

Received November 27, 2018, accepted December 8, 2018, date of publication December 11, 2018, date of current version January 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2886315

# Deriving Probabilistic SVM Kernels From Flexible Statistical Mixture Models and Its Application to Retinal Images Classification

SAMI BOUROUIS<sup>1</sup>, ATEF ZAGUIA<sup>1</sup>, NIZAR BOUGUILA<sup>2</sup>, (Senior Member, IEEE), AND ROOBAEA ALROOBAEA<sup>1</sup>

<sup>1</sup>College of Computers and Information Technology, Taif University, Ta'if 21974, Saudi Arabia

<sup>2</sup>The Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal, QC H3G 1T7, Canada

Corresponding author: Sami Bourouis (s.bourouis@tu.edu.sa)

This work was supported by the Deanship of Scientific Research at Taif University, Saudi Arabia, under Grant 1-438-5730.

**ABSTRACT** This paper aims to propose a robust hybrid probabilistic learning approach that combines appropriately the advantages of both the generative and discriminative models for the challenging problem of diabetic retinopathy classification in retinal images. We build new probabilistic kernels based on information divergences and Fisher score from the mixture of scaled Dirichlet distributions for support vector machines (SVMs). We also investigate the incorporation of a minimum description length criterion into the learning model to deal with the common problems of determining suitable components and also selecting the best model that describes the dataset. The developed hybrid model is introduced in this paper as an effective SVM kernel able to incorporate prior knowledge about the nature of data involved in the problem at hand and, therefore, permits a good data discrimination. Our approach has been shown to be a better alternative to other methods, which is able to describe the intrinsic nature of datasets and to be of a significant value in a variety of applications involving data classification. We demonstrate the flexibility and the merits of the proposed framework for the problem of diabetic retinopathy detection in eye images.

**INDEX TERMS** Retinal images, scaled Dirichlet mixture, SVM, generative-discriminative learning, MDL, probabilistic kernels.

## I. INTRODUCTION

### A. PROBLEM STATEMENT

The problem of diabetic retinopathy lesions detection in eye images is addressed in this paper. In fact, the World Health Organization (WHO) cited that diabetes mellitus (DM) is the most common endocrine disease in the World.<sup>1</sup> WHO estimated in 2014 that about 347 million people worldwide have diabetes. In 2012, 1.5 million deaths were directly caused by diabetes. Based on many studies in the world,<sup>2</sup> the most serious complication of long-term diabetes is the diabetic retinopathy (DR). Diabetic retinopathy is a serious disease which causes vision loss in the working-age population. For instance, based on some studies in [1]–[5], visual loss can be prevented if the DR is detected and treated earlier. However, this is can be difficult as the disease often shows few symptoms until it is too late to provide effective treatment. Unfortunately, the protocol followed by trained specialist

is time consuming and exhaustive. Consequently, there is an urgent need for efficient DR screening tools allowing timely treatment and early revealing of diseases and critical risks. This challenging topic has gained importance from the research community [4], [6]–[8] and several computer-aided screening systems have been developed in the last decay in which image processing helps to diagnose diabetic retinopathy. Their common purpose is to detect the most common signs of DR such as exudates, hemorrhages (HA), cotton wool spots (CWS) and microaneurysms (MA) [9].

### B. RELATED WORKS

In the literature there are numerous studies related to the detection of DR lesions such as exudate [10]–[12] microaneurysm and hemorrhage detection [13]–[15]. Among the successful approaches, the authors in [16] focused on the identification and staging of DR by applying a combination of SVM classifier and a set of features of higher-order spectra. An extension of this work was also developed in [17] that

<sup>1</sup>[http://www.who.int/topics/diabetes\\_mellitus/en/](http://www.who.int/topics/diabetes_mellitus/en/)

<sup>2</sup><https://www.diabetes.co.uk/references.html>

exploits texture features which are calculated by run length and co-occurrence matrices. In [18] authors proposed a relevant features extraction (AM-FM) technique and a hierarchical clustering to detect DR lesions and to grade their severity. Other relevant works are proposed to detect the microaneurysms (MA) lesion with different techniques like wavelet transform-based template matching [19]; dynamic thresholding-based method followed by a multi-scale correlation filtering [15]; detector framework [20] that uses the principle of selecting the best optimal combination (pre-processing and candidate extraction steps). In [21] MAs are identified based on the local rotating cross-section profile. Some features like shape, size, and height are extracted to analyze the profile. False detections are also removed through a classification step with Naive Bayes classifier. On the other hand, exudates pathologies are also treated and extracted with other approaches in retinopathy images such as fuzzy FCM clustering algorithm and morphological operators [22] and computational intelligence technique [10]. For the latter, segmented regions are classified into 2 classes: exudates and non-exudates. Classification process is based on the selection of a subset of features like the size, color, texture, and edge strength. Another interesting method was proposed in this context that uses generic contextual information [12]. In particular, a step of spatial relation modeling was performed between the exudates and the lesions in 2D retinal images. Other types of lesions “hemorrhages” are also examined and detected with a process of a splat feature classification method [14] that involves a step of extraction and selection of best subset of visual features (i.e spatial location, shape, and color). Other studies are also proposed in the literature for screening diabetic retinopathy through for example segmenting the vessels in the eye images [23]–[26]. Even several approaches have been proposed to solve the problem of diabetic retinopathy detection, however, the major problems facing researchers in developing accurate methods are related to the inter- and intra- variability in appearance that causes a lot of false positives. Furthermore, a lot of works ignore the richness of low-level visual content of eye images which can be very helpful. Thus, the need for developing more intelligent and flexible tool for DR detection able to achieve high performance rate is highly desirable.

### C. MOTIVATION AND CONTRIBUTION

Diabetic retinopathy detection can be viewed as the problem of reliably distinguishing between normal and abnormal eye images. Comparing classification algorithms is a too difficult problem as many efforts have been achieved in developing methods taking into account different assumptions (like the nature of data to be classified). It is possible to divide these approaches into two main categories: generative probabilistic-based approaches (known also as model-based approaches) and discriminative approaches. Generative models have the advantage of offering a principled effective way for handling uncertainty and they are practical especially when dealing with missing and incomplete

data [27]–[30] by learning the joint probability model  $p(x,y)$  of input  $x$  and class label  $y$ . Nevertheless, the classification performance of these models is generally lower than discriminative models [31]–[33]. The latter are particularly used in the construction of flexible decision boundaries and recognition problems. They have been applied with success in various applications such as speech recognition, image classification, text classification, and bio-informatics, etc. SVM is a well-known example of discriminative classifiers [31], [33], [34]. In many settings traditional discriminative or generative approaches fail to achieve accurate classification results. For example, SVMs-based discriminative classifiers are often relying on classic kernels which are not generally suitable for proportional data (such as the case of DR classification). One major restriction of such kernels is that they do not take into consideration the nature of data. Therefore, selecting an appropriate kernel is really a challenging task for machine learning and data mining applications. To cope with these limitations, there has been a growing attention in the so-called “hybrid generative-discriminative approaches” that can efficiently combine the strengths of both categories [35]–[38]. Several hybrid schemes have been developed and proven to be helpful for the above mentioned applications such as scene classification [39]. In general, the designed hybrid pipeline allows the estimation of the component model via a generative technique and the classification process is performed through a discriminative method. On the other hand, statistical methods have received a lot of attention in several domains such as medical image analysis, pattern recognition, and data mining and proved to be helpful in generating well-defined models. In particular, finite Gaussian mixture models (GMM) have been widely used and have played an important role in data classification and machine learning [27], [40], [41]. The main advantage of GMM is that it requires a small number of parameters for learning which can be accurately estimated by adopting the EM-algorithm [42]–[44] to maximize the log-likelihood function. Although GMMs present some flexibility for data analysis, their effectiveness is limited to the presence of outliers and to applications where underlying modeling assumptions (e.g. the per-components densities are Gaussian) are reasonably satisfied. Moreover, the tail of the Gaussian distribution is shorter than what is required. Consequently, much research efforts have been devoted to developing better alternatives to tackle the problem of data classification. Our work is motivated by the need to address the challenge of finding a more flexible distribution that can model sub-populations or clusters inherent in our observed eye images. Indeed, we focus here on constructing new statistical models able to take into account the complexity of the distributions of real proportional data. This novel mixture model is based on the scaled Dirichlet mixture model (SDMM) [45]. Our choice for scaled Dirichlet mixture model is justified by its high flexibility in data modelling and good results [45]. In fact, this new derived mixture model is considered as a more flexible model than the classic and

conventional ones [46]. On the other hand, many developed approaches for data classification and clustering are facing a common problem that concerns the automatic determination of accurate components (or classes) which best describe the data. In order to deal with this issue, some approaches have been suggested with relative success (see, for example, [27], [47]). In particular, the minimum description length (MDL) criterion was applied to solve such problem [27], [47]. Motivated by all these observations, we introduce in this work an alternative flexible mixture model based on the scaled Dirichlet mixture model which incorporates the MDL criterion to overcome the cited previous limitations and to improve data modeling and classification. In this study, we tackle the problem of modeling, classification and learning of retinal images by developing a new hybrid generative/discriminative framework able to better differentiate between normal and abnormal retina with possible DR lesions. The ultimate idea is to build accurate kernels from scaled Dirichlet distributions taking into account MDL criterion that maintain the balance between model complexity and goodness of fit. Thus, scaled Dirichlet mixture models will be our generative part and support vector machines as the discriminative counterpart. It is noteworthy that the proposed probabilistic hybrid approach has never been proposed before for DR detection and classification. It should be pointed out also that some different preliminary results were already published in our previous work [48]; however, we provide in this paper a more complete and detailed work including more contributions and more extensive experiments. In particular, previous work does not take into account MDL criterion through the developed generative mixture model which can lead to more stable and accurate data modelling and so better results. In addition, we develop here several SVM-based kernels to study and evaluate their robustness and also we perform more extensive experiments and comparison study for evaluation purposes.

The rest of this paper is organized as follows. The next section will introduce our statistical generative-discriminative framework and all related details. Then in section 3, we present and discuss the obtained experimental results. Finally, we conclude this work in Section 4.

## II. GENERATIVE/DISCRIMINATIVE LEARNING: A HYBRID OF SCALED DIRICHLET MIXTURE MODELS AND SVM

Support Vector Machine (SVM) is one of the most successful learning algorithms for binary classification and for many pattern recognition applications thanks to its good discrimination and generalization capabilities [49]. We choose to apply SVM for Diabetic Retinopathy detection since this problem can be viewed as a binary classification problem between normal and abnormal retine images. However, the challenging issue related to using SVM is the choice of the kernel function  $\mathcal{K} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  which is a measure of similarity between input vectors. To achieve better classification results, such kernel must capture the intrinsic properties of the input data and should consider the prior knowledge of the problem in hand. In reality, several kernel

functions have been developed in the past such as the linear, the polynomial, the radial basis function RBF, and the multi layer perceptron [50]. Nevertheless, these standard kernels ignore, in general, the intrinsic structure of the data and are used independently of the problem. On the other hand, it was shown that in order to have better results, the selected kernel function must be generated directly from data [37]. Among the most successful developed methods in this field that address certain practical shortcomings of standard kernels are the Fisher kernel [37] and the Kullback-Leibler divergence-based kernel [51]. In the following subsections, we derive different probabilistic kernels from mixture of scaled Dirichlet distributions for Support Vector Machines (SVMs) to tackle the problem of DR classification.

### A. THE SCALED DIRICHLET MIXTURE MODEL (SDMM)

If we suppose that  $\vec{Y}_n = (y_{n1}, \dots, y_{nD})$  is D-dimensional random vector distributed according to a scaled Dirichlet Mixture Model (SDMM) with parameters  $\theta_k = (\vec{\alpha}_k, \vec{\beta}_k)$ , then the corresponding density function  $p(\vec{Y}_n|\theta_k)$  is defined as:

$$p(\vec{Y}_n|\theta_k) = \frac{\Gamma(\alpha_{k+})}{\prod_{d=1}^D \Gamma(\alpha_{kd})} \frac{\prod_{d=1}^D \beta_{kd}^{\alpha_{kd}} y_{nd}^{\alpha_{kd}-1}}{(\sum_{d=1}^D \beta_{kd} y_{nd})^{\alpha_{k+}}} \quad (1)$$

where  $\Gamma$  denotes the Gamma function,  $\alpha_{k+} = \sum_{d=1}^D \alpha_{kd}$ ,  $\vec{\alpha}_k = (\alpha_{k1}, \dots, \alpha_{kD})$  and  $\vec{\beta}_k = (\beta_{k1}, \dots, \beta_{kD})$  are the shape and the scale parameters, respectively. The role of latter is to describe how the density is plotted, however, the role of the shape parameter is to control the distribution's form and to discover patterns inherent in a dataset. The formula of the corresponding likelihood is given as:

$$p(\mathcal{Y}|\Theta) = \prod_{n=1}^N \sum_{k=1}^K p_k p(\vec{Y}_n|\theta_k) \quad (2)$$

where  $\Theta = (\vec{p}, \theta)$ , is the complete parameter of the mixture.  $\theta = \{\theta_k\}$ 's are parameter vectors of all components and  $\vec{p} = (p_1, \dots, p_K)$  denotes their mixing parameters ( $\pi_j \geq 0$  and  $\sum_{j=1}^K \pi_j = 1$ ). Each sample vector  $\vec{Y}_n$  is drawn from one specific component  $k$  and we must estimate its membership.

For estimating the scaled Dirichlet mixture model's parameters, we employ here one of the widely used methods which is the maximum likelihood (ML). ML finds out the parameters values that maximize the probability of the vectors  $\vec{Y}_i$  given that we have unobserved latent (missing) vectors  $\vec{Z}_i = (Z_{i1}, \dots, Z_{iK})$ ,  $n = 1, \dots, N$ , where  $Z_{nk} = 1$  if  $\vec{Y}_n$  belongs to class  $j$  and 0, otherwise. In practice, the ML estimates are computed using the EM algorithm that takes into account the latent variables  $\mathcal{Z} = \{\vec{Z}_1, \vec{Z}_2, \dots, \vec{Z}_N\}$ . Once the data  $\mathcal{Y}$  are combined with the latent variables  $\mathcal{Z}$  (i.e the likelihood if  $\mathcal{Z}$  was observed), we can find the model parameters as follow:

$$\begin{aligned} \Theta_{MLE} &= \arg \max_{\Theta} \mathcal{L}(\mathcal{Y}, \mathcal{Z}|\Theta) \\ &= \arg \max_{\Theta} \log(p(\mathcal{Y}, \mathcal{Z}|\Theta)) \\ &= \arg \max_{\Theta} \left\{ \sum_{n=1}^N \sum_{k=1}^K \hat{Z}_{nk} \log[p_k p(\vec{Y}_n|\theta_k)] \right\} \quad (3) \end{aligned}$$

**B. ESTIMATION ALGORITHM AND SELECTION OF THE NUMBER OF CLUSTERS**

We recall that the maximum likelihood (ML) favors generally higher values of the number of components and this issue leads easily to over fitting. Thus, in order to deal with the difficult problem of finding the optimal number of clusters (M) and also selecting the best model that describes accurately the dataset, we investigate in this work the integration of a minimum description length (MDL) criterion [47] into the model learning step. The minimum description length (MDL) principle is a basic information-theoretic method in statistical model selection (and machine learning) for re-encoding the training dataset. Principally, it states that one must pick the model that gives you the shortest description of the data (or the most data (or the most compact description of the data while the model’s complexity is also considered). Thus, MDL assists methods to select the most suitable having the smallest description length while taking into account the given data. As proposed by Rissanen [47], MDL is the description length of a specified model applied to a specific training data w.r.t that model and it is determined by minimizing the sum of the code-length for describing the data (w.r.t the model fitting the data) and the code-length for describing the model (i.e. model’s algorithmic complexity).

Therefore, the optimal number of components in the developed scaled Dirichlet mixture model is obtained by minimizing the following function:

$$MDL \approx -\log p(Y|\Theta) + \frac{1}{2} N_l \log(N) \tag{4}$$

where  $N_l$  denotes the number of free parameters in the scaled Dirichlet mixture model. The complete parameter estimation steps for the mixture model are summarized in Algorithm 1.

**C. FISHER KERNEL**

The Fisher kernel was developed initially in [37] and it is based on the principle of mapping each individual sequence into a single feature vector, defined in the gradient log-likelihood space. The kernel is computed at the estimated  $\Theta$  on the resulting statistical manifold as follows:

$$\begin{aligned} \mathcal{K}(\vec{Y}, \vec{Y}_n) &= U_{\vec{Y}}^{tr}(\Theta) F(\Theta)^{-1} U_{\vec{Y}_n}(\Theta) \\ &= U_{\vec{Y}}^{tr}(\Theta) I(\Theta)^{-1} U_{\vec{Y}_n}(\Theta) \end{aligned} \tag{7}$$

where  $F(\Theta)$  denotes the Fisher information matrix whose role is less significant and then can be approximated by the identity matrix  $I(\Theta)$  [37]. Each component of  $U_{\vec{Y}}(\Theta) = \nabla \log(p(\mathcal{Y}|\Theta)) = \frac{\partial \log(p(\mathcal{Y}|\Theta))}{\partial \Theta}$  is the Fisher score information which is the derivative of the log-likelihood (i.e the gradient of  $\log(p(\mathcal{Y}|\Theta))$ ) with respect to  $\Theta$ . In the following, we develop the Fisher kernels for our generative mixture model (the scaled Dirichlet model) by computing the gradient of log-likelihood with respect to the parameters  $(p_k, \alpha_k, \beta_k)$

**Algorithm 1** SDMM-Based EM Algorithm and Model Selection

```

Input : Retine image to be classified  $\mathcal{Y}$ 
Output: Retine image classified as pathology (DR) or normal

1 Initialization algorithm :
2 Model’s parameters  $\Theta$  are initialized using the K-means algorithm.
3 foreach candidate value of  $M$  do
4   repeat
5     {E-step} : Compute the posterior probability.
6
7     
$$\hat{z}_{nj} = \frac{p_j p(\vec{Y}_n | \theta_j)}{\sum_{l=1}^K p_l p(\vec{Y}_n | \theta_l)} \tag{5}$$

8     {M-step} : Update the following parameters:
9     Update the weights as:
10    
$$p_j = \frac{1}{N} \sum_{n=1}^N \hat{z}_{nj} \tag{6}$$

11    Update the parameters mean, covariance and the scale parameter as in [38], [45]
12  until Algorithm convergence
13 Calculate the associated MDL criterion.
14 end
15 Select the optimal model  $M^*$  such that:
16  $M^* = \arg \max_M MDL(M)$ 

```

of each component  $k$ , as the following:

$$\begin{aligned} \frac{\partial \log(\mathcal{Y}|\Theta)}{\partial \alpha_{kd}} &= \sum_{n=1}^N \hat{Z}_{nk} \left( \Psi(\alpha_+) - \Psi(\alpha_d) + \log \beta_d \right. \\ &\quad \left. + \log y_{nd} - \log \left( \sum_{d=1}^D \beta_d y_{nd} \right) \right) \end{aligned} \tag{8}$$

$$\frac{\partial \log(\mathcal{Y}|\Theta)}{\partial \beta_{kd}} = \sum_{n=1}^N \hat{Z}_{nk} \left( \frac{\alpha_d}{\beta_d} \frac{\alpha_+ y_{nd}}{\sum_{d=1}^D \beta_d y_{nd}} \right) \tag{9}$$

$$\frac{\partial \log(\mathcal{Y}|\Theta)}{\partial p_n} = \sum_{n=1}^N \frac{\hat{Z}_{nk}}{p_k} \tag{10}$$

such as the function  $\Psi(\alpha) = \frac{\partial \log \Gamma(\alpha)}{\partial \alpha}$  represents the so-called “digamma function”.

**D. BHATTACHARYYA KERNEL**

Bhattacharyya kernel has been proposed in [52] by replacing the kernel computation in the original sequence space by computation in the probability density function (PDF) space as following:

$$\begin{aligned} \mathcal{K}(\mathcal{Y}_1, \mathcal{Y}_2) &\Rightarrow \mathcal{K}_{\frac{1}{2}}(p(\vec{Y}|\Theta_1), q(\vec{Y}|\Theta_2)) \\ &= \int_0^1 p(\vec{Y}|\Theta_1)^{1/2} q(\vec{Y}|\Theta_2)^{1/2} d\vec{Y}. \end{aligned} \tag{11}$$

We note that it is not possible to compute the Bhattacharyya kernel in a closed form for scaled Dirichled mixture density (SDMM). Thus, in the absence of closed forum for SDMM, we can approximate our Bhattacharyya kernel using Monte Carlo simulations [53].

$$\begin{aligned} \mathcal{K}_{\frac{1}{2}}(p(\vec{Y}|\theta_1), q(\vec{Y}|\theta_2)) \\ \approx \frac{\beta}{N_1} \sum_{i=1}^{N_1} \frac{p^{1/2}(\vec{Y}_i|\Theta_1)}{Z_1} p^{1/2}(\vec{Y}_i|\Theta_1) \\ + \frac{1-\beta}{N_2} \sum_{i=1}^{N_2} \frac{q^{1/2}(\vec{Y}_i|\Theta_2)}{Z_2} q^{1/2}(\vec{Y}_i|\Theta_2) \end{aligned} \quad (12)$$

Where  $\beta \in [0, 1]$  and  $Z_1, Z_2$  are normalized factors for  $p$  and  $q$  densities.

### E. INFORMATION DIVERGENCE KERNELS

More attractive alternatives of SVM kernels that take into consideration the advantages of generative models are well documented previously in [38] and [53]. In particular the one based on the information divergence distance such as the Kullback-Leibler kernel [51]. The later has been applied with success for several applications like visual recognition and speaker identification [54]. Let  $\vec{Y}_1$  and  $\vec{Y}_2$  be two sequences of feature vectors representing two objects or images and modeled respectively by two finite mixture models  $p(\vec{Y}|\Theta_1)$  and  $q(\vec{Y}|\Theta_2)$ . The Kullback-Leibler kernel is based on the the symmetric Kullback-Leibler divergence (KL) that measure the dissimilarity between two probability distributions and is defined as follow:

$$\mathcal{K}_{KL}(p(\vec{Y}|\Theta_1), q(\vec{Y}|\Theta_2)) = e^{\alpha F(p(\vec{Y}|\Theta_1), q(\vec{Y}|\Theta_2))} \quad (13)$$

Where  $\alpha$  is a kernel parameter used for numerical stability, and  $F$  is the symmetric KL divergence between the two distributions:

$$\begin{aligned} F(p(\vec{Y}|\Theta_1), q(\vec{Y}|\Theta_2)) = \int_{\Omega} \left( p(\vec{Y}|\Theta_1) \log \frac{p(\vec{Y}|\Theta_1)}{q(\vec{Y}|\Theta_2)} \right. \\ \left. + q(\vec{Y}|\Theta_2) \log \frac{q(\vec{Y}|\Theta_2)}{p(\vec{Y}|\Theta_1)} \right) \end{aligned} \quad (14)$$

In the case of the scaled Dirichlet distribution, we cannot find a closed-form expression for the KL-divergence, thus like the previous kernel we consider Monte Carlo simulation.

### F. RÉNYI AND JENSEN-SHANNON KERNELS

We will consider in this work two other special probabilistic kernels: the Rényi and Jensen-Shannon kernels. Indeed, these kernels have been introduced in [53] as a generalization of the symmetric Kullback-Leibler kernel. The Rényi kernel is based on the symmetric Rényi divergence [53]

$$\begin{aligned} \mathcal{K}(\mathcal{Y}_1, \mathcal{Y}_2) \Rightarrow \mathcal{K}_R(p(\vec{Y}|\Theta_1), q(\vec{Y}|\Theta_2)) \\ = \exp \left[ -AR(p(\vec{Y}|\Theta_1), q(\vec{Y}|\Theta_2)) \right] \end{aligned} \quad (15)$$

where

$$\begin{aligned} R(p(\vec{Y}|\Theta_1), q(\vec{Y}|\Theta_2)) \\ = \frac{1}{\sigma-1} \log \int_0^1 p(\vec{Y}|\Theta_1)^\sigma q(\vec{Y}|\Theta_2)^{1-\sigma} d\vec{Y} \\ + \frac{1}{\sigma-1} \log \int_0^1 q(\vec{Y}|\Theta_2)^\sigma p(\vec{Y}|\Theta_1)^{1-\sigma} d\vec{Y} \end{aligned} \quad (16)$$

where  $\sigma > 0$  and  $\sigma \neq 1$  is the order of Rényi divergence. By substituting Eq. 16 into Eq. 15, we obtain the following

$$\begin{aligned} \mathcal{K}_R(p(\vec{Y}|\Theta_1), q(\vec{Y}|\Theta_2)) \\ = \left[ \int_0^1 p(\vec{Y}|\Theta_1)^\sigma q(\vec{Y}|\Theta_2)^{1-\sigma} d\vec{Y} \right. \\ \left. \times \int_0^1 q(\vec{Y}|\Theta_2)^\sigma p(\vec{Y}|\Theta_1)^{1-\sigma} d\vec{Y} \right]^{\frac{1}{1-\sigma}} \end{aligned} \quad (17)$$

The second kernel is the Jensen-Shannon (JS) Kernel, generated according to the Jensen-Shannon divergence [55], and is given by [53]

$$\begin{aligned} \mathcal{K}_{JS}(\mathcal{Y}_1, \mathcal{Y}_2) \Rightarrow \mathcal{K}(p(\vec{Y}|\Theta_1), q(\vec{Y}|\Theta_2)) \\ = \exp \left[ -AJ S_\omega(p(\vec{Y}|\Theta_1), q(\vec{Y}|\Theta_2)) \right] \end{aligned} \quad (18)$$

where

$$\begin{aligned} JS_\omega(p(\vec{Y}|\Theta_1), q(\vec{Y}|\Theta_2)) \\ = H[\omega p(\vec{Y}|\Theta_1) + (1-\omega)q(\vec{Y}|\Theta_2)] \\ - \omega H[p(\vec{Y}|\Theta_1)] - (1-\omega)H[q(\vec{Y}|\Theta_2)] \end{aligned}$$

where  $\omega$  is a parameter and

$$H[p(\vec{Y}|\Theta_1)] = - \int_0^1 p(\vec{Y}|\Theta_1) \log p(\vec{Y}|\Theta_1) d\vec{Y} \quad (19)$$

is the Shannon entropy.

## III. EXPERIMENTAL RESULTS

In this section, experiments have been carried out to evaluate the capabilities of the developed hybrid framework in retinal images classification problem. In particular, we compare our proposed model with other state of the art mixture models that have been widely used in the literature.

### A. DESIGN OF EXPERIMENTS

The following measures have been used to evaluate obtained results:

- **Accuracy:** is a performance metric that gives an indication of overall well classified elements and is defined as:  $Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$ , where  $TP$ ,  $FP$ ,  $TN$ ,  $FN$  denote: true positive, false positive, true negative, and false negative, respectively.
- **AUC = Area Under the ROC Curve:** measures the performance of the multi-class classification problem under the ROC (Receiver Operating Characteristics) curve. It is one of the most interesting assessment measures for examining classification model's performance.

**TABLE 1.** Results using DIARETDB1 data set based on kernels generated from finite scaled Dirichlet (SDMM), Dirichlet (DMM), and Gaussian mixtures (GMM).

	AUC	Accuracy
SDMM + Fisher Kernel	0.98	98.72
SDMM + Bhattacharyya Kernel	0.97	96.88
SDMM + Kullback-Leibler Kernel	0.98	98.58
SDMM + Renyi Kernel	0.85	92.51
SDMM + Jensen-Shannon Kernel	0.85	92.94
DMM + Fisher Kernel	0.95	97.90
DMM + Bhattacharyya Kernel	0.94	96.11
DMM + Kullback-Leibler Kernel	0.95	97.33
DMM + Renyi Kernel	0.82	92.12
DMM + Jensen-Shannon	0.82	92.21
DMM + Fisher Kernel	0.90	94.59
DMM + Bhattacharyya Kernel	0.91	94.41
DMM + Kullback-Leibler Kernel	0.90	96.93
GMM + Renyi Kernel	0.82	90.88
GMM + Jensen-Shannon	0.80	90.81

**TABLE 2.** Results using DRIVE data set based on kernels generated from finite scaled Dirichlet (SDMM), Dirichlet (DMM), and Gaussian mixtures (GMM).

	AUC	Accuracy
SDMM + Fisher Kernel	0.87	90.87
SDMM + Bhattacharyya Kernel	0.90	91.33
SDMM + Kullback-Leibler Kernel	0.85	88.14
SDMM + Renyi Kernel	0.83	86.97
SDMM + Jensen-Shannon	0.82	86.39
DMM + Fisher Kernel	0.84	88.54
DMM + Bhattacharyya Kernel	0.86	90.67
DMM + Kullback-Leibler Kernel	0.84	88.01
DMM + Renyi Kernel	0.81	86.19
DMM + Jensen-Shannon	0.80	85.87
GMM + Fisher Kernel	0.81	87.84
GMM + Bhattacharyya Kernel	0.81	89.02
GMM + Kullback-Leibler Kernel	0.81	87.11
GMM + Renyi Kernel	0.80	85.93
GMM + Jensen-Shannon	0.80	85.22

**TABLE 3.** Results using e-optha data set based on kernels generated from finite scaled Dirichlet (SDMM), Dirichlet (DMM), and Gaussian mixtures (GMM).

	AUC	Accuracy
SDMM + Fisher Kernel	0.95	96.07
SDMM + Bhattacharyya Kernel	0.94	95.91
SDMM + Kullback-Leibler Kernel	0.90	94.33
SDMM + Renyi Kernel	0.88	91.01
SDMM + Jensen-Shannon	0.89	91.18
DMM + Fisher Kernel	0.92	95.42
DMM + Bhattacharyya Kernel	0.91	93.08
DMM + Kullback-Leibler Kernel	0.88	93.77
DMM + Renyi Kernel	0.87	90.79
DMM + Jensen-Shannon	0.84	90.85
GMM + Fisher Kernel	0.90	94.84
GMM + Bhattacharyya Kernel	0.89	92.81
GMM + Kullback-Leibler Kernel	0.85	92.53
GMM + Renyi Kernel	0.86	89.12
GMM + Jensen-Shannon	0.83	88.90

**TABLE 4.** Results using VDIS data set based on kernels generated from finite scaled Dirichlet (SDMM), Dirichlet (DMM), and Gaussian mixtures (GMM).

	AUC	Accuracy
SDMM + Fisher Kernel	1.00	100
SDMM + Bhattacharyya Kernel	1.00	100
SDMM + Kullback-Leibler Kernel	1.00	100
SDMM + Renyi Kernel	0.95	97.12
SDMM + Jensen-Shannon	0.94	97.08
DMM + Fisher Kernel	1.00	100
DMM + Bhattacharyya Kernel	1.00	100
DMM + Kullback-Leibler Kernel	1.00	100
DMM + Renyi Kernel	0.93	97.06
DMM + Jensen-Shannon	0.94	97.01
GMM + Fisher Kernel	1.00	100
GMM + Bhattacharyya Kernel	1.00	100
GMM + Kullback-Leibler Kernel	1.00	100
GMM + Renyi Kernel	0.93	96.13
GMM + Jensen-Shannon	0.89	96.50

**TABLE 5.** Results using MESSIDOR data set based on kernels generated from finite scaled Dirichlet (SDMM), Dirichlet (DMM), and Gaussian mixtures (GMM).

	AUC	Accuracy
SDMM + Fisher Kernel	1.00	100
SDMM + Bhattacharyya Kernel	1.00	100
SDMM + Kullback-Leibler Kernel	0.84	100
SDMM + Renyi Kernel	0.84	92.91
SDMM + Jensen-Shannon	0.83	92.45
DMM + Fisher Kernel	1.00	100
DMM + Bhattacharyya Kernel	1.00	100
DMM + Kullback-Leibler Kernel	0.82	100
DMM + Renyi Kernel	0.82	92.66
DMM + Jensen-Shannon	0.82	92.13
GMM + Fisher Kernel	1.00	100
GMM + Bhattacharyya Kernel	1.00	100
GMM + Kullback-Leibler Kernel	0.83	100
GMM + Renyi Kernel	0.81	91.85
GMM + Jensen-Shannon	0.81	91.78

**TABLE 6.** Results using HRIS data set based on kernels generated from finite scaled Dirichlet (SDMM), Dirichlet (DMM), and Gaussian mixtures (GMM).

	AUC	Accuracy
SDMM + Fisher Kernel	1.00	100
SDMM + Bhattacharyya Kernel	1.00	100
SDMM + Kullback-Leibler Kernel	1.00	100
SDMM + Renyi Kernel	0.95	97.43
SDMM + Jensen-Shannon	0.95	97.38
DMM + Fisher Kernel	1.00	100
DMM + Bhattacharyya Kernel	1.00	100
DMM + Kullback-Leibler Kernel	1.00	100
DMM + Renyi Kernel	0.93	96.95
DMM + Jensen-Shannon	0.93	97.05
GMM + Fisher Kernel	1.00	100
GMM + Bhattacharyya Kernel	1.00	100
GMM + Kullback-Leibler Kernel	1.00	100
GMM + Renyi Kernel	0.91	95.77
GMM + Jensen-Shannon	0.90	96.13

**TABLE 7.** Results using HRF data set based on kernels generated from finite scaled Dirichlet (SDMM), Dirichlet (DMM), and Gaussian mixtures (GMM).

	AUC	Accuracy
SDMM + Fisher Kernel	0.95	96.22
SDMM + Bhattacharyya Kernel	0.93	95.91
SDMM + Kullback-Leibler Kernel	0.92	95.17
SDMM + Renyi Kernel	0.84	90.19
SDMM + Jensen-Shannon	0.83	90.03
DMM + Fisher Kernel	0.90	95.82
DMM + Bhattacharyya Kernel	0.90	95.70
DMM + Kullback-Leibler Kernel	0.90	94.92
DMM + Renyi Kernel	0.82	89.91
DMM + Jensen-Shannon	0.81	89.94
GMM + Fisher Kernel	0.89	93.88
GMM + Bhattacharyya Kernel	0.88	94.77
GMM + Kullback-Leibler Kernel	0.87	93.66
GMM + Renyi Kernel	0.81	88.05
GMM + Jensen-Shannon	0.79	88.31

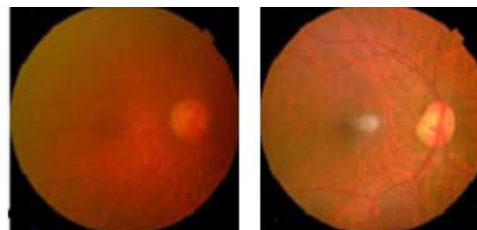
**TABLE 8.** Accuracy when deploying directly the three generative models to the different data sets.

	SDMM	DMM	GMM
e-optha	85.34	84.95	81.45
HRIS	91.22	90.73	88.98
MESSIDOR	90.09	89.11	88.53
DIARETDB1	85.55	85.12	84.60
VDIS	86.61	86.12	85.80
DRIVE	84.99	84.79	84.01
HRF	86.55	86.35	86.08

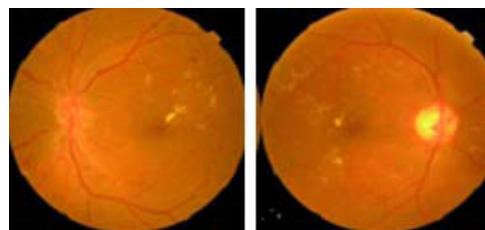
**TABLE 9.** AUCC when deploying directly the three generative models to the different data sets.

	SDMM	DMM	GMM
e-optha	0.83	0.83	0.81
HRIS	0.87	0.87	0.85
MESSIDOR	0.86	0.86	0.80
DIARETDB1	0.83	0.82	0.80
VDIS	0.90	0.90	0.88
DRIVE	0.75	0.72	0.70
HRF	0.73	0.72	0.70

We started by constructing a multi-class SVM for all our experiments. The developed probabilistic kernels are used to train the SVM. We divide the input data (retinal images) into two subsets: The first one is used for training purpose and the second one for testing. By applying the cross-validation technique, we pick out 70% from the training subset to achieve training step and the rest is used for validation. We apply also Monte Carlo approximation technique to generate data points which are used to evaluate the developed probabilistic kernels. This mechanism helps us to select the kernel appropriately model’s parameters. In practice, the kernel parameters are determined as follow:  $\omega$  is selected from  $\{0.1, 0.2, \dots, 0.8\}$  and  $A$  from  $\{2^{-10}, 2^{-9}, \dots, 2^4\}$ . On the other hand, we opt for  $C$ -support vectors machines such as  $C \in \{2^{-2}, 2^{-1}, \dots, 2^{12}\}$ .



**FIGURE 1.** Examples of fundus images showing healthy eyes.



**FIGURE 2.** Examples of images containing exudates and manifestations of retinopathy.

Our main objectives through this study are to take advantage, to evaluate and to compare between the probabilistic kernels and so that the proposed generative-discriminative framework in the same manner as for the Gaussian, scaled Dirichlet, and Dirichlet mixtures. Moreover, our focus here is to compare between our developed hybrid framework and the generative mixture models.

**B. RESULTS**

1) DATA SETS AND PREPROCESSING

In our experiments, we have considered publicly available data sets such as e-optha, HRIS, MESSIDOR, DIARETDB1, VDIS, DRIVE, and HRF [5]. Figure 3 shows an example of one of the images that have been processed. Some pre-processing steps were applied to the original images as described in [5] in order to simplify the overall classification process. Then, a crucial step of visual features extraction is performed for diabetic Retinopathy classification purpose. To tackle this problem, we adopted the so-called “speeded up robust (SURF)” visual local features [56]. Indeed, each input retinal image is modeled using the implemented finite scaled Dirichlet mixture model constrained with the MDL criterion. After that each image is described through a set of SURF features. We recall that we start by representing our training images with the different tested mixture models (Gaussian, Dirichlet, scaled Dirichlet), then, we perform classification over the developed kernels which are deployed within the multi-class SVM classifier.

2) RESULTS AND DISCUSSIONS

Table 1 shows the results concerning DIARETDB1 data set. Quantitative performances are obtained based on the ground truth and in terms of AUC and accuracy metrics when deploying different kernels generated from scaled Dirichlet,

**TABLE 10.** Comparison results: AUC and Accuracy for normal/abnormal retinal image classification on MESSIDOR dataset.

Method	AUC	Accuracy
HCF/CNN [59] (2018)	0.893	-
Pedro Costa et al. [60] (2017)	0.90	-
Zoom-in-Net [61] (2017)	0.921	90.5
Nandy et al. [62] (2016)	0.921	-
Vo and Verma [63] (2016)	0.862	-
Quellec et al. [64] (2016)	0.89	-
Roychowdhury et al. [65] (2014)	0.90	-
Proposed Method1: SDMM+ Fisher Kernel	1.00	100
Proposed Method2: SDMM+ Kullback-Leibler Kernel	0.84	100

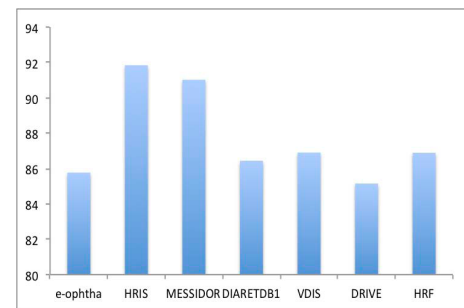
**TABLE 11.** Comparison results: Accuracy for normal/abnormal retinal image classification on DRIVE dataset.

Method	Accuracy
Fleming et al. [66]	89.80
Garcia et al. [67]	73.55
Li and Chutatape [68]	85.50
Wang et al. [69]	85.00
Proposed Method1: SDMM+ Fisher Kernel	90.87
Proposed Method2: SDMM+ Bhattacharyya Kernel	91.33

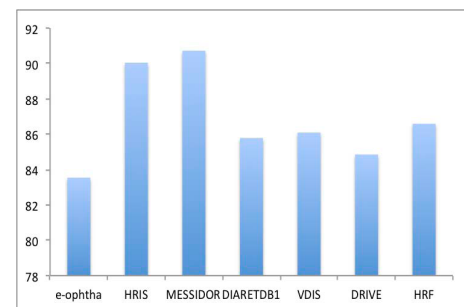


**FIGURE 3.** An example from the set of images that have been processed from DRIVE database.

Dirichlet, and Gaussian finite mixtures. According to this table, we can see clearly that the SVM based on Fisher kernel generated from the scaled Dirichlet mixture model provides the best result. This result is actually very close to the one reached by the SVM approach based on Kullback-Leibler kernel. Table 2 displays the classification results for the DRIVE data set. For this data set, the Bhattacharyya kernel generated from the scaled Dirichlet clearly outperforms the other kernels. We can notice also that the that the scaled Dirichlet-based kernels provide the best results followed by the Dirichlet and then the Gaussian-based ones. This can be explained by the flexibility of this recently introduced finite mixture model and its ability to fit well the extracted features vectors. The results related to sets are summarized in Tables 3 and 4, respectively. For both data sets, it is clear that the Fisher kernel provides the best results for the three tested generative models. It is noteworthy also that the Fisher, the Bhattacharyya, and the Kullback-Leibler kernels classification accuracy for the VDIS data set. The same conclusion is valid for the MESSIDOR and HRIS data sets as shown in Tables 5 and 6, respectively. For these two datasets we can notice that the three generative models perform slightly similarly if we take into account the kernels



**FIGURE 4.** SVM accuracy for all tested data sets.



**FIGURE 5.** KNN accuracy for all tested data sets.

generated from them. The performances of the hybrid generative discriminative tested frameworks are given in Table 7 in the case of the HRF data set. Again for this data set the kernels generated from the scaled Dirichlet model outperform the Dirichlet and the Gaussian mixtures-based kernels. It is worth mentioning that for all the tested data sets, we have obtained good and promising results as compared to state of the art techniques [5]. For instance, Figures 4 and 5, show the results using SVM with RBF kernel and KNN approaches, respectively.



According to the obtained results, we can notice also that the different kernels perform comparably despite the fact that the Fisher kernel performs generally slightly better. This can be explained by the main idea of the Fisher kernel which is to exploit the geometric structure on the statistical manifold by mapping a given individual sequence of vectors into a single feature vector, defined in the gradient log-likelihood space. The main motivation is that the gradient of the log-likelihood of the model captures the generative process of the data. It has been shown that the Fisher asymptotically at least as good as the generative model from which is developed [37]. This is actually obvious from Tables 8 and 9 that show the accuracy and AUC when deploying directly scaled Dirichlet, Dirichlet, and Gaussian mixture models. In order to deploy directly the generative models we have used the bag of visual words approach [57]. The main idea is to construct a visual vocabulary based on a quantization step and then to represent each image as a vector of frequencies that we have normalized. According to these two tables, we can see also that our proposed hybrid frameworks are able to provide strongly acceptable results as compared to their generative counterparts.

A comparative study is also provided in tables 10 and 11. Indeed, the accuracy and AUC (the area under the receiver operating curve) metrics are used to quantify the performance. We present obtained results compared with other methods and studies from the literature for normal and abnormal retinal image classification on MESSIDOR and DRIVE datasets. We can notice that we attain the highest metrics.

#### IV. CONCLUSION

Diabetic retinopathy classification is a challenging problem that has received some attention recently. In this paper, we have developed a hybrid generative discriminative framework to tackle this problem. The main goal was to use the recently proposed scaled Dirichlet mixture model to generate kernels that take into account the structure of the extracted features. The choice of the scaled Dirichlet mixture is mainly motivated by the fact that our vectors of features are proportional since they are based on the bag of visual words approach. Indeed, the scaled Dirichlet has been shown to be a relevant natural choice for this kind of vectors. In order to learn the parameters of the scaled Dirichlet mixture model, we have developed an EM-based algorithm for parameters estimation while the model selection part was based on MDL criterion. The obtained results have shown that the developed kernels are promising and could be then applied for other classification problems that generates proportional vectors. According to the results, we can say also that the scaled Dirichlet has better modeling capabilities than the Gaussian and the Dirichlet mixtures. Future works could be devoted to develop a feature selection approach like the one in [69]. Indeed, in our framework we suppose that the extracted features have the same weight which is not necessarily the case in our application. Adding weights to the features could improve the results further. Another promising future work, is to

extend the learning approach to online settings to generate incremental kernels that could be updated when new training data are introduced. The learning approach could be actually purely Bayesian in order to handle problems related to EM-based approaches such as dependency on initialization and convergence to local maxima.

#### REFERENCES

- [1] L. Dai et al., "Clinical report guided retinal microaneurysm detection with multi-sieving deep learning," *IEEE Trans. Med. Imag.*, vol. 37, no. 5, pp. 1149–1161, May 2018.
- [2] S. S. Kar and S. P. Maity, "Automatic detection of retinal lesions for screening of diabetic retinopathy," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 3, pp. 608–618, Mar. 2018.
- [3] K. M. Adal, P. G. van Etten, J. P. Martinez, K. W. Rouwen, K. A. Vermeer, and L. J. van Vliet, "An automated system for the detection and classification of retinal changes due to red lesions in longitudinal fundus images," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 6, pp. 1382–1390, Jun. 2018.
- [4] S. Joshi and P. T. Karule, "A review on exudates detection methods for diabetic retinopathy," *Biomed. Pharmacotherapy*, vol. 97, pp. 1454–1460, Jan. 2018.
- [5] J. Amin, M. Sharif, M. Yasmin, H. Ali, and S. L. Fernandes, "A method for the detection and classification of diabetic retinopathy using structural predictors of bright lesions," *J. Comput. Sci.*, vol. 19, pp. 153–164, Mar. 2017.
- [6] B. Dashtbozorg, J. Zhang, F. Huang, and B. M. ter Haar Romeny, "Retinal microaneurysms detection using local convergence index features," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3300–3315, Jul. 2018.
- [7] M. D. Abràmoff, M. Niemeijer, M. S. Suttorp-Schulten, M. A. Viergever, S. R. Russell, and B. van Ginneken, "Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes," *Diabetes Care*, vol. 31, no. 2, pp. 193–198, 2008.
- [8] S. Philip et al., "The efficacy of automated 'disease/no disease' grading for diabetic retinopathy a systematic screening programme," *Brit. J. Ophthalmol.*, vol. 91, no. 11, pp. 1512–1517, 2007.
- [9] M. D. Abràmoff et al., "Automated early detection of diabetic retinopathy," *Ophthalmology*, vol. 117, no. 6, pp. 1147–1154, 2010.
- [10] A. Osareh, B. Shadgar, and R. Markham, "A computational-intelligence-based approach for detection of exudates in diabetic retinopathy images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 4, pp. 535–545, Jul. 2009.
- [11] A. Sopharak et al., "Machine learning approach to automatic exudate detection in retinal images from diabetic patients," *J. Modern Opt.*, vol. 57, no. 2, pp. 124–135, 2010.
- [12] C. I. Sánchez et al., "Contextual computer-aided detection: Improving bright lesion detection in retinal images and coronary calcification identification in CT scans," *Med. Image Anal.*, vol. 16, no. 1, pp. 50–62, 2012.
- [13] A. Sopharak, B. Uyyanonvara, and S. Barman, "Simple hybrid method for fine microaneurysm detection from non-dilated diabetic retinopathy retinal images," *Computerized Med. Imag. Graph.*, vol. 37, nos. 5–6, pp. 394–402, 2013.
- [14] L. Tang, M. Niemeijer, J. M. Reinhardt, M. K. Garvin, and M. D. Abràmoff, "Splat feature classification with application to retinal hemorrhage detection in fundus images," *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 364–375, Feb. 2013.
- [15] B. Zhang, X. Wu, J. You, Q. Li, and F. Karray, "Detection of microaneurysms using multi-scale correlation coefficients," *Pattern Recognit.*, vol. 43, no. 6, pp. 2237–2248, 2010.
- [16] R. Acharya, C. K. Chua, E. Y. K. Ng, W. Yu, and C. Chee, "Application of higher order spectra for the identification of diabetes retinopathy stages," *J. Med. Syst.*, vol. 32, no. 6, pp. 481–488, 2008.
- [17] U. R. Acharya, E. Y. K. Ng, J.-H. Tan, S. V. Sree, and K.-H. Ng, "An integrated index for the identification of diabetic retinopathy stages using texture parameters," *J. Med. Syst.*, vol. 36, no. 3, pp. 2011–2020, 2012.
- [18] C. Agurto et al., "Multiscale AM-FM methods for diabetic retinopathy lesion detection," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 502–512, Feb. 2010.
- [19] G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener, and C. Roux, "Optimal wavelet transform for the detection of microaneurysms in retina photographs," *IEEE Trans. Med. Imag.*, vol. 27, no. 9, pp. 1230–1241, Sep. 2008.

- [20] B. Antal and A. Hajdu, "An ensemble-based system for microaneurysm detection and diabetic retinopathy grading," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 6, pp. 1720–1726, Jun. 2012.
- [21] I. Lazar and A. Hajdu, "Retinal microaneurysm detection through local rotating cross-section profile analysis," *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 400–407, Feb. 2013.
- [22] A. Sopharak, B. Uyyanonvara, S. Barman, and T. H. Williamson, "Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods," *Computerized Med. Imag. Graph.*, vol. 32, no. 8, pp. 720–727, 2008.
- [23] E. Ricci and R. Perfetti, "Retinal blood vessel segmentation using line operators and support vector classification," *IEEE Trans. Med. Imag.*, vol. 26, no. 10, pp. 1357–1365, Oct. 2007.
- [24] D. Marín, A. Aquino, M. E. Gegúndez-Arias, and J. M. Bravo, "A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features," *IEEE Trans. Med. Imag.*, vol. 30, no. 1, pp. 146–158, Jan. 2011.
- [25] M. M. Fraz et al., "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2538–2548, Sep. 2012.
- [26] M. M. Fraz et al., "Blood vessel segmentation methodologies in retinal images—A survey," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 407–433, 2012.
- [27] G. McLachlan and D. Peel, *Finite Mixture Models*. Hoboken, NJ, USA: Wiley, 2004.
- [28] F. Najar, S. Bourouis, N. Bouguila, and S. Belghith, "A comparison between different Gaussian-based mixture models," in *Proc. IEEE/ACS 14th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Hammamet, Tunisia, Oct./Nov. 2017, pp. 704–708.
- [29] W. Fan, H. Sallay, N. Bouguila, and S. Bourouis, "A hierarchical Dirichlet process mixture of generalized Dirichlet distributions for feature selection," *Comput. Electr. Eng.*, vol. 43, pp. 48–65, Apr. 2015.
- [30] I. Channoufi, S. Bourouis, N. Bouguila, and K. Hamrouni, "Image and video denoising by combining unsupervised bounded generalized Gaussian mixture modeling and spatial information," *Multimedia Tools Appl.*, vol. 77, no. 19, pp. 25591–25606, 2018.
- [31] V. N. Vapnik and V. Vapnik, *Statistical Learning Theory*, vol. 1. New York, NY, USA: Wiley, 1998.
- [32] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [33] H. Sallay, A. Ammar, M. B. Saad, and S. Bourouis, "A real time adaptive intrusion detection alert classifier for high speed networks," in *Proc. IEEE 12th Int. Symp. Netw. Comput. Appl.*, Cambridge, MA, USA, Aug. 2013, pp. 73–80.
- [34] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. Workshop Comput. Learn. Theory*, 1992, pp. 144–152.
- [35] N. Bouguila, "Bayesian hybrid generative discriminative learning based on finite Liouville mixture models," *Pattern Recognit.*, vol. 44, no. 6, pp. 1183–1200, 2011.
- [36] R. Raina, Y. Shen, A. McCallum, and A. Y. Ng, "Classification with hybrid generative/discriminative models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 545–552.
- [37] T. Jaakkola and D. Haussler, "Exploiting generative models in discriminative classifiers," in *Proc. Adv. Neural Inf. Process. Syst.*, 1999, pp. 487–493.
- [38] N. Bouguila, "Hybrid generative/discriminative approaches for proportional data modeling and classification," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 12, pp. 2184–2202, Dec. 2012.
- [39] A. Bosch, A. Zisserman, and X. Muñoz, "Scene classification using a hybrid generative/discriminative approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 712–727, Apr. 2008.
- [40] X. Yang and S. M. Krishnan, "Image segmentation using finite mixtures and spatial information," *Image Vis. Comput.*, vol. 22, no. 9, pp. 735–745, 2004.
- [41] A. K. Jain, R. P. W. Duin, and J. C. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.
- [42] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. Roy. Stat. Soc. B, Methodol.*, vol. 39, no. 1, pp. 1–38, 1977.
- [43] M.-S. Yang, C.-Y. Lai, and C.-Y. Lin, "A robust EM clustering algorithm for Gaussian mixture models," *Pattern Recognit.*, vol. 45, no. 11, pp. 3950–3961, 2012.
- [44] F. Najar, S. Bourouis, A. Zaguia, N. Bouguila, and S. Belghith, "Unsupervised human action categorization using a Riemannian averaged fixed-point learning of multivariate GGMM," in *Proc. Int. Conf. Image Anal. Recognit. (ICIAR)*, Póvoa de Varzim, Portugal, Jun. 2018, pp. 408–415.
- [45] B. S. Oboh and N. Bouguila, "Unsupervised learning of finite mixtures using scaled Dirichlet distribution and its application to software modules categorization," in *Proc. IEEE Int. Conf. Ind. Technol. (ICIT)*, Mar. 2017, pp. 1085–1090.
- [46] N. Bouguila and D. Ziou, "Unsupervised selection of a finite Dirichlet mixture model: An MML-based approach," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 8, pp. 993–1009, Aug. 2006.
- [47] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.
- [48] S. Bourouis, A. Zaguia, and N. Bouguila, "Hybrid statistical framework for diabetic retinopathy detection," in *Proc. Int. Conf. Image Anal. Recognit. (ICIAR)*, Póvoa de Varzim, Portugal, Jun. 2018, pp. 687–694.
- [49] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
- [50] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer science & business media, 2013.
- [51] P. J. Moreno, P. P. Ho, and N. Vasconcelos, "A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 1385–1392.
- [52] T. Jebara and R. Kondor, "Bhattacharyya and expected likelihood kernels," in *Proc. 16th Annu. Conf. Kernel Mach. (COLT)*. Berlin, Germany: Springer, 2003, pp. 57–71.
- [53] A. B. Chan, N. Vasconcelos, and P. J. Moreno, "A family of probabilistic kernels based on information divergence," Univ. California, San Diego, CA, USA, Tech. Rep. SVCL-TR-2004-1, 2004.
- [54] N. Vasconcelos, P. Ho, and P. Moreno, "The Kullback-Leibler kernel as a framework for discriminant and localized representations for visual recognition," in *Computer Vision—ECCV*. Berlin, Germany: Springer, 2004, pp. 430–441.
- [55] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 145–151, Jan. 1991.
- [56] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
- [57] D. Singh, A. Bhure, S. Mamtani, and C. K. Mohan, "Fast-bow: Scaling bag-of-visual-words generation," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Newcastle, U.K., Sep. 2018, p. 287.
- [58] J. I. Orlando, E. Prokofyeva, M. del Fresno, and M. B. Blaschko, "An ensemble deep learning based approach for red lesion detection in fundus images," *Comput. Methods Programs Biomed.*, vol. 153, pp. 115–127, Jan. 2018.
- [59] P. Costa and A. Campilho, "Convolutional bag of words for diabetic retinopathy detection from eye fundus images," *IPSN Trans. Comput. Vis. Appl.*, vol. 9, no. 1, p. 10, 2017.
- [60] Z. Wang, Y. Yin, J. Shi, W. Fang, H. Li, and X. Wang, "Zoom-in-net: Deep mining lesions for diabetic retinopathy detection," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 267–275.
- [61] J. Nandy, W. Hsu, and M. L. Lee, "An incremental feature extraction framework for referable diabetic retinopathy detection," in *Proc. IEEE 28th Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2016, pp. 908–912.
- [62] H. H. Vo and A. Verma, "New deep neural nets for fine-grained diabetic retinopathy recognition on hybrid color space," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2016, pp. 209–215.
- [63] G. Quellec et al., "Automatic detection of referral patients due to retinal pathologies through data mining," *Med. Image Anal.*, vol. 29, pp. 47–64, Apr. 2016.
- [64] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "DREAM: Diabetic retinopathy analysis using machine learning," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 5, pp. 1717–1728, Sep. 2014.
- [65] A. D. Fleming, S. Philip, K. A. Goatman, G. J. Williams, J. A. Olson, and P. F. Sharp, "Automated detection of exudates for diabetic retinopathy screening," *Phys. Med. Biol.*, vol. 52, no. 24, p. 7385, 2007.
- [66] M. Garcia, R. Hornero, C. I. Sanchez, M. I. López, and A. Díez, "Feature extraction and selection for the automatic detection of hard exudates in retinal images," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Aug. 2007, pp. 4969–4972.
- [67] H. Li and O. Chutatape, "A model-based approach for automated feature extraction in fundus images," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 394–399.

- [68] H. Wang, W. Hsu, K. G. Goh, and M. L. Lee, "An effective approach to detect lesions in color retinal images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2000, pp. 181–186.
- [69] M. A. Mashrgy, T. Bdiri, and N. Bouguila, "Robust simultaneous positive data clustering and unsupervised feature selection using generalized inverted Dirichlet mixture models," *Knowl.-Based Syst.*, vol. 59, pp. 182–195, Mar. 2014.



**SAMI BOUROUIS** received the engineering and M.Sc. degrees in computer science from the National School for Computer Science, Tunisia, in 2003 and 2005, respectively, and the Ph.D. degree from the National Engineering School of Tunis, Tunisia, in 2011. He is currently an Assistant Professor with the College of Computers and Information Technology, Taif University, Saudi Arabia. His research interests include image processing, machine learning, computer vision, and pattern recognition.



**ATEF ZAGUIA** received the bachelor's degree in computer engineering from Ottawa University and the M.S. and Ph.D. degrees in computer science from the Ecole de Technologie Superieure (E.T.S.), University of Quebec, Montreal, Canada. He held a post-doctoral position at E.T.S. for one year. He is currently an Assistant Professor with the College of Computers and Information Technology, Taif University, Saudi Arabia. He was working on developing application for newborn

cry-based diagnosis system with the integration of interaction context, supported by the Bill and Melinda Gates Foundation. including a multimodal system, pervasive and ubiquitous computing and context-aware systems. He has published papers in national and international conferences and journals. He was in the Program Committee for the Tenth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, Venice, Italy, in 2016.



**NIZAR BOUGUILA** received the engineering degree from the University of Tunis, Tunis, Tunisia, in 2000, and the M.Sc. and Ph.D. degrees from Sherbrooke University, Sherbrooke, QC, Canada, in 2002 and 2006, respectively, all in computer science. He is currently a Professor with the Concordia Institute for Information Systems Engineering, Concordia University, Montréal, QC, Canada. His research interests include image processing, machine learning, data mining, 3-D graphics, computer vision, and pattern recognition.



**ROOBAEA ALROOBAEA** received the bachelor's degree in computer science from King Abdulaziz University, Saudi Arabia, in 2008, and the master's degree in information system and the Ph.D. degree in computer science from the University of East Anglia, U.K., in 2012 and 2016, respectively. He is currently an Assistant Professor with the College of Computers and Information Technology, Taif University, Saudi Arabia. He is currently a Chair of support researches and system at the Deanship

of Scientific Research at Taif University. He has been honored by HRH Prince Mohammed bin Nawaf Al Saud, the Saudi Ambassador to the U.K., in recognition of his research excellence at the University of East Anglia. His research interests include human-computer interaction, cloud computing, and machine learning.

...