messidor-2	Grading Grading	– –	0.991 0.990	0.975 0.961	0.934 0.939	–	• Grading is performed with multiple binary classifications.
Residual CNN [26] Gargeya&Leng (2017)	Messidor-2 E-Optha	Detection Detection	– –	0.94 0.97	0.93 0.94	0.87 0.98	–	• Image *meta*-data is appended with a feature vector DT Classifier is used for classification

- Results not reported.

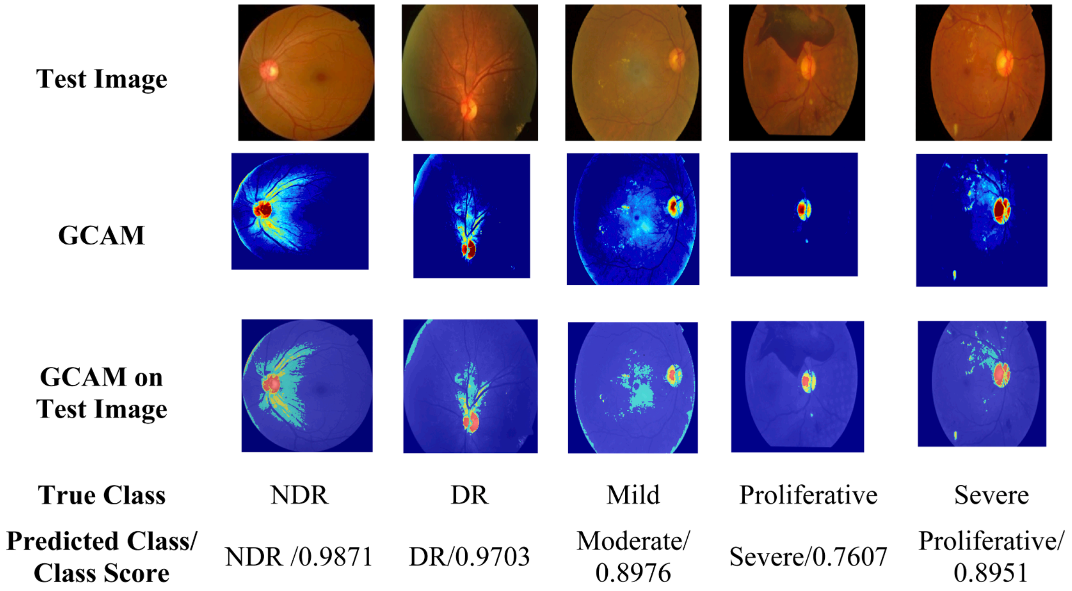


Fig. 5. XAI Analysis.

Table 7
Performance Metrics Under Ablation Study.

Classifier Type	**No. of**	**Metrics**					
	Episodes	**Accuracy in %**	**Sensitivity in %**	**Specificity in %**	**Precision in %**	**F1-Score**	**MCC**
Binary (DR Detection)	75	94.24	94.33	94.15	94.16	0.9424	0.9398
	50	89.76	89.84	89.67	89.68	0.8976	0.8951
Five- Class (DR Grading)	75	93.27	92.54	94.57	91.65	0.9208	0.9164
	50	87.87	87.18	89.10	86.34	0.8675	0.8633

3. Highly modularized Framework

With the modularization of the framework, the component embedding module and aggregated convolutional module can be replaced with other embedding and aggregated modules. A generic framework must meet this requirement in order to be flexible enough to be used across a wide range of applications.

4. Inbuilt attention mechanism

In deep learning models, attention mechanisms are usually implemented as separate units with CNNs. By integrating the underlying *meta*-learning approach and embedding module, the DRNet substantially reduces additional computational overhead.

All of these features of the DRNet are highly desirable for an intelligent automated diagnosis system, irrespective of the pathology.

It is evident from the experimental results and comparative analysis that the proposed model is more effective than the state-of-the-art approaches reported in the literature. The present study, however, has two limitations that should be addressed in future investigations. First, few-shot classification in DR detection and grading cannot be benchmarked due to a lack of benchmarked data sets. Consequently, the comparisons made in this paper are limited to the results reported on diverse datasets in the literature. The second limitation is that the episodes are not able to explore the entire dataset due to the maximum number of episodes. This is assumed to be 100. There may be instances in which some of the samples are not included in the training process during random episodic training. As a result, this effect of episodes needs to be considered in our future research. However, while the proposed system achieves better results than the hybrid model presented in [11], which utilizes the entire dataset, optimization of the training process by minimizing classification losses and episodes is expected to provide better generalization.

6. Conclusion

A novel framework for FSL-based detection called DRNet is presented in this paper as a new prototypical grading and detection framework for DR. This *meta*-classifier exhibits superior performance characteristics and objective metrics on the APTOS2019 dataset when compared to baseline classifiers and hybrid approaches. In order to achieve high detection, and grading accuracy, the proposed framework makes use of the aggregated transformation capabilities of ResNext to construct image embeddings to train the base classifiers. In addition, the system can be easily adapted to screen and stage various kinds of ocular disorders such as glaucoma, age-related macular degeneration, and dry eye disease. There may be opportunities to expand on the results of this work by developing an objective grading scheme based on disc morphologies in an attempt to better define the relationship between disc morphologies and class scores, thus providing a more precise interpretation of disc morphology as the disease advances.

CRediT authorship contribution statement

M. Murugappan: Conceptualization, Data curation, Formal analysis, Investigation, Methodology. **N.B. Prakash:** Data curation, Formal analysis. **R. Jeya:** Investigation. **A. Mohanarathinam:** Validation, Visualization. **G.R. Hemalakshmi:** Validation, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] IDF Diabetes Atlas. Global estimates of the prevalence of diabetes for 2011 and 2030. International Diabetes Federation Diabetes Atlas, 2019, 9th Edition. [Internet]. Available from: http://www.diabetesatlas.org/ Accessed on 13 October 2021.

[2] Early Treatment Diabetic Retinopathy Study Research Group. (1991). Early Treatment Diabetic Retinopathy Study design and baseline patient characteristics: ETDRS report number 7. *Ophthalmology*, *98*(5), 741-756.

[3] J. Nayak, P.S. Bhat, R. Acharya, C.M. Lim, M. Kagathi, Automated identification of diabetic retinopathy stages using digital fundus images, J. Med. Syst. 32 (2) (2008) 107–115.

[4] T. Teng, M. Lefley, D. Claremont, Progress towards automated diabetic ocular screening: a review of image analysis and intelligent systems for diabetic retinopathy, Med. Biol. Eng. Compu. 40 (1) (2002) 2–13.

[5] J.H. Hipwell, F. Strachan, J.A. Olson, K.C. McHardy, P.F. Sharp, J.V. Forrester, Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool, Diabet. Med. 17 (8) (2000) 588–594.

[6] J.A. Olson, F.M. Strachan, J.H. Hipwell, K.A. Goatman, K.C. McHardy, J. V. Forrester, P.F. Sharp, A comparative evaluation of digital imaging, retinal photography and optometrist examination in screening for diabetic retinopathy, Diabet. Med. 20 (7) (2003) 528–534.

[7] M.D. Abràmoff, J.M. Reinhardt, S.R. Russell, J.C. Folk, V.B. Mahajan, M. Niemeijer, G. Quellec, Automated early detection of diabetic retinopathy, Ophthalmology 117 (6) (2010) 1147–1154.

[8] S. Roychowdhury, D.D. Koozekanani, K.K. Parhi, DREAM: diabetic retinopathy analysis using machine learning, IEEE J. Biomed. Health. Inf. 18 (5) (2014) 1717–1728.

[9] G.T. Reddy, S. Bhattacharya, S.S. Ramakrishnan, C.L. Chowdhary, S. Hakak, R. Kaluri, M.P.K. Reddy, in: February). An ensemble based machine learning model for diabetic retinopathy classification, IEEE, 2020, pp. 1–6.

[10] Y. Miao, S. Tang, P. Du, Z. Li, in: September). Research on Deep Learning in the Detection and Classification of Diabetic Retinopathy, IEEE, 2021, pp. 107–113.

[11] Narayanan, B. N., Hardie, R. C., De Silva, M. S., & Kueterman, N. K. (2020). Hybrid machine learning architecture for automated detection and grading of retinal images for diabetic retinopathy. *Journal of Medical Imaging*, *7*(3), 034501.

[12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[13] A. Arrigo, M. Teussink, E. Aragona, F. Bandello, M.B. Parodi, MultiColor imaging to detect different subtypes of retinal microaneurysms in diabetic retinopathy, Eye 35 (1) (2021) 277–281.

[14] J. Guan, Z. Lu, T. Xiang, A. Li, A. Zhao, J.R. Wen, Zero and few shot learning with semantic feature synthesis and competitive learning, IEEE Trans. Pattern Anal. Mach. Intell. 43 (7) (2020) 2510–2523.

[15] APTOS 2019 blindness detection," https://www.kaggle.com/c/aptos2019-blindnessdetection/overview (Accessed 24 August 2021).

[16] R. Rashmi, U. Snekhalatha, P.T. Krishnan, Fat based studies for computer assisted screening of child obesity using thermal imaging based on deep learning techniques: a comparison with quantum machine learning approach, Soft. Comput. (2022), https://doi.org/10.1007/s00500-021-06668-3.

[17] S. Umapathy, P.T. Krishnan, Automated detection of Orofacial Pain from thermograms using machine learning and deep learning approaches, Expert systems 38 (7) (2021), https://doi.org/10.1111/exsy.12747.

[18] A. Banan, A. Nasiri, A. Taheri-Garavand, Deep learning-based appearance features extraction for automated carp species identification, Aquacult. Eng. 89 (2020), 102053.

[19] S. Shamshirband, T. Rabczuk, K.W. Chau, A survey of deep learning techniques: application in wind and solar energy resources, IEEE Access 7 (2019) 164650–164666.

[20] Y. Fan, K. Xu, H. Wu, Y. Zheng, B. Tao, Spatiotemporal modeling for nonlinear distributed thermal processes based on KL decomposition, MLP and LSTM network, IEEE Access 8 (2020) 25111–25121.

[21] W.L. Alyoubi, W.M. Shalash, M.F. Abulkhair, Diabetic retinopathy detection through deep learning techniques: a review, Inf. Med. Unlocked 20 (2020), 100377.

[22] H. Pratt, F. Coenen, D.M. Broadbent, S.P. Harding, Y. Zheng, Convolutional neural networks for diabetic retinopathy, Procedia Comput. Sci. 90 (2016) 200–205.

[23] V. Gulshan, L. Peng, M. Coram, M.C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P.C. Nelson, J.L. Mega, D.R. Webster, Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, JAMA 316 (22) (2016) 2402.

[24] J. Cuadros, G. Bresnick, EyePACS: an adaptable telemedicine system for diabetic retinopathy screening, J. Diabetes Sci. Technol. 3 (3) (2009) 509–516, https://doi.org/10.1177/193229680900300315.

[25] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, J.-C. Klein, Feedback on a publicly distributed image database: the Messidor database, Image Analysis & Stereology 33 (3) (2014) 231.

[26] R. Gargeya, T. Leng, Automated identification of diabetic retinopathy using deep learning, Ophthalmology 124 (7) (2017) 962–969.

[27] J. Sahlsten, J. Jaskari, J. Kivinen, L. Turunen, E. Jaanio, K. Hietala, K. Kaski, Deep learning fundus image analysis for diabetic retinopathy and macular edema grading, Sci. Rep. 9 (1) (2019) 1–11.

[28] X. Li, X. Hu, L. Yu, L. Zhu, C.W. Fu, P.A. Heng, CANet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading, IEEE Trans. Med. Imaging 39 (5) (2019) 1483–1493.

[29] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[30] Y.-P. Liu, Z. Li, C. Xu, J. Li, R. Liang, Referable diabetic retinopathy identification from eye fundus images with weighted path for convolutional neural network, Artif. Intell. Med. 99 (2019), 101694.

[31] J. Wang, J. Luo, B. Liu, R. Feng, L. Lu, H. Zou, Automated diabetic retinopathy grading and lesion detection based on the modified R-FCN object-detection algorithm, IET Comput. Vision 14 (1) (2020) 1–8.

[32] G.T. Zago, R.V. Andreão, B. Dorizzi, E.O. Teatini Salles, Diabetic retinopathy detection using red lesion localization and convolutional neural networks, Comput. Biol. Med. 116 (2020), 103537.

[33] Z. Wu, G. Shi, Y. Chen, F. Shi, X. Chen, G. Coatrieux, J. Yang, L. Luo, S. Li, Coarse-to-fine classification for diabetic retinopathy grading using convolutional neural network, Artif. Intell. Med. 108 (2020), 101936.

[34] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, F. Meriaudeau, Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research, Data 3 (3) (2018) 25.

[35] K. Shankar, A.R.W. Sait, D. Gupta, S.K. Lakshmanaprabu, A. Khanna, H.M. Pandey, Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model, Pattern Recogn. Lett. 133 (2020) 210–216.

[36] L. Dai, L. Wu, H. Li, C. Cai, Q. Wu, H. Kong, R. Liu, X. Wang, X. Hou, Y. Liu, X. Long, Y. Wen, L. Lu, Y. Shen, Y. Chen, D. Shen, X. Yang, H. Zou, B. Sheng, W. Jia, A deep learning system for detecting diabetic retinopathy across the disease spectrum, Nat. Commun. 12 (1) (2021).

[37] Q. Sun, Y. Liu, T.S. Chua, B. Schiele, Meta-transfer learning for few-shot learning, in: In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 403–412.

[38] Snell, J., Swersky, K., & Zemel, R. S. (2017). Prototypical networks for few-shot learning. *arXiv preprint arXiv:1703.05175*.

[39] F. Wu, J.S. Smith, W. Lu, C. Pang, B. Zhang, in: August). Attentive prototype few-shot learning with capsule network-based embedding, Springer, Cham, 2020, pp. 237–253.

[40] G. Quellec, M. Lamard, P.-H. Conze, P. Massin, B. Cochener, Automatic detection of rare pathologies in fundus photographs using few-shot learning, Med. Image Anal. 61 (2020), 101660.

[41] Pan, L., Ji, B., Xi, P., Wang, X., Chongcheawchamnan, M., & Peng, S. (2021). FEDI: Few-shot learning based on Earth Mover's Distance algorithm combined with deep residual network to identify diabetic retinopathy. *arXiv preprint arXiv:2108.09711*.

[42] Y. Shigeto, I. Suzuki, K. Hara, M. Shimbo, Y. Matsumoto, September). Ridge regression, hubness, and zero-shot learning, in: Joint European conference on machine learning and knowledge discovery in databases, Springer, Cham, 2015, pp. 135–151.

[43] Dhillon, G. S., Chaudhari, P., Ravichandran, A., & Soatto, S. (2019). A baseline for few-shot image classification. *arXiv preprint arXiv:1909.02729*.

[44] B. Kulis, Metric learning: a survey, Foundations and Trends® in Machine Learn. 5 (4) (2013) 287–364.

[45] Y. Wang, Q. Yao, J.T. Kwok, L.M. Ni, Generalizing from a few examples: A survey on few-shot learning, ACM Computing Surveys (CSUR) 53 (3) (2020) 1–34.

[46] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in: In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1492–1500.

[47] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: visual explanations from deep networks via gradient-based localization, in: In *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[48] Kavya, Snekhalatha U, Palani Thanaraj Krishnan (2021). Deep learning techniques for Automated classification of Autism using Thermal imaging. *Journal of Engineering in Medicine*. https://doi.org/10.1177%2F09544119211024778.

[49] U. Snekhalatha, K. Palani Thanaraj, K. Sangamithirai, Computer aided diagnosis of obesity detection based on thermal imaging using various convolutional neural networks, Biomed. Signal Processing and Control J. 63 (2020), 102233, https://doi.org/10.1016/j.bspc.2020.102233.