Prediction of Diabetic Retinopathy using Fusion of Hierarchically Reduced Color Texture Features

Holly Vo
Department of Computer Science
California State University
Fullerton, CA, USA
hhvo@csu.fullerton.edu

Abhishek Verma
Department of Computer Science
California State University
Northridge, CA, USA
abhishek.verma@csun.edu

***Abstract*—This paper proposes a novel hierarchical feature dimensionality reduction (FDR) methodology for fusion of features from multiple color spaces. The objective is to automatically recognize diabetic retinopathy (DR) stages; an eye disease that attacks one in three diabetes carriers in USA. Our work examines: (1) the role of FDR in extracting discriminatory features and (2) the discriminatory information carried in different color spaces on fundus images. Effectiveness of the proposed methodology is evaluated on the challenging fine grained diabetic retinopathy image dataset.**

Features are extracted from color retina images by local binary pattern descriptors at multiple scales, on three channels of each color space, and organized as a feature hierarchy. Proposed hierarchical FDR methodology performs feature reduction by PCA, LDA, random forest, and multi-layer perceptron on the hierarchy of features. Our experiments show (i) the hierarchical FDR methodology with multi-layer perceptron efficiently transforms features to lower dimensional space by achieving best accuracy. (ii) To investigate the discriminatory power of color spaces; FDR methodology and classification is evaluated on RGB, L*a*b*, HSV, YIQ, and rgb color spaces with best classification accuracy of 74.26% on RGB, and (iii) the novel combination of the proposed FDR method with fusion of multiple color features achieves 75.69% accuracy by one-vs-one SVM classifier.

***Index Terms*—diabetic retinopathy classification, local binary pattern, color feature fusion**

I. Introduction

Diabetic retinopathy (DR) is a very common eye disease, US Centers for Disease Control and Prevention estimates that DR affects 33% of the Americans with diabetes and may cause blurred vision and blindness. High glucose level in the blood could damage the blood vessels in the retina causing DR. The macula swells and becomes thick as blood, cholesterol, fluids, and fats leak from the damaged blood vessels of the retina. The retina stops receiving oxygen and nutrients due to damage to the capillaries. Intraretinal microvascular abnormalities (IrMAs) are new abnormal fragile blood vessels that grow in the retina so that it can deliver sufficient blood to the retina. The retina usually develops scar tissue that may get wrinkled or get detached and cause distortion of vision. As the DR progresses, it can cause increase in eye pressure leading to damage to the optic nerve [1].

Progression of DR may cause irreversible loss of vision in both eyes. In the early stages of DR the patient may not notice changes to the vision. If DR could be detected in its early stages it could help in early treatment and thereby reduce the chances of severe loss of vision by more than 90%. Detection of DR if done manually is a time-consuming process that requires ophthalmologists to evaluate retina images, perform a review, and follow up. This process could take multiple days and may cause delay in providing treatment [10], [11].

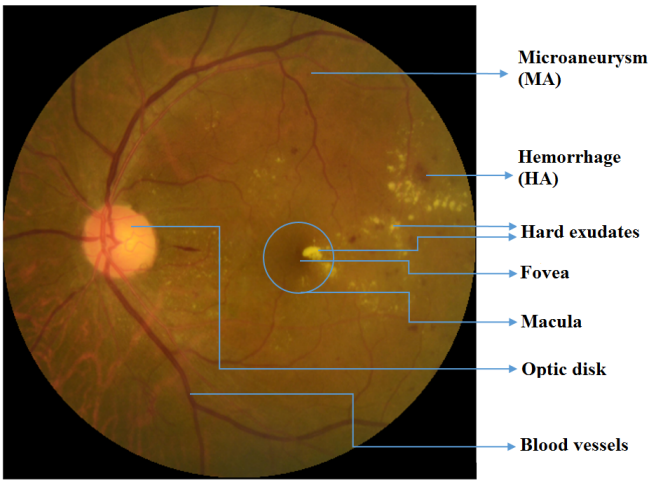

Fig. 1. Image of a retina showing signs of Diabetic Retinopathy.

Fig. 1 shows retina image and signs of diabetic retinopathy. A retina image has a macula, blood vessels, and an optic disc. Center part of the retina is called Macula, which contains rods that are sensitive to color. Fovea is the central point of the retina where visual acuity is the highest. Blood vessels develop tiny bulges that appear as red dots and called Microaneurysms (MAs). Tiny spots of blood discharge show as hemorrhages. Protein and lipid leakage into the retina causes hard exudates, appearing as lesions of white or cream in color. Microaneurysms and hard exudates could damage the macula and cause permanent loss of vision [10], [12].

An ophthalmologist conducts a retinal examination to distinguish between two types of diabetic retinopathy: nonproliferative (NPDR) and proliferative (PDR). NPDR is categorized according to the signs of diabetic retinopathy. In mild NPDR, manifestations include microaneurysms, exudates, and venous

loops. Moderate NPDR exhibits hemorrhages, multiple microaneurysms, and hard exudates. Severe NPDR is identified by the 4-2-1 rule: presence of intraretinal microvascular abnormalities (IrMAs) in one quadrant, venous beading in two quadrants, and in four quadrants the hemorrhages and microaneurysms. PDR represents an advanced stage characterized by the development of fragile new blood vessels that hemorrhage and leak, along with the formation of hard exudates in the vitreous. The vitreous is a gel-like substance located in the middle of an eye. Complications such as scar tissue formation and retinal wrinkles may occur, leading to increased pressure within the retina, damage to the macula, and optic nerve impairment, ultimately resulting in vision loss [4], [10].

Instead of isolating DR signs with parametric filters and segmentation on small datasets, our work targets a large-scale dataset and examines robust discriminatory texture features and their fusion in multiple color spaces. However, due to the high dimensionality of texture features, a good candidate method for feature dimensionality reduction must be considered for the retinopathy image dataset at multiple levels. Such a method should retain highly discriminative features even at lower dimensions at the time of fusion of texture scales and colors and thereby improve accuracy and speed of classification.

In section IV we describe uniform local binary pattern color (LBP) feature descriptors at multiple LBP scales and five selected color spaces. In section V we describe the feature dimensionality reduction techniques and proposes dimensionality reduction method by neural networks. Section VI presents our novel proposed methodology. Section VII discusses the results from two sets of experiments: the experiment set of dimensionality reduction on RGB color features and the experiment set of proposed dimensionality reduction on multiple color spaces and color feature fusion.

II. Related Work

Microaneurysms are considered as the early sign of diabetic retinopathy [15]. Hence, many researches in the past have focused on extracting features from the retina by localization and segmentation of various elements such as lesions, blood vessels, optic disks, and the macula. This process involves applying point operators to image pixels to balance and enhance local contrast. Localization and segmentation are achieved through linear filtering and neighborhood operations like morphology, median filtering, and Gaussian filtering, during the preprocessing stage, as noted in survey research [10], [16]. In [17], the authors utilize watershed transformation to address the challenge of segmenting beyond what is ideal due to thresholding. Additionally, in [16], the authors employ modeling contours actively and growing the region recursively (RRGT) to isolate blood vessels and other notable regions within the retina.

Texture extraction presents another intriguing avenue in the recognition of diabetic retinopathy (DR). Statistical texture extraction relies on the correlation of intensity values between image pixels. In this method, entropy, pixel contrast, and correlation are computed using the gray level co-occurrence matrix. In the study by [3], contrast of texture patterns is extracted alongside areas of microaneurysms (MAs) and hemorrhages (HAs) for DR identification. For DR identification on small datasets containing fewer than 100 images of the retina, [19], [20] utilize the Local Binary Pattern (LBP) texture feature descriptor. In other computer vision applications such as face and scene recognition, texture descriptors like local binary patterns have been demonstrated to significantly enhance performance [13].

A robust vision system needs to rely on an invariant color space to be able to handle images that are captured from different camera devices under varying conditions of light and intensity.HSI (Hue, Saturation, Intensity) analysis is applied to the Messidor and DB-rect datasets for the extraction of microaneurysms (MAs) and exudates, as discussed in [1], and for locating the fovea, as explored in [18]. The green channel of RGB is used for extracting the blood vessels in studies such as [2], [3]. In [4], each channel of the RGB color space is individually examined, and morphological operations are conducted to extract the total area and perimeter of blood vessels, hemorrhages (HAs), and microaneurysms (MAs). Researchers in [5] consider multiple color spaces. These studies primarily focus on the segmentation of diabetic retinopathy signs using small datasets ranging from a few hundred to a few thousand images.

Classification techniques such as support vector machine (SVM), artificial neural network (ANN), and random forest (RF) are widely employed in diabetic retinopathy (DR) research [1]–[4], [16], [19]–[21]. In addition traditional machine learning approaches along with neural networks have been successfully applied to image datasets for skin cancer recognition [28] and for detection of gene mutation [27].

III. Description of Diabetic Retinopathy Image Dataset

In this study, we conduct experiments using the EyePACS diabetic retinopathy image dataset, a publicly available resource for retinopathy screening [26]. The dataset comprises 35,126 images for training and 53,576 images for testing. These images are sourced from several camera models and types, resulting in variations in resolution. Each image has been manually reviewed by a human clinician for signs of diabetic retinopathy disease. The dataset is categorized into five classes representing DR stages from 0 to 4, corresponding to no DR, mild, moderate, severe, and proliferative DR. The training set comprises 74% of stage 0 (no DR) images, 7% of stage 1 (mild) images, 15% of stage 2 (moderate) images, and 2% each of stage 3 (severe) and stage 4 (proliferative DR) images. The test set follows the same distribution as the training set across the five stages.

The dataset presents several challenges, including variations in image resolution, intensity, and quality. Upon examining the training set, we observe that image heights range from 289 to 3,456 pixels, widths range from 400 to 5,184 pixels, and the

aspect ratio (height to width) falls between 0.66 and 1.00. The average image intensity ranges from 1 to 192, with a mean average intensity of 63 on a scale of 0 to 255. Depending on the image resolution, the image size on disk varies from 8KB for low-resolution images to 2MB for high-resolution images.

IV. Color Feature Extraction

A. Local Binary Pattern (LBP) descriptor

The Local Binary Pattern (LBP) texture descriptor, introduced by Ojala [6], [7], has been recognized as a robust feature extractor for texture information [8], [9]. It has demonstrated success in various applications including biometrics, face detection and recognition, and scene recognition [8], [9].

The basic version of the LBP descriptor focuses on each 3x3 neighborhood to generate an ordered 8-bit LBP code by comparing the gray values of the surrounding pixels with that of the center pixel. The local binary pattern around a pixel P with eight neighbors P_x is encoded as follows:

$$LBP = \sum_{x=1}^{8} s(x)2^{x-1} s(x) = \begin{cases} 1 & \text{if } P_x > P \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The LBP operator is known for its invariance to changes in average intensity. To achieve scale invariance, the original LBP operator has been extended to consider a circularly symmetric neighborhood of P pixels on a circle with radius R surrounding the center pixel. This extended operator is denoted as $LBP_{P,R}$. In this setup, the top-middle neighbor is considered the most significant bit in the LBP code, and the remaining neighbors are ordered clockwise. To ensure accuracy, the gray value of each neighbor point, which may not lie exactly at the center of pixel area, is estimated through interpolation.

The extension proposed by Ojala [7] introduces a uniform pattern denoted as $LBP^{riu2}_{P,R}$. This binary pattern is circularly chained and contains a maximum of two spatial transitions, namely "1-0" and "0-1". $LBP^{riu2}_{P,R}$ enhances rotation invariance and significantly reduces the dimensionality of LBP by consolidating all nonuniform patterns into a single bin.

In the diabetic retinopathy (DR) dataset utilized in our experiments, images are captured from diverse cameras under varying illumination and quality conditions. Additionally, the DR signs exhibit fine-grained characteristics with variations in granularity, rendering the classification task more challenging. Microaneurysms (MAs), characterized by tiny red dots, hemorrhages, and hard exudates, can manifest in various sizes and shapes. Moreover, fragile blood vessels may develop in unpredictable patterns. To address these challenges, we employ the uniform local binary pattern (LBP) descriptor at different scales to capture fine-grained discriminatory features. The multi-scale LBP features exhibit invariance to rotation, global intensity variations, and scales, enhancing the robustness of our classification approach.

We divide each 512 x 512 retina image into four quadrants. Texture information in the image is extracted at 10 different scales by $LBP^{riu2}_{8,1}$, $LBP^{riu2}_{8,2}$, $LBP^{riu2}_{12,3}$, $LBP^{riu2}_{16,4}$, $LBP^{riu2}_{20,5}$, $LBP^{riu2}_{24,6}$, $LBP^{riu2}_{28,7}$, $LBP^{riu2}_{32,8}$, $LBP^{riu2}_{32,9}$, and $LBP^{riu2}_{32,10}$ operators on four 256x256 quadrants to form a LBP feature block on each scale for each channel of a given color space. The extracted LBP features are normalized within its block. The feature dimension of each LBP block is shown in Table I. Note that feature dimension will grow to 928 for gray images and 2,784 for color images if feature dimensionality reduction techniques are not considered. Furthermore, dimension will continue to grow in case of fusion of LBP features from multiple color spaces.

TABLE I
LBP Feature Dimension at Multiple Scales on a Single Color Channel and on a Color Space without Dimensionality Reduction

LBP operators	Scale (R)	Neighbors (P)	Feature dimension
(8,1)	1	8	40
(8,2)	2	8	40
(12,3)	3	12	56
(16,4)	4	16	72
(20,5)	5	20	88
(24,6)	6	24	104
(28,7)	7	28	120
(32,8)	8	32	136
(32,9)	9	32	136
(32,10)	10	32	136
Concatenated feature dimension	single color channel		928
	3-channel color space		2784

B. Color Spaces

In this paper, discriminative color features are extracted for DR recognition by uniform LBP descriptors on five color spaces: RGB, YIQ, rgb, HSV, and L*a*b*.

V. Feature Dimensionality Reduction Techniques

Need for feature dimensionality reduction (FDR) has become increasingly important in recent years to efficiently classify processing of ever increasing real-world data adequately. Features from high resolution digital images are usually represented in high dimensional spaces. FDR improves classification in terms of efficiency and accuracy. FDR techniques have two categories: feature selection (FS) and feature extraction (FE). FS approaches aim at finding a subset of the original features, whereas FE transforms higher dimensional data observation space to lower dimension feature space [23].

Several nonlinear data transformation techniques for dimensionality reduction have been proposed to overcome the limitation of principle component analysis (PCA) on complex nonlinear data. Empirical results in [22] on five artificial datasets and five real-world datasets compare PCA and twelve other nonlinear transformation techniques. However, results show that most nonlinear techniques do not outperform PCA on real-world datasets. In our work, feature dimension from LBP descriptors grows to several thousand on a single color space and further increases as features from multiple color

spaces are fused together. Our proposed hierarchical feature dimensionality reduction methodology is described in detail in the proposed methodology section of this paper.

A. Feature Selection by Feature Importance

RF can be used for ranking the importance of features based on its recorded out-of-bag error (OOB). Approximately one-third of the training data is left out when building each bootstrap sub-sample. OOB is estimated by combining one-third of the classifiers in RF ensemble. A threshold θ is commonly defined for feature selection. Let m be the mean of feature importance in the feature space, features whose importance is greater than or equal to $\theta * m$ are selected while others are dropped.

Additionally, we run experiments using Principal Component Analysis (PCA) for dimensionality reduction. PCA is good for dimensionality reduction but it does not perform well for class prediction. We also perform experiments with Linear Discriminant Analysis (LDA). LDA seeks to reduce feature dimensionality while preserving the information needed for differentiating between classes. There are several limitations to LDA dimensionality reduction technique. First, the number of classes L in a given problem restricts the maximum reduced features to $L-1$. Second, LDA is a parametric model that assumes Gaussian distribution; thus, it is unable to preserve complex data where the distribution is significantly non-Gaussian. Lastly, LDA fails if the discriminatory information is not in the mean but in the variance of data [24].

B. Dimensionality Reduction by Multi-Layer Perceptron

The idea behind applying MLP for supervised dimensionality reduction is derived from autoencoder. An autoencoder is similar to MLP except its output layer has same number of neurons as that in the input layer. Autoencoder consists of two components, the encoder and the decoder. The two components share a hidden layer where compressed data can be extracted. Autoencoder is one of the nonlinear techniques to overcome limitations of traditional PCA on complex nonlinear data. Autoencoder is similar to PCA in the sense that it is an unsupervised dimensionality reduction technique. It compresses the input into a data representation that can be used for reconstructing the original input. However, with nonlinear transformation on hidden layers, the autoencoder can capture multi-modal aspects of the input and outperforms PCA [25].

For DR problem, the aim of feature dimensionality reduction technique is not about preserving the principal information for possible reconstruction but to extract the most discriminatory information for DR stage separation. Thus, supervised MLP data transformer is proposed as a candidate FDR technique in our work. MLP data transformer is created by dropping the output layer of a MLP classifier after training. Thereby, the last hidden layer of the original MLP classifier becomes output data with lower dimension than input data.

C. Classification Techniques

In our experiments we build models using Multi-Layer Perceptron (MLP), Support Vector Machines (SVM), and Random Forest (RF). This section briefly introduces these classification techniques along with their strengths and weaknesses.

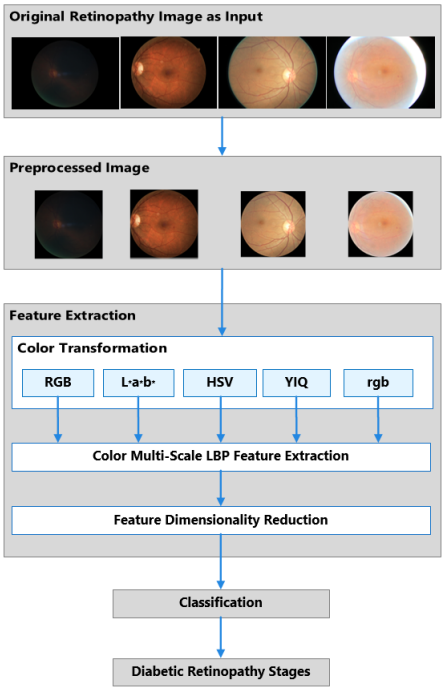

Fig. 2. Overview of proposed methodology.

In addition to classification, RF and MLP are selected for reducing feature dimensionality. Due to simplicity of parameter tuning in RF classifier, we select it as the primary classifier to evaluate dimensionality reduction methodology in our work.

VI. Proposed Methodology

In our work, DR classification performance is based on uniform LBP features that are extracted on multiple scales for each of the color space RGB, HSV, L*a*b*, YIQ, and rgb. Proposed methodology is structured in three main parts: image preprocessing, feature extraction, and classification. Three aforementioned classification techniques are used. Details of the first two parts are described in this section. Fig. 2 gives an overview of the proposed methodology.

A. Dataset Preprocessing and Augmentation

As mentioned in the dataset description section, there is large variation in size of input images. We crop images of various sizes to a tight square around the retina. Next, we scale the image to a 512x512 size by bicubic interpolation method. The dataset is imbalanced as there are many more images of DR stage 0, i.e., no disease, compared to DR stages 1-4. We augment the stages 1-4 by rotation and mirroring to make the dataset balanced and thereby avoid biasing training on stage 0.

B. Hierarchical Feature Dimensionality Reduction

In this work, LBP features at level-1 are extracted on individual color channels of a given color space by multiple

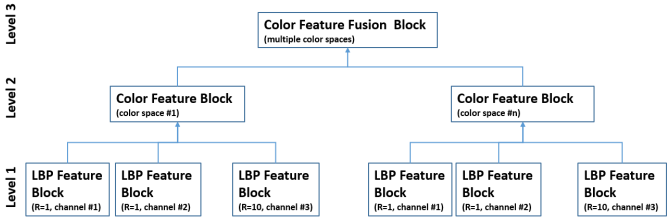

Fig. 3. Hierarchy of features at three levels (bottom-up) extracted with multiple LBP operators and its fusion in multiple color spaces.

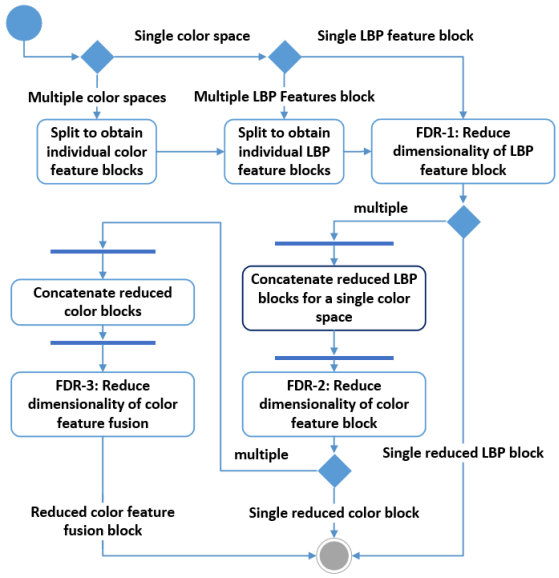

Fig. 4. Hierarchical feature dimensionality reduction methodology.

LBP operators and each one is named as a *LBP feature block*. Features at level-2 are named *color feature block*, it consists of multiple LBP feature blocks from multiple texture scales over three color channels of a single color space. Features at level-3 is named *color feature fusion block*, it consists of color feature blocks from multiple color spaces that are selected for color fusion. Fig. 3 illustrates hierarchy of features at three levels extracted with multiple LBP operators and its fusion in multiple color spaces.

A candidate FDR technique can be applied at one or more levels from bottom-up on the hierarchy of features. In a FDR-n, features blocks at and below the n-th level are reduced by the specified FDR technique. If there are any more levels above n-th level in the features hierarchy, no FDR technique will be applied at those levels. Example: FDR-1 reduces features at level-1 but does not reduce features at level 2 and 3. Thus, features block at 2nd level is the fusion of reduced features from level-1 and features block at level-3 is the fusion of non-reduced features from level-2. Fig. 4 illustrates the work-flow to obtain reduced features at multiple levels.

VII. Experiment Results

We structure our experiments into three sets, each toward particular objective. Experiments in the first set focus on RGB color space and evaluate the proposed FDR-1 and FDR-2 methodology, which comprises of four selected FDR techniques: PCA, LDA, RF feature selection, and MLP FDR. Classification is performed with RF in first set. The FDR technique that yields best performance is selected to be applied in second and third set of experiments. Second set of experiments apply FDR-2 MLP on features from RGB, YIQ, rgb, HSV, and L*a*b* color spaces. Classification is performed with RF, one-vs-rest, and one-vs-one SVM, and MLP. In order to further improve classification performance, third set of experiments is performed on fusion of features from multiple color spaces using FDR-2 and FDR-3 MLP methodology. Classification is performed with SVM.

Ideally class sizes should be close to equal. However, class sizes in DR dataset are extremely unequal. It is challenging to collect a large and balanced DR dataset simply due to much lower occurrence of DR stages 1-4 in the real world. Despite our efforts to balance the class sizes for training set by increasing the image count for specific classes through flipping and rotation, it still remains highly unbalanced. Considering the above limitation of DR dataset, experimental performance is primarily evaluated by total accuracy, i.e., mean correct classification rate (CCR) across all test images. However, for sake of completeness, we do provide average class accuracy, i.e., mean of average accuracy across five DR classes.

A. FDR-1 and FDR-2 Experiments on RGB Color Features

First set of experiments focus on level-2 feature hierarchy in RGB color space. Proposed FDR-1 and FDR-2 methods are applied for each of the four aforementioned FDR techniques. There are features blocks at 2 levels: LBP feature block and color feature block. Concatenation of features from level-1 results in 2,784 dimensional feature vector, which is partially reduced by FDR-1 and further reduced by FDR-2. All experiments are evaluated by RF classifier with 1,000 trees.

1) PCA vs. LDA Results: The reduced dimension of each features block is determined by the total variance to be retained in each PCA experiment. Total range of variance for retained PCA features is 95.00% to 99.95%. For LDA we use reduced features of 1 to 4 dimensions. These linear FDR approaches best perform on FDR-1 at 99.5% total variance for PCA (Fig. 5(a)) and 3 features for LDA (Fig. 5(b)). On FDR-2, both techniques drastically degrade on the fusion of reduced LBP feature blocks, possibly due to loss of discriminatory information. FDR-2 PCA is at least 1% behind FDR-1 PCA, while the degradation increases by more than 7% for FDR-2 LDA. Therefore, these linear FDR techniques are not the best candidates to be applied at higher levels of DR feature hierarchy.

In addition, empirical results on FDR-1 PCA improve from total accuracy of 71.03% at 95.0% variance to 72.28% at 99.5% variance. Furthermore, there is an increase of 3.0% in the average class performance across the variance ranging from 95.0% to 99.5%. Results indicate that DR class discriminatory information is better represented in the tailed variance of PCA rather than through principal components

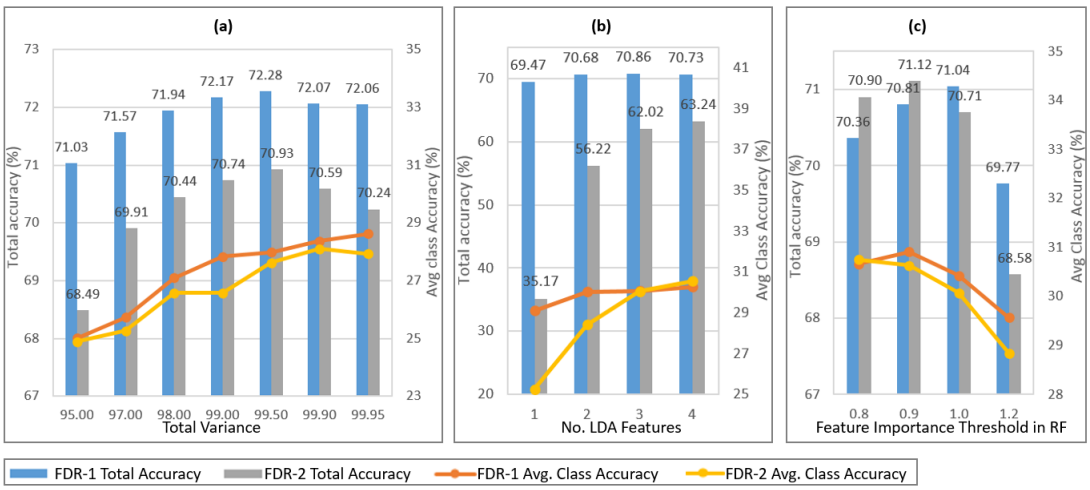

Fig. 5. RF classification accuracy on RGB color features using FDR-1 & 2 PCA, LDA, FS-RF feature dimensionality reduction techniques. Classification accuracy on (a) FDR-1 & 2 PCA at various total variance, (b) FDR-1 & 2 LDA at various LDA feature sizes, and (c) FDR-1 & 2 FS-RF at various feature importance thresholds.

themselves. This is one of the reasons why unsupervised FDR PCA outperforms the supervised FDR LDA on DR dataset.

2) Feature Selection Results: Feature selection by random forest (FS-RF) is applied at multiple values of θ. The best performance of 71.12% is achieved by FDR-2 FS-RF with $\theta = 0.9$ (Fig. 5(c)), which is approximately 1% less than FDR-1 PCA performance (Fig. 5(a)). After applying FDR-2 FS-RF at level-1, we obtain 1,723 features concatenated from 30 reduced LBP feature blocks from level-1. Importance of these features ranges from $0.32e$-3 to $6.02e$-3 with mean $m = 0.58e$-3. FDR-2 FS-RF at $\theta = 0.9$ discards all features whose importance is below $0.52e$-3. Highest ranked feature is 12.58 times more important than the lowest amongst 506 selected features. It implies that DR information is disseminated in almost all texture features and in order to reduce dimensionality FS-RF consequently discards what it considers as weak features even though it may carry partially significant information in our DR feature space. Thus, we conclude that supervised FDR FS-RF technique is surpassed by unsupervised technique such as FDR PCA.

3) Feature Reduction by MLP: In the hierarchical FDR MLP approach, features in a block at the i-th level are reduced by a 3-layer MLP denoted as $MLP_i[x_i, l_i]$ with x_i and l_i as nodes in its first and second layer respectively. l_i is the expected dimension of features at the i-th level and x_i is adjusted by the original feature dimension. Experiments are conducted with $MLP_1[50, 20]$ for LBP feature blocks using FDR-1 and FDR-2 MLP. Using FDR-2 MLP at level-2, $MLP_2[100, 50]$ is selected to reduce 600 concatenated features from 30 reduced LBP feature blocks at level-1 to 50 features for the reduced color feature block at level-2. Fig. 6 shows the classification accuracy across all four FDR techniques using FDR-1 & 2 approaches on RGB color features. It is observed that FDR-2 MLP outperforms rest of the FDR techniques both in terms of total and average class accuracy. Since, FDR-2 MLP outperforms FDR-1 MLP, we consider FDR MLP as an effective dimensionality reduction technique at multiple levels of feature hierarchy.

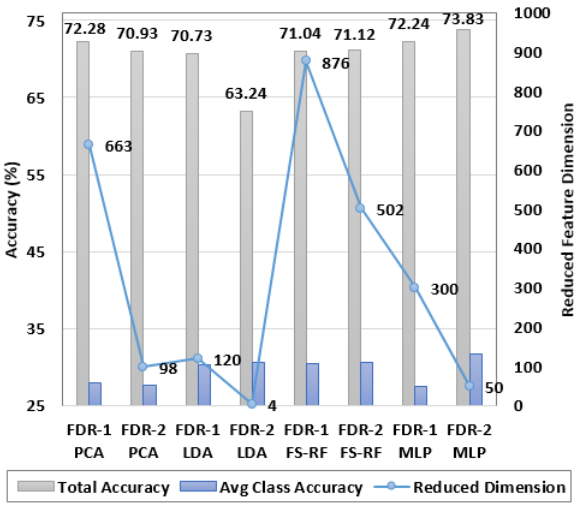

Fig. 6. RF classification accuracy on reduced features using FDR-1 & 2 techniques on RGB color features.

In addition, FDR MLP approach is flexible in terms of expected dimensionality of features on a single color space. Furthermore, in future, other feature descriptors can be scaled and fused with reduced multi-scale color LBP descriptor via hierarchical FDR MLP approach. Thereby, it can strengthen the discriminatory power of features on a single color space and ensure that the dimensionality of features stays within a reasonable upper bound.

B. FDR-2 MLP Experiments on Various Color Spaces

FDR-2 MLP approach is applied to reduce features from each of the color spaces: RGB, YIQ, rgb, HSV, and L*a*b*. $MLP_2[100, 50]$ reduces feature dimensionality to 50. Classi-

fication accuracy is evaluated by RF, one-vs-rest SVM, one-vs-one SVM, and MLP (softmax decision layer).

Table II shows the total accuracy and average class accuracy on various color spaces using four different classifiers. One-vs-one SVM classifier consistently outperforms other classifiers on each of the five color spaces in terms of total accuracy by approximately 0.4% on average.

We notice an approximate average gain of 1.4% both in total accuracy and class average accuracy with FDR-2 MLP upon FDR-1 MLP using MLP classifier consistently across each of the five color spaces. Thus, it illustrates the positive role that FDR plays in improving recognition of DR stages with LBP descriptors.

In DR problem, RGB is still the most discriminatory color space within the selected group. This is in contrast to empirical results in other domains such as scene and face recognition where several other color spaces perform better than RGB [9], [13], [14]. Labels in face or scene dataset represent true ground truth, i.e., a face or scene image is taken of a known subject or place respectively, which serves as ground truth and does not require inferencing on part of a human to label an image. However, DR stages are manually labeled by ophthalmologists by observing retina images in RGB color space, which could induce inference bias towards RGB color space in generating the ground truth. The extremely fine grained nature of DR signs within an image and across DR stages may further bias the manual labeling process towards RGB color space, it is reasonable to obtain the most accurate prediction on RGB. Thus, DR labels may not strictly represent ground truth.

Our experiments show both luminance and chrominance information is useful in DR recognition. A color space that lacks luminance information such as rgb is lower in terms of both total and average class accuracies in comparison to other color spaces. RGB surpasses rgb by 2.1% in total accuracy and 6.1% in average class accuracy as seen in Table II.

C. FDR-2 and FDR-3 MLP Experiments on Fusion of Multiple Color Spaces

Fusion of multiple color spaces is organized in a level-3 feature hierarchy. FDR-2 & 3 MLP approaches are applied on fusion of selected color spaces. $MLP_3[80, 50]$ reduces 100 concatenated color features from fusion of two color spaces and $MLP_3[100, 50]$ reduces fused color features from various combinations of three or more color spaces, reduced feature size is 50 dimensional. SVM one-vs-one classification accuracy on color fusion and on individual color spaces is summarized in Fig. 7.

Although, performance of FDR-3 MLP is slightly less than FDR-2 MLP on color fusion, there is a huge saving in terms of time. For instance, FDR-3 MLP on 4-color fusion of RGB, Lab, HSV, and YIQ loses merely 0.11% in accuracy but the SVM recognition task on 50 dimensional feature vector is 60 times faster than FDR-2 MLP. Results from both approaches show that color fusion clearly outperforms individual color spaces in terms of accuracy.

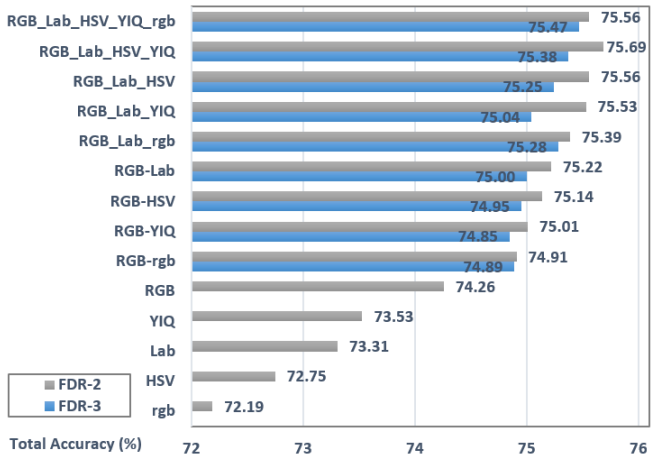

Fig. 7. SVM one-vs-one total classification accuracy on reduced features using FDR-2 & 3 MLP techniques on multiple color spaces and their fusion.

Performance on fused features from RGB and rgb color spaces improve upon RGB by approximately 0.6% in total accuracy and 1.0% in average class accuracy, as shown in Table III. Thus, information contained in rgb features is not completely redundant when combined with RGB. Discriminatory power of color fusion is best expressed in fusion of four color spaces RGB, Lab, HSV, and YIQ. It gains at least 1.4% in total accuracy and 2.7% in average class accuracy upon individual color spaces. This implies that there exists different discriminatory information in different color spaces in context of DR recognition task. Results in Table III indicate incremental accuracy gain across almost all DR stages by color fusion upon the strongest single color space RGB. Also, the table reveals that additional advanced techniques for feature descriptors are definitely required to improve performance on individual DR stages and more so for NPDR-mild stage.

VIII. Conclusion and Future Work

This paper proposes a novel hierarchical feature dimensionality reduction (FDR) methodology for fusion of features from multiple color spaces. Results of the experiments using grand challenge fine grained DR dataset show FDR with multi-layer perceptron surpasses other techniques in terms of accuracy and scalability. Feature descriptor on RGB improves performance on descriptors from other four color spaces. The novel combination of the proposed FDR method with fusion of multiple color features improves classification performance, which indicates various color descriptors contain complementary information.

The proven success of hierarchical FDR methodology in this paper offers higher accuracy, scalability, and time efficiency. Thereby, it could serve as an efficient mechanism for combining LBP features with other descriptors and extract more discriminatory features for DR recognition as future work.

TABLE II
CLASSIFICATION ACCURACY OF VARIOUS CLASSIFIERS ON FIVE COLOR SPACES. FEATURES ARE REDUCED BY FDR-2 MLP

Color Space	Dim	MLP (%)		Random Forest (%)		SVM one vs. rest (%)		SVM one vs. one (%)	
		Total Accuracy	*Average Class Accuracy*	*Total Accuracy*	*Average Class Accuracy*	*Total Accuracy*	*Average Class Accuracy*	*Total Accuracy*	*Average Class Accuracy*
RGB	50	73.39	34.20	73.83	32.58	73.75	33.60	**74.26**	33.06
YIQ	50	72.25	33.58	73.17	31.51	73.31	31.99	**73.53**	31.23
L*a*b*	50	71.99	33.80	73.01	31.54	72.52	33.02	**73.31**	32.63
HSV	50	71.44	30.65	72.47	27.83	72.69	29.12	**72.75**	28.82
rgb	50	70.38	29.14	72.09	26.25	71.80	28.03	**72.19**	27.18

TABLE III
CLASSIFICATION ACCURACY OF SVM ONE-VS-ONE ON RGB AND FUSED COLOR FEATURES ACROSS DR STAGES. FEATURES ARE REDUCED BY FDR-2 MLP

Color Space(s)	DR Stages Accuracy (%)					Accuracy (%)	
	non-DR (0)	*mild (1)*	*mod (2)*	*severe (3)*	*PDR (4)*	*Total*	*Avg. Class*
RGB	94.69	0.90	22.19	12.44	35.16	74.26	33.06
RGB-rgb	**94.89**	0.69	**25.51**	**15.98**	33.00	74.91	34.01
RGB-Lab-HSV-YIQ	**94.94**	0.53	**29.97**	**17.63**	**35.66**	75.69	35.75
5-color fusion	**95.15**	0.59	**27.48**	**15.49**	**35.57**	75.56	34.86

REFERENCES

[1] J. Lachure et al., "Diabetic retinopathy using morphological operations and machine learning," *IEEE Int. Advance Computing Conf. (IACC)*, Bangalore, India, 2015.

[2] U. R. Acharya et al., "Computer-based detection of diabetes retinopathy stages using digital fundus images," *Proc. Institution of Mechanical Engineers, Part H: J. Engineering in Medicine*, vol. 223, no. 5, pp. 545-553, 2009.

[3] J. Nayak et al., "Automated identification of diabetic retinopathy stages using digital fundus images," *J. Medical Syst.*, vol. 32, no. 2, pp. 107-115, 2008.

[4] W. L. Yun et al., "Identification of different stages of diabetic retinopathy using retinal optical images," *Inform. Sci.*, vol. 178, no. 1, pp. 106-121, 2008.

[5] K. Ram and S. Jayanthi, "Multi-space clustering for segmentation of exudates in retinal color photographs," *Annu. Int. Conf. of the IEEE Eng. in Medicine and Biology Society (EMBC)*, Minneapolis, MN, 2009.

[6] T. Ojala et al., "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51-59, 1996.

[7] T. Ojala et al., "Multiresolution Gray Scale and Rotation Invariant Texture Classification With Local Binary Patterns," *IEEE Tran. on Pattern Anal. and Machine Intell.*, vol. 24, no. 7, pp. 971-987, Jul. 2002.

[8] T. Ahonen et al., "Face description with local binary patterns: Application to face recognition," *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 28, no. 12 , pp. 2037-2041, 2006.

[9] S. Banerji et al., "Novel color LBP descriptors for scene and image texture classification," *Int. Conf. on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, Las Vegas, NV, 2011.

[10] O. Faust et al., "Algorithms for the automated detection of diabetic retinopathy using digital fundus images: a review," *J. Medical Syst.*, vol. 36, no. 1, pp. 145-157, 2012.

[11] S. Roychowdhury et al., "Dream: Diabetic retinopathy analysis using machine learning," *IEEE J. Biomed. and Health Informatics*, vol. 18, no.5, pp. 1717-1728, 2014.

[12] S. Ravishankar et al., "Automated feature extraction for early detection of diabetic retinopathy in fundus images," *IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pp. 210-217, 2009.

[13] C. Liu and H. Wechsler, "Robust coding schemes for indexing and retrieval from large face databases," *IEEE Trans. Image Processing*, vol. 9, no. 1, pp. 132-137, 2000.

[14] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Trans. Image Processing*, vol. 11, no. 4, pp. 467-476, 2002.

[15] B. Antal and H. Andras, "An ensemble-based system for microaneurysm detection and diabetic retinopathy grading," *IEEE Trans. Biomedical Eng.*, vol. 59, no. 6, pp. 1720-1726, 2012.

[16] M. R. K. Mookiah et al., "Computer-aided diagnosis of diabetic retinopathy: A review," *Comput. in Biology and Medicine*, vol. 43, no. 12, pp. 2136-2155, 2013.

[17] T. Walter and J. Klein, "Segmentation of color fundus images of the human retina: Detection of the optic disc and the vascular tree using morphological techniques," *Int. Symp. on Medical Data Analysis (ISMDA)*, Madrid, Spain, 2001.

[18] C. Sinthanayothin et al., "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images," *British J. Ophthalmology*, vol. 83, no. 8, pp. 902-910, 1999.

[19] J. de la Calleja et al., "LBP and Machine Learning for Diabetic Retinopathy Detection," *Intelligent Data Eng. and Automated Learning (IDEAL)*, Salamanca, Spain, pp. 110-117, 2014.

[20] M. N. Ashraf et al., "Texture Feature Analysis of Digital Fundus Images for Early Detection of Diabetic Retinopathy," *11th Int. IEEE Conf. Computer Graphics, Imaging and Visualization (CGIV)*, Singapore, pp. 57-62, 2014.

[21] R. Casanova et al., "Application of random forests methods to diabetic retinopathy classification analyses," *PloS one*, vol. 9, no.6, e98587, 2014.

[22] P. van der MLJP and J. van den HH, "Dimensionality reduction: A comparative review," Tilburg Centre for Creative Computing, Tilburg University, Tilburg, Netherlands, Rep. 2009-005, 2009.

[23] J. Yan et al, "Effective and efficient dimensionality reduction for large-scale and streaming data preprocessing," *IEEE Trans. Knowledge and Data Eng.*, vol. 18, no. 3, pp. 320-333, 2006.

[24] A. A. Farag and S. Y. Elhabian, "A Tutorial on Data Reduction Linear Discriminant Analysis (LDA)," University of Louisville, Tech. Rep., 2008.

[25] N. Japkowicz et al, "Nonlinear autoassociation is not equivalent to PCA," *Neural computation*, vol. 12, no. 3, pp. 531-545, 2000.

[26] (Accessed: 2021, February 1). [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection

[27] J. S. Mandava, F. Kocaman, D. Bein, A. Verma, et al., "Machine Learning for Classification of Cancer Dataset for Gene Mutation Based Treatment," the 19th International Conference on Information Technology: New Generations, Apr. 11-12, 2022, Virtual.

[28] P. Ly, D. Bein, and A. Verma, "New Compact Deep Learning Model for Skin Cancer Recognition," the 9th IEEE Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), Nov. 8-10, 2018, New York, NY, USA.