



Diabetic retinopathy screening using improved support vector domain description: a clinical study

Ali Karsaz¹

Accepted: 8 July 2022 / Published online: 30 July 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Diabetic retinopathy (DR) is the major cause of visual impairment among diabetic patients. Significant works have been done to hybrid a modified CNN architecture such as AlexNet with some of classifiers such as support vector machines (SVMs) or fuzzy C-Means (FCM) to improve the DR screening. This new hybrid innovative structure uses more efficient extracting features of a retinal images in both spatial and spectral domains. In spite the advantages of this innovative architecture, the different kernel functions affect the performance of the proposed algorithm. Using the appropriate transformed data into two- or three-dimensional feature maps and using an improved support vector domain description (ISVDD) can obtain more flexible and more accurate image description. To this end, the optimal degree values of different kernel functions can be extracted by using a particle swarm optimization (PSO) algorithm. Also, we compared the performance of our approach (modified-AlexNet-ISVDD) with the results obtained by hybrid modified AlexNet and some of classifiers such as K-Nearest Neighbors (KNN) and FCM clustering. We achieve the proposed CNN architecture using ISVDD on the DIARETDB1 and MESSIDOR datasets, with more than 99% sensitivity.

Keywords Diabetic retinopathy screening · Deep learning · Optimal kernel functions · Improved Support vector domain description (ISVDD) · Particle swarm optimization (PSO) · Clinical study

1 Introduction

Diabetic retinopathy (DR) is the most common cause of irreversible blindness in working-age populations. According to studies, by the year 2030, the number of people diagnosed with DR will increase from 126.6 million in 2010 to 191 million, and the number of people with vision-threatening DR (VTDR) will grow from 37.3 million to 56.3 million by the same time (Congdon et al., 2012). Evidence shows that by diagnosing DR in early stages it can be treated just by diabetes management and can be prevented from further damages to the retina (Antal and Hajdu, 2012, Kamadi et al., 2016).

Generally, DR is diagnosed by an experienced ophthalmologist using a detailed and highly accurate retinal fundus image. Ophthalmologist diagnoses the presence and

severity of DR by carefully investigating fundus images and finding the different symptoms of DR, such as micro-aneurysms, hemorrhages, neovascularization, and exudates. Finding DR signs is highly subjective, which makes it difficult to diagnose in early stages. The high cost of the physical examination and lack of professional experts are the other obstacles for early DR diagnosis. Therefore, large numbers of early stage DR cases are missing from early diagnosis and treatment (Hani and Nugroho, 2010).

The main purpose of a DR diagnostic method is to classify images into DR and NoDR classes (Niemeijer et al., 2010). Typically, these methods use hand-crafted features of retinal images for training their systems and classification base on machine learning approaches. Some of these machine learning algorithms are support vector machine (SVM), artificial neural network (ANN), K-nearest neighbor (KNN) and fuzzy C-Means (FCM) methodologies (Li et al, 2015; Osareh et al., 2009; Bhatkar and Kharat, 2015; Priya and Aruna, 2013; Saranya et al., 2012). Supervised classification methods such as SVM (Wang et al., 2012; Selvaraj et al., 2007), and KNN (Anbeek et al.,

✉ Ali Karsaz
karsaz@khorasan.ac.ir

¹ Khorasan Institute of Higher Education, Moalem 77, Mashha, Iran

2005; Cocosco et al., 2003) are trained from different labeled images for segmentation purpose. Because the supervised classification methodologies use the prior knowledge about the image classes, they can improve the classification accuracy. Some of the most important unsupervised classification methodologies in the field of DR screening include KM algorithm (Khalid et al., 2014), expectation–maximization (EM) (Lalaoui et al., 2015) and FCM methodologies (Bezdek et al., 1984). Medical images are often corrupted by some environment noises, artifacts are caused by operator performance (Singh et al., 2016). Since the standard FCM algorithm does not consider local spatial classification, it is very sensitive to noise. Many researchers have incorporated the local spatial information into the standard FCM to remove the effects of noise, such as Wang et al. proposed the FCM distance function as weight sum of distance influenced by local and nonlocal information (LNLFCM) (Wang et al., 2008). Using local spatial features increases the computational cost because it needs the computation for each pixel neighborhood. Gong et al. extended fuzzy local information C-Means algorithm by replacing the Euclidean distance in the object function of the FCM by kernel distance-based cost function (Gong et al., 2013). As an effort in comparing different available techniques, a comparative analysis of nine common classifier algorithms is implemented in the application of automatic screening of diabetic retinopathy cases (Mohammadian et al., 2017a).

Generally speaking, feature extraction techniques are complex tasks and require depth knowledge of the images and their differences. Therefore, recent studies are using the state-of-the-art convolutional neural network (CNN) for various fields, especially in medical image analysis (Hussain et al., 2018). One of the main reasons for implementing CNN in medical applications is its ability to extract features automatically by using deep multiple layers (Tajbakhsh et al., 2016). Therefore, there has been an increase in using CNN in medical diagnosis applications. For instance, CNNs were used for grading brain tumors in magnetic resonance imaging (MRI) scans (Pan et al., 2015; Menze et al., 2015; Nie et al., 2018). Another CNN-based method was conducted by (Wang et al., 2015) for feature extraction and ensemble classification for retinal blood vessel segmentation and a study related to severity DR diagnosis using CNN was addressed in (Pratt 2016). Also, two different comparative studies of two CNN structures for DR screening have been performed in (Vo and Verma, 2016) and (Mohammadian et al., 2017b).

Other examples of CNNs applications in medical image detection and classification, including correctly detecting and predicting polyp type during colonoscopy videos (Tajbakhsh et al., 2015a; Zhang et al., 2017), automatic detection of pulmonary embolism (PE) in computed

tomography (CT) images (Tajbakhsh and Liang, 2015b), computer-aided automatic detection of mitotic cells in histopathology datasets (Cireşan et al. 2013), detection of lymph nodes in CT images (Roth et al., 2014), namely thoraco-abdominal lymph node (LN) detection and interstitial lung disease (ILD) classification (Shin et al., 2016a), and automatic anatomy detection in CT volumes (Zheng et al., 2015). Applications of CNNs are not only limited to detection systems but also recently CNNs have been used for medical image measurement and segmentation such as pancreas segmentation in CT scans (Roth et al., 2018; Vo and Lee, 2018), multimodality isointense infant brain image segmentation (Zhang et al., 2015), neuronal membrane segmentation in electron microscopy images (Ciresan et al., 2012), knee cartilage segmentation in MRI scans (Prasoon et al., 2013), and carotid intima-media thickness measurement in ultrasound images (Shin et al., 2016b).

The CNNs are a class of deep learning models that can learn a complex hierarchy of features by building high-level features from low-level ones. Also, the validation of the performance of the final trained network on clinical data is an important step in performance analysis of the work. Requiring a large amount of medical training images, extensive computational and memory devices, complications about training of a deep CNN such as overfitting and convergence issues made the full training of a CNN (or training from scratch) tedious and in some cases impractical (Erhan et al., 2009; Razavian et al., 2014). In medical imaging and diagnosis field, it is relatively rare to have an image dataset of sufficient size to completely train a CNN from scratch (Tajbakhsh et al., 2016). In addition, the state-of-the-art CNNs included in the GitHub or Keras core library demonstrate a strong ability to be generalized to images outside the ImageNet dataset via transfer learning, such as feature extraction and fine-tuning¹ (Simonyan and Zisserman, 2014; Motamedi et al., 2016). Therefore, it is very common to fine-tune a CNN that has been trained using a large labeled dataset from a different application to avoid training networks for many general features (Zhang et al., 2015). For example, Kaggle dataset is one of the largest diabetic retinopathy database which consists of only 35,126 retina images with different various qualities.² Various qualities of the images in the dataset make the feature extraction approaches more difficult to implement. Each image is rated by a clinician for the presence of diabetic retinopathy on a scale of 0 to 4. The scales of 0, 1, 2, 3, and 4 correspond to NoDR, mild, moderate, severe, and proliferative DR, respectively. A novel automatic recognition system for the five severity level of diabetic retinopathy (SLDR) was developed through learning of

¹ <https://github.com/fchollet/deep-learning-models>.

² <https://www.kaggle.com/diabeticretinopathy-detection>.

deep visual features (DVF) (Abbas et al., 2017). To test and evaluate the performance of the SLDR system, the 750 digital images were collected from three online public and one private sources.

The main contribution of our work in this paper is to improve the performance of the previous proposed architecture where a modified AlexNet structure consisting of both spatial and spectral domains image feature extraction mixed with a support vector domain description (SVDD) (modified AlexNet-SVDD) for a retinal image DR or NoDR recognition. In modified-AlexNet-SVDD algorithm, a pre-trained AlexNet architecture employed and enforced to classify fundus images of a clinical dataset into cases of DR patient or healthy. The first layers of the typical CNNs are mostly related to extracting general information from the images such as the edges, while their last layers are specifically trained to extract more detailed features related to the images dataset specifically. Therefore, the type of the input retinal image and the weights of the last layers of CNNs can be modified to adapt the networks for our application and increase the performance accuracy. This new innovative architecture uses two pathways extracting features of the retinal images in both spatial and spectral domains. In addition, the SVDD is a domain description method inspired by SVM algorithm that tries to find the sphere with minimum volume containing almost all objects.

The rest of the paper is organized as follows. Section 2 briefly explains some contributions of our work. A standard AlexNet architecture and the respective concepts of the FCM and the SVM algorithms are explained in Sect. 3. The proposed modified AlexNet-ISVDD architecture is presented in Sect. 4. The results of the simulations and the comparative analysis of the other classifiers are presented in Sect. 5, while Sect. 6 concludes the paper.

2 Contributions

There are a few number of medical images to train a CNN which has a lot of weights needed to be trained. This limitation of the image samples, usually leads to the overfitting problem (Bar et al., 2015; Nie et al., 2018). We propose a new modified CNN architecture using an improved support vector machine classification (ISVDD) for DR screening, making the following contributions:

- In this paper, we proposed a modification to the standard SVDD to achieve optimum kernels using metaheuristic approaches such as particle swarm optimization algorithm that can obtain more flexible and more accurate differences between the DR and NoDR retinal images.
- Traditional CNN-based approaches use only object features or spatial features for image category recognition (Tajbakhsh et al., 2016; Doshi et al., 2016; Nie et al., 2018). Our proposed structure is based on a modified AlexNet which uses two pathways extracting retinal image features.
- We demonstrated how optimization of a pre-trained AlexNet with metaheuristic algorithm without any fine-tuning, leads to incremental performance improvement in DR or NoDR classification. This unique approach distinguishes our work from the full training (Pan et al., 2015; Menze et al., 2015) and even fine-tuning approaches of a pre-trained CNN methods (Tajbakhsh et al., 2016; Vo and Verma, 2016; Razavian et al., 2014; Margeta et al., 2015).
- We present our results with consistent advantages especially for retinal image classification, but almost all above mentioned works in Sect. 1, were focused on other different medical or nonmedical imaging modalities involving classification, detection and segmentation in the field of CNN application.
- There are multiple approaches in the literature using various algorithms to implement diabetic retinopathy classification based on segmentation and detection of different exudates, hemorrhage and blood vessels using some CNN architectures. To our knowledge, the implementation of a typical CNN architecture with an improved SVDD (ISVDD) as a classifier does not have history in the field of automatic screening of DR.

3 Standard alexnet architecture and classification algorithms

The standard AlexNet achieved significantly improved performance over the other non-deep learning methods for ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2012 (Krizhevsky et al., 2012). ImageNet is a large database (over 1.2 million images) that is used for visual object recognition (1000 separate object categories). The standard AlexNet has approximately 60 million parameters (about 5 million parameters in its convolution layers and approximately 55 million parameters in fully connected layers). So we use this CNN architecture with some necessary modifications to obtain the main features of retinal images. The AlexNet computes 11×11 , 5×5 , 3×3 , 3×3 and 3×3 convolutions within the same layers of the Maxpool and concatenates the output of the whole process to pass it to the Softmax layer as the latest layer of the network. Figure 1 demonstrates schematic diagram of the standard AlexNet consists of 12 main layers

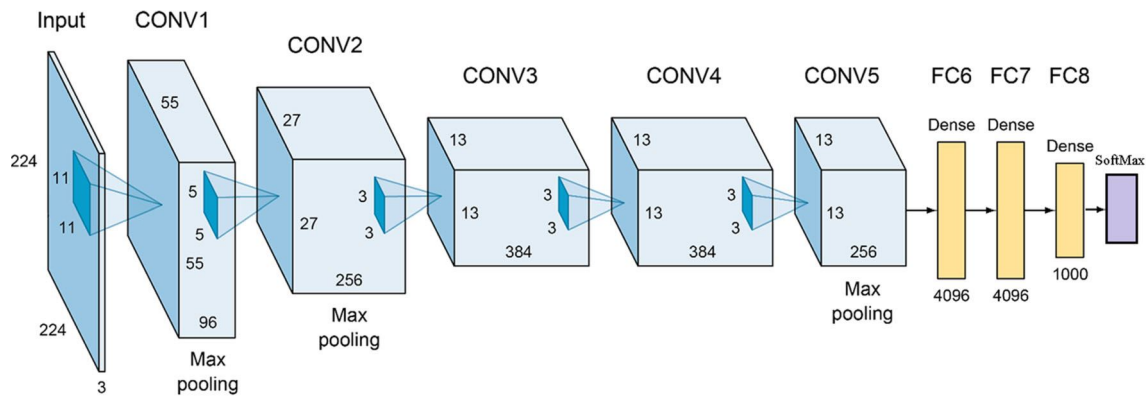


Fig. 1 Schematic diagram of the standard AlexNet (Shin et al., 2016a)

(5 convolutional layers, 3 pooling layers, 3 fully connected layers and 1 Softmax layer).

3.1 Standard fuzzy C-means

FCM as an unsupervised classification approach also known as clustering algorithm, first introduced in (Dunn, 1973) and extended by (Bezdek et al., 1984) which is an improvement of KM classification algorithm. It can group image features x_i into C_{fcm} clusters where x_i is a n -dimensional feature vector,

$$X = (x_1, x_2, \dots, x_i, \dots, x_{N_s}) \quad (1)$$

where X is a data matrix with the size of $n \times N_s$, n represents the dimension of each x_i feature vector and N_s represents the number of feature vectors (the output dimension of networks). For a training dataset of N_s data **objects** and by minimizing the J_m as the following quadratic objective function,

$$J_m(u, v) = \sum_{i=1}^{N_s} \sum_{j=1}^{C_{fcm}} u_{ij}^m \|x_i, v_j\|_2 \quad (2)$$

where

$$\sum_{j=1}^{C_{fcm}} u_{ij} = 1, i = 1, 2, \dots, N_s \quad (3)$$

And u_{ij} is the membership of i^{th} data point in the j^{th} cluster v_j . m is the weighting exponent or the index of fuzziness. When the high values for the m are selected, the membership functions tend to be equal, whereas for $m = 1$, the FCM transferred to the KM algorithm, where the memberships are crisp (in our study, we explore for the best m). $\|x_i, v_j\|_2$ is the L_2 or Euclidean norm similarity measured between a feature vector x_i and the cluster center v_j in the feature space. When the data points are located close to the centroid of their clusters, high membership values are assigned for them, and low membership values

are assigned to those are located farther from the centroid. By using the first derivative of $J_m(u, v)$ with respect to u and v equal to zero yields the two following necessary conditions for minimizing $J_m(u, v)$:

$$u_{ij} = \left(\frac{C_{fcm}}{\sum_{k=1}^{C_{fcm}} \left(\frac{\|x_i, v_j\|_2}{\|x_i, v_k\|_2} \right)^{\frac{2}{m-1}}} \right)^{-1} \quad (4)$$

and

$$v_j = \frac{\sum_{i=1}^{N_s} u_{ij}^m x_i}{\sum_{i=1}^{N_s} u_{ij}^m} \quad (5)$$

The FCM algorithm is based on an iteratively process where the cluster centers v_j and the membership values u_{ij} will be updating to minimizing the cost function $J_m(u, v)$ (Wang et al., 2008).

3.2 Standard support vector machine

The standard SVM is a supervised learning method that has the aim of determining the location of decision boundaries or hyperplanes that provide the optimal separation of the classes based on statistical theory (Cortes and Vapnik, 1995; Vapnik, 1995). In the case of a two-class image classification, the SVM selects the hyperplane that has the greatest margin between the two classes. The maximization problem of the margin is usually solved using the quadratic programming optimization algorithm. The margin is defined as the distance between the hyperplane and the nearest data sample point on each side of the classes. The sample points that are located in the nearest distance of the each hyperplane are named as *support vectors*. For our two-class classification problem and for a training dataset of N_s points of the form (x_i, y_i) , where the y_i is either 1 or -1 , indicates the DR or NoDR classes, respectively. Any hyperplane can be written as the set of point x_i satisfying,

$$Wx_i - b = 0 \quad (6)$$

where the W is the normal vector to the hyperplane and determines the orientation of the discrimination plane. The parameter $\frac{b}{\|W\|}$ determines the distance of the hyperplane from the origin along the normal vector W . The SVM tries to maximize the perpendicular distance between two hyperplanes by solving the following constrained problem,

$$J_m(W, b) = \min_{W, b} \frac{1}{2} \|W\|^2 \quad (7)$$

The trade-off between maximization of the margin and the number of misclassification errors can be controlled by a constant parameter C_{svm} ,

$$J_m(W, b, \xi_i) = \min_{W, b, \xi_1, \dots, \xi_{N_s}} \left(\frac{1}{2} \|W\|^2 + C_{svm} \sum_{i=1}^{N_s} \xi_i \right) \quad (8)$$

where $0 < C_{svm} < \infty$, the variable $\xi_i \geq 0$ is a slack variable that allows some data samples out of the sphere.

4 Proposed method

Three main stages constitute our proposed DR detection algorithm: (1) Image preparation (2) Spatial and spectral features extraction (3) improved SVDD classification.

Image preparation section itself consists of rescaling, normalizing and finally obtaining a 2D color histogram of a given retinal image. Also, we use the AlexNet in our proposed algorithm as a multi-layer feature extractor with some modification to obtain more spatial and frequency domain complementary information of a retinal image. The last section of our proposed scheme is the improved SVDD classification algorithm which determines the image category for a color fundus image.

The overall structure of the proposed scheme is shown in Fig. 2. After the modification, the obtained features can

be used in conjunction with classifier such as SVM, FCM or ISVDD. The input of our system unlike other traditional CNNs application is a 2D histogram of the retinal image, and the output of the system is the DR or NoDR label of the image. After several layers of convolution and pooling, the 2D histogram of the input image can be converted into a 2D or 3D feature vector, which contains the spatial and spectral features within the image. For a deep understanding of the new scheme performance, the performances of the 2D and 3D classification will be calculated separately, which are called *2D classification* or *3D classification* case studies.

4.1 Image preprocessing

Medical images usually contain noise and shading artifacts due to interference and other phenomena that affect the process of classification in screening systems (Reddy et al., 2017). Artifacts due to non-uniform illumination which is a general problem in retinal imaging degrade the efficiency of image classification as well as the effects of camera variations. Preprocessing is an essential step to reduce the image variation by normalizing and equalization of the irregular illuminations of a color fundus image. To decrease the variation among images due to different camera resolutions and settings, an image preprocessing algorithm is applied to the images. The first step of the algorithm is rescaling the images such that all the input images have the same size. In the next step, the color of each pixel is subtracted by the local average, mapping the average to 50% gray. Using this approach, the sharpness of the images will be more unified. Finally, for a given image the color histogram is extracted by counting the number of times each color occurs in the image array. Histograms are invariant to rotation about an axis perpendicular to the image plane, and change slowly under change in scale and change of angle of view. Adaptive histogram equalization

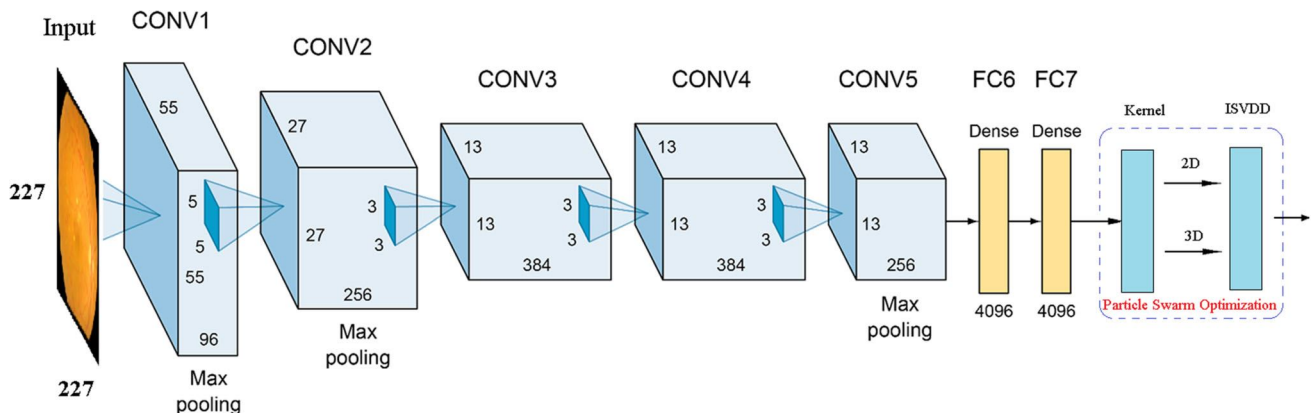


Fig. 2 Schematic diagram of the proposed modified AlexNet and ISVDD

(AHE) can be used to improve the contrast on retinal images (Andonová et al. 2017).

4.2 Spatial and spectral features extraction

The modification begins with transferring the weights from a pre-train AlexNet as a part of our final network, and is completed with only two fully connected layers, i.e., Fc6 and Fc7, instead of three FC layers typically used in the standard AlexNet (so the Fc8 is replaced by a modified Fc8). In our study, we deal with 2-class classification tasks, in two different 2D or 3D feature space (spatial and spectral domain pathways as shown in Fig. 2). In the spatial domain pathway, only the last layer (Fc8) is substituted with a typical fully connected summation as an activation function. Our proposed method processes the data in the frequency domain to attain greater accuracy besides to the spatial feature processing. By separating the image feature into different sub-bands, important difference occurs over varying low to high frequencies. When digital images are handled at multiple resolutions, the discrete Fourier transformation (DFT) is viable mathematical tool. So the spectral pathway in Fig. 2, returns the discrete Fourier transform of the Fc7 output, computed with a fast Fourier transform (FFT). In this section of our algorithm, we integrate the spectral and spatial features together to construct a powerful framework using 2D or 3D classification.

4.3 Improved SVDD classification

Improvement of a standard SVDD is the main part of our algorithm. The SVDD is a domain description method inspired by SVM algorithm that tries to find the sphere with minimum volume containing almost all objects (Tax and Duin, 1999). Since the output features obtained by the modified CNN architecture are not normally spherically distributed in the input space of the classifier data, the SVDD algorithm uses a nonlinear transformation ($\varphi(\cdot)$) to transform the data from input space to a new high dimensional feature space. Let the x_i be a dataset containing N_s sample points as mentioned in (1). For a sphere, described by center a and radius R , the SVDD minimizes the following equation:

$$F(R, a, \xi_i) = R^2 + C_{\text{svdd}} \sum_i \xi_i \quad (9)$$

where the variable C_{svdd} gives the trade-off between the volume of the sphere and the number of target objects rejected. This equation should be minimized under the following constraints:

$$\|\varphi(x_i - a)\|^2 \leq R^2 + \xi_i \quad \forall i, \xi_i \geq 0 \quad (10)$$

Incorporating this constrain in (9), the following *Lagrangian* can be introduced

$$L(R, a, \alpha_i, \xi_i) = R^2 + C_{\text{svdd}} \sum_i \xi_i - \sum_i \alpha_i \{R^2 + \xi_i - \|\varphi(x_i - a)\|^2\} - \sum_i \gamma_i \xi_i \quad (11)$$

where $\alpha_i \geq 0$ and $\gamma_i \geq 0$ are two Lagrange multipliers. The following constraints can be obtained by setting the partial derivatives of the Lagrangian (40) with respect to R , α and ξ_i to zero, new constraints are obtained:

$$\sum_i \alpha_i = 1, a = \frac{\sum_i \alpha_i \varphi(x_i)}{\sum_i \alpha_i} = \sum_i \alpha_i \varphi(x_i) \quad (12)$$

$$C_{\text{svdd}} - \alpha_i - \gamma_i = 0 \quad \forall i \quad (13)$$

By remove the γ_i from (13), the $0 \leq \alpha_i \leq C_{\text{svdd}}, \forall i$, will be obtained. By resubstituting (13) in Lagrangian, the following *Wolfe* dual form, which is a maximization problem respect to α_i will be obtained:

$$W(\alpha_i) = \sum_i \alpha_i \langle \varphi(x_i), \varphi(x_i) \rangle - \sum_{i,j} \alpha_i \alpha_j \langle \varphi(x_i), \varphi(x_j) \rangle \quad (14)$$

where the $\langle \varphi(x_i), \varphi(x_j) \rangle$ is the inner product can be replaced with an appropriate kernel function $K(x_i, x_j)$ such that satisfies the *Mercer's* theorem (Niazmardi et al., 2013). There are different kernel functions; however, the *Gaussian* kernel function is shown to have better performance than the others (Tax and Duin, 1999);

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right), \sigma \in \mathbb{R}^+ \quad (15)$$

The different kernel functions result in different description boundaries in the original input space of the SVM. The generic Gaussian kernels in (44) regard each component of x_i with equal emphasis in their effects into feature space. The problem is to find a suitable nonlinear transformation for each component of x_i to have a larger effect in feature space. To this end, the nonlinear transformation $\varphi(x_i) = x_i^n$ corresponding to each modified AlexNet outputs (i.e., spatial and spectral features) is used to scale each feature before mapping it into feature space. In our proposed algorithm, the Gaussian kernel function with above nonlinear pre-transformation ($\varphi(\cdot)$) is considered to transform the data from input space to a new high nonlinear feature space. So we consider two kernel functions, the first one is a two-dimensional kernel (2D) and the second one is a three-dimensional kernel function (3D) as follows:

$$K_{2D}(\varphi(x_i), \varphi(x_j)) = \exp\left(-\frac{\|\varphi_{2D}(x_i) - \varphi_{2D}(x_j)\|^2}{\sigma^2}\right) \quad (16)$$

$$K_{3D}(\varphi(x_i), \varphi(x_j)) = \exp\left(-\frac{\|\varphi_{3D}(x_i) - \varphi_{3D}(x_j)\|^2}{\sigma^2}\right) \quad (17)$$

In this paper, we present a modified AlexNet model with two pathways retinal image recognition, which extract both spatial and spectral features of images respectively. Experiments show that the two kinds of features contain complementary information for the category recognition of an image as the two-pathway model always achieve better performance than single pathway models. So each feature space x_i has two main parts, spatial feature space x_{ispa} , obtained by the spatial pathway and spectral feature space x_{ispe} , obtained by the spectral pathway, respectively (as shown in Fig. 2). The two kernel functions (16) and (17) are considered with 2D and 3D nonlinear transformations as follows:

$$(x_{ispa}, x_{ispe})^{\varphi_{2D}(x_i)} (x_{ispa}^{n_1}, |x_{ispe}|^{n_2}) \quad (18)$$

$$(x_{ispa}, x_{ispe})^{\varphi_{3D}(x_i)} (x_{ispa}^{n_1}, \text{Re}(x_{ispe})^{n_2}, \text{Im}(x_{ispe})^{n_2}) \quad (19)$$

where the free parameters n_1 and n_2 are the degree of the polynomials, and $\text{Re}(x_{spe})$ and $\text{Im}(x_{spe})$ represent the real and imaginary parts of the spectral feature. Tax and Duin (1999) consider L1 loss in the formulation of SVDD as in (38). In SVDD, L2 loss is a common alternative to L1 loss (Fan et al., 2005). The performance of L2 loss SVDD (ISVDD) can outweigh L1 loss SVDD in some circumstances and has some subtle differences. The ISVDD minimizes the following equation:

$$F(R, a, \xi_i) = R^2 + C_{svdd} \sum_i \xi_i^2 \quad (20)$$

subject to:

$$\|\varphi(x_i - a)\|^2 \leq R^2 + \xi_i; \quad \forall i \quad (21)$$

Note that the constraint, $\xi_i \geq 0; \forall i$ is not necessary for L2 loss SVDD, because if at an optimum $\xi_i < 0$ for some i , we can then replace ξ_i with zero so that,

$$\|\varphi(x_i - a)\|^2 \leq R^2 + \xi_i < R^2 + 0 \quad (22)$$

5 Experimental results and analysis

5.1 Evaluation criterion for performance analysis

To estimate the accuracy of classification or clustering for the clinical data, all clustering validity criterion can be used. These indices are divided into three categories: (1)

internal criteria, (2) relative criteria (3) external criteria (Niazmardi et al., 2013). The internal index uses some metrics that are based on the database and clustering methodology. To compare the performance of the proposed modified AlexNet-SVDD for the clinical data, four performance indices are calculated. These indices are accuracy, precision or prediction value, sensitivity or recall, specificity (Franklin and Rajan, 2014; Olson and Delen, 2008). The precision or predictive value is the probability of the retinal image that has been classified as DR is really DR itself (Franklin and Rajan, 2014). Sensitivity (Zhou et al., 2017) or recall is the true positive rate (TPR) and specificity is the true negative rate (TNR) (Liu and Tang, 2014). The precision criterion, keeps the balance between the sensitivity and specificity to evaluate the effectiveness of the proposed algorithm (Zhou et al., 2017). P and N represent the labeled DR and NoDR samples, respectively, and $P + N$ demonstrates the total number of samples or $P + N = TP + FP + TN + FN$.

The relative criteria are based on evaluation of clustering results by comparing them with other clustering methods. In a wide range of medical image classification, the free response operating characteristic (FROC) curve is used (Edwards et al., 2002) as a fundamental index for diagnostic test evaluation (Tajbakhsh et al., 2016). In a FROC curve the true positive rate is plotted in function of the false positive rate for different threshold parameter. Sensitivity shows the probability that a test will be positive when the disease is present. Therefore, each point on FROC represents a sensitivity–specificity pair corresponding to a particular decision cut-off point.

To evaluate our proposed scheme, Kappa coefficient of agreement has been used as an external criterion (Kuncheva, 2011; Cohen, 1960). Kappa error relations are used to gain insights about who much a clustering method is better than another on a specific dataset (Niazmardi et al., 2013; Pasolli et al., 2014). Diversity between two classifiers is measured by κ represents the Kappa coefficient as

$$\kappa = (OA - AC)/(1 - AC) \quad (23)$$

where OA is the observed agreement or accuracy and AC is the agreement by chance.

5.2 Preparing dataset

The main problem in the research of DR screening is the non-availability of a suitable standard datasets for training, testing and evaluation of a developed algorithm. Every academic research groups uses their own databases for evaluation and testing with different number of samples; therefore, a general comparison with similar studies would not be possible. The well establish public dataset DIA-RETDB1 consists 89 retinal images, of which 84 images

are labeled as DR and remaining five images are considered as NoDR, has been used by some papers such as (Kauppi, 2007; Franklin and Rajan, 2014). These images were acquired by using a digital fundus camera with 50° field of view (FOV). Our proposed algorithm is evaluated with the fundus images available in the DIARETDB1 dataset as input to our two-pathway modified AlexNet architecture. We further augmented some NoDR images to this dataset to create a balanced database of 168 retinal images in which both DR and NoDR classes were represented equally (84 samples for each class). Also, our algorithm is applied to diagnose DR cases from real fundus images that are captured from the Navid-Didegan ophthalmology clinic, Iran. The labeling was performed by an experienced independent ophthalmologist. All images are in different compressed formats such as JPEG, JPG and PNG with two different sizes 3872×2592 , 3060×2580 pixels for the Navid-Didegan database. This private dataset contains 94 retinal images, in which 47 of the images are labeled as NoDR and 47 images are labeled as DR class. Finally, our proposed algorithm is applied to the free dataset MESSIDOR, where 1200 images are collected (Patry et al., 2016). Therefore, by applying the labeled images of these three datasets to the modified AlexNet-ISVDD in the test step, the performance of the network is examined for the test images, in which all the test images are independent of the train data. We hold out 70% of the dataset for training in classification procedure, while 30% is used to test the performance of the methodology. As the train and test datasets include different fundus images taken with different devices to remove the variations in images, preprocessing algorithm described in Sect. 4.1 is implemented on both train and test points.

The result of image preprocessing steps are shown in Fig. 3a–c. Figure 3 shows two selected examples of DR (the left column) and NoDR (the right column) fundus images and their experimental results of rescaling, illumination and normalizing. Figure 4 shows the obtained 2D histograms for these DR and NoDR samples.

The image preprocessing level as mentioned before, consists of three main parts: *the image rescaling* (Fig. 3b), *the RGB equalization* (Fig. 3c) and *the 2D histogram extraction* (Fig. 4). The task for *image rescaling* is performed such that all the input images have the same size of $2592 \times 2592 \times 3$. At *the normalization* level, the non-uniform brightness of the retinal fundus image will be removed by dividing the three components of a color pixel by its intensity. The third step of preprocessing level in our proposed algorithm is the choosing red and green components of the retinal image, because these channels contain most information with blood and vessels in a retinal fundus image. In this stage, the color histogram is obtained by

counting the number of times each red and green colors occur in the image array.

The next step is to apply the preprocessed images to the proposed network that has already been modified and discussed in Sects. 4.2 and 4.3. Figure 5 shows spatial and spectral feature maps obtained by the modified AlexNet containing 47×2 retinal images from Navid-Didegan dataset (47 as DR and 47 as NoDR images). Figure 5a shows the spatial feature maps (i.e., $x_{ispa}^{n_1}$ in (18)) obtained by the modified AlexNet, Fig. 5b shows the spectral features (i.e., $|x_{ispe}|^{n_2}$ in (18)) and finally Fig. 5c, shows the real and imaginary parts of the spectral feature maps (i.e., $(Re(x_{ispe}))^{n_2}$ and $(Im(x_{ispe}))^{n_2}$ mentioned in (19)).

5.3 Optimum degree values of ISVDD's kernels

As mentioned in Sect. 4.3, the different kernel functions result in different description boundaries in the original input space of the SVDD. These kernels map the features into the high nonlinear features space (Niazmardi et al., 2013). To find the optimal values of the kernel parameters (Wen et al., 2008), the particle swarm optimization (PSO) has been used. The PSO approach has the advantage of being less computationally expensive compared with other well-known evolutionary algorithms such as genetic algorithm (GA), firefly algorithm (FFA), gray wolf optimization (GWO). After 300 iterations, the applied PSO found the best values for $n_1 = 0.81$, $n_2 = 4.2$ involved in the kernel functions. By setting $n_1 = 0.81$, $n_2 = 4.2$, as shown in Fig. 6, the location of the feature maps in input space will be changed in order that the ISVDD can easily classify them. Navid-Didegan dataset is used in this scenario to evaluate the proposed modified AlexNet-ISVDD algorithm. The effects of these optimum kernel values are shown in Fig. 6 with both 2D and 3D feature maps obtained by the modified AlexNet-ISVDD containing 47×2 retinal images from Navid-Didegan dataset (47 as DR and 47 as NoDR images). The first row of this figure shows the results obtained by using the 2D kernel function, so the inputs of the ISVDD algorithm would be the nonlinear transformation over the spatial pathway (i.e., $x_{ispa}^{n_1}$) and spectral pathway (i.e., $|x_{ispe}|^{n_2}$) as mentioned in (18), respectively. In Fig. 6a, the spatial pathway outputs respect to spectral pathway outputs are illustrated. Figure 6b shows the classification results obtained by ISVDD with 2D pathway. The second row of Fig. 6 shows the results obtained by using 3D kernel function, where the inputs of the ISVDD algorithm would be $x_{ispa}^{n_1}$ from spatial pathway and $(Re(x_{ispe}))^{n_2}$ and $(Im(x_{ispe}))^{n_2}$ from the spectral pathway as in (19), respectively. Figure 6d shows the classification results obtained by ISVDD with 3D pathway. Unlike the traditional SVDD algorithms that the predetermined kernel

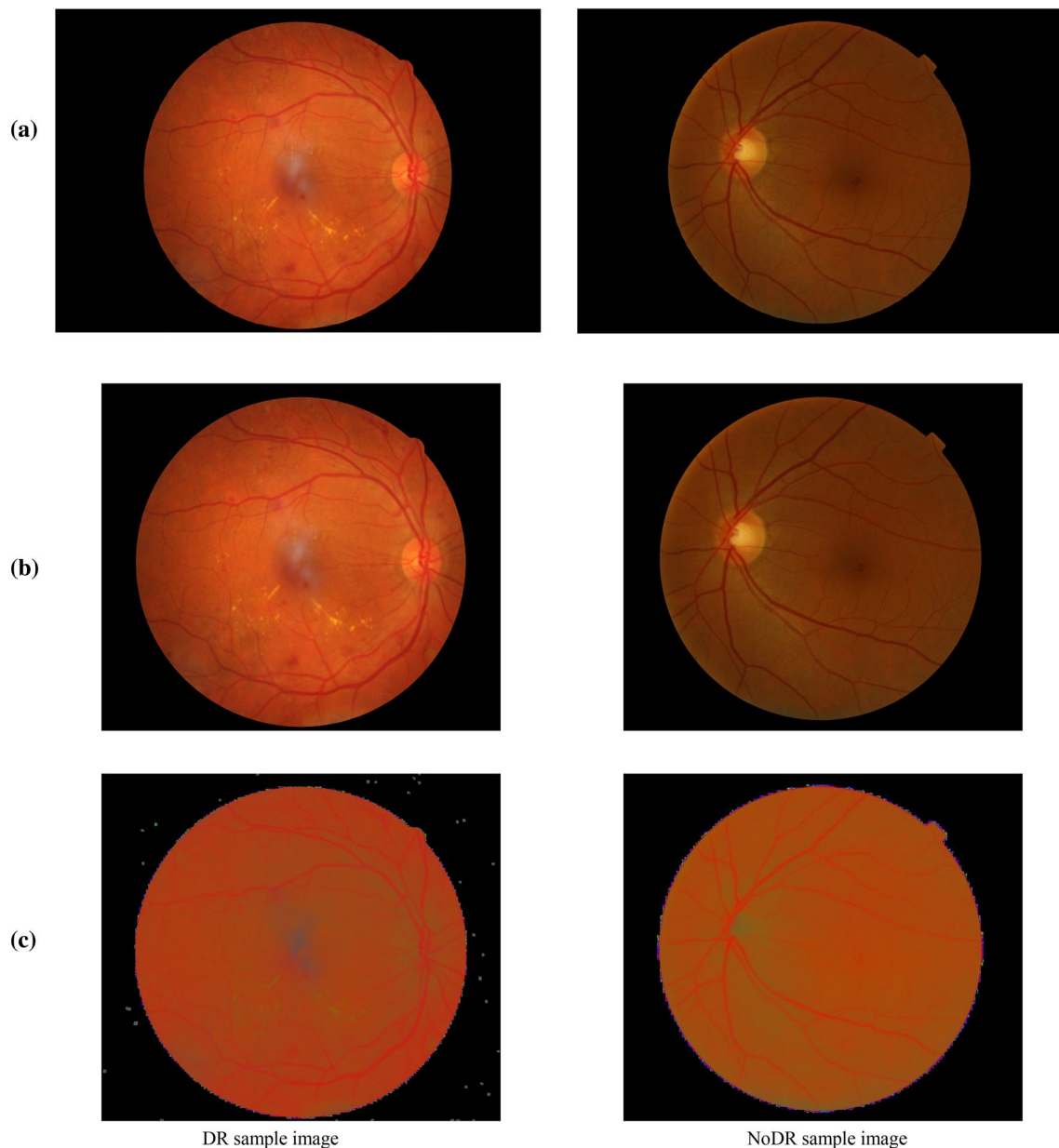


Fig. 3 Image preprocessing level of our proposed algorithm (Sect. 4.1) **a** the original DR (the left column) and NoDR (the right column) retinal fundus images **b** rescaled images (2592×2592 patches) **c** RGB normalized images

function acts as a measure to optimize the Lagrange cost function, we find the optimal kernel values by PSO algorithm on some nonlinear transformation directly on input features in (18) and (19). Therefore, a simple kernel function can be replaced by a kernel function with an optimum nonlinear pre-transformation into 2D or 3D feature spaces. Our experimental results, show that the C_{svdd} parameter in (9) does not significant impact on the ISVDD performance as reported in (Niazmardi et al., 2013), thus throughout this paper $C_{svdd} = 1.2$ is considered. The optimum kernels and C_{svdd} parameters are obtained by PSO algorithm $n_1 = 0.81$ and $n_2 = 4.2$ and $C_{svdd} = 1.2$ for

Navid-Didegan dataset and $n_1 = 1.2$ and $n_2 = 4$ and $C_{svdd} = 1.2$ for the DIARETDB1 dataset.

5.4 Comparison to other clustering algorithms

In this section, the results of the proposed algorithm are compared with the KNN, K-Means, subtractive clustering and FCM algorithms as the most frequently used clustering algorithms for DIARETDB1 and MESSIDOR datasets. Therefore, we evaluate and compare the performance of the proposed modified AlexNet with five different classification algorithms and with both 2D and 3D proposed

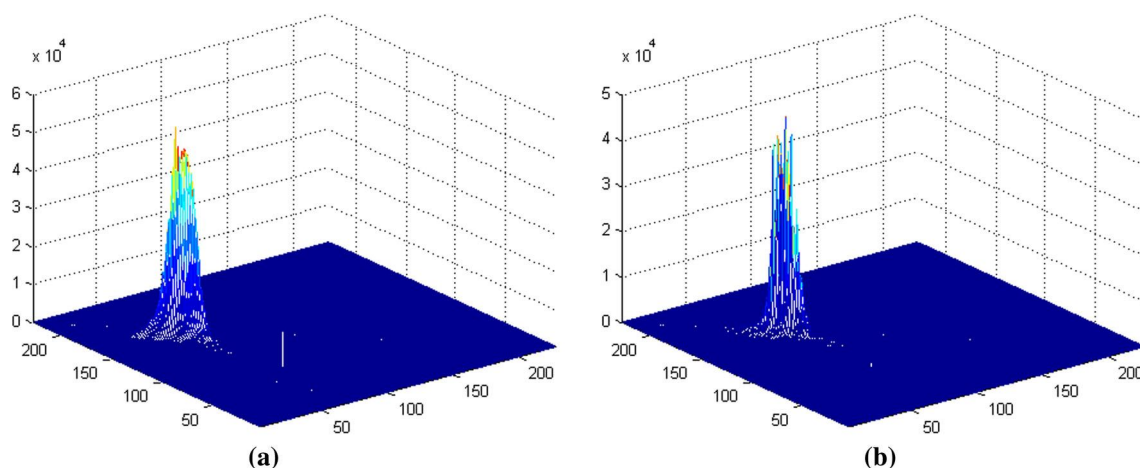


Fig. 4 Two-dimensional histograms (227×227 patches), **a** DR image sample, **b** NoDR image sample

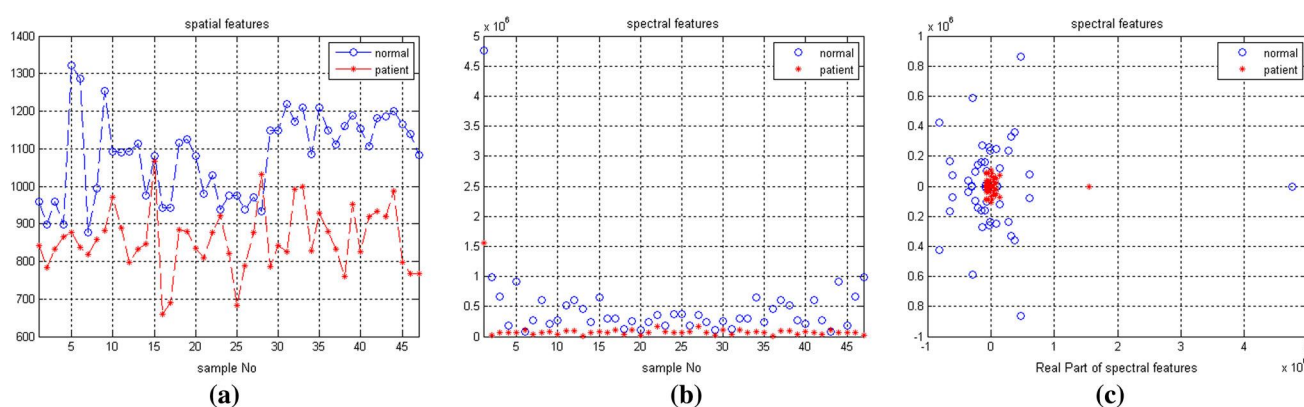


Fig. 5 Spatial and spectral feature maps obtained by the modified AlexNet containing 47 DR sample data (the red stars) and 47 sample data as NoDR (the blue circle), **a** spatial features **b** spectral features **c** the real and imaginary parts of the spectral features

nonlinear transformation over the input space (i.e., Modified AlexNet-K-Means, Modified AlexNet-KNN, Modified AlexNet-Subtractive, Modified AlexNet-FCM, Modified AlexNet-ISVDD). In this experimental scenario, all the available samples are split in two training and test sets. Our experimental results show that the modified AlexNet using the FCM as a classifier (Modified AlexNet-FCM) has the comparable performance. Thus, for keeping the results comparable, the optimum parameters for the FCM classification algorithm should be obtained. The modified AlexNet-FCM has two different parameters which affect the classification results, the number of training samples and the fuzzifier parameter (m). Since the parameter m in (4) and (5) affects the fuzziness of the standard FCM classification algorithm, so the best value for this parameter should be selected based on PSO algorithm. In some paper, the different initial cluster centers of FCM are considered to be study (Niazmardi et al., 2013), but we let that the initial cluster centers are chosen in a completely random way, and the best initial values for the cluster centers are considered. As mentioned before, choosing the high values

for m as the fuzzifier parameter, the FCM classifier represents an increased sharing of the samples among the clusters, whereas for $m = 1$, the FCM transferred to the K-Means algorithm. To show the effects of this parameter in our experiment, the algorithm is run with five different values for m . For our dataset with more similar classes, increasing in the fuzzifier value will cause a decrease in classification result. As Fig. 7a shows, the value $m = 5$ for this parameter seems to be an appropriate value in our proposed methodology.

To study the effects of the number of training data sample on the results, the algorithm is run with six different number of training sets. In this scenario, the selected value for the fuzzifier parameter is set to $m = 5$ as the best value estimated from the previous practice. Based on this scenario, the proposed methodology seems to be sensitive to the number of training data. An incremental performance improvement is observed when the algorithm included more training data samples, as seen in Fig. 7b.

The subtractive clustering method was first introduced in the field of extracting fuzzy rules (Chiu 1994). The

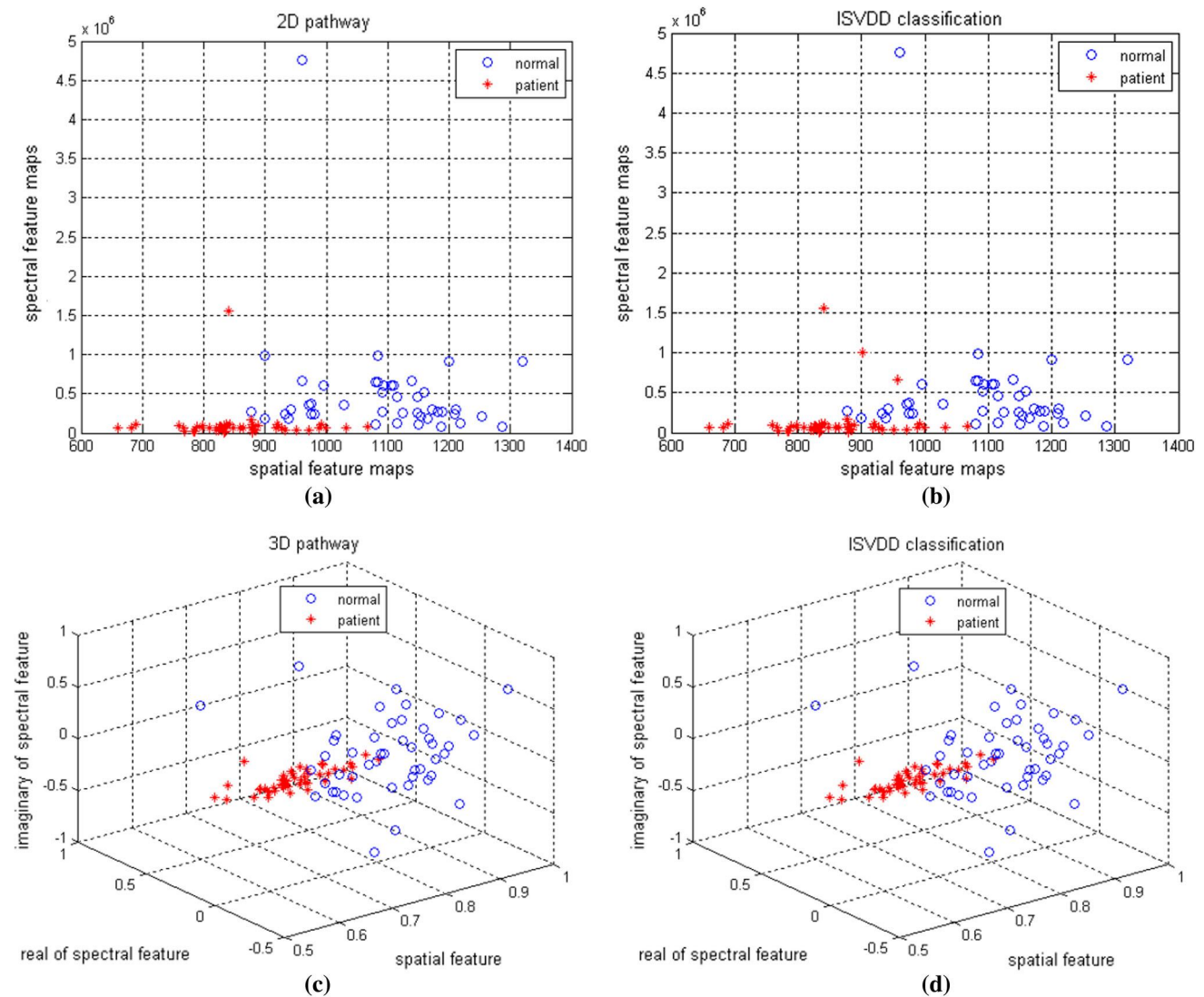


Fig. 6 Classification results obtained by proposed modified-ISVDD using the 2D (upper row) and 3D (lower row) kernel functions **a** the spatial features respect to spectral features **b** the classification result with optimal degrees $n_1 = 0.81, n_2 = 4.2$ obtained by PSO (2D

pathway) **c** the spatial features respect to the real and imaginary parts of spectral features **d** the classification result obtained by ISVDD (3D pathway)

subtractive clustering method considers each data point as a potential cluster center and defines a measure of the potential of a data point to serve as a cluster center. The potential of a data point is a function of its distance to all other data points. Thus, a high potential data sample would be a data with many neighboring data samples. By selecting *the range of influence*, the data points with the highest potential are determined so that the feature space is covered. Therefore, the alignment parameter for subtractive clustering method is *the range of influence* in each of the data dimensions. In this scenario, the optimum range of influence for subtractive clustering algorithm has been obtained and implemented. The initial training samples are selected randomly from the available sets and the number of training data is set to 70% (i.e., 33 samples per class for

each DR and NoDR classes for the Navid-Didegan dataset) and 30% of the remained data are selected as the test dataset. The testing and training sets are kept the same for all methods, keeping the results comparable. Since the clustering algorithms are sensitive to the initial clusters, the best initial conditions are considered for all methodologies. Figure 8 shows the classification results obtained by the modified AlexNet using K-Means, KNN, subtractive and FCM with 3D kernel functions.

To show the performance of the different classification methods, the performance indices mentioned in Sect. 5.1 are used. Table 1 exhibits the numerical results of TP , FP , TN , FN , accuracy, precision, sensitivity, specificity and Kappa coefficient criteria for DIARETDB1 dataset. To compare the results in both the DIARETDB1 and Navid-

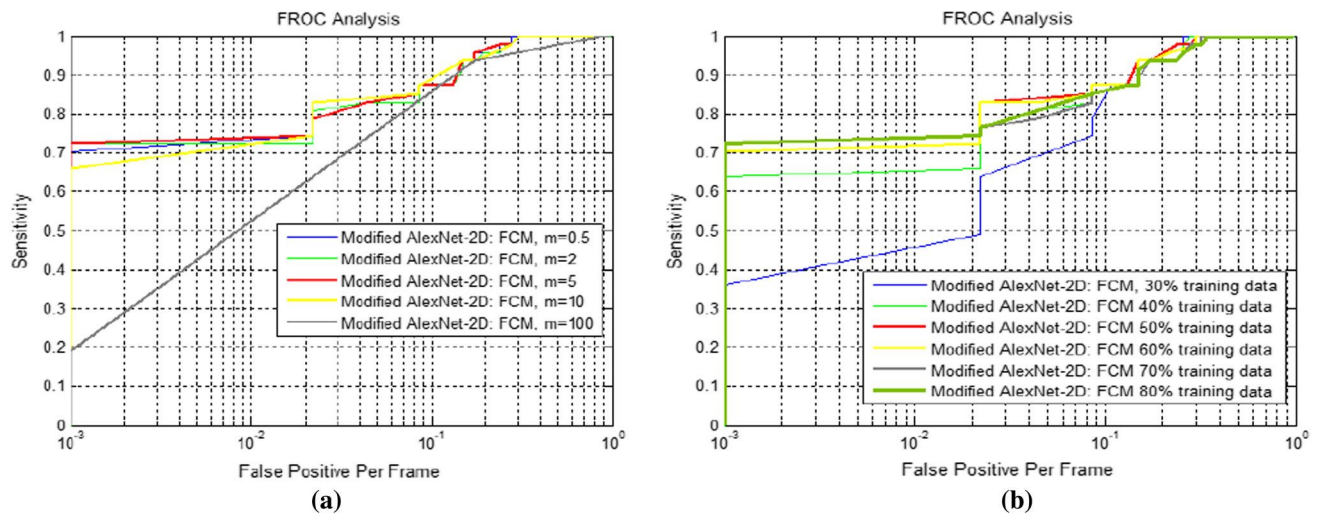


Fig. 7 FROC analysis for retinal DR or NoDR detection by the modified AlexNet-2D with FCM classification algorithm **a** with different values of the fuzzifier parameter m **b** with different numbers of training data

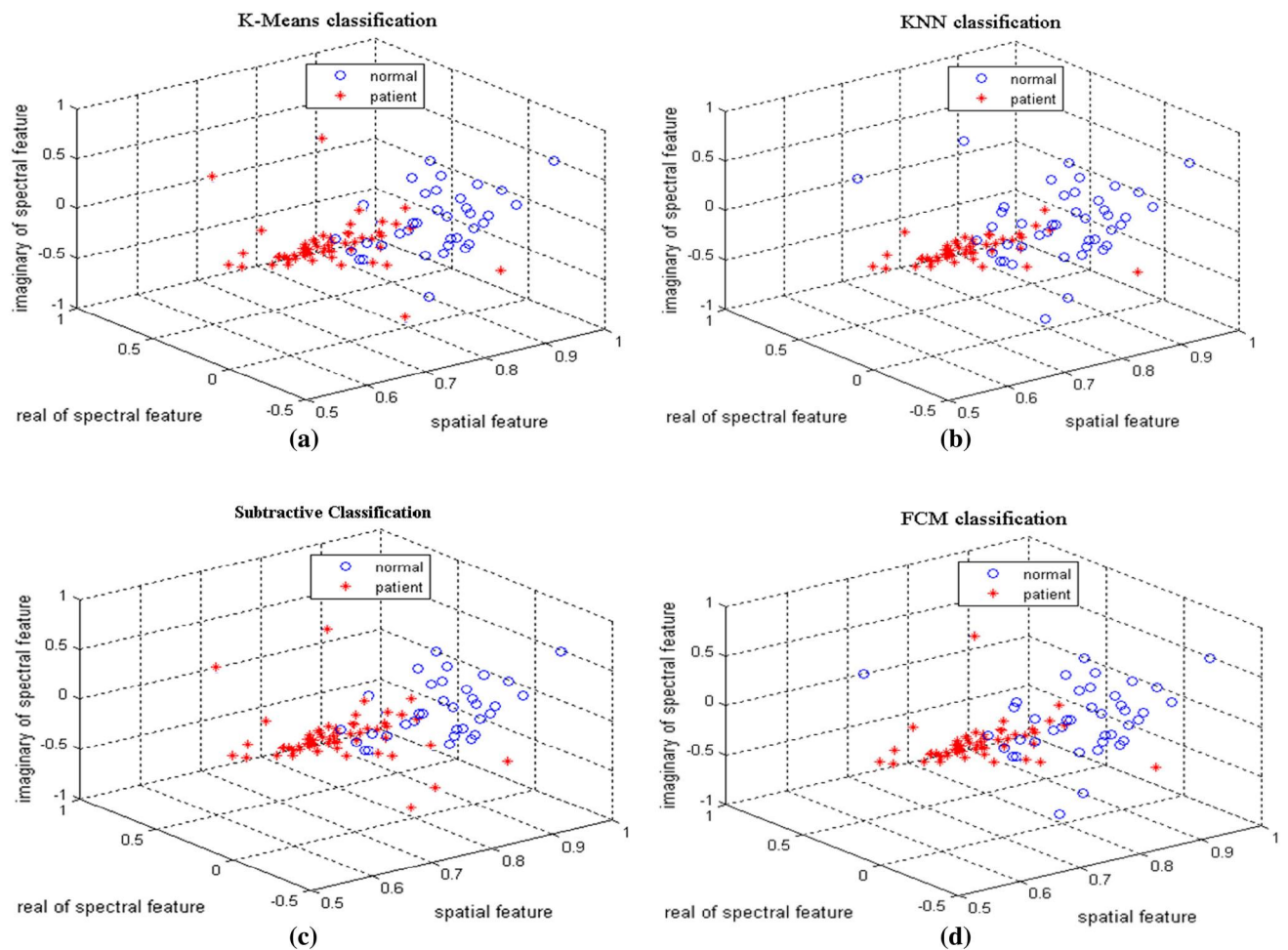


Fig. 8 Classification results obtained in 3D feature maps in Navid-Didegan dataset by the modified AlexNet using **a** K-Means, **b** KNN, **c** Subtractive and **d** FCM

Table 1 Performance indices for DIARETDB1 dataset obtained by using different 2D, 3D kernel functions and different classification methods

Method	Feature space	TP	FP	TN	FN	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	κ (%)
Modified AlexNet-K means	2D	46	21	26	1	76.60 (5.95)	68.66 (7.45)	97.87 (3.45)	55.32 (6.12)	53.19 (5.02)
Modified AlexNet-KNN	2D	46	2	45	1	91.00 (2.95)	95.83 (2.45)	97.87 (1.45)	95.74 (2.12)	93.61 (2.02)
Modified AlexNet-Subtractive	2D	47	17	30	0	81.91 (3.4)	73.44 (2.09)	100.00 (0)	63.8 (3.20)	63.83 (5.96)
Modified AlexNet-FCM	2D	42	7	40	5	87.23 (1.74)	85.71 (1.63)	89.36 (1.36)	85.11 (2.03)	74.47 (2.48)
Modified AlexNet-ISVDD	2D	47	1	46	0	98.94 (1.31)	97.92 (1.21)	100.00 (0)	97.87 (1.41)	97.87 (1.82)
Modified AlexNet-K means	3D	45	16	31	2	80.85 (5.23)	73.77 (7.12)	95.74 (5.34)	65.96 (6.12)	61.70 (4.95)
Modified AlexNet-KNN	3D	46	1	45	2	91.00 (2.95)	97.87 (1.45)	95.83 (2.12)	97.82 (1.67)	93.61 (2.02)
Modified AlexNet-Subtractive	3D	41	23	24	6	69.15 (3.2)	64.06 (2.54)	87.23 (2.98)	51.06 (3.12)	38.30 (4.32)
Modified AlexNet-FCM	3D	46	12	35	1	86.17 (1.23)	79.31 (2.12)	97.87 (2.17)	74.47 (3.12)	72.34 (2.78)
Modified AlexNet—ISVDD	3D	47	1	46	0	98.94 (1.31)	97.92 (1.21)	100.00 (0)	97.87 (1.41)	97.87 (1.82)
Modified AlexNet—ISVDD	(Spatial pathway)	43	15	32	4	79.79 (2.18)	74.14 (3.16)	91.49 (2.98)	68.09 (2.89)	72.13 (3.08)
Modified AlexNet—ISVDD	(Spectral pathway)	45	3	44	2	94.68 (2.28)	93.75 (3.89)	95.74 (3.19)	93.62 (2.08)	93.62 (3.98)

Didegan datasets, we augmented and randomly missing some NoDR and DR images to the DIARETDB1 dataset, respectively. Therefore, a balanced DIARETDB1 database of 94 retinal images is created in which both DR and NoDR classes were represented equally (47 samples for each class). One feature of the modified AlexNet-ISVDD structure is that the Fc7 of the standard AlexNet is forwarded to two separate pathways for better classification. Similar methodology was used in GoogleNet where the auxiliary classifiers connected to intermediate layers. To have a deep understanding of the performance of the modified AlexNet-ISVDD due to its two-pathway structure, we also measured the performance of the two pathways model separately (i.e., the spatial and spectral pathway models). So the last two rows of Table 1 show the numerical results of modified AlexNet-ISVDD in spatial and spectral pathway structures. Since the two pathways contain complementary information of a retinal fundus image, the two-pathway structure performs much better than the individual pathways. Table 1 exhibits the rounded mean values of accuracy, precision, sensitivity, specificity

and Kappa coefficient criterions and their standard deviations in parentheses.

We have obtained a mean accuracy 98.94%, a mean precision 97.92%, a mean sensitivity 100%, a mean specificity 97.87% and Kappa coefficient 0.97 by using the proposed modified AlexNet-ISVDD classification method. The sensitivity and specificity values obtained by the proposed algorithm are very high. Therefore, the modified AlexNet-ISVDD classification method states that our proposed algorithm does not misclassify a DR as NoDR image. Also, the proposed ISVDD algorithm in both 2D and 3D kernels has the same results and does not affect all performance metrics results.

5.5 Comparison with the most related works

In Table 2, we compare the proposed method performance with the most related works in the literature. There are the works of Andonová et al., (2017), Qureshi et al., (2021), Franklin and Rajan (2014), Ghosh et al., (2017) and Kwasigroch et al., (2018). To compare the results in both

Table 2 Performance comparisons with state-of-the-art binary and five-level of DR algorithms

Research study	Dataset	Images	method	Accuracy (%)	Sensitivity (%)	Specificity (%)
Andonová et al., (2017)	MESSIDOR	29	Binary	82.51	–	–
Qureshi et al., (2021)	EyePACS	54,000	Five-level DR	98	92.2	95.1
Franklin and Rajan, (2014)	DIARETDB1	57	Binary	99.7	96.3	99.8
Ghosh et al., (2017)	Kaggle	3000	Binary	95	–	–
			Five-level DR	85	–	–
Kwasigroch et al., (2018)	EyePACS	88,000	Binary	82	–	–
			Five-level DR	51	–	–
Proposed Modified AlexNet- ISVVD	MESSIDOR	94	Binary	85	93	85
Proposed Modified AlexNet- ISVVD	DIARETDB1	94	Binary	98.94	100	97.87

DIARETDB1 and MESSIDOR datasets, also we use a balanced MESSIDOR dataset of 94 retinal images (47 samples for each class) as shown in Table 2.

In Andonová et al., (2017), the retinal images have been used from the publicly available database, MESSIDOR. The accuracy criteria 79.92% and the cross-validation criteria 82.51% were reported in this work to evaluate the classification efficiency. As mentioned before, every academic research groups uses their own databases for evaluation and testing with different number of samples and different classification in two or more classes. In Franklin and Rajan (2014) an algorithm to detect the presence of exudates by using an artificial neural network has been presented. The proposed approach was based on feature extraction and clustering technique. They have evaluated their works by using 57 color retinal images of DIARETDB1 which contains 5137 objects for training and testing the neural network. As summarized in Table 2, Franklin and Rajan reported a mean accuracy 99.7, mean sensitivity 96.3 and mean specificity 99.8. As during the screening stage, the goal is to save time for the physician while reducing the number of test images and labeling the ones which are suspicious of DR as well as the ones which are close. Having a high recall or sensitivity score means that most of the patients will be screened correctly and their images will be labeled for ophthalmologist consideration. Although Franklin and Rajan have obtained a better performance indices in terms of accuracy and specificity than the performance of our methodology, the sensitivity performance index of our algorithm is higher than their results.

Ghosh et al., (2017) reported 95% accuracy for two class classification and 85% accuracy for five class classification on around 3000 validation images of Kaggle dataset. In Kwasigroch et al., (2018), a special class coding technique was proposed to include the information about relation between predicted and true level of disease. The utilized

models were trained using dataset containing over 88,000 retina photographs, and the best tested model achieved an accuracy of about 82% in detecting the retinopathy and 51% in assessing its stage for five class classification.

Utilizing a modified pre-trained CNNs for classification makes Graphics Processing Unit (GPU) and external memories unnecessary. The software package that was used in this paper was MATLAB 2017b. All the steps of the proposed method were done by an Intel i7 core CPU, with 8 GB memory, which is considerably advantageous comparing to common CNN training hardware requirements.

6 Conclusion and further work

In this paper, a modified AlexNet improved SVDD optimized by PSO was proposed in the diagnosis of DR and NoDR retinal images. The reasoning behind modification of a pre-trained CNN network is to avoid the time complexity of the training process for the convolutional systems. The algorithm uses the ISVDD algorithm with suitable kernel functions to classify the CNN data. A comparative study on different kernel parameters and different classifiers were presented. The comparative study was performed to demonstrate the effect of different degrees of kernel functions on the performance in diagnosing screening diabetic retinopathy cases.

To demonstrate the performance of the proposed modified and AlexNet-ISVDD in clinical applications, Navid-Didegan, DIARETDB1 and MESSIDOR datasets were applied and evaluated. The results of the study can be helpful to determine the proposed architecture for screening diabetic retinopathy cases in real clinical cases.

As shown in Table 1, the modified AlexNet-ISVDD had the best performance results among the other classification methods, considering all performance indices. Also, the

proposed ISVDD algorithm in both 2D and 3D kernels has the same results and does not affect all performance metrics results.

In Table 2, we compare the performance of the proposed method with other state-of-the-art methods. From the clinical usage perspective of the automatic DR screening approach, the sensitivity or recall, which demonstrates the correctness of DR diagnosis, is the most important factors. Our proposed algorithm outperforms other methodologies by considering sensitivity performance index.

The proposed scheme in this paper is not limited to the PSO optimization algorithm, so the other existing meta-heuristic algorithms can be used.

Acknowledgements The author would like to thank Dr. M. Ansari and Khatam-al-Anbia eye hospital employees for making available the data sets used in this paper. The author would also like to thank Dr. Hamid Khakshur and Navid-Didegan Clinic employees for their participation in preparing and labeling the retinal images to use in this study. This work was supported by [Iranian Society of Ophthalmology] (Grant number [ISO-G341397101]) and by [Khorasan Institute of Higher Education] (Grant number [KIHE-13971230]).

Funding None

Data availability Enquiries about data availability should be directed to the authors.

Declarations

Conflict of interest The author has no relevant financial or non-financial interests to disclose.

References

- Abbas Q et al (2017) Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features. *Biol Eng Comput* 55(11):1959–1974
- Anbeek P, Vincken KL, Bochove GS, Osch MJ, Grond J (2005) Probabilistic segmentation of brain tissue in MR imaging. *Neuroimage* 27(4):795–804
- Andonová M, et al (2017) Diabetic retinopathy screening based on CNN. In: *Proceedings of IEEE international symposium ELMAR*, pp 51–54
- Antal B, Hajdu A (2012) An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. *IEEE Trans Biomed Eng* 59(6):1720–1726
- Bar Y, Diamant I, Wolf L, Greenspan H (2015) Deep learning with non-medical training used for chest pathology identification. In: *Proceedings SPIE medical imaging*, p 94140V
- Bezdek JC, Ehrlich R, Full W (1984) FCM: the fuzzy C-Means clustering algorithm. *Comput Geosci* 10(2–3):191–203
- Bhatkar AP, Kharat GU (2015) Detection of diabetic retinopathy in retinal images using MLP classifier. In: *Proceedings of the international symposium on nanoelectronic and information systems*
- Chiu S (1994) Fuzzy model identification based on cluster estimation. *J Intell Fuzzy Syst* 2(3):267–278
- Ciresan DC, Giusti A, Gambardella LM, Schmidhuber J (2012) Deep neural networks segment neuronal membranes in electron microscopy images. In: Pereira F, Burges C, Bottou L, Weinberger K (eds), *Advances in neural inf. process. syst.*, Red Hook, NY: Curran, 25, pp. 2843–2851.
- Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J (2013) Mitosis detection in breast cancer histology images with deep neural networks. In: *Proceedings of the MICCAI*, pp 411–418
- Cocosco CA, Zijdenbos AP, Evans AC (2003) A fully automatic and robust brain MRI tissue classification method. *Med Image Anal* 7(4):513–527
- Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20(1):37–46
- Congdon N, Zheng Y, He M (2012) The worldwide epidemic of diabetic retinopathy. *Indian J Ophthalmol* 60(5):428
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
- Doshi D, Shenoy A, Sidhpura D, Gharpure P (2016) Diabetic retinopathy detection using deep convolutional neural networks. In: *Proceedings of the international conference on computer analysis security Trends (CAST)*, pp 261–266
- Dunn JC (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J Cybern* 3(3):32–57
- Edwards DC, Kupinski MA, Metz CE, Nishikawa RM (2002) Maximum likelihood fitting of FROC curves under an initial-detection- and-candidate-analysis model. *Med Phys* 29(12):2861–2870
- Erhan, D., Manzagol, P. A., Bengio, Y., Bengio, S., Vincent, P., 2009. The difficulty of training deep architectures and the effect of unsupervised pre-training. In: *Proceedings of the international conference on artificial intelligence and statistics*, pp 153–160
- Fan RE, Chen PH, Lin CJ (2005) Working Set Selection Using Second Order Information for Training SVM. *J Mach Learn Res* 6:1889–1918
- Franklin S, Rajan S (2014) Diagnosis of diabetic retinopathy by employing image processing technique to detect exudates in retinal images. *IET Image Process* 8(10):601–609
- Ghosh R, Ghosh K, Maitra S (2017) Automatic detection and classification of diabetic retinopathy stages using CNN. In: *IEEE international conference on Signal Processing and Integrated Networks (SPIN)*, pp 550–554
- Gong M, Liang Y, Shi J, Ma W, Ma J (2013) Fuzzy C-Means clustering with local information and kernel metric for image segmentation. *IEEE Trans Image Process* 22(2):573–584
- Hani AFM, Nugroho HA (2010) Gaussian bayes classifier for medical diagnosis and grading: application to diabetic retinopathy. In: *Proceedings of conference on biomedical engineering science (EMBS)*
- Hussain S, Anwar S, Majid M (2018) Segmentation of glioma tumors in brain using deep convolutional neural network. *Neurocomputing* 282:248–261
- Kamadi V et al (2016) A computational intelligence technique for the effective diagnosis of diabetic patients using principal component analysis (PCA) and modified fuzzy SLIQ decision tree approach. *Appl Soft Comput* 49:137–145
- Kauppi T, et al (2007) The DIARETDB1 diabetic retinopathy database and evaluation protocol. In: *Proceedings of the british conference on machine vision*, pp 252–261
- Khalid M, Pal N, Arora K (2014) Clustering of image data using K-means and fuzzy K-means. *Int J Adv Comput Sci Appl* 5(7):160–163
- Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: *Proceedings of the conference on neural information processing systems (NIPS)*
- Kuncheva LI (2011) A bound on Kappa-error diagrams for analysis of classifier ensembles. *IEEE Trans Knowl Data Eng* 25(3):494–501

- Kwasigroch A, Jarzembinski B, Grochowski M (2018) Deep CNN based decision support system for detection and assessing the stage of diabetic retinopathy. In: Proceedings of the IEEE international interdisciplinary PhD workshop (IIPhDW), pp 111–116
- Lalaoui L, Mohamadi T, Djaalab A, Abdelghani H (2015) A modified expectation of maximization method and its application to image segmentation. *Current Med Image Rev* 11(2):132–137
- Li W, Du Q, Zhang F, Hu W (2015) Collaborative-representation-based nearest neighbor Classifier for Hyperspectral imagery. *IEEE Geosci Remote Sens Lett* 12(2):389–393
- Liu X, Tang J (2014) Mass classification in mammograms using selected geometry and texture features, and a new SVM-based feature selection method. *IEEE Syst J* 8(3):910–920
- Margeta J, Criminisi A, Lozoya RC, Lee DC, Ayache N (2015) Fine-tuned convolutional neural nets for cardiac MRI acquisition plane recognition. *Comput Methods Biomechan Biomed Eng Image vis* 5(5):1–11
- Menze B, Reyes M, Leemput KV (2015) The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans Med Imag* 34(10):1993–2024
- Mohammadian S, Karsaz A, Roshan YM (2017a) A comparative analysis of classification algorithms in diabetic retinopathy. In: Proceedings of the international conference on software engineering and knowledge engineering
- Mohammadian S, Karsaz A, Roshan YM (2017b) Comparative study of fine-tuning of pre-trained convolutional neural networks for diabetic retinopathy screening. Presented at the 2nd international conference on biomedical engineering, Iran, to be published
- Motamedi M, Gysel P, Akella V, Ghiasi S (2016) Design space exploration of FPGA-based deep convolutional neural network. In: Proceedings of the IEEE conference on design automation
- Niazmardi S, Homayouni S, Safari A (2013) An improved FCM algorithm based on the SVDD for unsupervised hyperspectral data classification. *IEEE J Sel Top Appl Earth Observ Remote Sens* 6(2):831–839
- Nie, D., Wang, L., Adeli, E., Lao, C., Lin, W., Shen, D., 2018. 3-D fully convolutional networks for multimodal isointense infant brain image segmentation. *IEEE Trans. Cybern. PP* (99), pp. 1–14.
- Niemeijer M et al (2010) Retinopathy online challenge: Automatic detection of microaneurysms in digital color fundus photographs. *IEEE Trans Med Imag* 29(1):185–195
- Olson DL, Delen D (2008) Advanced data mining techniques. *Choice Rev* 45(12):45–6838
- Osareh A, Shadgar B, Markham R (2009) A computational-intelligence-based approach for detection of exudates in diabetic retinopathy images. *IEEE Trans Inf Technol Biomed* 13(4):535–545
- Pan Y et al (2015) Brain tumor grading based on neural networks and convolutional neural networks. In: Proceedings of the conference on engineering in medicine and biology Society (EMBS)
- Pasolli E, Melgani F, Tuia D, Pacifici F, Emery W (2014) SVM active learning approach for image classification using spatial information. *IEEE Trans Geosci Remote Sens* 52(4):2217–2233
- Patry G, Gauthier G, Lay B, Roger J, Elie D (2016) ADCIS download third party: Messidor database. ADCIS S.A., 2016. [Online]. Available: <http://messidor.crihan.fr>. Accessed: Nov. 16, 2016
- Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M (2013) Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: Proceedings of the MICCAI, pp 246–253
- Pratt H, et al (2016) Convolutional neural networks for diabetic retinopathy. In: Proceedings of the conference on medical imaging understanding and anal
- Priya R, Aruna P (2013) Diagnosis of diabetic retinopathy using machine learning techniques. *ICTACT J Soft Comput* 03(4):563–575
- Qureshi I, Ma J, Abbas Q (2021) Diabetic retinopathy detection and stage classification in eye fundus images using active deep learning. *Multimedia Tools Appl* 80(8):11691–11721
- Razavian AS, Azizpour H, Sullivan J, Carlsson S (2014) CNN features off-the-shelf: an astounding baseline for recognition. In: proceedings of the IEEE conference on computer vision and pattern recognition Workshops, pp 512–519
- Reddy YMS, Ravindran RE, Kishore KH (2017) Diabetic retinopathy through retinal image analysis: A review. *Int J Eng Technol* 7(1–5):19
- Roth H, et al (2014) A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. In: Goll P, Hata N, Barillot C, Hornegger J, Howe R, (eds), Proceedings of the MICCAI, 8673, LNCS, pp 520–527
- Roth H, Lu L, Lay N, Harrison A, Farag A, Sohn A, Summers R (2018) Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. *Med Image Anal* 45:94–107
- Saranya K, Ramasubramanian B, Kaja S, Mohideen G (2012) A novel approach for the detection of new vessels in the retinal images for screening diabetic retinopathy. In: Proceedings of the international conference on communication and signal processing
- Selvaraj H, Selvi ST, Selvathi D, Gewali L (2007) Brain MRI slices classification using least squares support vector machine. *Int J Intell Comput Med Sci Image Process* 1(1):21–33
- Shin H, Roth H, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers R (2016a) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imag* 35(5):1285–1298
- Shin JY, Tajbakhsh N, Hurst JY, Kendall CB, Liang J (2016b) Automating carotid intima-media thickness video interpretation with convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV.
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. <http://arxiv.org/abs/1409.1556>
- Singh C, Ranade SK, Singh K (2016) Invariant moments and transform-based unbiased nonlocal means for denoising of MR images. *Biomed Signal Process Control* 30:13–24
- Tajbakhsh N et al (2016) Convolutional neural networks for medical image analysis: fine tuning or full training? *IEEE Trans Med Imag* 35(5):1299–1312
- Tajbakhsh N, Gurudu SR, Liang J (2015a) A comprehensive computer-aided polyp detection system for colonoscopy videos. In: Information processing in medical imaging, pp 327–338
- Tajbakhsh N, Liang J (2015b) Computer-aided pulmonary embolism detection using a novel vessel-aligned multi-planar image representation and convolutional neural networks. In: Proceedings of the MICCAI
- Tax DMJ, Duin RPW (1999) Support vector domain description". *Pattern Recognit Lett* 20(11–13):1191–1199
- Vapnik VN (1995) The nature of statistical learning theory. Springer, NewYork
- Vo D, Lee S (2018) Semantic image segmentation using fully convolutional neural networks with multi-scale images and multi-scale dilated convolutions. *Multimedia Tools Appl* 77:1–19
- Vo HH, Verma A (2016) New deep neural nets for fine-grained diabetic retinopathy recognition on hybrid color space. In: Proceedings of the IEEE international symposium on multimedia

- Wang C, Zhang X, Yang H, Bu J (2012) A pixel-based color image segmentation using support vector machine and fuzzy C-means. *Neural Netw* 33:148–159
- Wang J, Kong J, Lu Y, Qi M, Zhang B (2008) A modified FCM algorithm for MRI brain image segmentation using both local and non-local spatial constraints. *Comput Med Imag Graph* 32(8):685–698
- Wang S et al (2015) Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing* 149:708–717
- Wen X, Zhang H, Jiang Z (2008) Multiscale unsupervised segmentation of SAR imagery using the genetic algorithm. *Sensors* 8(3):1704–1711
- Zhang R, Zheng Y, Mak T, Yu R, Wong S, Lau, j., Poon, C. (2017) Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain. *IEEE J Biomed Health Inform* 21(1):41–47
- Zhang W et al (2015) Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *Neuroimage* 108:214–224
- Zheng Y, Liu D, Georgescu B, Nguyen H, Comaniciu D (2015) 3d deep learning for efficient and robust landmark detection in volumetric data. In: *Proceedings of the MICCAI*, pp 565–572
- Zhou W, Wu C, Chen D, Wang Z, Yi Y, Du W (2017) Automatic microaneurysm detection of diabetic retinopathy in fundus images. In: *Proceedings of the IEEE conference on control and decision (CCDC)*

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.