





## Article

# Discriminative Learning Approach Based on Flexible Mixture Model for Medical Data Categorization and Recognition

Fahd Alharithi <sup>1</sup> \*, Ahmed Almulihi <sup>1</sup>, Sami Bourouis <sup>1</sup> , Roobaea Alroobaea <sup>1</sup>  and Nizar Bouguila <sup>2</sup> 

<sup>1</sup> College of Computers and Information Technology, Taif University, Taif, P.O. Box 11099, Taif 21944, Saudi Arabia; a.almulihi@tu.edu.sa (A.A.); s.bourouis@tu.edu.sa (S.B.); r.robai@tu.edu.sa (R.A.)

<sup>2</sup> The Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal, QC H3G 1T7, Canada; nizar.bouguila@concordia.ca

\* Correspondence: f.alshalawi@tu.edu.sa

**Abstract:** In this paper, we propose a novel hybrid discriminative learning approach based on shifted-scaled Dirichlet mixture model (SSDMM) and Support Vector Machines (SVMs) to address some challenging problems of medical data categorization and recognition. The main goal is to capture accurately the intrinsic nature of biomedical images by considering the desirable properties of both generative and discriminative models. To achieve this objective, we propose to derive new data-based SVM kernels generated from the developed mixture model SSDMM. The proposed approach includes the following steps: the extraction of robust local descriptors, the learning of the developed mixture model via the expectation–maximization (EM) algorithm, and finally the building of three SVM kernels for data categorization and classification. The potential of the implemented framework is illustrated through two challenging problems that concern the categorization of retinal images into normal or diabetic cases and the recognition of lung diseases in chest X-rays (CXR) images. The obtained results demonstrate the merits of our hybrid approach as compared to other methods.

**Keywords:** shifted-scaled Dirichlet distribution; mixture model; SVM kernels; data categorization and recognition; medical image analysis



**Citation:** Alharithi, F.; Almulihi, A.; Bourouis, S.; Alroobaea, R.; Bouguila, N. Discriminative Learning Approach Based on Flexible Mixture Model for Medical Data Categorization and Recognition. *Sensors* **2021**, *21*, 2450. <https://doi.org/10.3390/s21072450>

Academic Editor: Yen-Wei Chen

Received: 4 March 2021

Accepted: 30 March 2021

Published: 2 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Unsupervised data categorization and recognition are widely used tools in statistical data analysis and are progressively being applied to complex datasets allowing the discovery of similar statistical patterns. They have been applied in a diversity of applications, ranging from the fields of image processing, data mining, to those of biomedicine, security and social media. Nowadays, the trend of data mining applications in healthcare is remarkable as there are huge datasets in this sector that require a deep analysis by effective model-based techniques derived from artificial intelligence and machine learning areas. Large scale artificial intelligence and machine learning tools are increasingly successful in image-based diagnosis and have been employed for medical decision making [1–5] and other complex problems like scene and web pages categorization [6–8], retinal images classification [9] and action recognition [10]. The manual processing of these tasks is difficult, tedious and time consuming and so it is important to move to automatic methods, which are able to learn models from labeled and non-labeled data and allow faster and more accurate decisions. Most of the model-based techniques applied to classification and recognition problems are approached through finite mixture models and the commonly used mixtures are based on Gaussian assumption [11]. Mixture models are designed as well principled statistical models and have the advantage of using different distributions in order to describe its components. They are used for effectively modeling visual features thanks to their capability of describing multidimensional distributions and heterogeneous data in a finite number of classes [11–13]. They focus on modeling a given distribution by

weighted sum (i.e., a mixture) of several basic distributions (combination of two or more probability density functions).

It is noted that while the data follows a non-Gaussian distribution in nature, the Gaussian model can give a weak performance. Noticing this fact, various mixture models have been proposed in the literature and some distinguished ones are based on the generalized Gaussian [14], Dirichlet and generalized Dirichlet [15], Beta-Liouville [15], t-student distributions, etc. For instance, the Dirichlet mixture and its extensions (like generalized Dirichlet) have been successfully employed and can often outperform the Gaussian model for data clustering, categorization and action recognition [9,16–18,18]. In this context, the work in this paper is based on recent research outcomes that have shown the importance of some specific distributions for complex visual vectors modeling based on Dirichlet and scaled Dirichlet mixture distributions [16,19]. In particular, some recent studies have shown that the derived model, the called shifted-scaled Dirichlet mixture model (SSDMM), can be applied successfully for a variety of applications [20,21]. SSDMM is presented as a powerful generalization of both Dirichlet and scaled Dirichlet where the term shifted, here, means a perturbation in the simplex.

It is noteworthy that several generative probabilistic models have a lot of benefits in terms of their capability to categorize similar data and analyze complex data, but when the data are heavily corrupted due to noise and outliers, these models sometimes fail. To cope with these disadvantages, it is possible to envisage applying discriminative classifiers, in particular Support Vector Machines (SVM), instead. However, most of the conventional classifiers do not take into account the nature of input data and therefore fail to reach very good results. Consequently, it is better to think of a hybrid approach that considers both the benefits of generative and discriminative models in order to reach better performance. This objective can be achieved for instance by generating powerful mixture-based probabilistic SVM kernels.

#### *Motivations and Contributions*

In this paper, we propose to investigate recent research outcomes that have shown the importance of the shifted-scaled Dirichlet mixture model (SSDMM) and then we go a step further by developing a discriminative learning approach. First of all, we address in this work an important step in mixture modeling problems, which is the model complexity problem, and we propose to solve it in order to avoid the over-fitting issue. Indeed, in order to determine the optimal statistical mixture model with less complexity we investigate an effective criterion named Minimum Message Length (MML) [22]. This criterion has the advantage of providing better generalization capabilities. Our second contribution is to develop a family of SVM kernels generated from the finite mixture of SSDMM and its approximation. In particular, we propose to derive some kernels on the basis of probabilistic distances in order to tackle the problem of data classification using SVM. It is noted that classical SVM kernel functions are used to compute the similarity between vectors and not between mixture of distributions, which limit their use in terms of capturing the intrinsic properties of the data to classify. In this work, we propose to address certain practical shortcomings of standard kernels by taking into account the prior knowledge of the problem at hand via mixture models. Thus, we propose to derive new kernels for SVM from generative models, which are able to take into account the complexity of data, in order to enhance its modeling and categorization capabilities. To the best of our knowledge, this is the first time a hybrid approach based on SSDMM and some derived probabilistic SVM kernels, such as Fisher kernel, Kulback–Leibler kernel, and Bhattacharyya-based kernel [23–25], has been developed. As a result, we expect an improvement in terms of image modeling and categorization, which can be achieved by our hybrid framework.

This paper is structured as follows. A brief introduction that describes our motivation is given in Section 1. The generative model based on the shifted-scaled Dirichlet mixture and its learning algorithm is presented in Section 2. The discriminative hybrid learning approach is developed in Section 3. In Section 4, we summarize our complete algorithm.

Section 5 shows the merits of our work through extensive experiments related to two important applications namely lung disease recognition and retinopathy detection. Finally, Section 6 summarizes this manuscript and emphasizes some potential future works.

## 2. Finite Shifted-Scaled Dirichlet Mixture Model

As a part of our research, we will demonstrate that the shifted scaled Dirichlet distribution can be applied with conjunction of discriminative classifiers to model and discriminate complex biomedical multidimensional data. Let us define  $\vec{Y} = (Y_1, \dots, Y_D)$  a random vector. We say that  $\vec{Y}$  follows a shifted scaled Dirichlet distribution with parameters  $\theta = (\vec{\alpha}, \vec{\beta}, b)$  such that  $\vec{\alpha} = (\alpha_1, \dots, \alpha_D) \in \mathbb{R}_+^D$ ,  $\vec{\beta} = (\beta_1, \dots, \beta_D) \in \mathcal{S}^D$  and  $b \in \mathbb{R}_+$  if its density function is defined as [26]:

$$p(\vec{Y}|\theta) = \frac{\Gamma(\alpha_+)}{\prod_{i=1}^D \Gamma(\alpha_i)} \frac{1}{b^{D-1}} \frac{\prod_{i=1}^D \beta_i^{-(\alpha_i/b)} y_i^{(\alpha_i/b)-1}}{(\sum_{i=1}^D (y_i/\beta_i)^{(1/b)})^{\alpha_+}} \quad (1)$$

where  $\vec{\beta}$  denotes a location parameter.  $\alpha_+ = \sum_{d=1}^D \alpha_d$ .  $\Gamma(\cdot)$  is the Gamma function. The shifted-scaled Dirichlet distribution has  $2D$  parameters. If the parameter  $a = 1$ , then this model is reduced to a scaled Dirichlet model.

A mixture of  $K$ -components of SSDMM distributions can be written as:

$$p(\vec{Y}|\Theta) = \sum_{k=1}^K \pi_k p(\vec{Y}|\theta_k) \quad (2)$$

where  $\Theta$  is the set of all model's parameters:  $\Theta = (\pi_k, \theta_k)$ . Here, we denote  $\pi_k, k = 1, \dots, K$  by the mixing proportion that must satisfy the unity condition (are positive and their sum is equal to one). Let  $\mathcal{Y}$  be a set of vectors,  $\mathcal{Y} = \{\vec{Y}_1, \vec{Y}_2, \dots, \vec{Y}_N\}$ , such that each vector is a realization from a  $K$ -component mixture. Each sample vector  $\vec{Y}_n = (y_{n1}, \dots, y_{nD})$  is  $D$ -dimensional. The log-likelihood function of the complete dataset is as:

$$L(\mathcal{Y}|\Theta) = \prod_{n=1}^N \sum_{k=1}^K \pi_k p(\vec{Y}_n|\theta_k) \quad (3)$$

where  $p(\vec{Y}_n|\theta_k)$  is the shifted-scaled Dirichlet probability with parameter  $\theta_k = (\vec{\alpha}_k, \vec{\beta}_k, b_k)$ .

In order to estimate the parameters of the mixture  $\Theta$ , we consider here a widely applied technique, which is the Maximum Likelihood Estimate (MLE) via the well known algorithm expectation-maximization (EM) [27]. This algorithm is able to provide several estimates  $\{\Theta^t, t = 0, 1, 2, \dots\}$  by alternating between the following two steps (E-step) and (M-step) until convergence based on certain criteria:

1. **Initialization-step:** Apply K-means algorithm to initialize the parameters of the mixture.
2. **E-step:** Calculate the posterior probability  $\hat{Z}_{ij}$  as :

$$\hat{Z}_{ij} = \frac{p(\vec{Y}_i|\theta_j)\pi_j}{\sum_{l=1}^M p(\vec{Y}_i|\theta_l)\pi_l} \quad (4)$$

3. **M-step:** Update the model's parameter by maximizing the log-likelihood function as:

$$\begin{aligned} \hat{\Theta} &= \arg \max_{\Theta} L(\Theta, \mathcal{Z}, \mathcal{Y}) \\ &= \arg \max_{\Theta} \prod_{i=1}^N \sum_{j=1}^M Z_{ij} \log(p(\vec{Y}_i|\theta_j)\pi_j) \end{aligned} \quad (5)$$

where the membership  $\mathcal{Z} = \{\vec{Z}_1, \dots, \vec{Z}_N\}$ ,  $\vec{Z}_i = (Z_{i1}, \dots, Z_{iM})$  with:

$$Z_{ij} = \begin{cases} 1 & \text{if } \vec{Y}_i \in \text{component } j \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

In the current study, the implemented learning statistical model takes into account the problem of model complexity. Indeed, determining the optimal model  $M$  is an important step in mixture modeling problems, which is able to provide better generalization capabilities. To evaluate statistical models, we proceed with a criterion derived from information theory named “Minimum Message Length criterion (MML)” as it was successfully applied previously in [22]. It is noted that MML has the advantage to accurately estimate the model complexity and returns the optimal number of components in the mixture by minimizing this function:

$$\begin{aligned} \text{MML}(\Theta, Y) &\simeq -\log(p(\Theta)) - L(\Theta, Y) + \frac{1}{2} \log|F(\Theta)| + \frac{N_p}{2} + \frac{N_p}{2} \log(K_{N_p}) \\ &\simeq -\log(p(\Theta)) - L(\Theta, Y) + \frac{1}{2} \log|F(\Theta)| + \frac{N_p}{2} + \frac{N_p}{2} \log(12) \end{aligned} \quad (7)$$

where  $p(\Theta)$  is a prior probability for the model,  $|F(\Theta)|$  is the determinant of the Fisher information matrix, and  $N_p$  is the number of parameters where it is in our case equal to  $K(2D + 1) - 1$ . Here, the parameter  $K_{N_p}$  is named as the optimal quantization lattice constant for  $R^{N_p}$ . It is noted that  $K_1 = 1/12 \simeq 0.083$ , for  $N_p = 1$ . Moreover, as  $N_p$  increases,  $K_{N_p}$  tends to an asymptotic value equal to  $\frac{1}{2\pi e} \simeq 0.05855$ , which can be approximated by  $\frac{1}{12}$  [13,22,28] (since  $K_{N_p}$  does not vary much).

In the following, we determine an appropriate a prior distribution,  $p(\Theta)$ , and we conduct an expression for  $|F(\Theta)|$ . As there is no prior knowledge about the model’s parameters (i.e., all mixture’ components are independent), thus we proceed by modeling these parameters as follows:

$$p(\Theta) = p(\alpha)p(\beta)p(b)p(\pi) \quad (8)$$

It is noted that both the location parameter  $\beta$  is defined on the simplex such that  $\sum_{d=1}^D \beta_d = 1$ . In the same way, the mixing weight  $\pi$ , it is defined on the simplex and we have  $\sum_{j=1}^K \pi_j = 1$ . For such reason, it is a natural choice to choose a Dirichlet prior with different hyperparameters  $Dir(\pi|u)$  and  $Dir(\beta|v)$  for both location and mixing parameters, respectively. These priors are defined as:

$$p(\pi) = Dir(\pi|u) = \frac{\Gamma(\sum_{i=1}^K u_i)}{\prod_{i=1}^K \Gamma(u_i)} \prod_{i=1}^K \pi^{u_i-1} \quad (9)$$

$$p(\beta) = Dir(\beta|v) = \frac{\Gamma(\sum_{d=1}^D v_d)}{\prod_{d=1}^D \Gamma(v_d)} \prod_{d=1}^D \beta^{v_d-1} \quad (10)$$

For the scale parameter, we determine it experimentally and the good choice for this scalar is found equal  $p(b) = \frac{1}{10}$ . Finally, given there is no knowledge regarding the shape parameter  $\alpha$ , we suppose that all  $\alpha_j$  are independent. Thus, we proceed by taking a simple uniform prior. According to Ockham’s razor, it has been previously proven that this choice is able to give stable results [29]. We have:

$$p(\alpha) = \prod_{j=1}^K p(\alpha_j) = \prod_{j=1}^K \prod_{d=1}^D p(\alpha_{jd}) \quad (11)$$

where  $p(\alpha_{jd}) = e^6 \frac{\alpha_{jd}}{\|\alpha_{jd}\|}$ . Here  $\alpha_{jd}$  denotes the estimated shape vector and  $\|\alpha_{jd}\|$  corresponds to its norm.

Now, regarding the Fisher information matrix, it is noted that calculating this quantity for the case of mixture models is very complex since there is no analytical form. For such reason, we approximate this quantity by adopting the determinant of Fisher Information matrix as

$$|F(\Theta)| = |F(\pi)| \prod_{j=1}^K |F(\theta_j)| = |F(\pi)| \prod_{j=1}^K |F(\alpha_j)| |F(\beta_j)| |F(b_j)| \quad (12)$$

where  $|F(\theta_j)|$  is the determinant of the Fisher information of the parameters  $\theta_j = (\alpha_j, \beta_j, b_j)$  to be estimated. The Fisher information related to the mixing weights is given as:

$$|F(\pi)| = \frac{N}{\prod_{j=1}^K \pi_j} \quad (13)$$

For  $|F(\theta_j)|$ , it is calculated by taking the determinant of the Hessian matrix of the negative log-likelihood function (i.e., second derivative of the log-likelihood).

### 3. Discriminative Learning Approach Based on SSDMM

To deal with the disadvantages of generative models, one could apply instead discriminative classifiers to improve expected results for data categorization and or recognition. Actually, many classifiers such as SVM show great potential compared to generative models for several applications [30,31]. However, in most applications, the conventional SVMs kernels (i.e., linear, polynomial, RBF) [30] are not able to consider the nature of the data and it was noted that choosing conventional SVM kernels was not the right choice [14]. This disadvantage limits their performance. This problem has been well solved, for example, by combining a discriminative classifier (such as SVM) with a generative learning model into the same framework [9]. Indeed, building new SVM kernels directly from data, using for instance information divergence or Fisher score [23,24] between distributions, may lead to better performance. Thus, a hybrid approach could be a good choice, which is designed as a compromise between generative and discriminative ones. In this current work, we suggest to construct new data-based probabilistic classifiers through the shifted scaled Dirichlet mixture model (SSDMM). It is noted that the resulting hybrid model has the advantage of merging the strengths of both generative and discriminative models and thus of increasing performance given that we will take into account the intrinsic structure of input data (since we measure the similarity between input vectors). In this section, we focus on generating three probabilistic kernels named as Fisher kernel, Kulback–Leibler kernel, and Bhattacharyya-based kernel [23–25]. These kernels are developed as follows.

**The Fisher Kernel (FK):** The key intuition behind the Fisher kernel is to exploit the geometric structure on the statistical manifold and it is defined in the gradient log-likelihood space [23], which is obtained by calculating the gradient of sample log-likelihood as regards to the model parameters. Therefore, similar mixtures involve similar log-likelihood gradients. The similarity between SSDMM mixtures ( $Y$  and  $Y'$ ) based on the Fisher kernel is defined as:

$$\mathcal{FK}(Y, Y') = U_Y^{tr}(\Theta) I^{-1}(\Theta) U_{Y'}(\Theta') \quad (14)$$

The derivatives of the log-likelihood (i.e., the gradient of  $\log(p(\mathcal{Y}|\Theta))$ ) with respect to parameters  $\Theta$  are expressed as:

$$U_Y(\Theta) = \nabla \log(p(\mathcal{Y}|\Theta)) = \frac{\partial \log(p(\mathcal{Y}|\Theta))}{\partial \Theta} \quad (15)$$

To solve the previous equation, we need to compute the gradient with respect to parameters  $\pi_j$  and  $\alpha_j, j = 1, \dots, K$ . The resulting derivative equations are given as:

$$\frac{\partial \log(p(\mathcal{Y}|\Theta))}{\partial \pi_j} = \sum_{i=1}^N \left[ \frac{\hat{Z}_{ij}}{\pi_j} - \frac{\hat{Z}_{i1}}{\pi_1} \right] \quad (16)$$

$$\frac{\partial \log(p(\mathcal{Y}|\Theta))}{\partial \alpha_j} = \sum_{i=1}^N \hat{Z}_{ij} \left( \psi(\alpha_+) - \psi(\alpha_j) + \frac{\log(Y_n) - \log(\beta_j)}{b_j} - \log \left( \sum_{d=1}^D \left( \frac{Y_{nd}}{\beta_{jd}} \right)^{\frac{1}{b_j}} \right) \right) \quad (17)$$

where  $\psi$  is the digamma function.

**The Symmetrized Kullback–Leibler Kernel (SKK):** We propose here to compute the symmetrized Kullback–Leibler distance (SKK) [24] to measure the degree of similarity

between two mixture models. This symmetrized distance has the advantage of offering a more balanced measurement than the asymmetrized one and makes it more appropriate for the classification task. It has been shown that, in speaker recognition, the Gaussian mixture model (GMM) and SVM with Kullback–Leibler (KL) kernel [32] can give good performance. It is also noted that, in a conventional hybrid system, the SKK measure has often been applied to calculate the distance between two distributions and not between mixtures. Subsequently, we propose here to develop new generative SVM-SKK kernels based on SSDMM mixture models. The dissimilarity (SKK) between two mixtures  $p_1$  and  $p_2$  is given as:

$$\begin{aligned} SKK(p_1(\vec{Y}|\Theta_1), p_2(\vec{Y}'|\Theta_2)) = & \\ & \mathcal{KL}(p_1(\vec{Y}|\Theta_1), p_2(\vec{Y}'|\Theta_2)) + \\ & \mathcal{KL}(p_2(\vec{Y}'|\Theta_2), p_1(\vec{Y}|\Theta_1)) \end{aligned} \quad (18)$$

$$\mathcal{KL}(p_1(\vec{Y}|\Theta_1), p_2(\vec{Y}'|\Theta_2)) = e^{-BF(p_1(\vec{Y}|\Theta_1), p_2(\vec{Y}'|\Theta_2))} \quad (19)$$

where  $B$  is a real positive factor used for computational stability purpose, and

$$\begin{aligned} F(p_1(\vec{Y}|\Theta_1), p_2(\vec{Y}'|\Theta_2)) = & \\ & \int_{\omega} p_1(\vec{Y}|\Theta_1) \log \frac{p_1(\vec{Y}|\Theta_1)}{p_2(\vec{Y}'|\Theta_2)} + p_2(\vec{Y}'|\Theta_2) \log \frac{p_2(\vec{Y}'|\Theta_2)}{p_1(\vec{Y}|\Theta_1)} \end{aligned}$$

Unfortunately, a closed form expression does not exist in the case of the SSDMM mixture. Therefore, we propose the use of a sampling approach based on the Monte Carlo numerical approximation method [33]:

$$SKK(p_1(\vec{Y}|\Theta_1), p_2(\vec{Y}'|\Theta_2)) \approx \frac{1}{L} \sum_{i=1}^L \log \frac{p_1(\vec{Y}_i|\Theta_1)}{p_2(\vec{Y}'_i|\Theta_2)} \quad (20)$$

**The Bhattacharyya kernel (BK):** In this study, we exploit another kernel distance derived from the family of probability product kernels called the Bhattacharyya kernel [33] (known also as Bhattacharyya's symmetry measure of affinity between distributions). Within this kernel we can inject the domain knowledge and invariance of the SSDMM generative model to the SVM classifier. Here, a general inner product is evaluated as the integral of the product of pairs of distributions (or mixtures) and defined as:

$$BK_{\frac{1}{2}}(\vec{Y}_1, \vec{Y}_2) = \int_0^{\infty} p(\vec{Y}|\Theta_1)^{1/2} q(\vec{Y}|\Theta_2)^{1/2} d\vec{Y} \quad (21)$$

In the absence of a closed form for the generative SSDMM mixture, it is possible to approximate BK using the Monte Carlo simulation method [33].

$$\begin{aligned} BK_{\frac{1}{2}}(\vec{Y}_1, \vec{Y}_2) \approx & \frac{\beta}{N_1} \sum_{i=1}^{N_1} \frac{p^{1/2}(\vec{Y}_i|\Theta_1)}{Z_1} p^{1/2}(\vec{Y}_i|\Theta_1) \\ & + \frac{1-\beta}{N_2} \sum_{i=1}^{N_2} \frac{q^{1/2}(\vec{Y}_i|\Theta_2)}{Z_2} q^{1/2}(\vec{Y}_i|\Theta_2) \end{aligned} \quad (22)$$

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

#### 4. Complete Algorithm

The proposed hybrid approach includes different steps that must be taken in order to achieve optimum performance. In this work, a first preprocessing step is performed to extract robust visual features from each image in the dataset. Indeed, image description is one of the crucial steps in many medical image processes, and extracting informative

patterns (color, shape, or /and texture) help further steps such as image interpretation and classification. For instance, many image modalities (like X-rays) are difficult to interpret directly by radiologists; thus, it is important to extract relevant details and patterns for better image understanding and decisions. Thus, each image in the dataset was represented as a bag of feature vectors (i.e., the Speeded Up Robust Features (SURF) or the Haralick texture features). The second step (i.e., the generative stage of the hybrid framework) is performed by fitting the generative SSDMM model to the feature vectors extracted from images. Consequently, each image in our dataset is modeled by a finite mixture model of distributions SSDMM. We start by initializing the mixing weight  $\pi$  and model parameters  $\Theta$  with the conventional K-Means algorithm. After that, the statistical model is learned using the maximum likelihood principle and the parameters are estimated through the expectation–maximization (EM). The last step (the discriminative stage of the hybrid framework) is dedicated to compute different probabilistic distances between each of these mixture models that give us kernel matrices to feed the SVM classifier. In particular, we focus on deriving some effective measures based on Fisher, Kullback–Leibler and Bhattacharyya kernels for the mixtures of shifted scaled Dirichlet distributions. The goal is to calculate the dissimilarity between mixtures that generates the different kernels (matrices). The resulting matrices are fed to the SVM classifier to classify images (vectors). Finally, the implemented algorithm will be trained with the computed kernel matrices using the one-versus-all training approach and perform classification results using 10-fold cross-validation. The implemented hybrid framework is summarized in Algorithm 1 and the different steps are illustrated in Figure 1.

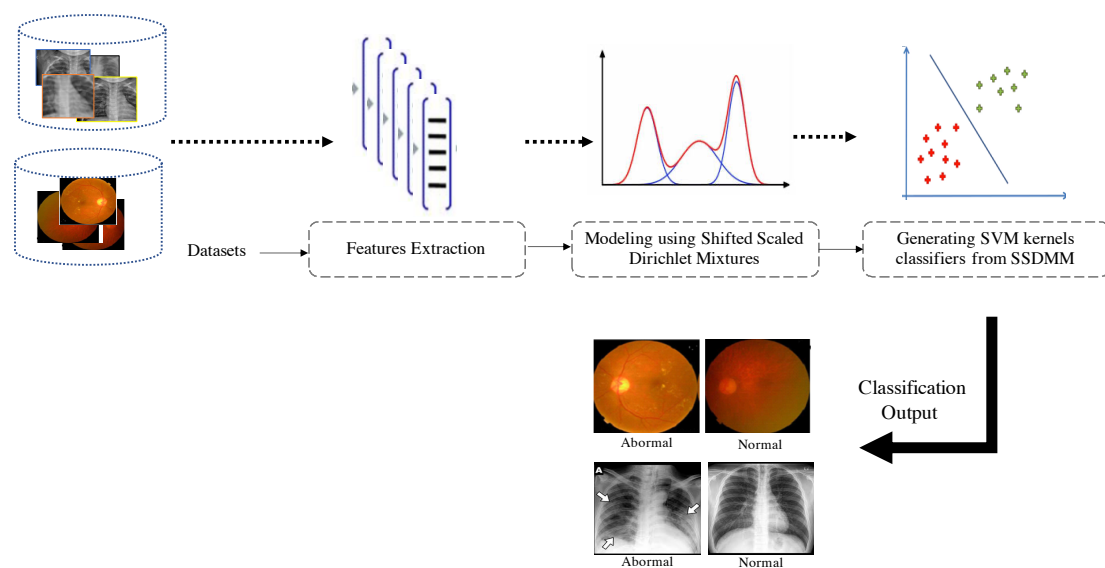
---

**Algorithm 1:** Discriminative learning approach based on SSDMM.
 

---

**Input** : Data set  $\mathcal{Y}$ ;  $M_{max}$   
**Output**: Data set categorized  
**foreach**  $1 \leq M \leq M_{max}$  **do**  
  **Initialization algorithm** :  
  Apply K-Means to initialize the parameters of each component.  
  **repeat**  
    **for**  $0 \leq j \leq M$  **do**  
      **E-step** :  
      Compute posterior probabilities using the following:  
      
$$\hat{Z}_{ij} = p(j | \vec{Y}_i) = \frac{\pi_j p(\vec{Y}_i | \theta_j)}{\sum_{j=1}^M \pi_j p(\vec{Y}_i | \theta_j)}$$
  
      **M-step** :  
      Update all parameters for the model.  $\Theta_j$  using Equation (5).  
    **end**  
    Allocate all vectors to their appropriate clusters using maximum posterior probability (MAP) estimate.  
  **until** *Convergence of Likelihood*  
  Calculate the associated message length using Equation (7).  
  Save  $\Theta$ ,  $M$  and the MML of each model  
**end**  
Return the optimal model  $M_{opt}$  with the minimum MML criterion. **foreach**  
 $1 \leq i \leq M_{opt}$  **do**  
  **foreach**  $i + 1 \leq j \leq M_{opt}$  **do**  
    Calculate all kernels' matrices between each component  $i$  and  $j$  as:  
    - Compute Fisher measure using Equation (14).  
    - Compute Symmetrized Kullback–Leibler measure using Equation (20).  
    - Compute Bhattacharyya measure using Equation (22).  
  **end**  
**end**  
Feeding the SVM classifier by the Kernel matrices.  
Return classified images.

---



**Figure 1.** Steps of the developed approach. After extracting local features from each image, we move to the modeling step using the flexible mixture model (SSDMM). Finally, we feed the Support Vector Machines (SVM) Kernel matrices, which are built to classify images as normal or abnormal.

## 5. Experimental Results

The purpose of this section is to validate the proposed approach. We have considered some real medical databases for performance evaluation.

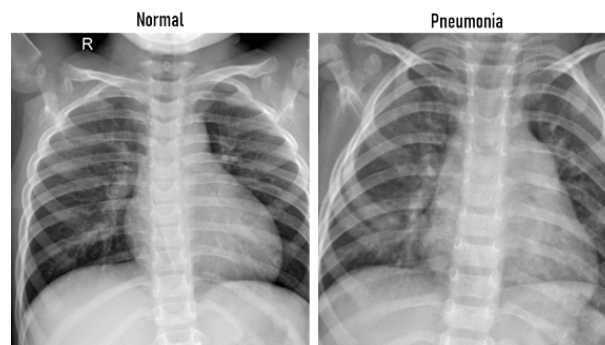
### 5.1. Lung Disease Recognition

Lung disease is one of the most common diseases worldwide. It may include pneumonia, tuberculosis, asthma, and chronic obstructive pulmonary disease. Pneumonia is one of the most serious respiratory diseases, which can be due to the causative virus of COVID-19 [34,35] or may be caused by bacterial or viral infection in the lungs [36]. In developing countries, the danger of pneumonia is enormous as thousands of people face air pollution and poverty. In December 2019, a new disease, named the COVID-19 pandemic, appeared in China and was rapidly spread worldwide [34]. This virus is a highly contagious disease. It primarily occurs by a severe acute respiratory syndrome [37]. The World Health Organization (WHO) has declared the novel virus outbreak a global pandemic of our time. Nowadays, the study of COVID-19 in particular and pneumonia in general has attracted a lot of attention due to its strong spread and quite high mortality. Early diagnosis of this disease is of great importance as it will significantly reduce its death rate. Recently, chest X-ray (CXR) radiography diagnosis is assumed to be one of the effective methods for the detection of pneumonia since it is less sensitive than CT-scan in diagnosis of lung diseases [38]. The rapid and accurate diagnosis of these diseases remains a difficult problem. To achieve this goal, it is important to develop new tools for effective infection screening and recognition. However, the lack of contrast at the boundary of lungs prevent the precise analysis of medical images. Many image processing and machine learning models have been developed for this purpose. In this work, the proposed framework is applied to identify which among the radiographic images are suspected or not of pneumonia.

Our focus here is to evaluate the developed hybrid framework with some challenging pneumonia and COVID-19 datasets. The first application is related to COVID-19 detection. For this purpose, we considered the available dataset in [39,40] (<https://github.com/ieee8023/covid-chestxray-dataset>, accessed on 20 January 2021). This dataset contains 542 chest X-ray (CXR) images. A subset of 434 CXR images are positive for COVID-19 and the rest are normal cases. We also conduct our experiments to recognize pneumonia infections from CXR images. To achieve this purpose, we consider the publicly repository “Kaggle”



([www.kaggle.com/paultimothymooney/chest-xray-pneumonia](https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia), accessed on 20 January 2021). It contains 5856 images divided into Pneumonia and Normal categories. There are 1583 normal images and 4273 pneumonia cases. An example of both normal and abnormal images from this dataset is given in Figure 2.



**Figure 2.** Examples of Chest X-Rays images. (Left) Normal patient, (Right) patient with pneumonia.

The proposed approach is built to discriminate between normal and abnormal lungs. First, a step of features extraction is computed using the Haralick feature [41,42]. The extraction step involves the calculation of the GLCM matrix [43]. In our experiments, we picked 70% of the dataset to be used for training purposes and the rest for testing and validation.

To evaluate the performance of our hybrid approach for lung diseases classification, we used the average accuracy (ACC), detection rate (DR), and false-positive rate (FPR) metrics. The accuracy (ACC) defines the overall well-classified images. The DR is the proportion of positive instances (vectors) in the data that are classified (or identified) correctly. The FPR is the proportion of negative instances that are classified incorrectly. A summary of the performance is displayed in Tables 1 and 2. These tables show the obtained results using different methods based on Gaussian, Gamma, generalized Gamma, Dirichlet, scaled Dirichlet, and shifted scaled Dirichlet mixtures, respectively. In these experiments, we fed the SVM classifier with probabilistic kernels generated from different generative mixture models. According to these results, we can notice that our hybrid framework (shifted scaled Dirichlet mixture + SVM-kernels) achieves superior performance with ACC equal to 88.81% for the CXR-COVID dataset and 94.83% for the CXR-Pneumonia dataset. It also reaches a high detection rate (DR) and low false positive rate (FPR). It is clear that these values are considered very encouraging given that we approach the detection problem in an unsupervised manner. The Gaussian-based framework has the worst accuracy equal to 83.25% for CXR-COVID and 88.18% for CXR-Pneumonia. This degradation in performance confirms the fact that Gaussian mixtures are not flexible enough for complex multidimensional data. Furthermore, the accuracies increase as the size of the dataset increases (here, performance is better for the larger dataset: CXR-Pneumonia). According to these tables, when using shifted scaled Dirichlet mixture the average classification accuracy was better than the accuracy achieved by scaled Dirichlet using different kernels. Compared with Dirichlet and scaled Dirichlet, the lung disease recognition accuracy of shifted scaled Dirichlet was improved by 1% for CXR-COVID, and more than 1.5% for the CXR-Pneumonia dataset. This proves that the shifted scaled Dirichlet is a good choice and flexible enough to be applied in medical classification problems. We report that the developed framework does not take into account any background subtraction or pre-segmentation step and for this reason the accuracy is considered quite high. To improve more obtained findings, it may be a good idea to simultaneously incorporate a feature selection mechanism into the developed statistical discriminative framework.

**Table 1.** Overall accuracy for the chest x-rays (CXR)-COVID dataset.

Approach/Metrics	ACC(%)	DR(%)	FPR(%)
<b>Generative Models</b>			
Gaussian Mixture	82.11	81.02	0.18
Gamma Mixture	85.22	83.76	0.16
Dirichlet Mixture	87.80	85.92	0.13
Scaled Dirichlet Mixture	87.96	86.02	0.13
Shifted Scaled Dirichlet Mixture	88.01	86.12	0.12
<b>Hybrid Models</b>			
Gaussian Mixture + Fisher Kernel	83.43	82.29	0.17
Gaussian Mixture + Kullback–Leibler Kernel	83.27	82.20	0.17
Gaussian Mixture + Bhattacharyya Kernel	83.25	82.18	0.17
Gamma Mixture + Fisher Kernel	86.01	84.11	0.16
Gamma Mixture + Kullback–Leibler Kernel	85.99	84.08	0.16
Gamma Mixture + Bhattacharyya Kernel	85.94	84.03	0.16
generalized Gamma Mixture + Fisher Kernel	87.01	87.90	0.12
generalized Gamma Mixture + Kullback–Leibler Kernel	87.71	87.01	0.12
generalized Gamma Mixture + Bhattacharyya Kernel	87.67	86.96	0.12
Dirichlet Mixture + Fisher Kernel	87.80	85.92	0.13
Scaled Dirichlet Mixture + Fisher Kernel	87.96	86.02	0.13
<b>Shifted Scaled Dirichlet Mixture + Fisher Kernel</b>	<b>88.81</b>	<b>86.91</b>	<b>0.11</b>
<b>Shifted Scaled Dirichlet Mixture + Kullback–Leibler Kernel</b>	<b>88.77</b>	<b>86.85</b>	<b>0.11</b>
<b>Shifted Scaled Dirichlet Mixture + Bhattacharyya Kernel</b>	<b>88.74</b>	<b>86.82</b>	<b>0.11</b>

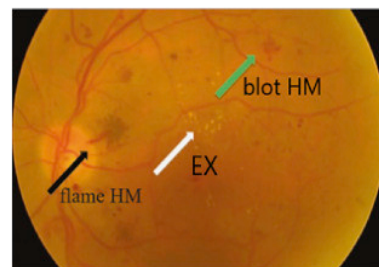
**Table 2.** Overall accuracy for CXR-Pneumonia dataset.

Approach/Metrics	ACC(%)	DR(%)	FPR(%)
<b>Generative Models</b>			
Gaussian Mixture	87.66	85.80	0.13
Gamma Mixture	90.54	88.54	0.10
Dirichlet Mixture	93.01	90.94	0.07
Scaled Dirichlet Mixture	93.33	91.90	0.07
Shifted Scaled Dirichlet Mixture	93.62	92.14	0.07
<b>Hybrid Models</b>			
Gaussian Mixture + Fisher Kernel	88.25	86.90	0.12
Gaussian Mixture + Kullback–Leibler Kernel	88.22	86.83	0.12
Gaussian Mixture + Bhattacharyya Kernel	88.18	86.79	0.12
Gamma Mixture + Fisher Kernel	90.88	88.60	0.10
Gamma Mixture + Kullback–Leibler Kernel	90.85	88.53	0.10
Gamma Mixture + Bhattacharyya Kernel	90.84	88.51	0.10
generalized Gamma Mixture + Fisher Kernel	91.98	91.11	0.09
generalized Gamma Mixture + Kullback–Leibler Kernel	91.77	91.05	0.09
generalized Gamma Mixture + Bhattacharyya Kernel	91.75	91.02	0.09
Dirichlet Mixture + Fisher Kernel	93.01	90.94	0.07
Scaled Dirichlet Mixture + Fisher Kernel	93.33	91.90	0.07
<b>Shifted Scaled Dirichlet Mixture + Fisher Kernel</b>	<b>94.83</b>	<b>93.99</b>	<b>0.06</b>
<b>Shifted Scaled Dirichlet Mixture + Kullback–Leibler Kernel</b>	<b>94.51</b>	<b>93.82</b>	<b>0.06</b>
<b>Shifted Scaled Dirichlet Mixture + Bhattacharyya Kernel</b>	<b>94.48</b>	<b>93.77</b>	<b>0.06</b>

## 5.2. Retinopathy Detection

Diabetic Retinopathy (DR) is a human eye disease and one of the most aggressive complications of diabetes, which causes damage to the retina. It is known that the augmentation of the quantity of glucose in the blood due to the lack of insulin leads to diabetes (<https://www.idf.org/aboutdiabetes/what-is-diabetes.html>, accessed on 20 January 2021). A previous study shows that diabetes affects the retina and nerves and can attack millions of people worldwide [44]. It has been found to be the principal leading cause of blindness among working elderly people worldwide, if not detected early [45]. It is reported by the International Diabetes Federation that Saudi Arabia and the Arab

world have the largest amount of affected people (in 2014, 3.8 million diabetics in Saudi Arabia). Detection of diabetic retinopathy in early stages is one of the essential challenges and is very important to avoid complete blindness and for treatment success. Unfortunately, the accurate DR detection is time- and cost-consuming, known to be difficult and requires a trained specialist to identify the presence of lesions and abnormalities in digital retina photographs. Early and automated regular screening of DR is essential for good prognosis and can help speed up the process of decision-making and can notably reduce the risk of blindness of millions of people. Previous efforts have progressed well towards a comprehensive DR screening using image processing, data mining, and pattern recognition techniques. It is possible to detect DR by analyzing the presence of different types of lesions such as microaneurysms (MA), exudates (EX), and hemorrhages (HM) [46]. For instance, there are two types of HMs called blot (deeper HM) and flame (superficial HM). This lesion looks like larger spots on the image of the retina (see Figure 3). DR is usually distinguished into four stages: mild, moderate, severe, and proliferative. Generally, diseases begin with little changes in blood vessels (Mild DR) and it increases further to achieve severe and/or proliferative DR. At this final stage, if proper care is not taken, it will lead to blindness.



**Figure 3.** Types of hemorrhage (HM) [47].

In the literature, various works have been carried out for retinal image classification and DR detection with interesting results. The study in [48] integrated a set of features of higher-order with SVM to classify eye images as DR or not-DR. A CNN model is developed in [49] for DR detection and macular edema (DME). The work in [50] introduced a method based on features extraction (AM-FM) to detect some kind of lesions and then try to measure their severity. Some works focus on classifying microaneurysms (MA) lesions by applying different methods like dynamic thresholding [51], wavelet transform [52], and a detector framework [53]. In [54], the authors proposed a method based on a convolution neural network to classify DR. Exudates lesions are detected in [55] by integrating fuzzy FCM and morphological operators into the same method. In [56], authors applied different machine learning algorithms (Decision Trees, SVM, ANN) to evaluate their performance in terms of DR prediction. Segmenting vasculature structures using robust segmentation techniques in retinal photographs may help in predicting DR at an early stage, as shown in [57], where an ensemble system and decision trees are investigated. The current problem of DR screening and classification is addressed in the present study. Indeed, we demonstrate the use of the developed hybrid generative discriminative framework for fundus images classification and DR detection.

The first step, even before applying our statistical model, is to extract relevant details (patterns) for retinal image classification. To address this step, we used “Speeded Up Robust Features (SURF)” extractor local features [58]. It is able to provide accurate description of input images. Each retinal image is then modeled through the developed shifted scaled Dirichlet mixture model. The last step is to classify the resulting descriptors on the basis of three constructed probabilistic kernels, which are deployed within an SVM classifier. These kernels are used to train the SVM. For this application, input observed vectors are divided into two subsets: training and testing subsets. We adopt a 10 fold cross-validation technique such that 70% of the vectors are considered during the training phase and the rest for testing. Our objective regarding the implementation of three different kernels is to evaluate their robustness and to analyze the stability of the SSDMM model.

It should be noted that we keep the same methodology for other models in order to conduct a correct comparative study. These models are Gaussian, Dirichlet, scaled and shifted scaled Dirichlet. The quantitative performance study is reported in terms of AUC (Area Under the Curve) and accuracy (ACC) metrics. Here AUC and ACC are performance measurements for the classification problem. They indicate how well the method is able to distinguish between patients with disease and those without disease (i.e., between different classes). Indeed, the higher the ACC and AUC, the better the method. In this study, we have considered two publicly available datasets named e-ophtha and DRIVE.

- E-ophtha [59]: this first dataset contains 47 images with EX and 35 normal images and includes 148 images with MA and 233 normal images.
- DRIVE [60]: This dataset includes 40 images with the size of  $565 \times 584$  pixels where 7 are mild DR images, and the rest are normal retinal images.

The obtained results are given in Tables 3 and 4. According to these results, it is clear that our discriminative framework based on the SSDMM model outperforms the other frameworks and gives the highest accuracy scores for both datasets. In particular, for the DRIVE dataset, the average accuracy of classifying retinal images using our method (SSDMM + Bhattacharyya kernel) is 91.65%, which is the best score. By contrast, the Gaussian-based classifier has obtained the worst score. For the second dataset (e-ophtha), we obtain the highest accuracy with our framework (SSDMM + Fisher kernel) with 96.88% compared to 96.07% for SDMM + Fisher kernel, 95.42% for DMM + Fisher kernel, and 94.84% for GMM + Fisher kernel. Here, the performance differences are statistically significant if we consider the Student's t-test. Clearly, these results confirm the efficiency of our framework for modeling and classifying complex data. On the other side, the performance of the hybrid learning approaches outperform the generative mixture models (i.e., their counterparts) and also other methods from the literature for both datasets. It is noteworthy that the SSDMM mixture would be preferred here since it is more flexible than the rest of the models. It is also noted that when analyzing the reported values for the two datasets (DRIVE and e-ophtha), the accuracies of e-ophtha are better since its size is larger than DRIVE.

**Table 3.** Classification performance (%) comparison using different approaches for the DRIVE dataset.

Approach/Metrics	AUC	ACC
<b>Generative Models</b>		
Gaussian Mixture	0.70	84.01
Dirichlet Mixture	0.72	84.79
Scaled Dirichlet Mixture	0.75	84.99
Shifted Scaled Dirichlet Mixture	0.77	85.36
<b>Hybrid Models</b>		
Gaussian Mixture + Fisher Kernel	0.81	87.84
Gaussian Mixture + Bhattacharyya Kernel	0.81	89.02
Gaussian Mixture + Kullback–Leibler Kernel	0.81	87.11
Dirichlet Mixture + Fisher Kernel	0.84	88.54
Dirichlet Mixture + Bhattacharyya Kernel	0.86	90.67
Dirichlet Mixture + Kullback–Leibler Kernel	0.84	88.01
Scaled Dirichlet Mixture + Fisher Kernel	0.87	90.87
Scaled Dirichlet Mixture + Bhattacharyya Kernel	0.90	91.33
Scaled Dirichlet Mixture + Kullback–Leibler Kernel	0.85	88.14
Shifted Scaled Dirichlet Mixture + Fisher Kernel	0.88	91.13
Shifted Scaled Dirichlet Mixture + Bhattacharyya Kernel	0.91	91.65
Shifted Scaled Dirichlet Mixture + Kullback–Leibler Kernel	0.91	88.98
<b>Other Methods</b>		
Fleming et al. [61]		89.80
Garcia et al. [62]		73.55
Li and Chutatape [63]		85.50
Wang et al. [64]		85.00

**Table 4.** Classification performance (%) comparison using different approaches for the e-ophtha dataset.

Approach/Metrics	AUC	ACC
<b>Generative Models</b>		
Gaussian Mixture	0.81	81.45
Dirichlet Mixture	0.83	84.95
Scaled Dirichlet Mixture	0.83	85.34
Shifted Scaled Dirichlet Mixture	0.84	86.10
<b>Hybrid Models</b>		
Gaussian Mixture + Fisher Kernel	0.90	94.84
Gaussian Mixture + Bhattacharyya Kernel	0.89	92.81
Gaussian Mixture + Kullback–Leibler Kernel	0.85	92.53
Dirichlet Mixture + Fisher Kernel	0.92	95.42
Dirichlet Mixture + Bhattacharyya Kernel	0.91	93.08
Dirichlet Mixture + Kullback–Leibler Kernel	0.88	93.77
Scaled Dirichlet Mixture + Fisher Kernel	0.95	96.07
Scaled Dirichlet Mixture + Bhattacharyya Kernel	0.94	95.91
Scaled Dirichlet Mixture + Kullback–Leibler Kernel	0.90	94.33
Shifted Scaled Dirichlet Mixture + Fisher Kernel	0.96	96.88
Shifted Scaled Dirichlet Mixture + Bhattacharyya Kernel	0.96	96.72
Shifted Scaled Dirichlet Mixture + Kullback–Leibler Kernel	0.93	95.12
<b>Other Methods</b>		
linear-SVM [65]	0.89	85.33
RBF-SVM [65]	0.92	87.96
Random Forests [65]	0.92	95.08
Gaussian Processes [65]	0.93	87.62

## 6. Conclusions

In this paper, we proposed an effective hybrid approach that considered the advantage of both generative and discriminative models. This approach utilizes a variant of a mixture model named the shifted-scaled Dirichlet mixture model (SSDMM), which is quite flexible to fit different shapes of observed data. In order to make our proposed framework particularly appropriate for image classification and abnormality detection problems, we derived new discriminative classifiers based on SVM kernels such as Fisher, Kullback–Leibler and Bhattacharyya for SSDMM. This strategy makes the developed framework more effective for complex and noisy data. Experiment results demonstrated that our approach has improved the accuracy when compared to other related methods in terms of accuracy over real CXR and retinal images. It has been found that our proposed algorithms outperformed many other methods. In particular, the highest accuracies obtained with our framework are 88.81% and 94.83% for CXR-COVID and CXR-Pneumonia datasets, respectively. On the other side, for the case of retinopathy detection, we achieved superior accuracies equal to 91.65% and 96.88% for both DRIVE and e-ophtha datasets, respectively. A potential future work could be devoted to extending the generative model to a non-parametric Bayesian principle in order to address the issue of accurate estimation of the number of components. We also plan to address other tasks such as image segmentation by classifying smaller regions in order to improve classification and categorization tasks.

**Author Contributions:** Conceptualization, F.A. and S.B.; methodology, A.A. and N.B.; software, R.A.; validation, F.A., S.B. and N.B.; formal analysis, A.A.; investigation, R.A.; writing—original draft preparation, F.A., S.B. and A.A.; writing—review and editing, N.B.; visualization, R.A.; supervision, N.B.; project administration, F.A.; funding acquisition, F.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received Fund from Taif University: grant number: 1-441-138.

**Data Availability Statement:** Data available in a publicly accessible repository: COVID-19 dataset available at: <https://github.com/ieee8023/covid-chestxray-dataset>, accessed on 20 January 2021; Pneumonia dataset available at: <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>, accessed on 20 January 2021; E-optha dataset [59]; DRIVE dataset [60].

**Acknowledgments:** The authors would like to thank the Deanship of Scientific Research at Taif University, Kingdom of Saudi Arabia, for their funding support under grant number: 1-441-138.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Razzak, M.I.; Naz, S.; Zaib, A. Deep Learning for Medical Image Processing: Overview, Challenges and Future. In *Classification in BioApps*; Springer: Berlin/Heidelberg, Germany, 2018.
- Ker, J.; Wang, L.; Rao, J.; Lim, T.C.C. Deep Learning Applications in Medical Image Analysis. *IEEE Access* **2018**, *6*, 9375–9389. [[CrossRef](#)]
- de Bruijne, M. Machine learning approaches in medical image analysis: From detection to diagnosis. *Med. Image Anal.* **2016**, *33*, 94–97. [[CrossRef](#)] [[PubMed](#)]
- Wernick, M.N.; Yang, Y.; Brankov, J.G.; Yourganov, G.; Strother, S.C. Machine Learning in Medical Imaging. *IEEE Signal Process. Mag.* **2010**, *27*, 25–38. [[CrossRef](#)] [[PubMed](#)]
- Alroobaea, R.; Rubaiee, S.; Bourouis, S.; Bouguila, N.; Alsufyani, A. Bayesian inference framework for bounded generalized Gaussian-based mixture model and its application to biomedical images classification. *Int. J. Imaging Syst. Technol.* **2020**, *30*, 18–30. [[CrossRef](#)]
- Zhu, R.; Dornaika, F.; Ruichek, Y. Learning a discriminant graph-based embedding with feature selection for image categorization. *Neural Netw.* **2019**, *111*, 35–46. [[CrossRef](#)]
- Zhou, H.; Zhou, S. Scene categorization towards urban tunnel traffic by image quality assessment. *J. Vis. Commun. Image Represent.* **2019**, *65*, 102655. [[CrossRef](#)]
- Sánchez, D.L.; Arrieta, A.G.; Corchado, J.M. Visual content-based web page categorization with deep transfer learning and metric learning. *Neurocomputing* **2019**, *338*, 418–431. [[CrossRef](#)]
- Bourouis, S.; Zaguia, A.; Bouguila, N.; Alroobaea, R. Deriving Probabilistic SVM Kernels From Flexible Statistical Mixture Models and its Application to Retinal Images Classification. *IEEE Access* **2019**, *7*, 1107–1117. [[CrossRef](#)]
- Najar, F.; Bourouis, S.; Bouguila, N.; Belghith, S. Unsupervised learning of finite full covariance multivariate generalized Gaussian mixture models for human activity recognition. *Multim. Tools Appl.* **2019**, *78*, 18669–18691. [[CrossRef](#)]
- McLachlan, G.J.; Peel, D. *Finite Mixture Models*; John Wiley & Sons: Hoboken, NJ, USA, 2004.
- Khan, A.M.; El-Daly, H.; Rajpoot, N.M. A Gamma-Gaussian mixture model for detection of mitotic cells in breast cancer histopathology images. In Proceedings of the 21st International Conference on Pattern Recognition, ICPR 2012, Tsukuba, Japan, 11–15 November 2012; pp. 149–152.
- Bourouis, S.; Channoufi, I.; Alroobaea, R.; Rubaiee, S.; Andejany, M.; Bouguila, N. Color object segmentation and tracking using flexible statistical model and level-set. *Multim. Tools Appl.* **2021**, *80*, 5809–5831. [[CrossRef](#)]
- Najar, F.; Bourouis, S.; Bouguila, N.; Belghith, S. A new hybrid discriminative/generative model using the full-covariance multivariate generalized Gaussian mixture models. *Soft Comput.* **2020**, *24*, 10611–10628. [[CrossRef](#)]
- Fan, W.; Bouguila, N. Expectation propagation learning of a Dirichlet process mixture of Beta-Liouville distributions for proportional data clustering. *Eng. Appl. Artif. Intell.* **2015**, *43*, 1–14. [[CrossRef](#)]
- Oboh, B.S.; Bouguila, N. Unsupervised learning of finite mixtures using scaled dirichlet distribution and its application to software modules categorization. In Proceedings of the IEEE International Conference on Industrial Technology, ICIT, Toronto, ON, Canada, 22–25 March 2017; pp. 1085–1090.
- Bourouis, S.; Al-Osaimi, F.R.; Bouguila, N.; Sallay, H.; Aldosari, F.M.; Mashrgy, M.A. Bayesian inference by reversible jump MCMC for clustering based on finite generalized inverted Dirichlet mixtures. *Soft Comput.* **2019**, *23*, 5799–5813. [[CrossRef](#)]
- Fan, W.; Sallay, H.; Bouguila, N.; Bourouis, S. Variational learning of hierarchical infinite generalized Dirichlet mixture models and applications. *Soft Comput.* **2016**, *20*, 979–990. [[CrossRef](#)]
- Bourouis, S.; Mashrgy, M.A.; Bouguila, N. Bayesian learning of finite generalized inverted Dirichlet mixtures: Application to object classification and forgery detection. *Expert Syst. Appl.* **2014**, *41*, 2329–2336. [[CrossRef](#)]
- Alsuroji, R.; Zamzami, N.; Bouguila, N. Model Selection and Estimation of a Finite Shifted-Scaled Dirichlet Mixture Model. In Proceedings of the 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018, Orlando, FL, USA, 17–20 December 2018; pp. 707–713.
- Bourouis, S.; Alharbi, A.; Bouguila, N. Bayesian Learning of Shifted-Scaled Dirichlet Mixture Models and Its Application to Early COVID-19 Detection in Chest X-ray Images. *J. Imaging* **2021**, *7*, 7. [[CrossRef](#)]
- Baxter, R.A.; Oliver, J.J. Finding overlapping components with MML. *Stat. Comput.* **2000**, *10*, 5–16. [[CrossRef](#)]
- Jaakkola, T.S.; Haussler, D. Exploiting Generative Models in Discriminative Classifiers. In Proceedings of the Advances in Neural Information Processing Systems 11, NIPS Conference, Denver, CO, USA, 30 November–5 December 1998; Kearns, M.J., Solla, S.A., Cohn, D.A., Eds.; pp. 487–493.

24. Moreno, P.J.; Ho, P.; Vasconcelos, N. A Kullback-Leibler Divergence Based Kernel for SVM Classification in Multimedia Applications. In Proceedings of the Advances in Neural Information Processing Systems 16 Neural Information Processing Systems, NIPS, Vancouver, BC, Canada, 8–13 December 2003; Thrun, S.; Saul, L.K.; Schölkopf, B., Eds.; ; pp. 1385–1392.
25. Jebara, T.; Kondor, R. Bhattacharyya and Expected Likelihood Kernels. In *Learning Theory and Kernel Machines. In Proceedings of the 16th Annual Conference on Learning Theory (COLT), Graz, Austria, 9–12 July 2002*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 57–71.
26. Monti, G.S.; Mateu i Figueras, G.; Pawlowsky-Glahn, V.; Egozcue, J.J. The shifted-scaled Dirichlet distribution in the simplex. In Proceedings of the 4th International Workshop on Compositional Data Analysis, Girona, Spain, 10–13 May 2011.
27. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. Methodol.* **1977**, *39*, 1–38.
28. Wallace, C.S.; Freeman, P.R. Estimation and inference by compact coding. *J. R. Stat. Soc. Ser. B* **1987**, *49*, 240–265. [[CrossRef](#)]
29. Bdiri, T.; Bouguila, N. Positive vectors clustering using inverted Dirichlet finite mixture models. *Expert Syst. Appl.* **2012**, *39*, 1869–1882. [[CrossRef](#)]
30. Vapnik, V.N.; Vapnik, V. *Statistical Learning Theory*; Wiley: New York, NY, USA, 1998; Volume 1.
31. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006.
32. Campbell, W.M.; Sturim, D.E.; Reynolds, D.A. Support vector machines using GMM supervectors for speaker verification. *IEEE Signal Process. Lett.* **2006**, *13*, 308–311. [[CrossRef](#)]
33. Chan, A.B.; Vasconcelos, N.; Moreno, P.J. *A family of Probabilistic Kernels Based on Information Divergence*; Technical Report; Tech. Rep. SVCL-TR-2004-1; Statistical Visual Computing Laboratory: San Diego, CA, USA, 2004.
34. Wang, L.S.; Wang, Y.R.; Ye, D.W.; Liu, Q.Q. Review of the 2019 Novel Coronavirus (COVID-19) based on current evidence. *Int. J. Antimicrob. Agents* **2020**, *55*, 105948. [[CrossRef](#)]
35. Bourouis, S.; Sallay, H.; Bouguila, N. A Competitive Generalized Gamma Mixture Model for Medical Image Diagnosis. *IEEE Access* **2021**, *9*, 13727–13736. [[CrossRef](#)]
36. Bharati, S.; Podder, P.; Mondal, M.R.H. Hybrid deep learning for detecting lung diseases from X-ray images. *Inform. Med. Unlocked* **2020**, *20*, 100391. [[CrossRef](#)]
37. Vellingiri, B.; Jayaramayya, K.; Iyer, M.; Narayanasamy, A.; Govindasamy, V.; Giridharan, B.; Ganesan, S.; Venugopal, A.; Venkatesan, D.; Ganesan, H.; et al. COVID-19: A promising cure for the global panic. *Sci. Total Environ.* **2020**, *725*, 138277. [[CrossRef](#)] [[PubMed](#)]
38. Jacobi, A.; Chung, M.; Bernheim, A.; Eber, C. Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review. *Clin. Imaging* **2020**, *64*, 35–42. [[CrossRef](#)]
39. Cohen, J.P.; Morrison, P.; Dao, L. COVID-19 image data collection. *arXiv* **2020**, arXiv:2003.11597.
40. Cohen, J.P.; Morrison, P.; Dao, L.; Roth, K.; Duong, T.Q.; Ghassemi, M. COVID-19 Image Data Collection: Prospective Predictions Are the Future. *arXiv* **2020**, arXiv:2006.11988.
41. Haralick, R.M.; Shanmugam, K.S.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *3*, 610–621. [[CrossRef](#)]
42. Sallay, H.; Bourouis, S.; Bouguila, N. Online Learning of Finite and Infinite Gamma Mixture Models for COVID-19 Detection in Medical Images. *Computers* **2021**, *10*, 6. [[CrossRef](#)]
43. Pourghassem, H.; Ghassemian, H. Content-based medical image classification using a new hierarchical merging scheme. *Comput. Med. Imaging Graph.* **2008**, *32*, 651–661. [[CrossRef](#)]
44. Taylor, R.; Batey, D. *Handbook of Retinal Screening in Diabetes*; Wiley Online Library: Hoboken, NJ, USA, 2006.
45. Bourne, R.R.; Stevens, G.A.; White, R.A.; Smith, J.L.; Flaxman, S.R.; Price, H.; Jonas, J.B.; Keeffe, J.; Leasher, J.; Naidoo, K.; et al. Causes of vision loss worldwide, 1990–2010: A systematic analysis. *Lancet Glob. Health* **2013**, *1*, e339–e349. [[CrossRef](#)]
46. Scanlon, P.H.; Sallam, A.; Van Wijngaarden, P. *A Practical Manual of Diabetic Retinopathy Management*; John Wiley & Sons: Hoboken, NJ, USA, 2017.
47. Bandello, F.; Zarbin, M.A.; Lattanzio, R.; Zucchiatti, I. *Clinical Strategies in the Management of Diabetic Retinopathy*; Springer: Berlin/Heidelberg, Germany, 2016.
48. Acharya, R.; Chua, C.K.; Ng, E.; Yu, W.; Chee, C. Application of higher order spectra for the identification of diabetes retinopathy stages. *J. Med. Syst.* **2008**, *32*, 481–488. [[CrossRef](#)] [[PubMed](#)]
49. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Webster, D.R. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *J. Am. Med. Assoc. JAMA* **2016**, *316*, 2402–2410. [[CrossRef](#)] [[PubMed](#)]
50. Agurto, C.; Murray, V.; Barriga, E.; Murillo, S.; Pattichis, M.; Davis, H.; Russell, S.; Abramoff, M.; Soliz, P. Multiscale AM-FM methods for diabetic retinopathy lesion detection. *IEEE Trans. Med. Imaging* **2010**, *29*, 502–512. [[CrossRef](#)]
51. Zhang, B.; Wu, X.; You, J.; Li, Q.; Karray, F. Detection of microaneurysms using multi-scale correlation coefficients. *Pattern Recognit.* **2010**, *43*, 2237–2248. [[CrossRef](#)]
52. Quellec, G.; Lamard, M.; Josselin, P.M.; Cazuguel, G.; Cochener, B.; Roux, C. Optimal wavelet transform for the detection of microaneurysms in retina photographs. *IEEE Trans. Med. Imaging* **2008**, *27*, 1230–1241. [[CrossRef](#)] [[PubMed](#)]
53. Antal, B.; Hajdu, A. An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 1720–1726. [[CrossRef](#)] [[PubMed](#)]

54. Wu, Z.; Shi, G.; Chen, Y.; Shi, F.; Chen, X.; Coatrieux, G.; Yang, J.; Luo, L.; Li, S. Coarse-to-fine classification for diabetic retinopathy grading using convolutional neural network. *Artif. Intell. Med.* **2020**, *108*, 101936. [[CrossRef](#)]
55. Sopharak, A.; Uyyanonvara, B.; Barman, S.; Williamson, T.H. Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods. *Comput. Med. Imaging Graph.* **2008**, *32*, 720–727. [[CrossRef](#)]
56. Tsao, H.; Chan, P.; Su, E.C. Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms. *BMC Bioinform.* **2018**, *19*, 195–205. [[CrossRef](#)] [[PubMed](#)]
57. Fraz, M.M.; Remagnino, P.; Hoppe, A.; Uyyanonvara, B.; Rudnicka, A.R.; Owen, C.G.; Barman, S.A. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 2538–2548. [[CrossRef](#)] [[PubMed](#)]
58. Bay, H.; Ess, A.; Tuytelaars, T.; Gool, L.V. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [[CrossRef](#)]
59. Decencière, E.; Cazuguel, G.; Zhang, X.; Thibault, G.; Klein, J.C.; Meyer, F.; Marcotegui, B.; Quelled, G.; Lamard, M.; Danno, R.; et al. TeleOphta: Machine learning and image processing methods for teleophthalmology. *Irbm* **2013**, *34*, 196–203. [[CrossRef](#)]
60. Staal, J.; Abràmoff, M.D.; Niemeijer, M.; Viergever, M.A.; van Ginneken, B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **2004**, *23*, 501–509. [[CrossRef](#)] [[PubMed](#)]
61. Fleming, A.D.; Philip, S.; Goatman, K.A.; Williams, G.J.; Olson, J.A.; Sharp, P.F. Automated detection of exudates for diabetic retinopathy screening. *Phys. Med. Biol.* **2007**, *52*, 7385. [[CrossRef](#)] [[PubMed](#)]
62. Garcia, M.; Hornero, R.; Sanchez, C.I.; López, M.I.; Díez, A. Feature extraction and selection for the automatic detection of hard exudates in retinal images. In Proceedings of the 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France, 22–26 August 2007; pp. 4969–4972.
63. Li, H.; Chutatape, O. A model-based approach for automated feature extraction in fundus images. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; p. 394.
64. Wang, H.; Hsu, W.; Goh, K.G.; Lee, M.L. An effective approach to detect lesions in color retinal images. In Proceedings of the Computer Vision and Pattern Recognition, Hilton Head, SC, USA, 15 June 2000; Volume 2, pp. 181–186.
65. Colomer, A.; Igual, J.; Naranjo, V. Detection of Early Signs of Diabetic Retinopathy Based on Textural and Morphological Information in Fundus Images. *Sensors* **2020**, *20*, 1005. [[CrossRef](#)]