TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABREVIATIONS

CCR	Correct classification rate
DWT	Discrete wavelet transform
EMD	Empirical mode decomposition
IMF	Intrinsic mode function
k-NN	*k*-nearest neighbour algorithm
MA	Microaneurysm
PCA	Principal component analysis
PNN	Probabilistic neural networks
QDA	Quadratic discriminant analysis
Std.dev	Standard deviation
SVM	Support vector machines

INTRODUCTION

Medical image analysis plays an important role in eye disease identification in the field of ophthalmology. For instance, conventional retinal disease identification techniques are based on manual observation by physicians. Indeed, the conventional-based approach for retina diagnosis is time consuming. Therefore, the automatic analysis of medical images has received a large scientific attention with the purpose of providing computational and intelligent tools to assist quantification and/or visualization of pathologies in the texture of digitized medical images.

The purpose of this work is to design an automatic screening system which aims to detect Artery sheath (Coat's disease), blot hemorrhage, retinal degeneration (circinates), age-related macular degeneration (drusens), and diabetic retinopathy (microaneurysms and exudates).
Coat's disease (or Artery sheath, see Coat's disease website given in references) is a very rare condition where there is abnormal development in the blood vessels behind the retina. It is characterized by abnormal retinal vascular development which results in massive intra-retinal and sub-retinal lipid accumulation. Coats Disease progresses gradually and affects central vision. Complete loss of vision can occur if not treated until its later stages.

Blot hemorrhage (see Digital Reference of Ophthalmology given in references) lies deeper in the retina. Usually blood accumulates in the outer plexiform or inner nuclear layers, or more easily seen at peripheral retina where the nerve fiber layer is thin.

Circinate retinopathy (see Medical dictionary webpage given in references) is a retinal degeneration characterized by circle of white spots encircling the macula that causes complete foveal blindness. For instance, exudates often appear as large circinate structures surrounding clusters of microaneurysms (Sopharak et al., 2008).

Drusens are the primary manifestation of age-related macular degeneration, the leading cause of late age blindness. Drusens are abnormal white-yellow deposits that appear on the retina.

They also appear as cloudy blobs, which exhibit no specific size or shape. In addition, they are heterogeneous and not as dense as exudates. The automatic detection of drusens was invesitaged by Brandon and Hoover (2003), Khademi and Krishnan (2007), Niemeijer et al., (2007), and Soliz et al., (2008).

Finally, microaneurysms (MAs) are small reddish spots on the retina (Martins et al., 2009). Spots are also present in normal retinas but with less contrast compared to MAs (Martins et al., 2008). For instance, MAs are swellings of the capillaries caused by a weakening of the vessel wall (Martins et al., 2009). Thus, MAs are considered as the earliest sign of diabetic retinopathy. Therefore many papers were concerned with the detection of this category of lesions (Martins et al., 2008, 2009; Quellec et al., 2008).

A large body of literature was concerned with the automatic detection of Age-related macular degeneration (drusens) and diabetic retinopathy (microaneurysms and exudates) in retina digital images. Therefore, the automatic detection of other pathologies such as Coat's disease, blot hemorrhage and circinate is considered in our study.

In order to extract texture features from normal and pathological retina images two approaches are considered in this work. The first approach is based on the discrete wavelet transform (DWT), and the second approach is based on the empirical mode decomposition (EMD). Indeed, the DWT was largely employed for retina processing in the literature. However, to the best of our knowledge at this time; no previous work has considered the EMD to extract retina features.

The EMD is a multi-resolution decomposition technique introduced by Huang et al., (1998) to perform a joint space-spatial frequency representation of the signal. The EMD is adaptive – fully data driven method- and is suitable for non-linear, non-stationary data analysis (Nunes et al., 2003). In addition, the EMD does not use any pre-determined filter or wavelet functions (Liu et al., 2007; Chen et al., 2009). The major advantage of EMD is that the basis functions are derived directly from the signal itself. On the other hand, DWT has two major

drawbacks (Chen et al., 2009). First, ‘’the extracted features are only limited to the energy of the textures on the moving and scaling of a single function’’ (Chen et al., 2009). Second, ‘’the number and temporal resolution of each component is defined a priori’’ (McGonigle et al., 2010). For instance, Fourier and wavelet analysis use a pre-determined filter or basis functions (Liu et al., 2007). In other words, the selected basic wavelet is used to analyze the whole frequency range (Yuping et al., 2006).

Finally, in order to classify normal versus pathological retina images, support vector machines are employed as the main classifier because of their scalability and ability to avoid local minima (Vapnik, 1995) and also because of their high classification performance shown in biomedical image recognition (Fu et al., 2005; Xu and Luo, 2009). The SVM performance has also been compared to those of quadratic discriminant analysis (QDA), *k*-nearest neighbor algorithm (*k*-NN), and probabilistic neural networks (PNN). Finally, the performance of the classifiers has been examined with regards to a reduced set of features obtained by applying principal component analysis to the original feature sets.

The contribution of our study is the following: We compare the EMD performance versus that of the DWT in the detection of several pathologies in retina digital images using four classifiers.

The thesis is organized as follow. The review of literature is presented in section 2. The methodology is given in section 3. The data and experimental results are shown in section 4. Finally, section 5 concludes.

CHAPTER 1

THE DISCRETE WAVELET TRANSFORM AND THE EMPIRICAL MODE DECOMPOSITION

The purpose of this section is to introduce the discrete wavelet transform (DWT) and the empirical decomposition mode (EMD) respectively in sections 1.1 and 1.2; including the block diagram of the DWT and the algorithm behind the EMD. In addition, examples of analyzed signals will be provided for each multiresolution technique.

1.1 The discrete wavelet transform

The wavelet transform (Daubechies, 1988; Mallat, 1999) is a multi-resolution analysis of a signal that has the advantage of great ability to identify and extract signal details at several resolutions. For instance, the wavelet transform decomposes a signal into a number of frequency subbands. In the case of two dimensional (2D) signals –for example a digital image-, the series of wavelet subbands give a spatial view of the image details at various resolutions and orientations (horizontal, vertical, diagonal). The one-dimensional wavelet transform of a function f is defined by:

$$f(x) = \sum_{i,j} c_{i,j} \psi_{i,j}(x) \tag{1.1}$$

$$c_{i,j} = \int_{-\infty}^{+\infty} f(x) \psi_{i,j}(x) \tag{1.2}$$

where $\psi_{i,j}(x)$ stands for the wavelet functions and $c_{i,j}$ are the corresponding coefficients for $f(x)$. A mother wavelet $\psi(x)$ is used to generate the wavelet basis functions by using translation and dilation operations:

$$\psi_{i,j}(x) = 2^{-i/2}\psi\left(2^{-i}x - j\right) \tag{1.3}$$

where *j* and *i* are the translation and dilation parameters, respectively. The one-dimensional wavelet decomposition can be extended to two-dimensional signals by separating the row and column decompositions. For instance, the 2D-DWT of a digital image is implemented by a low-pass filter which is convolved with the image rows, and a high pass filter which is convolved with the image columns. The convolutions are followed by down sampling by a factor of two. As a result, the 2D-DWT hierarchically decomposes a digital image into a series of successively lower resolution images and their associated detail images. For instance, the result of filtering process is four subbands: the approximation subband (LL), the horizontal detail subband (LH), the vertical detail subband (HL), and the diagonal detail subband (HH). The LL, LH, HL, and HH subband are respectively low frequencies for both directions, low frequencies for the horizontal direction and high frequencies for the vertical direction, high frequencies for the horizontal direction and low frequencies for the vertical direction, and high frequencies for both directions. Then, the obtained approximation images (LL) are decomposed again to obtain second-level detail and approximation images. This decomposition process can be represented in the common square scheme which is depicted as in Figure 1.1.

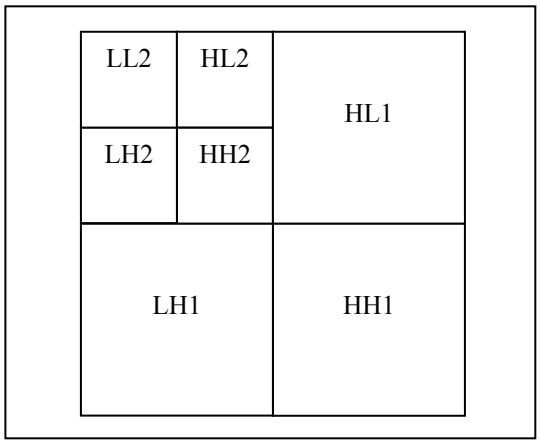

Figure 1.1 Two level 2D-DWT decomposition of an image

Another detailed representation of the decomposition process can be shown in Figure 1.2 that illustrates the first-level 2D-DWT of an image $F(x,y)$ based on filtering and down-sampling the row vectors of the image using one-dimensional low-pass filter L and one-dimensional high-pass filter H, whose respective impulse response components are the low and high-pass coefficients of a specific wavelet basis. Next, vertical filtering and down-sampling are performed again by L and H on each column of the obtained $F_L(x,y)$ and $F_H(x,y)$ results, yielding sub-images $F_{LL1}(x,y)$, $F_{LH1}(x,y)$, $F_{HL1}(x,y)$ and $F_{HH1}(x,y)$. Level two decomposition, repeats the process, starting from sub-image $F_{LL1}(x, y)$ to yields sub-images $F_{LL2}(x, y)$, $F_{LH2}(x, y)$, $F_{HL2}(x, y)$ and $F_{HH2}(x, y)$ which correspond respectively to LL2, LH2, HL2, and HH2 images in Figure 1.1. An example of brain magnetic resonance image (MRI) decomposition by wavelet transform as shown in El-Dahshan et al., (2010) is given in Figure 1.3 where g(n) and h(n) are respectively low pass and high pass filters.

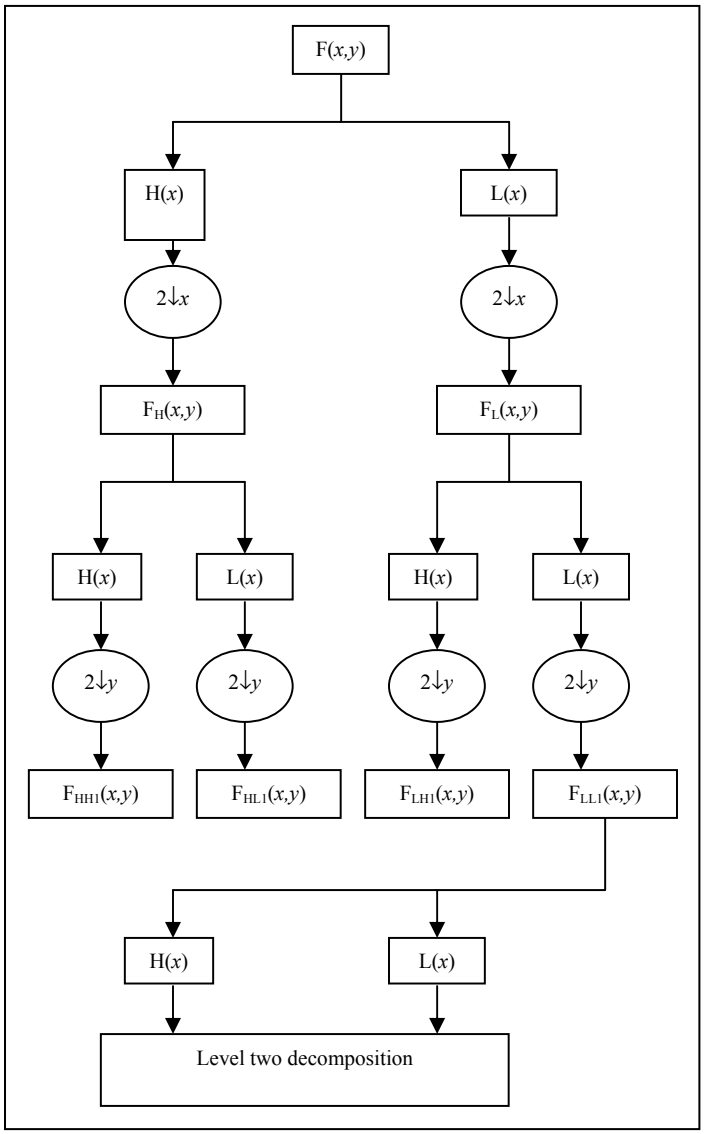

Figure 1.2 Block diagram of a 2D-DWT decomposition of an image

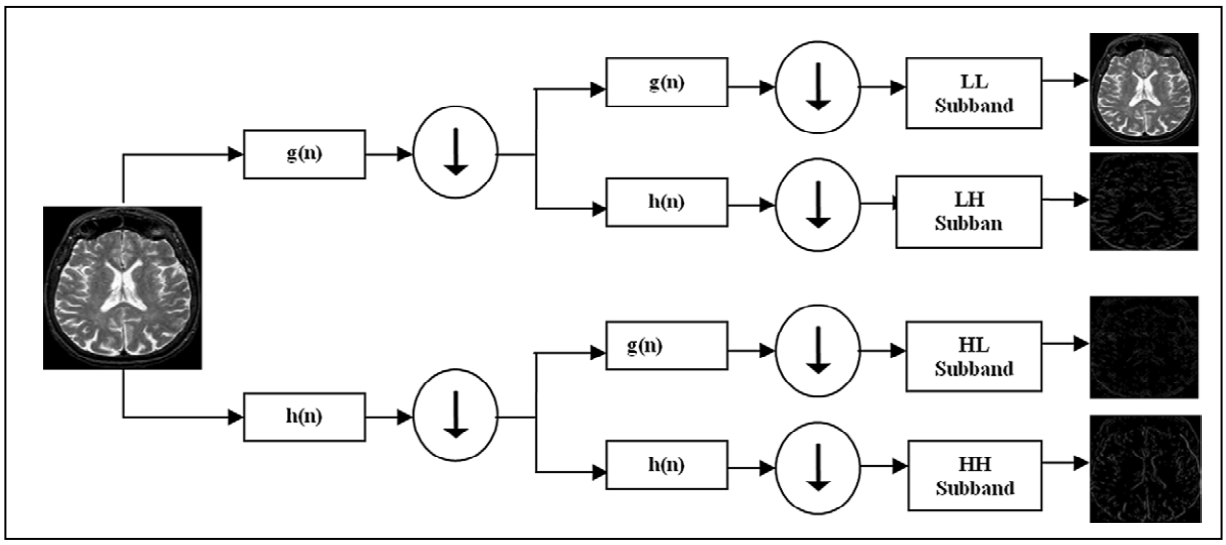

Figure 1.3 Block diagram of a 2D-DWT decomposition of a brain MRI

1.2 The empirical mode decomposition

The EMD is a general nonlinear, nonstationary signal processing method that was introduced by Huang et al., (1998). The major advantage of EMD is that the analysis is adaptive. In other words, the basis functions are derived directly from the signal itself. The key feature of the EMD is to decompose a signal into a sum of functions. Each of these functions (1) have the same numbers of zero crossings and extrema, and (2) is symmetric with respect to its local mean. These functions are called Intrinsic Mode Functions (IMF). The IMF are found at each scale going from fine to coarse by an iterative procedure called sifting algorithm. For a signal $s(t)$, the EMD decomposition is performed as follows (Liu, Xu, and Li, 2007):

a) find all the local maxima, $M_i, i = 1,2,...,$ and minima, $m_k, k = 1,2,...,$ in s(t),

b) compute by interpolation -for instance by cubic Spline- the upper and lower envelopes of the signal: $M(t) = f_M(M_i, t)$ and $m(t) = f_m(m_i, t)$,

c) compute the envelope mean e(t) as the average of the upper and lower envelopes: $e(t) = (M(t) + m(t))/2$,

d) compute the details as: $d(t) = s(t) - e(t)$,

e) check the properties of d(t):

- If d(t) meets the above conditions (1) and (2) given previously, compute the ith IMF as :

$IMF(t) = d(t)$ and replace s(t) with the residual $r(t) = s(t) - IMF(t)$,

- If d(t) is not an IMF, then replace s(t) with the detail: $s(t) = d(t)$,

f) return to step (a) and iterate to step (e) until the residual r(t) satisfies a given stopping criteria.

Finally, the signal $s(t)$ can be expressed as follows:

$$s(t) = \sum_{j=1}^{N} IMF_j(t) + r_N(t) \tag{1.4}$$

where N is the number of IMF which are nearly orthogonal to each other, and all have nearly zero means; and $r_N(t)$ is the final residue which is the low frequency trend of the signal s(t). Usually, the standard deviation (SD) computed from two consecutive sifting results is used as criteria to stop sifting process by limiting the SD size (Nunes et al., 2003; Qin et al., 2008; McGonigle et al., 2010) as:

$$SD(k) = \frac{\sum_{t=0}^{T} \left| d_{k-1}(t) - d_k(t) \right|^2}{\sum_{t=0}^{T} d_{k-1}^2(t)} < \varepsilon \tag{1.5}$$

where k is the index of the kth difference between the signal $s(t)$ and the envelope mean $e(t)$. The term ε is a pre-determined stopping value. The 2D signal EMD follows the same process as 1D signal EMD and 2D IMF are defined in the same manner (Guangtao et al., 2008; Demir and Ertürk, 2009, 2010; Xu et al., 2011).

In this study, two examples of signals analyzed with empirical mode decomposition are presented. The first example is a random signal with mean one and standard deviation equals to two, and the second example is a sine wave signal -w(t)- generated as follows:

$$w(t) = \sin(1000t) + 3\sin(2000t) + 5\sin(4000t) \tag{1.6}$$

where $t \in [0,100]$ with 0.1 step. Figure 1.4 shows the simulated random signal $s(t)$ and its first seven intrinsic mode functions (IMF): IMF_1 to IMF_7.
Figure 1.5 shows the original random signal $s(t)$ and its intrinsic mode functions in the form of color scale 2D amplitude representation where $s(t)$ and each IMF is depicted as a rectangular area in the image.

One may notice how the low frequency component; for instance the seventh (last) IMF; looks like an almost uniform color channel. The image in Figure 1.5 is associated with its color map. Along y-axis, rectangle 1 is the random signal, rectangle 2, 3, 4, 5, 6, 7, and 8 are respectively the first, second, third, fourth, fifth, sixth, and seventh intrinsic mode functions.

On the other hand, Figure 1.6 shows the simulated sine wave and all IMF, whilst Figure 1.7 shows simultaneously the results of the decomposition process: the original sine wave and its IMF. All these signals are shown in the form of color scale 2D amplitude representation where $s(t)$ and each IMF is depicted as a rectangular area in the image in Figure 1.8. Along y-axis, rectangle 1 is the original sine wave signal and rectangle 2, 3, 4, 5, 6, 7, and 8 are respectively the first, second, and third, fourth, fifth, sixth, and seventh intrinsic mode functions. One may notice how the low frequency component; for instance the seventh (last) IMF; looks like an almost uniform color channel. Finally, Figure 1.9 gives a real example of a biomedical image analyzed with empirical mode decomposition. It shows a red cell slice biomedical image, which was converted from true color format to gray format before decomposition (Liu, Xu, and Li, 2007).

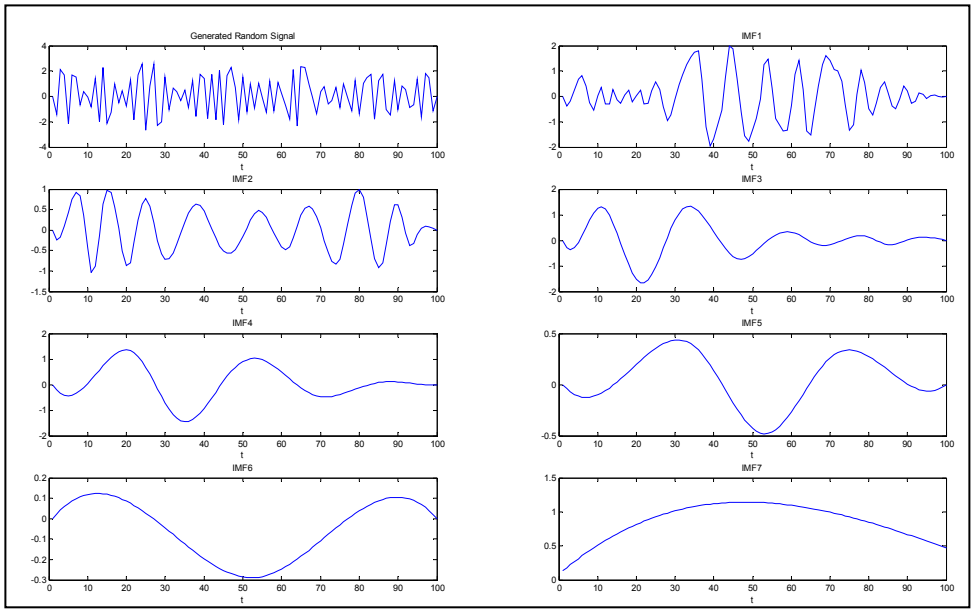

Figure 1.4 Generated signal s(t) and its seven empirical mode functions

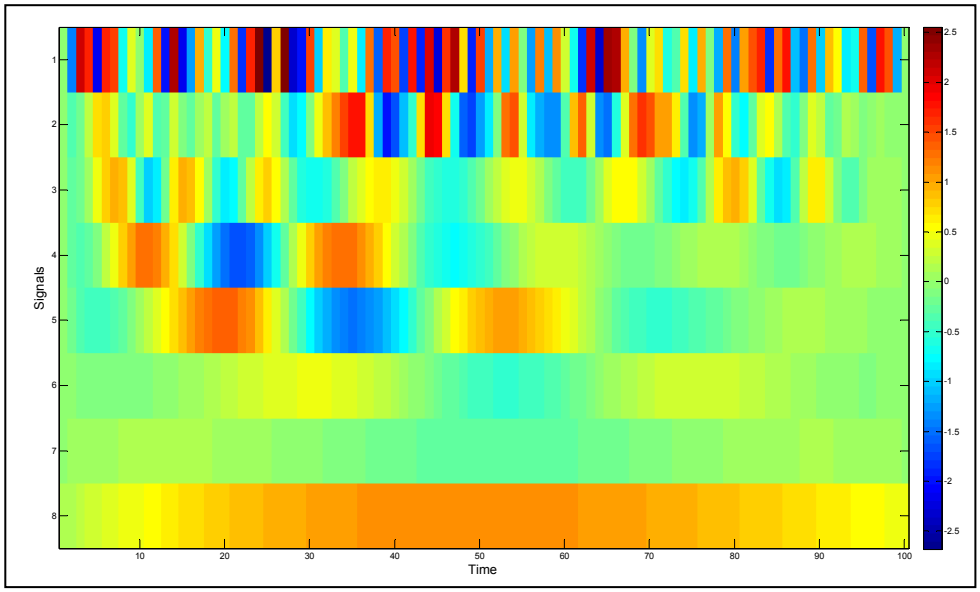

Figure 1.5 Generated signal s(t) and its seven empirical mode functions as an image

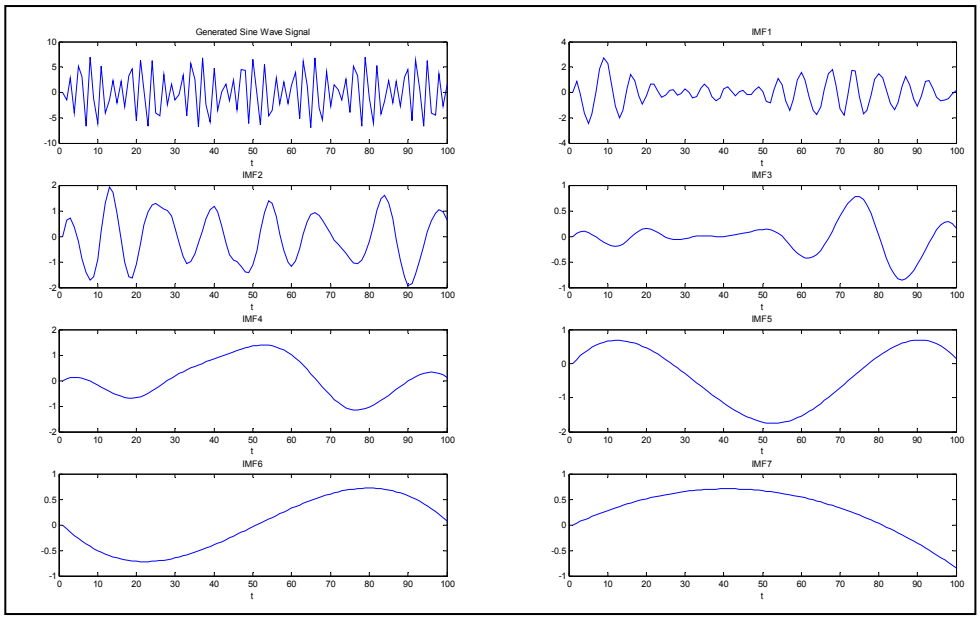

Figure 1.6 Generated sine wave signal and its seven IMF
w(t) = sin(1000.t) + 3.sin(2000.t) + 5.sin(4000.t)

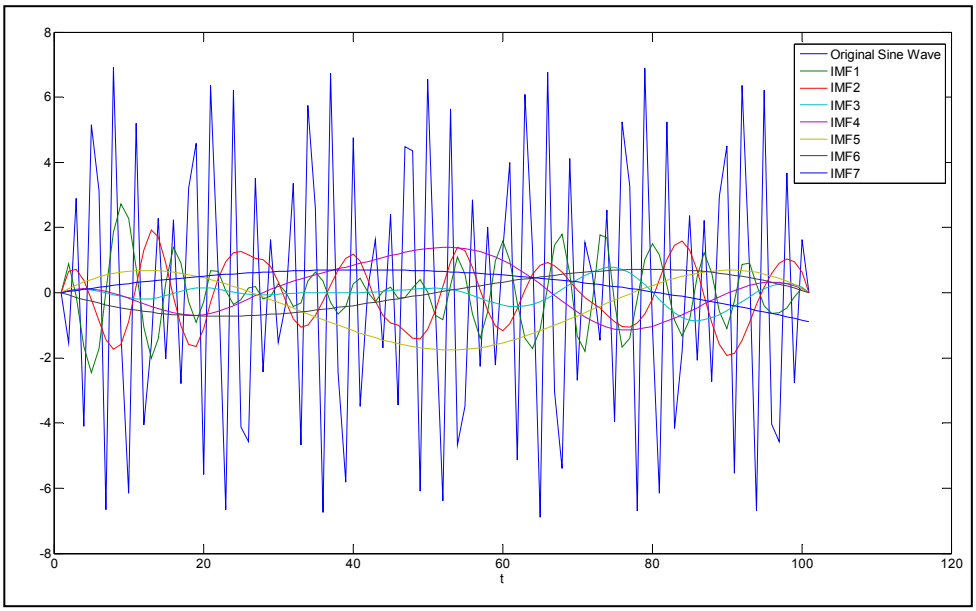

Figure 1.7 Generated sine wave signal w(t) and its seven IMF all togather

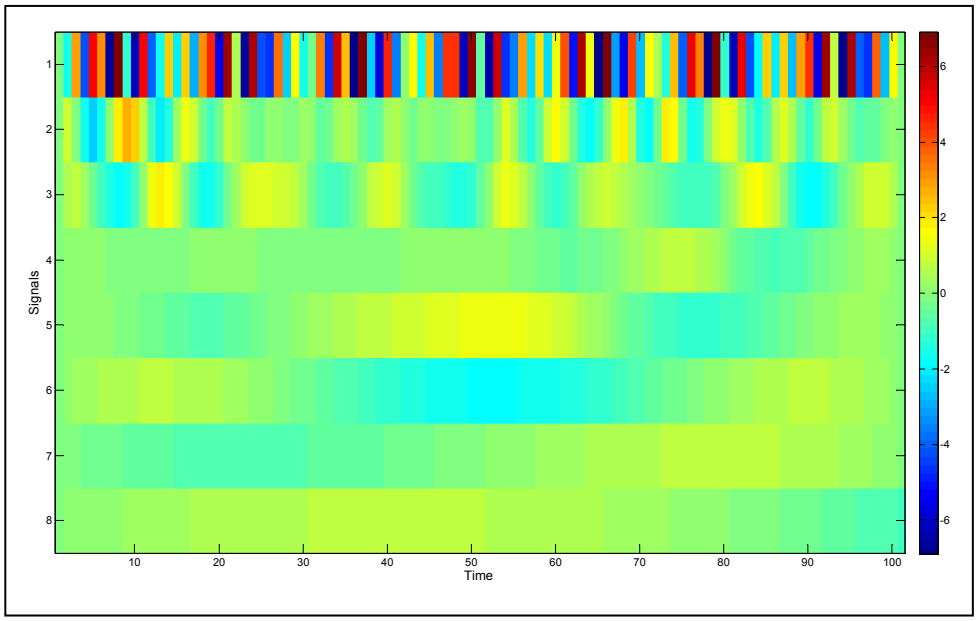

Figure 1.8 Generated sine wave w(*t*) and its seven empirical mode functions as an image

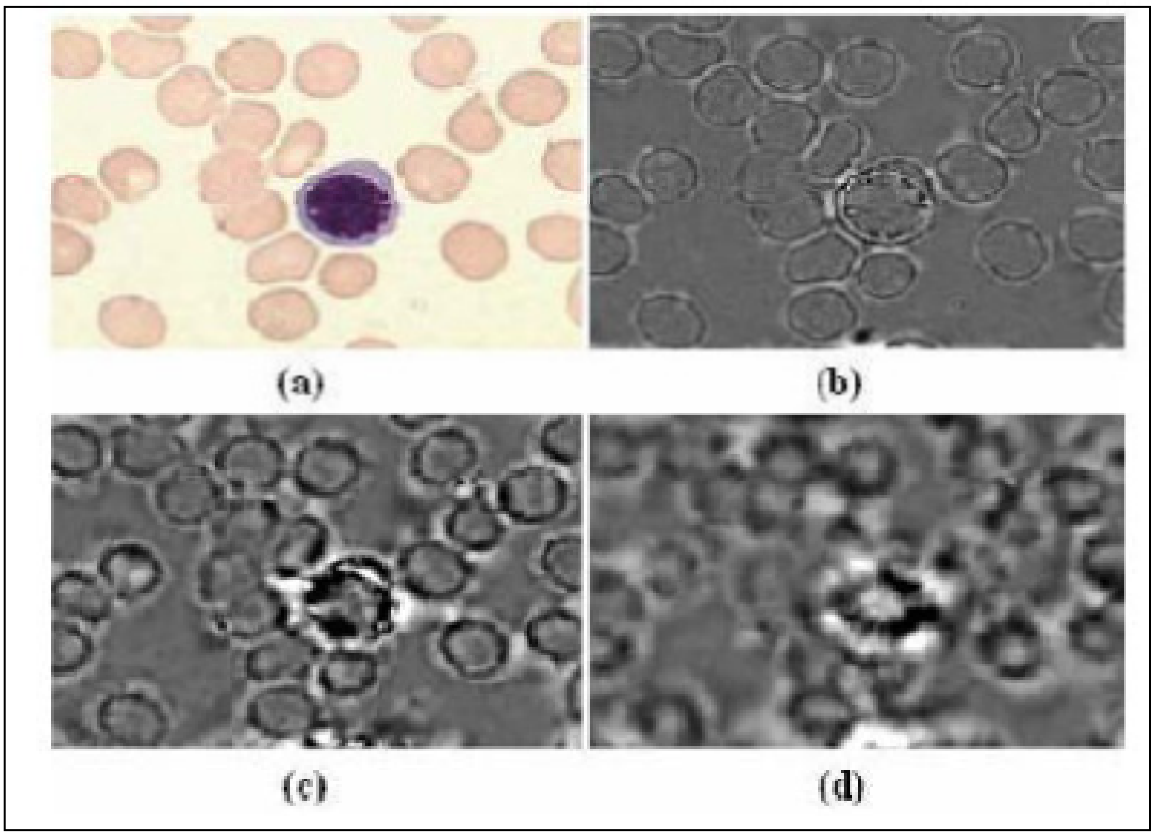

Figure 1.9 Red cell slice biomedical image decomposition by the EMD
As shown in Liu, Xu, and Li (2007) where (a) is a red cell slice, (b) is the first IMF, (c) is the second IMF, and (d) is the third IMF

CHAPTER 2

HISTORICAL BACKGROUND

In this paper, the applications of DWT and the EMD in biomedical signal processing are briefly reviewed. In addition, the results of other approaches used in the literature are presented.

2.1 The discrete wavelet transform in retina photographs processing

Lamard et al., (2007) introduced an algorithm based on the translation invariant wavelet transform and template matching to detect Retina microaneurisms. Then, they considered the horizontal and vertical sub-bands at several scales plus the approximation of the signal to extract features. Finally, specificity and sensitivity statistics are used to evaluate the classification system for different wavelet mothers and decomposition levels. The simulation results show that the best combination is obtained with the Haar wavelet and second level of decomposition with 96.18% specificity and 87.94% sensitivity. Khademi and Krishnan (2007) employed the 2-D version of Belkyns's shift-invariant DWT (SIDWT) to classify normal against abnormal (exudates, large drusens, fine drusens, choroidal neovascularization, central vein and artery occlusion, arteriosclerotic retinopathy, histoplasmosis, hemi-central retinal vein occlusion) retina images. In order to capture texture directional features, normalized gray level co-occurrence matrices (GLCMs) are computed at 0◦ in HL, 90◦ in LH, 45◦, 135◦ in HH and 0◦, 45◦, 90◦ and 135◦ in the LL sub-band. Then, homogeneity and entropy statistics were computed from LH, HL, HH and LL sub-bands for each decomposition level. For instance, the decomposition level was set to four. Finally, linear discriminant analysis (LDA) is used as classifier in conjunction with the Leave One Out Method (LOOM). The obtained classification accuracy was 82.2%.

In order to detect Microaneurysms in retina photographs, Quellec et al., (2008) employed 3-level wavelet decomposition and used genetic algorithms to find the best discriminative

wavelet mother (Haar, biorthogonal, orthogonal) function and coefficients from HH, HL, and LH sub-bands.

Color photographs, green filtered photographs, and angiographs were three image modalities used in the study, and template matching technique was employed for classification purpose. The simulation results showed that Microaneurysms were detected with sensitivities of 89.62%, 90.24%, and 93.74% and positive predictive values of 89.50%, 89.75%, and 91.67% for color photographs, green filtered photographs, and angiographs respectively. They concluded that optimal sub-band combinations weakly depend on the wavelet filter. In addition, optimal wavelet sub-bands improved the results obtained by conventional wavelet filters. Yagmur et al., (2008) used discrete wavelet transform and back-propagation neural networks for automatic classification of diabetic retinopathy, hypertensive retinopathy, macular degeneration, vein branch oclusion, vitreus hemorrhage, and normal retina. All color images were transformed to grey scale and the extracted features that were fed to the neural networks were the coefficients of all wavelet sub-band decompositions at level one. The recognition rates were respectively 50%, 70%, 83%, 90%, 93% and 95% for testing five retinopathy cases and normal images. Xu and Luo (2009) combined features obtained from stationary discrete wavelet transform (SWT) and grey level co-occurrence matrix (GLCM) to characterize retina images with hard exudates. The mean and standard deviation were extracted from the approximate and detail regions of SWT decomposed images at scale one and two. Finally, support vector machines (SVM) with Gaussian radial basis function were employed to classify hard exudates images. Based on leave-one-out method (LOOM), the obtained classification accuracy, sensitivity, and specificity were 84%, 88% and 80% respectively.

2.2 Other approaches for retina digital image processing

Baroni et al., (2007) suggested a computer approach based on co-occurrence matrices for the analysis of retinal texture and artificial neural networks (ANN) to classify single retinal layers. The obtained accuracy was 79%, specificity about 71% and sensibility was 87%. Meier et al., (2007) used four approaches to extract features from retina to automatically classify glaucoma images. The first set of features is obtained by taking the pixel intensities as input to principal component analysis. The second features are obtained from Gabor texture filter responses. The third set of features is computed from the coefficients of the Fast Fourier Transform. The fourth set of features is obtained from the histogram of the intensity distribution of the image. Finally, support vector machines were employed for classification purpose. The performance of the classifications using one feature set only was 73% with the histogram features, 76% with Fast Fourier Transform coefficients, 80% with the Gabor textures and 83% with the pixel intensities. Franzco et al., (2008) used Fourier transform to evaluate the optical degradation of a retinal image of a cataractous eye. The experimental results showed that Fourier analysis of retinal images is significantly correlated with LOCS III (*R-squared* = 0.59) in grading cataract severity. In addition, it is found that Fourier analysis shows a comparable correlation with visual acuity (*R-squared* = 0.39) as LOCS III (*R-squared* = 0.44). They concluded that Fourier analysis is a useful automated method in grading of cataract severity; but it cannot determine the anatomic type of cataract. Lee et al., (2008) employed probabilistic boosting algorithm for nonhomogenous texture discrimination. In particular, the main purpose was to detect drusen in retina texture. They used morphological scale space analysis and grey level co-occurrence matrices (GLCM) to extract texture features. Using four different test samples, the detection rate of normal images varied between 81.3% and 92.2%, and the detection rate of abnormal images varied between 71.7% and 85.2%.

In order to detect glaucoma in digital images, Nayak et al., (2009) used morphological measures such as the cup to disc ratio, ratio of the distance between optic disc center and optic nerve head to diameter of the optic disc, and ratio of blood vessels area in inferior-

superior side to area of blood vessel in the nasal-temporal side; all as features to be fed to neural networks classifier. The system automatically classified glaucoma with a sensitivity and specificity of 100% and 80% respectively. The overall average classification rate was 90%.

2.3 The empirical mode decomposition in biomedical image processing

The Empirical Mode Decomposition (EMD) technique has been successfully applied in biomedical engineering problems. In particular, it has been largely employed in cardiovascular signal processing; including classification of EEG (Tafreshi et al., 2008), ECG denoising (Pan et al., 2007), ECG, BCG, PPG, and IPG processing (Pinheiro et al., 2011). The EMD has also been applied to two dimension signals. For instance, Nunes et al., (2003) applied the 2D-EMD to synthetic and brain magnetic resonance images to extract features at multiple spatial frequencies. The study showed the effectiveness of the EMD to represent images. They concluded that the 2D EMD offers a new and promising way to decompose images and extract texture features without parameter.

In late 2000s, the EMD has been employed in medical image processing. For instance, Qin et al., (2008) employed 2D-EMD to enhance medical images. Experiments show that good results are obtained by using 2D-EMD than using linear gray-level, transformation, piecewise linear gray-level transformation, logarithmic transformation, exponential transformation, and histogram equalization transformation. In particular, details of medical images were more definite and distinct after enhancement. McGonigle et al., (2010) employed a Multi-EMD approach to analyze signals obtained from functional neuroimages.

In particular, the purpose was to find which Intrinsic Mode Function (IMF) from each voxel should be used to represent the data at each scale. Finally, *k*-means clustering was performed on multi-EMD components to discover regions that behave synchronously at each temporal scale. They found that Multi-EMD based cluster analysis discriminate much better temporal

scales than wavelet-based cluster analysis (WCA). McGonigle et al., (2010) concluded that Multi-EMD based clustering is a promising approach to explore functional brain imagery.
Liu et al., (2007) applied Bidimensional Empirical Mode Decomposition (BEMD) to the problem of biomedical images retrieval using *k*-means clustering. In other words, EMD was employed to analyze texture of images and the mean and standard deviation of the amplitude matrix, phase matrix and instantaneous frequency matrix of the Intrinsic Mode Functions (IMFs) and their Hilbert transformations. The experimental results show that retrieval results of Gabor features are higher than that of BEMD-based features. On the other hand, BEMD performs much better than wavelet-fractal based approach. Jai-Andaloussi et al., (2010) employed the BEMD to obtain characteristic signatures of images for Content Based Medical Image Retrieval (CBIR). Two approaches were considered. The first approach called BEMD-GGD was based on the application of BEMD to medical images. Then, the distribution of coefficients derived from BIMF was characterized using Generalized Gaussian Density (GGD). In the second approach called BEMD-HHT, the Huang-Hilbert transform (HHT) was applied to each Bidimensional Empirical Mode Function (BIMF). Then, the mean and standard deviation were extracted from the amplitude matrix, phase matrix and instantaneous frequency matrix of each transformed (BIMF). Genetic algorithms were employed to generate adapted BIMF distance weights for each image in the database. The previous approaches were tested on the three databases including a diabetic retinopathy, a mammography and a faces database. The experimental results show that BEMD-GGD gives globally better results than BEMD-HHT. In addition, the retrieval efficiency is higher than 95% for some cases.

Finally, the EMD method was compared to the DWT approach in the problem of biomedical time series classification. For instance, Ai et al., (2011) used singular value decomposition (SVD) to extract the features of intrinsic mode functions (IMFs) and support vector machine (SVM) to classify two types of tremor, namely essential tremor (ET) and Parkinson's disease (PD). Besides, they employed the singular value features of discrete wavelet transform (DWT) as input to the SVM. The results obtained from cross-validation tests show that the

EMD-SVD features outperform DWT-SVD features in terms of the accuracy (98%), sensitivity (97.5%), and specificity (98.33%).

CHAPTER 3

CONTRIBUTION AND METHODOLOGY

3.1 The contribution of our study

The discrete wavelet transform (DWT) was widely applied in the processing of retina digital images for pathologies detection (Khademi and Krishnan, 2007; Quellec et al., 2008; Yagmur et al., 2008; Xu and Luo, 2009). This approach allowed obtaining an accuracy that varies from 50% to 96%. For instance, DWT perfomed much better than the Fourier transform that leads to 76% obtained by Meier et al., (2007) and 59% obtained by Franzco et al., (2008). However, to the best of our knowledge there no paper that considered the empirical mode decomposition (EMD) for retina digital image processing for features extraction and classification. Indeed, EMD was very successful in the processing of one dimension biomedical signals and in the context of biomedical images retrieval. For example, Ai et al., (2011) found that EMD features perform much better than DWT features in the analysis and classification of biological time series. Therefore, we contribute to the literature by exploring the performance of EMD and DWT in the case of two dimensional biological signals.

The main purpose of our study is to compare the performance of the EMD and DWT in the problem of retina features extraction and classification. The main hypothesis is straightforward: since the EMD is an adaptive signal processing technique; in other word is a data-driven method, it could be more appropriate to analyze retina texture than the DWT which is not an adaptive procedure. As a result, the EMD-based statistical features are expected to outperform the DWT-based statistical features.

Another hypothesis made in our study: the statistical features extracted from EMD and DWT-processed images are sufficient to describe the distribution of the pixels that characterize retina.

Finally, another contribution of our study is concerned with the comparison of a set of classifiers which are widely used in biomedical image processing to validate the outperformance of the EMD-based statistical features. In sum, our study is based on rigorous experiments to check the performance of the EMD over the DWT.

The next chapter deals with our methodology; including the choice of the wavelet mother, the description of the chosen empirical mode decomposition algorithm (Rato et al., 2008), statistical features to be extracted, the classifiers used to detect abnormal images, and the statistical measures considered to evaluate the performance of the classifiers.

3.2 The proposed approach

As stated in introduction, the purpose of our study is to compare the performance of the empirical mode decomposition (EMD) features with those obtained with discrete wavelet transform (DWT) in the problem of the classification of normal retina digital images versus abnormal retina photographs. The generic detection system contains four basic steps; namely image acquisition using appropriate ophthalmology equipment, gray scale conversion, pixels equalization, EMD or DWT processing, features extraction, principal component analysis (PCA) (optional), and classification. In our study, three detection systems (approach) are designed: the EMD, DWT, and EMD-PCA system. The latter approach is considered to check whether principal component analysis reduced (selected) features helps improving the accuracy of the EMD-based features.

In particular, PCA can be applied to EMD features via two routes: applying PCA jointly to first and second IMF obtained from retina images or applying PCA separately to them. Indeed, EMD features are extracted only from the first and the second IMF since the third IMF does not contain helpful information. For instance, the third IMD contains more low frequency component. We have already used its features to classify retina images and obtained low accuracy in comparison with features extracted from first and second intrinsic mode functions. The reader may look at Figure 4.5 to how the third IMF looks like. In our

study, PCA is not considered for DWT sub-images to reduce the number of experiments. For instance, DWT features are extracted from three sub-bands (LH, HL, HH) at three level of decomposition as in Quellec et al., (2008) study. LL sub-bands are ignored because they contain no discriminant information (Quellec et al., 2008). The EMD-based features approach and the DWT-based features approach are shown in Figures 3.1 and 3.2 respectively. On the other hand, the EMD-PCA systems are shown in Figures 3.3 (joint PCA) and 3.4 (separate PCA). The pixels equalization technique, choice of the mother wavelet, EMD processing algorithm, extracted features, principal component analysis (PCA) technique, and the classifiers are described in the following sub-sections.

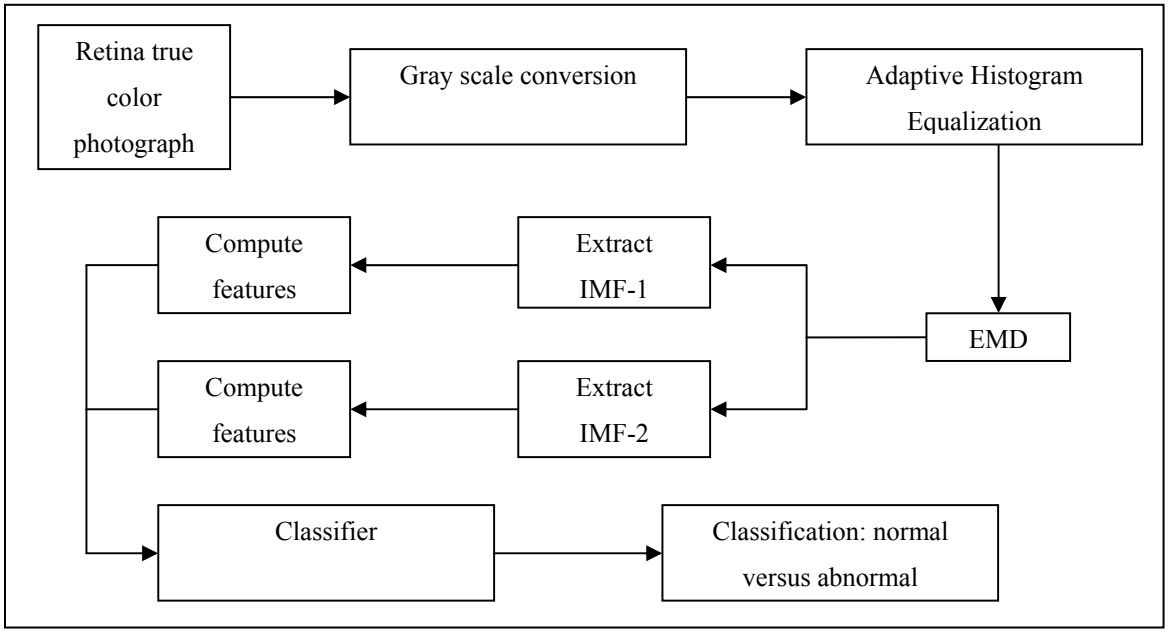

Figure 3.1 The EMD approach

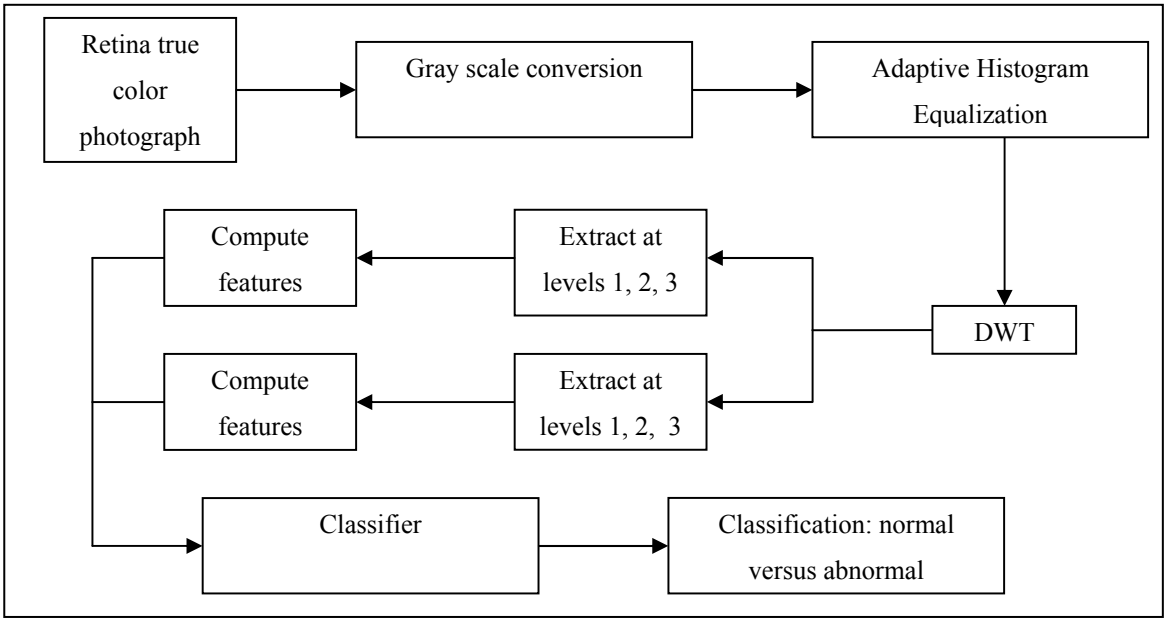

Figure 3.2 The DWT approach

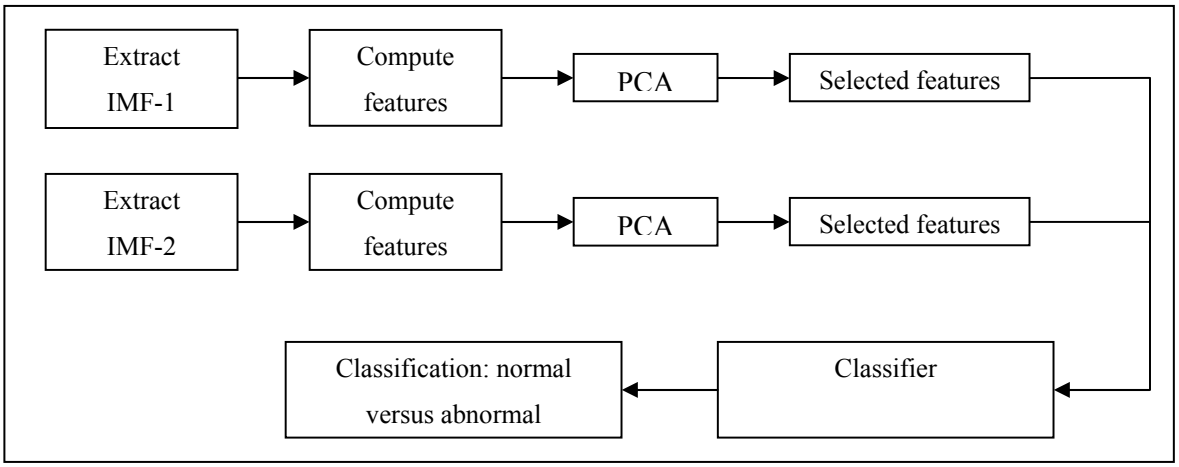

Figure 3.3 The EMD-PCA approach

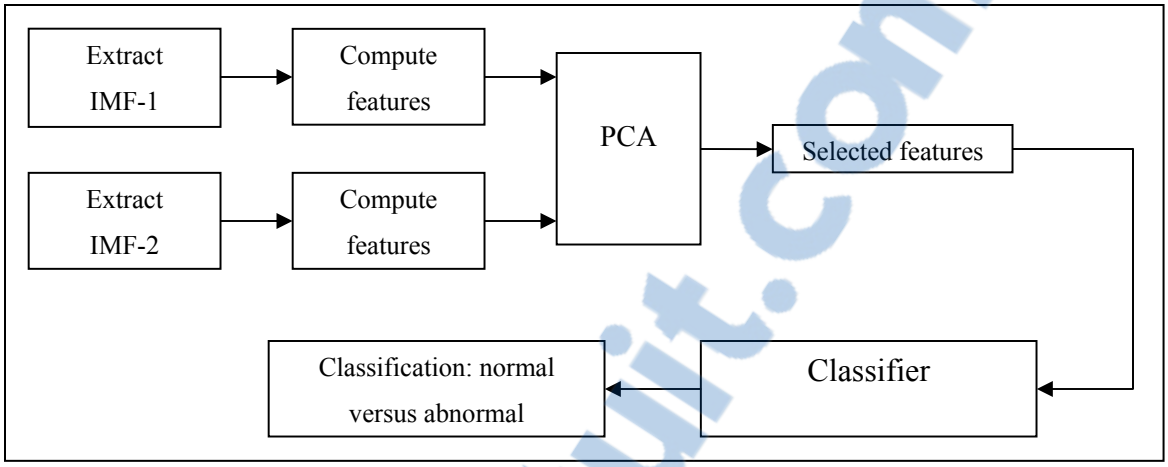

Figure 3.4 The EMD-PCA approach (joint PCA)

3.3 Details of the proposed approach

In this section, we expose in details our methodology; including image pre-processing, the mother wavelet, the empirical mode decomposition algorithm, features extraction, the classifiers, and measures of performance.

3.3.1 Image processing

The pre-processing of retina images is an important step in automated retinal pathology diagnosis; including contrast enhancement, foreground/background differentiation, and image de-noising. For instance, Foracchia et al., (2005) employed compensation based technique for eliminating the luminosity and contrast variations in retinal images. The compensation method is based on the estimation of the luminosity and contrast variability in the background part of the image. Then, it is used for the normalization of the whole image. Feng et al., (2007) employed the contourlet transform for edge enhancement in low contrast images.

Miri and Mahloojifar (2009) used the inverse of the second generation of Curvelet transform to obtain enhanced retina image. In comparison with standard contrast enhancement techniques, the results showed the improvement in visualization, contrast enhancement and performance. Anzalone et al., (2008) used the adaptive histogram equalization technique to minimize the non-uniform lighting. As in the paper of Anzalone et al., (2008), the adaptive histogram equalization technique is employed to pre-process retina images to minimize the non-uniform lighting. Indeed, our choice of the adaptive histogram equalization technique is based on its simplicity and a very low time processing needed to perform it. Moreover, this technique does not require a tuning parameters step.

The adaptive histogram equalization technique takes into account local image information, and does not cause contrast losses in small regions. This method adapts each pixel to its neighboring region. Thus, the high contrast can be obtained for all locations in the image. The three major steps for image enhancement using adaptive histogram equalization technique (Pizer et al.,1987; Zimmerman et al., 1988; Zhiming and Jianhua, 2006) are given as follows:

a) the image is partitioned *N* by *N* sub-regions,
b) apply equalization based on the local region surrounding each pixel,
c) each pixel is mapped to intensity proportional to its rank within the surrounding neighbourhood.

3.3.2 The mother wavelet

The Daubechies wavelet transform has better frequency properties (Daubechies, 1988). Indeed, the Daubechies wavelet gives remarkable results in image analysis and synthesis because it has a compact support interval and continuous derivatives.

In addition, they are real and continuous in nature and have least root-mean-square (RMS) error compared to other wavelets (Gonzalo and De La Cruz, 2004; Ma and Liu, 2005).

Moreover, the Daubechies wavelet exhibits a good trade-off between parsimony and information richness. Therefore, it has been largely used in signal and image processing with applications to biomedical signal analysis (Ayres and Rangayyan, 2005; Quellec et al., 2008; Buciu and Gacsadi, 2009). As a result, the Daubechies wavelet is considered in our study. In particular, the Daubechies wavelet of order four is adopted as the mother wavelet. For instance, the degree of the Daubechies wavelet smoothness increases with the order. Figure 3.5 exhibits the Daubechies wavelet of order 4. As mentioned in the previous section, the level of decomposition is set to three and DWT features are extracted from three sub-bands (LH, HL, HH) as in Quellec et al., (2008) study.

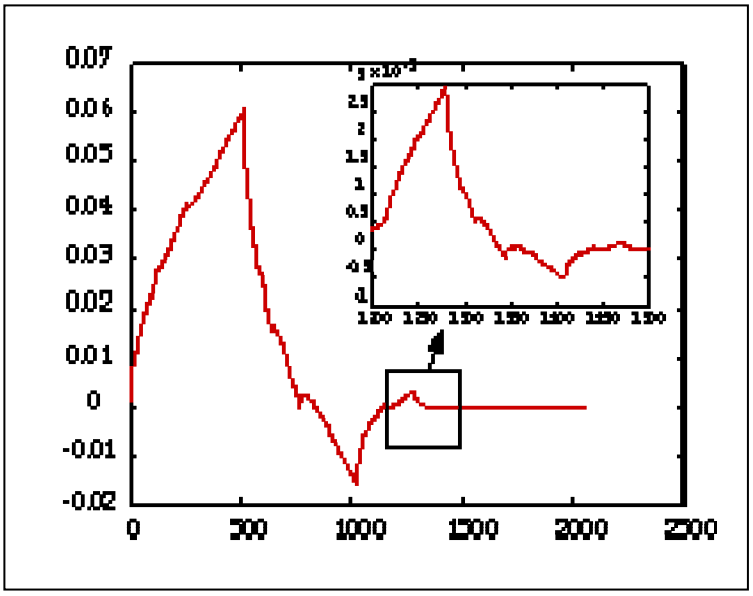

Figure 3.5 Zooming in a Daubechies 4
as in Amara's Wavelet Page (see references)

3.3.3 The empirical mode decomposition algorithm

The empirical mode decomposition (EMD) is simple and generally gives good time-frequency decomposition of the signal (Rato et al., 2008). However, the empirical mode decomposition technique is mainly a computational algorithm; thus it does not have an analytical formulation. As a result, for a given signal different EMD implementations lead to different intrinsic mode functions. But, the main problems with the EMD concern the extrema locations, extrema interpolation, end effects, sifting stopping criterion, and intrinsic mode function (IMF) removal (Rato et al., 2008). They proposed an algorithm to deal with these problems; in particular the goal was to determine correctly the extrema, to efficiently interpolate extremities, and to minimize computational errors. In our study, the EMD algorithm proposed by Rato et al., (2008) is adopted. This algorithm is designed for one dimensional signal. To analyze a digital retina image which is a two dimensional signal with Rato et al., (2008), the image is transformed to a one dimensional signal. For instance, let an image I be a two-dimensional $n \times n$ array of pixels. The corresponding image I_{new} is viewed as a vector with n^2 coordinates that results from a concatenation of successive rows of the image. Then, the alforithm of Rato et al., (2008) is applied to I_{new}. The improvements of the EMD that were proposed by Rato et al., (2008) are described in the following sub-sections.

3.3.3.1 Extrema location

In calculus, extrema definition is based on a continuous neighbourhood; thus with a discrete-time signal the location of an extremum may be obtained with error. Therefore, Rato et al., (2008) defined a classificatory function based on the relation of the actual signal and its left and right neighbours to assert whether it is an extremum or not. For instance, given a sample $x[n]$, they defined the classificatory function as follows:

If $((x[n] > x[n-1])$ AND $(x[n] \geq x[n+1]))$

OR $((x[n] \geq x[n-1])$ AND $(x[n] > x[n+1])$ Then Return (‘‘Is a Maximum’’)

Else if (($x[n] < x[n-1]$) AND ($x[n] \leq x[n+1]$))

OR (($x[n] \leq x[n-1]$) AND ($x[n] \leq x[n+1]$)) Then Return (‘‘Is a Minimum’’)

Else Return (‘‘Not an extrema’’);

3.3.3.2 Extrema interpolation

According to Rato et al., (2008), the sinc function is widely used as the interpolation kernel to interpolate limited signal; however it is computationally demanding and needs a lot of samples on the right and on the left. Therefore, they used parabolic interpolation to better locate the extremas. For instance, ‘’…parabolic interpolation comes into sight as a practical compromise between no interpolations at all and sinc based interpolations…’’ (Rato et al., 2008). The interpolating parabola function is defined as follows:

$$f(x) = a x^2 + b x + c \tag{3.1}$$

The sign of the parameter a determines the kind of extremum. For instance, if $a > 0$, a minimum is obtained; if $a < 0$, a maximum is obtained. If $a = 0$; the extremum is not found.

3.3.3.3 End effects

The end effect is about the decision to make regarding the first and last samples. According to Rato et al., (2008), first and last samples can be considered as (a) maxima and minima simultaneously, (b) as maxima or minima according to the nearest extremum, or (c) leave them free. The final decomposition result depends on the decisions (a), (b), and (c). In their paper, they adopted decision (a) and extrapolated the maxima and the minima. For instance, Rato et al., (2008) proposed the following algorithm to perform the extrapolation of the extrema:

a) find the first maximum, M_1, and minimum, m_1, and their time locations, T_1 and t_1. Assume, for example, that $T_1 > t_1$,

b) insert a new maximum, $M_0 = M_1$, located at $T_0 = -t_1$, and a new minimum, $m_0 = m_1$, located at $t_0 = -T_1$.

Rato et al., (2008) performed a similar extrapolation algorithm to the end of the signal.

3.3.3.4 Mean envelop removal

In order to avoid the problem of adding a non-existing component that can distort the actual IMF and at least will appear in one of the following IMF, Rato et al., (2008) introduced a step size α in the step (d) of the empirical mode decomposition algorithm (See section 1.2) in order to make the algorithm more reliable; but more time consuming. The modified step (d) is given by:

$$d(t) := s(t) - \alpha e(t) \quad where \quad 0 < \alpha < 1 \tag{3.2}$$

The value of the parameter α is chosen such that the energy of the resulting signal is minimized. For instance, the value of α is set equal to the correlation coefficient between $x(t)$ and $e(t)$.

3.3.3.5 Stopping criterion

In order to stop sifting procedure, Rato et al., (2008) defined a resolution factor by the ratio between the energy of the signal at the beginning of the sifting, $x(t)$, and the energy of average of the envelopes, $e(t)$. For instance, the resolution factor is set as follows:

$$\frac{Energy[x(t)]}{Energy[e(t)]} \leq resolution \tag{3.3}$$

As a result, the intrinsic mode functions computation stops when the ratio is higher than the predetermined resolution. It is interesting to notice that the number of obtained intrinsic mode functions increases with the resolution factor. In our study, the resolution was arbitrarily set to 50 as proposed by Rato et al., (2008).

3.3.4 Features extraction

The statistical parameters used to describe the processed images are the mean, standard deviation (std.dev), smoothness, third (3th) moment, uniformity, and entropy. They are chosen since they are widely used in pattern recognition and biomedical image processing (Sheshadri and Kandaswamy, 2005). The statistics are expressed as follows:

$$Mean = \sum_{i=0}^{L-1} z_i p(z_i) \tag{3.4}$$

$$std.dev = \sqrt{\mu_2(z)} = \sqrt{\delta^2} \tag{3.5}$$

$$Smoothness = 1 - \frac{1}{(1+\delta^2)} \tag{3.6}$$

$$3th.Moment = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i) \tag{3.7}$$

$$Uniformity = \sum_{i=0}^{L-1} p^2(z_i) \tag{3.8}$$

$$Entropy = -\sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i) \tag{3.9}$$

where z is a random variable indicating intensity, p is the probability density of the ith pixel in the histogram, and L is the total number of intensity levels. Finally, these characteristics will be fed to the classifiers presented in the coming section.

3.3.5 The classifiers

In order to check the stability of the results over different classifiers, four widely employed classifiers in the area of biomedical image classification are considered. They are support vector machines (SVM), quadratic discriminant analysis (QDA), *k*-nearest neighbor algorithm (*k*-NN), and probabilistic neural networks (PNN). They are presented in the following sub-sections.

3.3.5.1 Support vector machines

Support Vector Machines (SVM) is a supervised statistical learning technique introduced by Vapnik (1995). The support vector machine was developed from the theory of Structural Risk Minimization and was originally formulated for binary classification. The SVM seeks to implement an optimal marginal classifier that minimizes the structural risk in two steps. First, SVM transform the input to a higher dimensional space with a kernel (mapping) function. Second, SVM linearly combine them with a weight vector to obtain the output. As result, SVM provide very interesting advantages. The SVM avoid local minima in the optimization process. In addition, they offer scalability and generalization capabilities. For instance, to solve a binary classification problem in which the output $y \in \{-1,+1\}$ SVM seek for a hyper-plane $w.\Phi(x)+b = 0$ to separate the data from classes +1 and −1 with a maximal margin. Here, x denotes the input feature vector, w is a weight vector, Φ is the mapping function to a higher dimension, and b is the bias used for classification of samples. The maximization of the margin is equivalent to minimizing the norm of w (Cristianini and Taylor, 2000).
Following the notation of Lorena and Carvalho (2009), the value of w and b are found by solving the following optimization problem:

$$Minimize: \quad \|w\|^2 + C\sum_{i=1}^{n} \xi_i \tag{3.10}$$

Such that,

$$y_i\left(w\cdot\Phi\left(x_i\right)+b\right)\geq 1-\xi_i \quad \xi_i\geq 0 \quad i=1,\ldots,n \tag{3.11}$$

where C is a user-defined strictly positive parameter that determines the tradeoff between the maximum margin and the minimum classification error. For instance, a larger C means a higher penalty is assigned to empirical errors.

The value of the parameter C is set to its default value in Matlab ©; for example C = 1. The number n is the total number of samples, and generalization and ξ is the error magnitude of the classification. In other words, ξ is a measure of misclassification errors. The conditions above ensure that no training example should be within the margins. The number of training errors and examples within the margins is controlled by the minimization of the term:

$$\sum_{i=1}^{n}\xi_i \tag{3.12}$$

The solution to the previous minimization problem gives the decision frontier:

$$f\left(x\right)=\sum_{x_i} y_i\alpha_i\Phi\left(x_i\right)\cdot\Phi\left(x\right)+b \tag{3.13}$$

where each α_i is a Lagrange coefficient. As mentioned before, the role of the kernel function is to implicitly map the input vector into a high-dimensional feature space to achieve better separability. Finally, the classification is obtained as follows:

$$\text{If } f(x) > 0 \text{ then the predicted class } (y) \text{ is } +1, \tag{3.14}$$

$$\text{If } f(x) < 0 \text{ then the predicted class } (y) \text{ is } -1, \tag{3.15}$$

In this study the polynomial kernel is used since it is a global kernel. For instance, global kernels allow data points that are far away from each other to have an influence on the kernel values as well (Sun et al., 2009):

$$K(x, x_i) = \Phi(x_i)\Phi(x) = ((x_i \cdot x) + 1)^d \tag{3.16}$$

where the kernel parameter d is the degree of the polynomial to be used. In this study, d is set arbitrarily to 2. Finally, the optimal decision separating function can be obtained as follows:

$$y = sign\left(\sum_{i=1}^{n} y_i \alpha_i K(x_i, x) + b\right) \tag{3.17}$$

Figure 3.6 illustrated a binary linearly separable classification problem. There are two classes denoted by $y = 1$ (white dots) and $y = -1$ (black dots). H1 and H2 are parallel hyper-plane passing through the dots that are closest to the boundary. The distance between H1 and H2 is the margin. As mentioned before, the solution to the minimization problem (minimizing the norm of w) is about finding an optimal decision hyper-plane that maximizes the margin between the two classes $y = 1$ and $y = -1$. In this case, classes can be linearly separated by a linear kernel. For, non-linearly separated data, a non-linear kernel function should be introduced to map the data into higher dimensional space and thus allows their separation. In general, the kernel function can have different forms, including polynomial function and radial basis function and sigmoid function. In our experiments, the SVM with an order two polynomial kernel is chosen as mentioned before. Figure 3.7 shows the optimal hyper-plan using unsupervised SVM linear separation (top) and unsupervised SVM polynomial kernel of order 2 (down) of real retina data we have used in our study using STARE database which is presented in next chapter. In unsupervised learning, the machine (classifier) is given examples (patters and their correspondent outputs/classes). The machine learns the examples and classifies them. On the other hand, in supervised classification the machine learns to produce the correct output given a new input. All experiments in our study (See Methodology and Results) are conducted using supervised scheme.

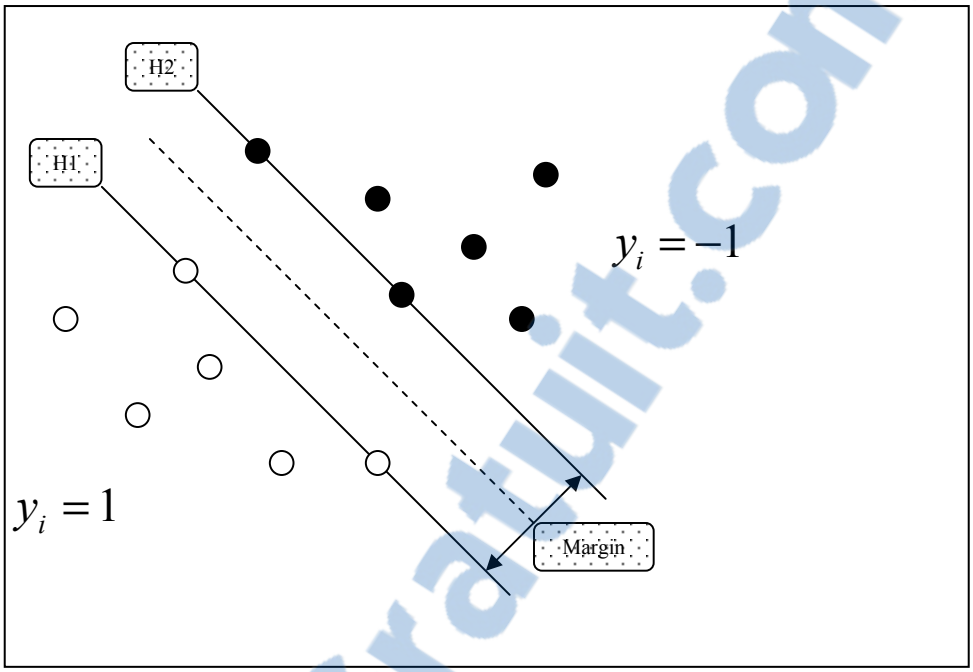

Figure 3.6 The optimal hyper-plan

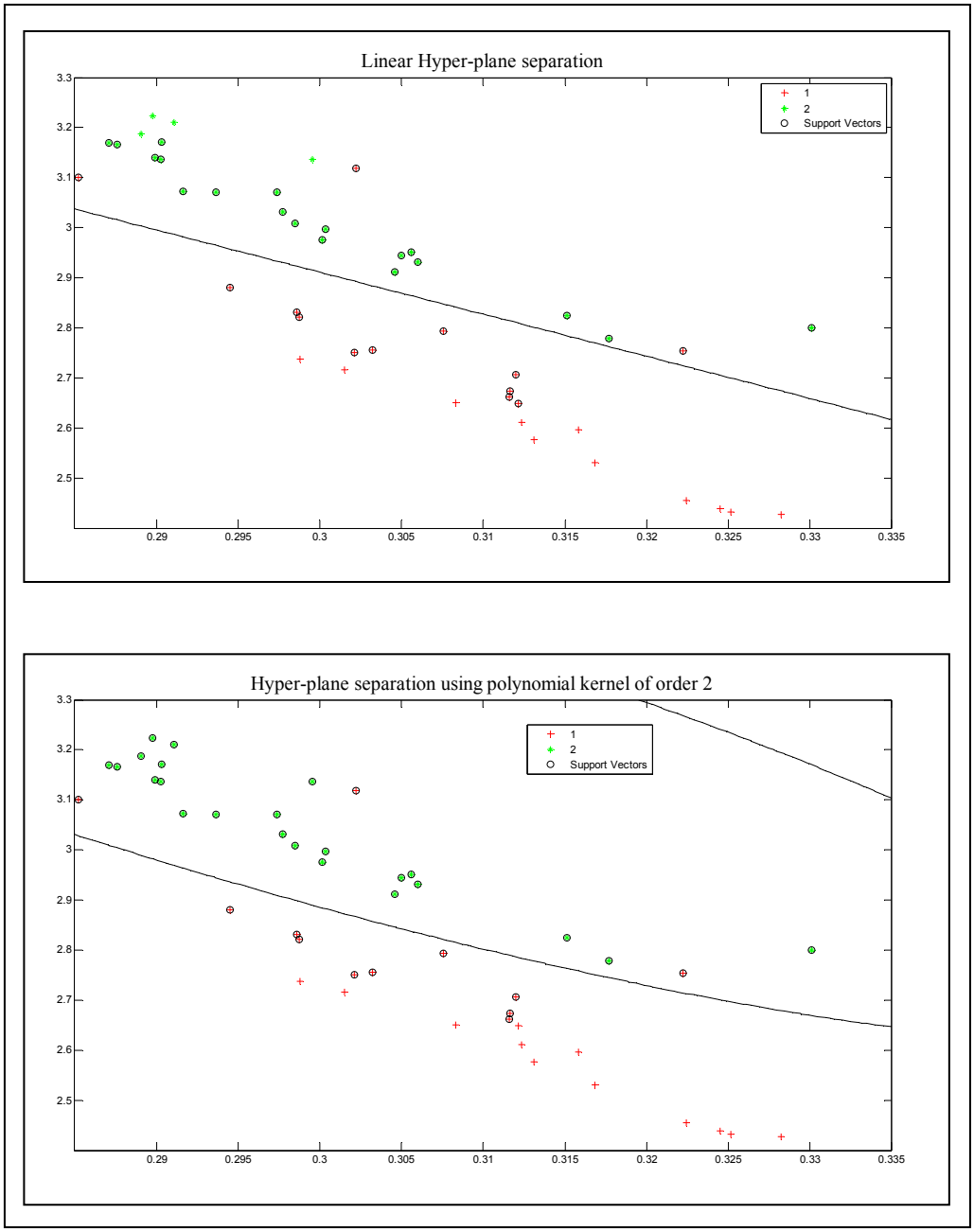

Figure 3.7 The optimal hyper-plan of a real retina data

3.3.5.2 Quadratic discriminant analysis

Discriminant analysis (Douda et al., 2000) is a parametric procedure that finds a set of optimal linear projections by simultaneously maximizing the between- class dissimilarity and minimizing the within-class dissimilarity. In other words, within-class and between-class scatter matrices are used to formulate the criteria of class separability. Unlike the linear discriminant analysis (LDA), the quadratic discriminant analysis (QDA) fits multivariate normal densities with covariance estimates stratified by group. Therefore, QDA is more suitable than LDA when the two classes have very different variance structures. In addition, unlike LDA; the boundary produced by QDA is a quadratic curve that may contain two separate sections of the boundary lines (Hastie et al., 2001). Now, let X be the vector of features which is supposed to be multivariate normally distributed with mean vector and group specific covariance matrix $\sum_i$. Let π_i be the prior probability of class i, $f_i(X)$ is the conditional density of X in class i, $g_i(X)$ is the classification function.

It is supposed that $g_i(X)>g_j(X)$ for $i \neq j$. Then, the conditional density of X and the classification function are computed as follows:

$$f_i(x) = (2\pi)^{-p/2} |\Sigma_i|^{-0.5} \exp\left\{-0.5(x-\mu_i)' \Sigma_i^{-1}(x-\mu_i)\right\} \tag{3.18}$$

$$g_i(x) = -0.5(X-\mu_i)' \sum_i^{-1}(X-\mu_i) - 0.5\log|\Sigma_i| + \log(\pi_i) \tag{3.19}$$

where p=2 is a dimension factor in the case of quadratic discriminant analysis. Figure 3.8 shows the unsupervised classification of real retina data using QDA.

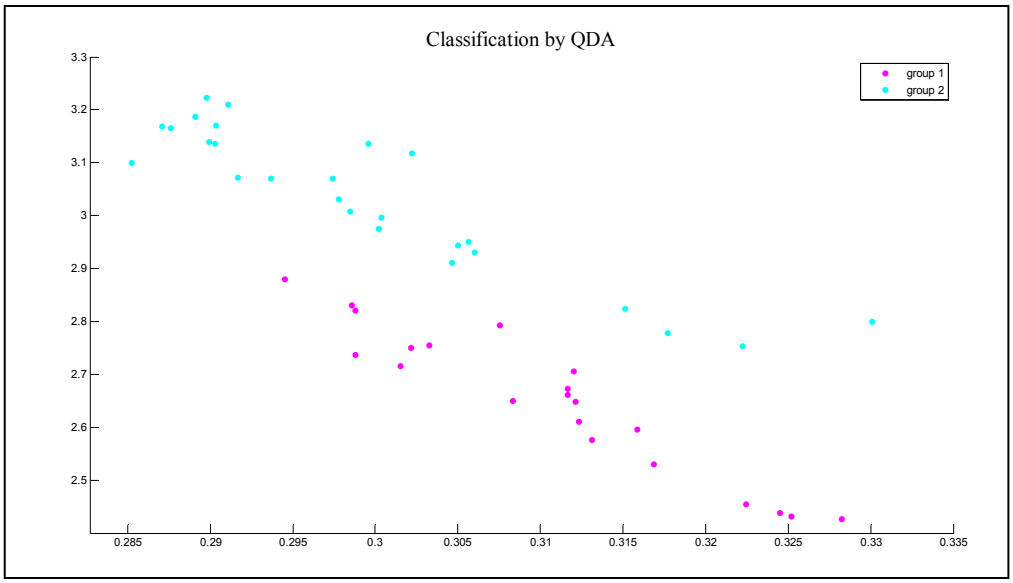

Figure 3.8 Unsupervised classification of normal and abnormal real retina data using QDA

3.3.5.3 The *k*-nearest neighbour algorithm

The *k*-nearest neighbor algorithm (Cover and Hart, 1967) is a nonparametric method that assigns query data to the class that the majority of its *k*-nearest neighbors belong to. For instance, the *k*-NN algorithm uses the data directly for classification without the need of an explicit model. The performance of *k*-NN depends on the number of the nearest neighbor *k*. In general, there is no solution to find the optimal *k*. However, trial and error approach is usually used to find its optimal value. The main objective is to find the value of *k* that maximizes the classification accuracy. The main advantage of *k*-NN algorithm is the ability to explain the classification results. On the other hand, its major drawback is the need to find the optimal *k* and to define the appropriate metric to measure the distance between the query instance and the training samples. In this paper, the distance metric chosen is the Euclidean distance and *k* was arbitrarily set to 1.

Let S be a given set of n annotated images, each image is a training example (x,y), where x is the image feature vector and y the class vector that specifies the class of the image. For instance, a set of training data $\{(x_1,y_1),(x_2,y_2),\ldots,(x_i,y_i),\ldots,(x_n,y_n)\}$ and the test data x, where x is the feature vector, and y_i is the class of data x_i. The distance measure is defined as:

$$D(x,x_i)=\sqrt{\sum_{j=1}^{d}(x_j-x_{ij})^2} \tag{3.20}$$

where d is the dimension of the feature vector. Finally, the nearest neighbor rule is defined as follows:

$$NN(x)=y_k \quad where \quad k=\arg\min_i D(x,x_i) \tag{3.21}$$

Since our classification problem is a binary one (normal versus abnormal images), each of the k nearest neighbors votes equally for its predicted class label, and x is assigned the class label with the majority of votes. The major steps of k-NN procedure are given as follows:

a) calculate Euclidean distances between an unknown features vector (x) and all the feature vectors in the training set,
b) select k feature vectors from the training set most similar to feature vector (x), according to the calculated distances,
c) classify feature vector ($\boldsymbol{x}$) with the group to which a majority of the K feature vectors belongs.

Figure 3.9 shows the unsupervised classification of real retina data using the k nearest neighbors machine.

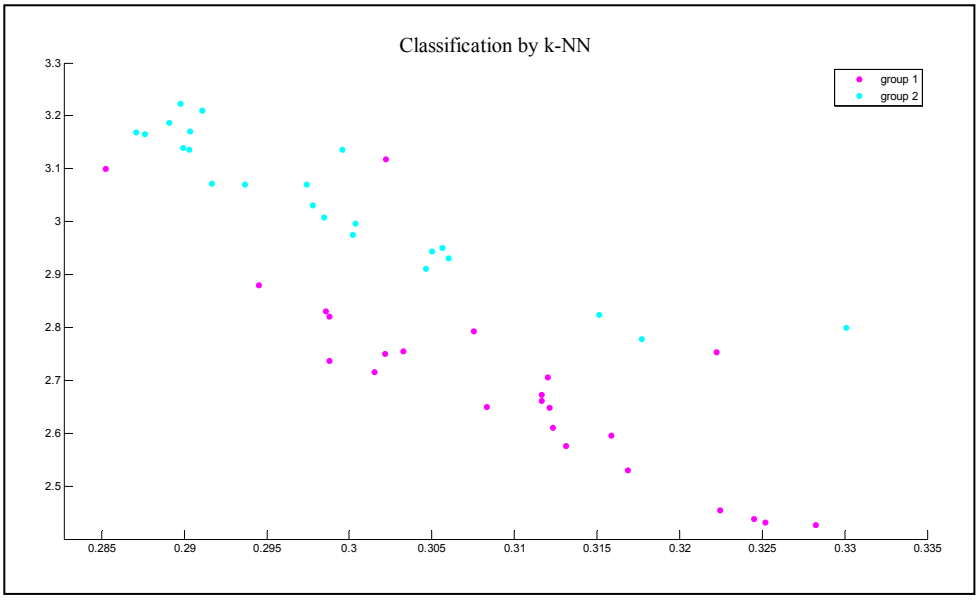

Figure 3.9 Unsupervised classification of normal and abnormal real retina data using *k*-NN

3.3.5.4 The probabilistic neural network

Unlike the standard backpropagation neural networks (Haykin, 2008), the probabilistic neural network (Specht, 1988; Poggio and Girossi, 1990) requires only a single presentation of each pattern. The PNN employs an exponential activation function rather than the sigmoid function that is commonly used in the standard backpropagation neural networks. Then, a PNN can identify nonlinear decision boundaries that approach the Bayes optimal (Specht, 1990). The basic network topology consists of four layers. The first layer is the inputs layer. In the second layer, the neurons are divided to A groups and the probability density function (PDF) of each group (a) of patterns is directly estimated from the set of training samples using (Parzen, 1962) window approximation method. The third layer performs the summation of all PDF. Finally, the Bayesian decision is made in the fourth layer. The PDF is assumed to follow a Gaussian distribution. Then, the PDF for a feature vector X to be of a certain category a, $1 \leq a \leq A$, is given by:

$$f_{a,i}(X) = \frac{1}{(2\pi)^{p/2}\sigma^{p}} \times \frac{1}{m} \times \sum_{i=1}^{m} \exp\left(-\frac{(X - X_{ai})'(X - X_{ai})}{2\sigma^{2}} \right) \tag{3.22}$$

Where, p is the number of patterns in X, m is the number of the training patterns of category a, i is the pattern number, and σ is the smoothing factor of the Gaussian curves used to construct the PDF. The value of σ is optimized during training based on the clearest separation of classes with the highest classification rate (Chen et al., 2007; Wang and Wen, 2008; Sert and Kalenderli, 2009). In the summation layer, the network computes the approximation of the conditional class probability functions by combining the previously computed PDFs as follows:

$$S_a(X) = \sum_{i=1}^{n_k} w_{ki} f_{a,i}(X) \tag{3.23}$$

where n_k is the number of pattern neurons of category k, $w_{k,i}$ are coefficients that satisfy both conditions:

$$\sum_{i=1}^{n_k} w_{k,i} = 1, \text{ and } w_{k,i} \geq 0, \forall k, \forall i \tag{3.24}$$

Finally, the feature vector X is associated to the class that corresponds to the summation layer with the maximum output according to:

$$C(X) = \arg\max_{1 \leq a \leq A} (S_a(X)) \tag{3.25}$$

The topology of the probabilistic neural networks is shown in Figure 3.10. Figure 3.11 shows the unsupervised classification of real retina data using the probabilistic neural networks.

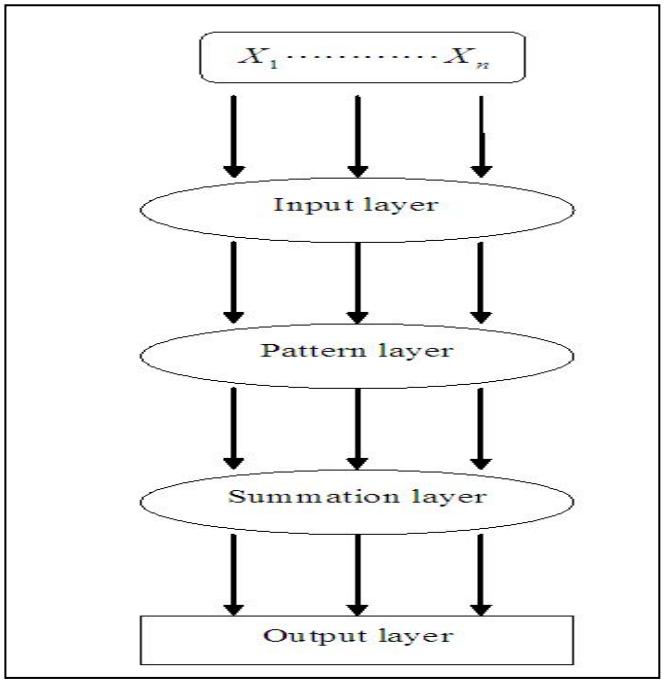

Figure 3.10 Topology of the PNN

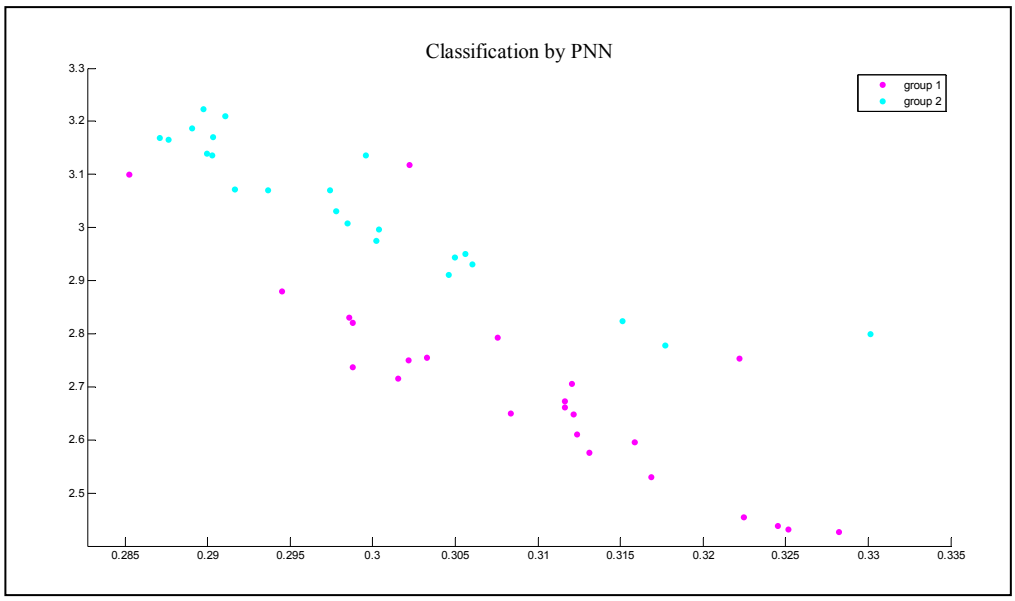

Figure 3.11 Unsupervised classification of normal and abnormal real retina data using PNN

3.3.6 The principal component analysis

Principal component analysis (Jolliffe, 2002) is one of the most popular linear projection statistical techniques for dimensionality reduction. It is also widely used for features selection in classification and detection problems (Malhi and Gao, 2004; Le and Satoh, 2005; Park and Choi, 2009). PCA produces a lower dimensional representation of the original data such that the correlation structure between the variables is preserved. Therefore, most of the information contained in the original data set is stored in the reduced data set. In other words, the original data set is reduced to a new small dimension set with low information loss. In sum, PCA performs a linear projection of the original data by finding linear combinations of the original variables which are uncorrelated and have maximum variance. The projected vectors capture the most expressive features of the original data. The transformation from the original set of features to principal components is carried out through the following process.

The covariance matrix of the original features set is computed. Then, its eigenvectors are extracted to represent the principal components. For instance, the eigenvectors specify a

linear mapping from the original space of to the new space of that contains uncorrelated features.

As a result, the principal components of the original features set are uncorrelated in the projected space. The obtained eigenvectors are ranked according to the level of variation in the original data that they account for. In particular, the first few eigenvectors account for most of the variation in the data set and; therefore; are kept. In particular, principal components which correspond to smaller eigenvalues are deleted based on a predetermined percentage of total cumulative variation.

Technically, the algorithm of principal component is described in our study as in Zhao and Lu (2007). Let M be an $N \times N$ matrix of real numbers. The first step of PCA is to calculate the covariance matrix V of M, which is defined as:

$$V = \sum_{i=1}^{N} \left(M_i - \overline{M}_R\right)^t \left(M_i - \overline{M}_R\right) \tag{3.26}$$

where M_i is the ith row vector in M, t denotes transpose operator, and $\overline{M}_i$ is a $1 \times N$ average vector of the row vectors in M given by:

$$\overline{M}_R = \frac{1}{N} \sum_{i=1}^{N} M_i \tag{3.27}$$

Then, eigenvalue decomposition is applied to V:

$$V = ULU^{-1} \tag{3.28}$$

where U^{-1} denotes the inverse matrix of U. L is a diagonal matrix with eigenvalues of V as its diagonal elements $\lambda_1, \lambda_2, \ldots, \lambda_N$, and the columns of U, $u_1, u_2, \ldots, u_N$, are the eigenvectors of V. Then, based on the theory of linear algebra; most of the information of V is in the larger

eigenvalues and corresponding eigenvectors (Zhao and Lu, 2007). For instance, the diagonal elements in L are sorted in descending order $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ (Zhao and Lu, 2007).

Then, the ratio of each of the eigen value (λ) to the total sum of all the eigen values indicates the proportion of variation explained by the corresponding principal component (u). As a result, the first principal component explains the maximum variance; the second principal component explains the second largest variance, etc. Therefore, using the eigenvectors corresponding to the largest eigenvalues would give the smallest error in representation (Malhi and Gao, 2004). In other words, only a subset of eigenvectors (principal vectors) is used as basis vectors of a feature space S. The coordination, Z_i, of the original data in the feature space S is obtained as follows:

$$Z_i = \left(M_i - \overline{M}_R\right) \cdot \left[u_1 u_2 \ldots u_m\right], \quad i = 1,2,\ldots,N \tag{3.29}$$

where Z_i are; in fact; principal components that can distinguish the original data. Finally, the first components that maximize the cumulative variance proportion (account for most of the variation in the data set) up to 80% are selected to be fed to the classifiers.

3.3.7 Performance measures

The performance of each classifier is measured using the following statistics:

$$correct\,classification\ \ rate = \frac{correctly\ \ classified\ \ samples}{classified\ \ samples} \tag{3.30}$$

$$sensitivity = \frac{correctly\ \ classified\ \ positive\ \ samples}{true\ \ classified\ \ samples} \tag{3.31}$$

$$specificity = \frac{correctly\ \ classified\ \ negative\ \ samples}{true\ \ negative\ samples} \tag{3.32}$$

CHAPTER 4

DATA AND RESULTS

In this chapter, the database is presented along with examples of retina images before and after pre-processing in the first section. The experimental results obtained with discrete wavelet transform, empirical mode decomposition, and principal analysis features selection are all given in the second section.

4.1 Database

A set of 133 color retina images from STARE (se references) database were employed to test EMD-based processing approach and DWT-based processing approach. The dataset includes 23 normal images, 20 images with drusens, 24 with microaneurysms (MA), and 22 with circinate, 21 artery sheath, and 23 small or medium blot hemmorhage. The mother wavelet used in the study is the Daubechies wavelet of order 4 at three level decomposition as in Quellec et al., (2008).

In each case, the classifiers (SVM, QDA, *k*-NN, PNN) were trained and test based on their ability to classify normal retina image and abnormal one. It is a one against one classification problem. All simulations were performed with ten folds cross-validations. For instance, the image data was split into ten random partitions of equal size; then, during each fold, one partition was kept out from training, and the correct classification rate (hit ratio), sensitivity and specificity were computed. Finally, then average and standard deviation of correct classification rate, sensitivity and specificity are computed. The Matlab (see references) Wavelet Toolbox © is used for wavelet analysis. The Matlab Bioinformatics Toolbox © is employed to train and test SVM, QDA, and *k*-NN. Besides, the Matlab Neural Networks Toolbox © is used to perform PNN simulations.

Finally, we used the EMD Matlab codes of Rato et al. (2008) to process retina digital images. Figure 4.1 exhibits examples of retina digital images used in the study. Grey scale images before adaptive histogram equalization are shown in Figure 4.2, grey scale images after adaptive histogram equalization are given in Figure 4.3, DWT sub-band images are shown in Figure 4.4, and first, second and third IMF obtained from EMD are given in Figure 4.5.

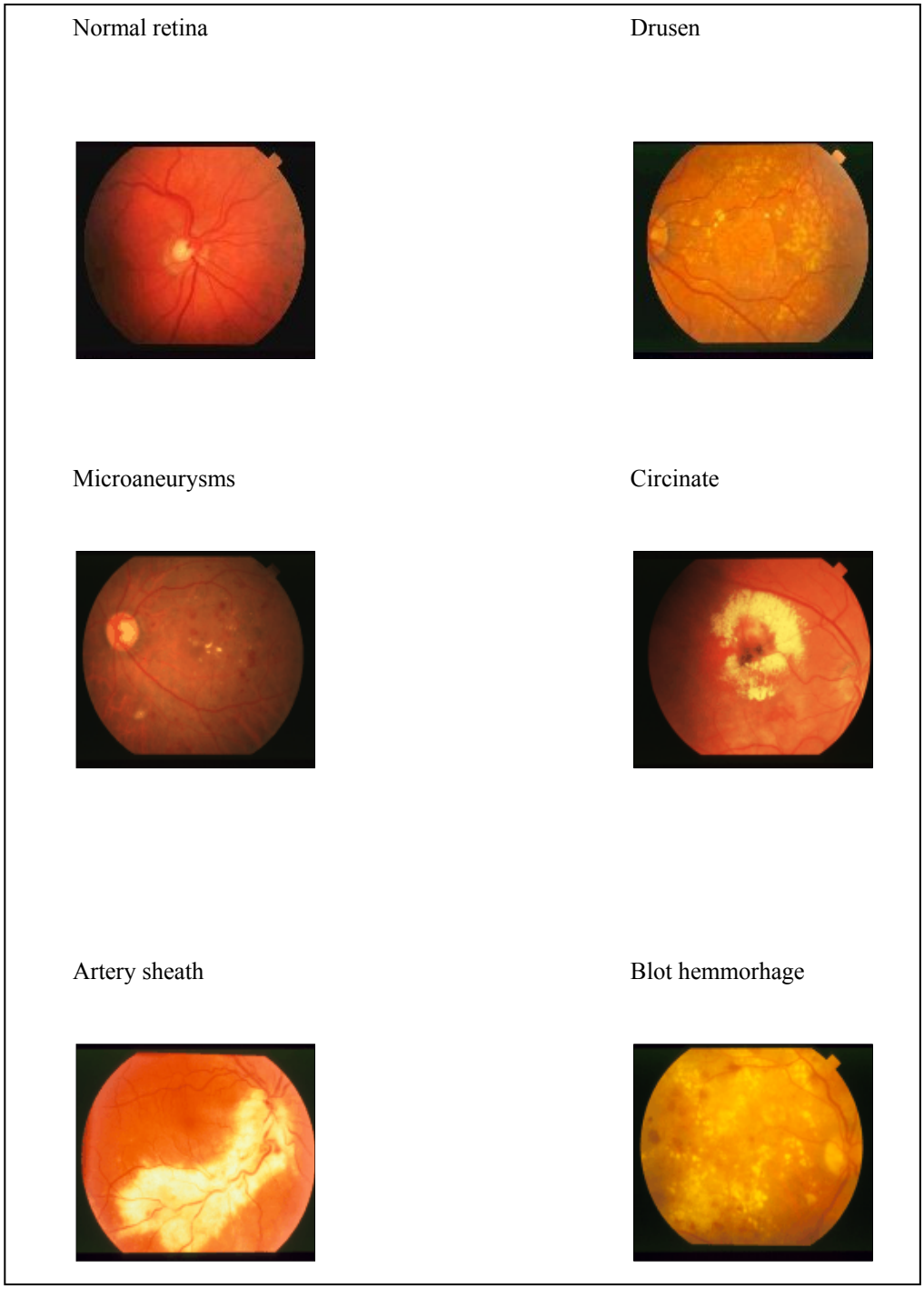

Figure 4.1 Examples of true color retina images

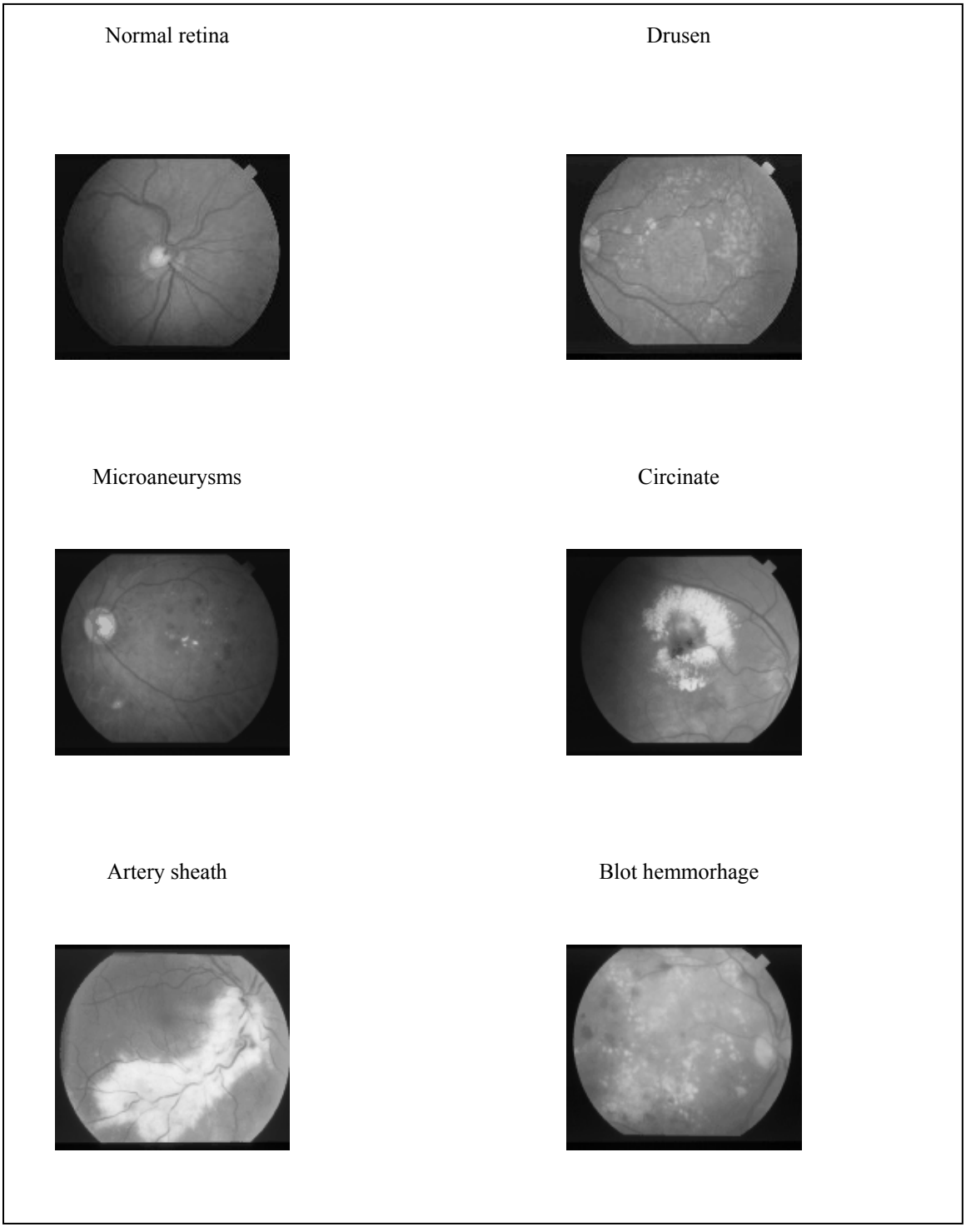

Figure 4.2 gray scale images before applying adaptive histogram Equalization

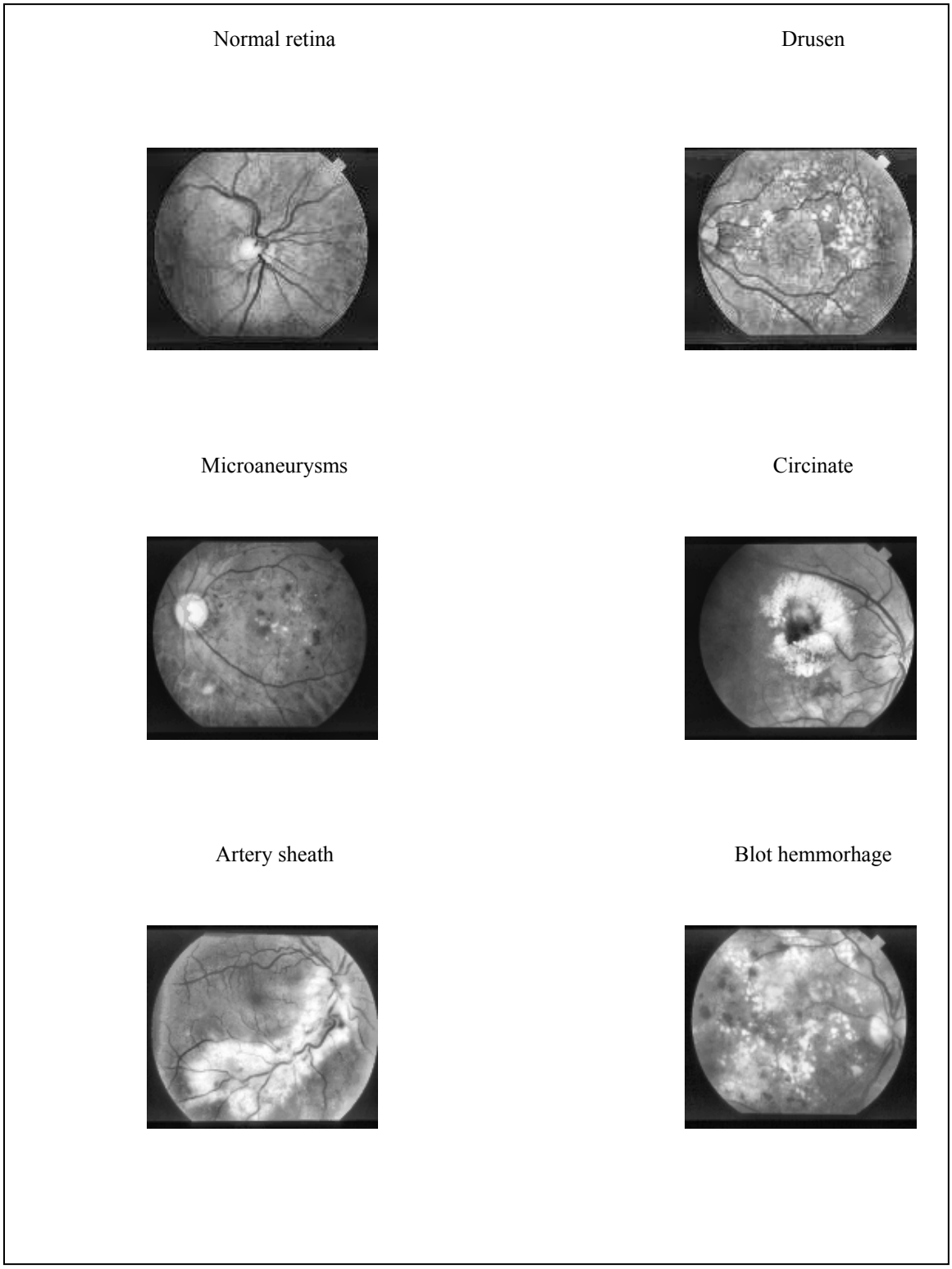

Figure 4.3 gray scale images after applying adaptive histogram equalization

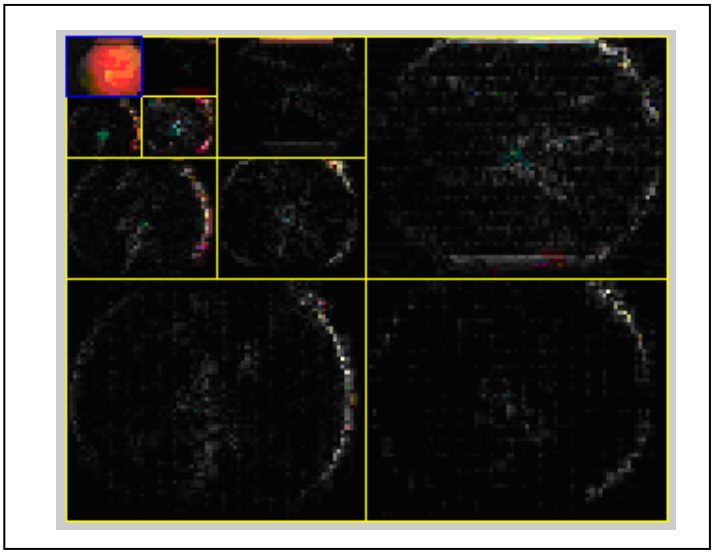

Figure 4.4 DWT sub-images of a normal retina

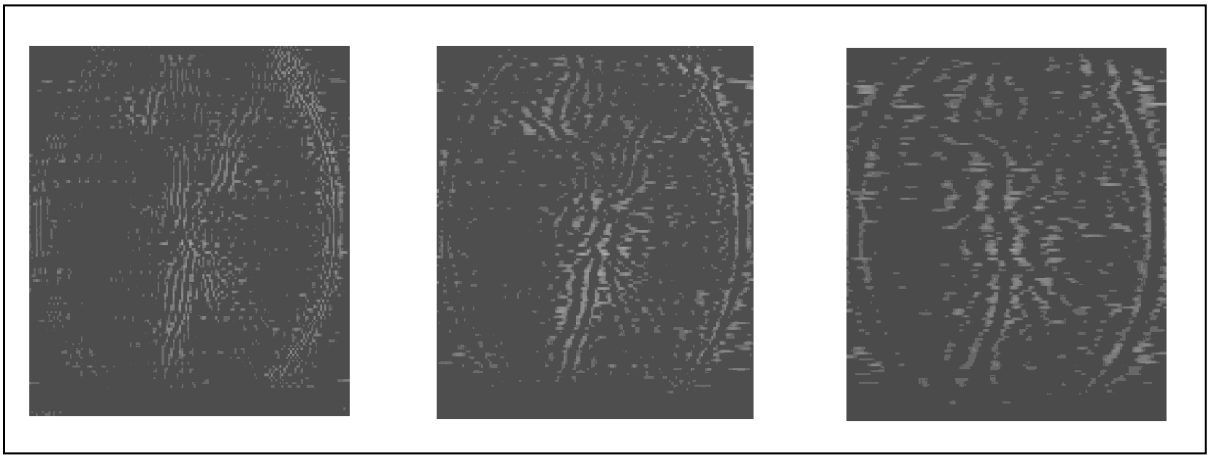

Figure 4.5 First (left), second (middle), and third IMF (right) obtain from EMD

4.2 Experimental results

Tests have been conducted with discrete wavelet transform and empirical mode decomposition for the support vector machines (SVM), quadratic discriminant analysis (QDA), k-nearest neighbor (*k*-NN), and probabilistic neural network (PNN) classifiers. These experiments were performed with and without principal component analysis selected features. The obtained results are given hereafter.

4.2.1 Discrete wavelet transform simulation results

As show in Figure 4.6, the average recognition (classification or detection) rates obtained with DWT-based features are 71.95% with SVM, 63.32% with QDA, 62.07% with *k*-NN, and 54.15% with PNN. Therefore, the simulations demonstrate the high effectiveness of SVM over the other classifiers.

The obtained average sensitivity are 99.08%, 70%, 67.58%, and 63.1% using PNN, SVM, QDA, and *k*-NN respectively (Figure 4.7). Therefore, the probabilistic neural network is much capable to correctly detect pathologies than the other classifiers with DWT-based features.

The overall average specificity with extracted features from DWT processed images are 74.28% using SVM, 60.73% using *k*-NN, 57.65% using QDA, and 1.57% using PNN (Figure 4.8). As a result, the SVM performs the best in the recognition of normal retina images with DWT-based features.

Tables 4.1 to 4.4 provide detailed results obtained with SVM, QDA, *k*-NN, and PNN.
Based on DWT features, the SVM is highly accurate in the detection of MA (92.9%) and circinate (89.91%). However, it fails to detect drusen (49.36%).

The QDA is highly accurate in the detection of MA (89.41%) and circinate (80.78%); and fails to detect drusen (32.45%) and blot (52.60%).

The *k*-NN is highly accurate in the detection of MA (89.41%) and circinate (80.78%); and fails to detect drusen (32.45%) and blot (52.60%).

Finally, the performance of PNN is very low in comparison with previous classifiers.
The best obtained correct classification rate is 60.24% (detection of artery) and the lowest classification arte is 48.72% (detection of blot) blot.

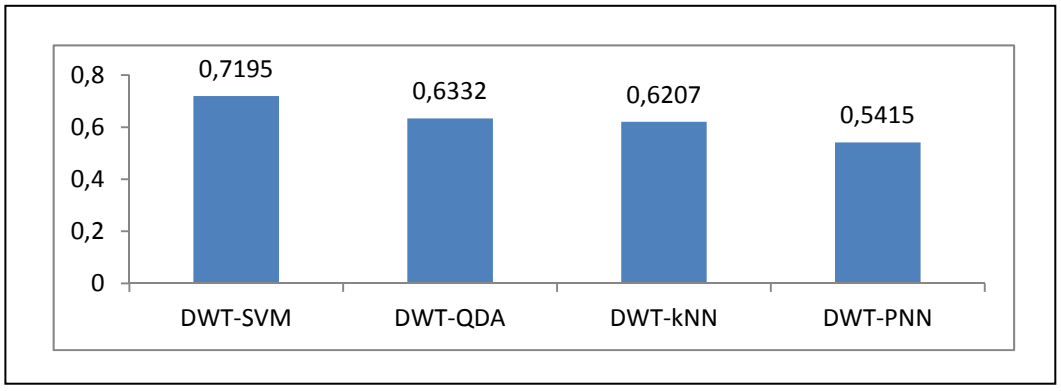

Figure 4.6 Average correct classification rate with DWT features

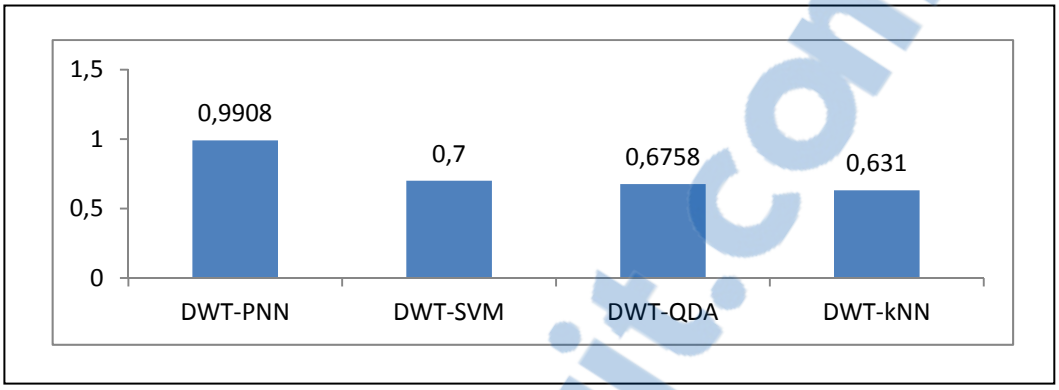

Figure 4.7 Average sensitivity with DWT features

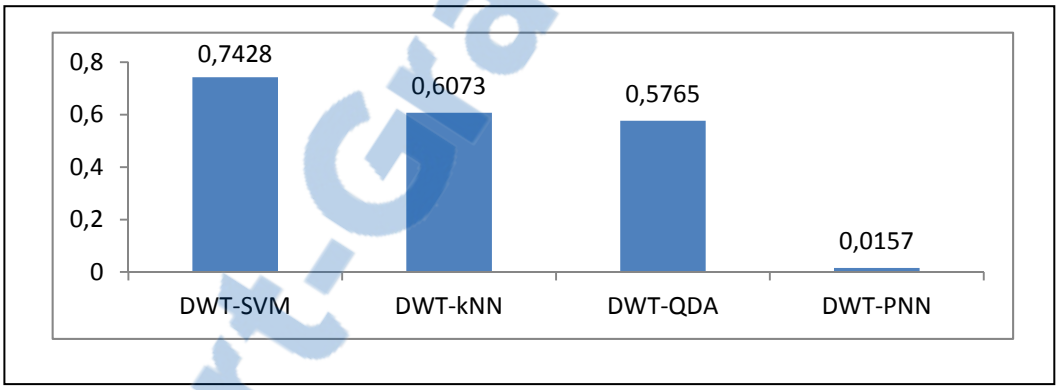

Figure 4.8 Average specificity with DWT features

Table 4.1 SVM classification results using DWT features

	Artery	**Artery**	**Blot**	**Blot**	**Circinate**	**Circinate**	**Drusen**	**Drusen**	**MA**	**MA**
	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.
C.C.R	**0.6718**	*0.0667*	**0.6041**	*0.1093*	**0.8991**	*0.0385*	**0.4936**	*0.1106*	**0.9290**	*0.0156*
Sensitivity	0.8769	*0.0675*	0.5244	*0.1796*	0.8460	*0.0724*	0.3748	*0.0674*	0.8781	*0.0234*
Specificity	0.4539	*0.1618*	0.6915	*0.1279*	0.9691	*0.0342*	0.6174	*0.2193*	0.9821	*0.0328*

Table 4.2 QDA classification results using DWT features

	Artery	**Artery**	**Blot**	**Blot**	**Circinate**	**Circinate**	**Drusen**	**Drusen**	**MA**	**MA**
	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.
C.C.R	**0.6135**	*0.0315*	**0.5260**	*0.0752*	**0.8078**	*0.0628*	**0.3245**	*0.1267*	**0.8941**	*0.0406*
Sensitivity	0.5162	*0.2032*	0.6641	*0.0595*	0.8440	*0.0636*	0.3874	*0.1526*	0.9673	*0.0293*
Specificity	0.6910	*0.1289*	0.3853	*0.1400*	0.7570	*0.1521*	0.2526	*0.1079*	0.7965	*0.1307*

Table 4.3 k-NN classification results using DWT features

	Artery	**Artery**	**Blot**	**Blot**	**Circinate**	**Circinate**	**Drusen**	**Drusen**	**MA**	**MA**
	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.
C.C.R	**0.5302**	*0.1237*	**0.4328**	*0.0693*	**0.9013**	*0.0465*	**0.4036**	*0.1614*	**0.8354**	*0.0502*
Sensitivity	0.4928	*0.2212*	0.4479	*0.0772*	0.8771	*0.0638*	0.4493	*0.1873*	0.8879	*0.0763*
Specificity	0.5619	*0.0636*	0.4235	*0.1305*	0.9333	*0.0498*	0.3513	*0.1393*	0.7666	*0.1130*

Table 4.4 PNN classification results using DWT features

	Artery	**Artery**	**Blot**	**Blot**	**Circinate**	**Circinate**	**Drusen**	**Drusen**	**MA**	**MA**
	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.
C.C.R	**0.6024**	*0.0139*	**0.4872**	*0.0169*	**0.5511**	*0.0317*	**0.5636**	*0.0239*	**0.5031**	*0.0188*
Sensitivity	1.0000	*0.0000*	0.9540	*0.0375*	1.0000	*0.0000*	1.0000	*0.0000*	1.0000	*0.0000*
Specificity	0.0783	*0.0566*	0.0000	*0.0000*	0.0000	*0.0000*	0.0000	*0.0000*	0.0000	*0.0000*

4.2.2 Empirical mode decomposition simulation results

As shown in Figure 4.9, the average recognition (classification or detection) rate obtained with EMD-based features is 83.34% with SVM, 76.39% with PNN, 75.27% with QDA, and 73.18% with *k*-NN. Therefore, the simulations demonstrate the high effectiveness of SVM over the other classifiers based on EMD features. Detailed results are shown in Tables 4.5 to 4.8 for SVM, QDA, *k*-NN, and PNN respectively.

The obtained average sensitivities are 86.83%, 82.93%, 76.66%, and 74.46% using SVM, PNN, *k*-NN, and QDA respectively (Figure 4.10). Therefore, the support vector machine is on average is much capable to correctly detect pathologies than the other classifiers with EMD-based features. Finally, the overall average specificities (Figure 4.11) given extracted features from DWT processed images are respectively 79.57% with SVM, 76.67% with QDA, 69.74% with *k*-NN, and 69.62% with PNN (Figure 4.8). As a result, the SVM perform the best in the recognition of normal retina images with EMD-based features.

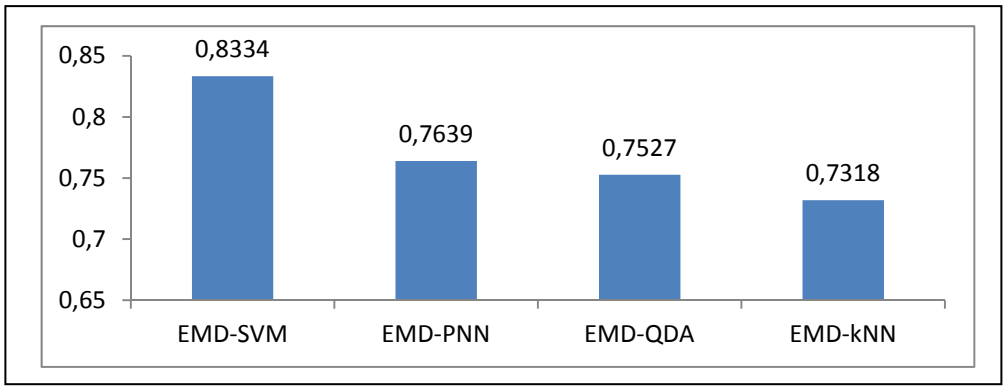

Figure 4.9 Average classification rate with EMD features

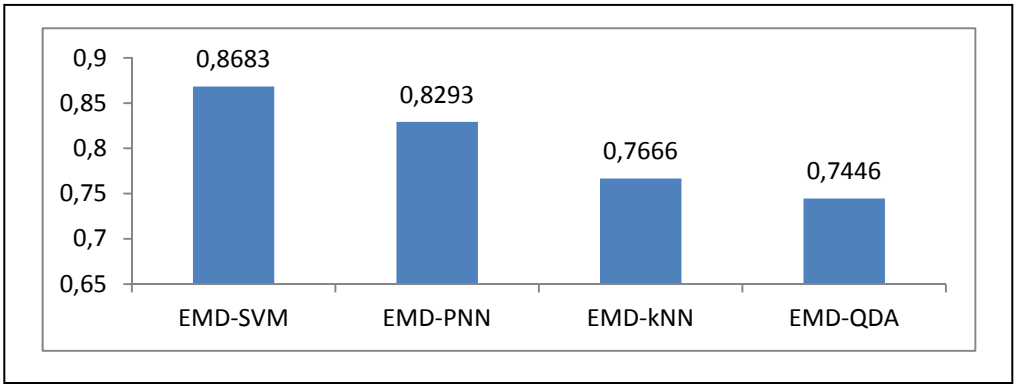

Figure 4.10 Average sensitivity with EMD features

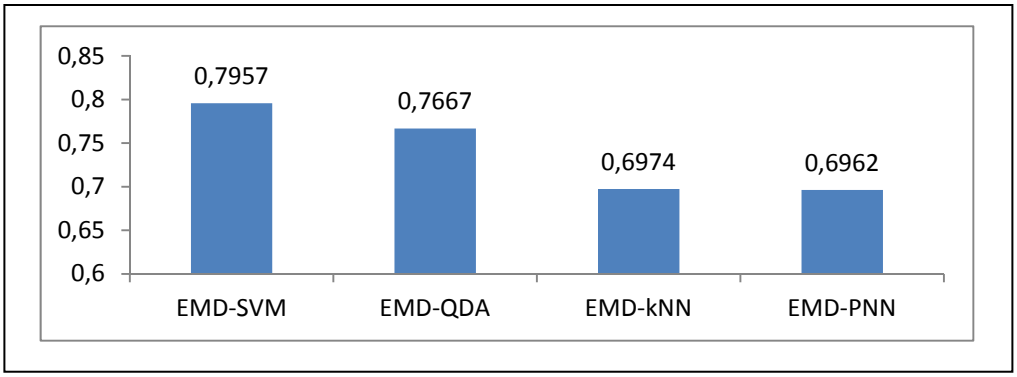

Figure 4.11 Average specificity with EMD features

The EMD-SVM system is highly accurate in the detection of MA (96.11% as recognition rate, 97.98% as sensitivity, and 94.27% as specificity) and circinate (95.11% as recognition rate, 97.31% as sensibility, and 92.56% as specificity). On the other hand, the EMD-SVM system on average detects artery, blot, and drusen with 73.33%, 73.45%, and 78.69% respectively.

The EMD-QDA system is highly accurate in the detection of circinate (93.84% as recognition rate, 92.09% as sensitivity, and 96.10% as specificity) and MA (92.84% as recognition rate, 100% as sensibility, and 85.59% as specificity). On the other hand, the

EMD-QDA system on average detects artery, blot, and drusen with 66.24%, 69.02%, and 54.39% respectively.

The EMD-*k*-NN system is highly accurate in the detection of circinate (95.08% as recognition rate, 100% as sensitivity, and 89.61% as specificity) and MA (91.13% as recognition rate, 92.69% as sensibility, and 89.56% as specificity). On the other hand, the EMD-*k*-NN system on average detects artery, blot, and drusen with 59.01%, 50.61%, and 70.05% respectively.

The EMD-PNN system is highly accurate in the detection of circinate (93.60% as recognition rate, 100% as sensitivity, and 88.10% as specificity) and MA (84.10% as recognition rate, 89.68% as sensibility, and 78.44% as specificity). On the other hand, the EMD-PNN system on average detects artery, blot, and drusen with 62.18%, 65.30%, and 76.75% respectively.

In summary, the experimental results show that in the EMD approach all classifiers obtain their highest accuracy to classify normal-versus-MA or normal-versus-circinate images using EMD-based features. In addition, they achieve interesting detection rate of images with drusen; except for the EMD-QDA system (54.39%). Indeed, the obtained average recognition rate of normal-versus-drusen digital images is 78.69% with EMD-SVM system, 70.05% with EMD-*k*-NN system, and 76.75% with the EMD-PNN system. Finally, they all obtain very moderate average recognition rate of artery and blot images.

Table 4.5 SVM classification results using EMD features

	Artery	**Artery**	**Blot**	**Blot**	**Circinate**	**Circinate**	**Drusen**	**Drusen**	**MA**	**MA**
	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.
C.C.R	**0.7333**	*0.0904*	**0.7345**	*0.1231*	**0.9511**	*0.0363*	**0.7869**	*0.0918*	**0.9611**	*0.0299*
Sensitivity	0.7616	*0.0504*	0.7810	*0.1352*	0.9731	*0.0290*	0.8461	*0.0717*	0.9798	*0.0264*
Specificity	0.7070	*0.1881*	0.6815	*0.1361*	0.9256	*0.0538*	0.7217	*0.1248*	0.9427	*0.0419*

Table 4.6 QDA classification results using EMD features

	Artery	**Artery**	**Blot**	**Blot**	**Circinate**	**Circinate**	**Drusen**	**Drusen**	**MA**	**MA**
	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.
C.C.R	**0.6624**	*0.0375*	**0.6902**	*0.0416*	**0.9384**	*0.0295*	**0.5439**	*0.1286*	**0.9284**	*0.0441*
Sensitivity	0.5625	*0.1102*	0.6717	*0.0697*	0.9209	*0.0542*	0.5678	*0.2262*	1.0000	*0.0000*
Specificity	0.7849	*0.1303*	0.7080	*0.1124*	0.9610	*0.0369*	0.5235	*0.0963*	0.8559	*0.0888*

Table 4.7 k-NN classification results using EMD features

	Artery	**Artery**	**Blot**	**Blot**	**Circinate**	**Circinate**	**Drusen**	**Drusen**	**MA**	**MA**
	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.
C.C.R	**0.5901**	*0.0369*	**0.5061**	*0.1372*	**0.9508**	*0.0402*	**0.7005**	*0.0371*	**0.9113**	*0.0712*
Sensitivity	0.5974	*0.2226*	0.5875	*0.0647*	1.0000	*0.0000*	0.7214	*0.1114*	0.9269	*0.0554*
Specificity	0.5746	*0.1738*	0.4398	*0.1961*	0.8961	*0.0860*	0.6807	*0.0795*	0.8956	*0.0967*

Table 4.8 PNN classification results using EMD features

	Artery	**Artery**	**Blot**	**Blot**	**Circinate**	**Circinate**	**Drusen**	**Drusen**	**MA**	**MA**
	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.	average	std. dev.
C.C.R	**0.6218**	*0.0703*	**0.6530**	*0.0588*	**0.9360**	*0.0543*	**0.7675**	*0.0989*	**0.8410**	*0.0545*
Sensitivity	0.6306	*0.0697*	0.7137	*0.0564*	1.0000	*0.0000*	0.9056	*0.0584*	0.8968	*0.0624*
Specificity	0.6074	*0.1627*	0.5904	*0.1008*	0.8810	*0.0880*	0.6178	*0.2303*	0.7844	*0.1236*

4.2.3 Principal component analysis based features results

In an attempt to improve the accuracy of the classifiers using EMD-based features, principal component analysis (PCA) was applied to the features set in order to reduce its dimensionality. Table 4.9 provides the selected features when principal component analysis is applied separately to intrinsic mode function one (IMF2) and to intrinsic mode function two (IMF2). On the other hand, Table 4.10 provide the selected features when principal

component analysis is applied jointly to intrinsic mode function one (IMF2) and to intrinsic mode function two (IMF2).

Table 4.9 shows that uniformity and entropy are the most dominant features that characterize IMF1 and IMF2 for all pathologies. Table 4.10 confirms the importance of these features since they are again selected by principal component analysis when it is applied jointly to IMF1 and IMF2; except for circinate pathology which characterized by the average statistic of its IMF1.

Table 4.9 PCA applied to IMF1 and IMF2 separately

	Selected features	**Cumulative variance proportion**
Artery		
PCA to IMF1	Uniformity, Entropy	0.9917
PCA to IMF2	Uniformity, Entropy	0.9916
Blot		
PCA to IMF1	Uniformity, Entropy	0.9877
PCA to IMF2	Uniformity, Entropy	0.9908
Circinate		
PCA to IMF1	Uniformity, Entropy	0.9895
PCA to IMF2	Uniformity, Entropy	0.9906
Drusen		
PCA to IMF1	Uniformity, Entropy	0.9844
PCA to IMF2	Uniformity, Entropy	0.9898
MA		
PCA to IMF1	Uniformity, Entropy	0.9878
PCA to IMF2	Uniformity, Entropy	0.9896

Table 4.10 PCA applied to IMF1 and IMF2 jointly

	Selected features	Cumulative variance proportion
Artery	Entropy (IMF1), Uniformity (IMF2)	0.9769
Blot	Entropy (IMF1), Uniformity (IMF2)	0.9670
Circinate	Mean (IMF1), Entropy (IMF2),	0.9752
Drusen	Entropy (IMF1), Uniformity (IMF2)	0.9580
MA	Entropy (IMF1), Uniformity (IMF2)	0.9654

Figure 4.12 exhibits the overall recognition rate for each classifier using empirical mode decomposition features that were selected by principal component analysis technique Detailed results for each classifier are given in tables 4.11 to 4.14. Figure 4.12 shows that principal analysis selected features do not help improving both the accuracy of support vector machines and the *k*-nearest neighbor algorithm. However, they do so for quadratic discriminant analysis; for instance the accuracy goes from 75.3% to 81.1%.

Although, PCA improves significantly the overall average detection rate of the *k*-nearest neighbor algorithm (81.1%) it still remains below the overall average accuracy of support vector machines (83.3%) before employing PCA selected inputs.

On the other hand, the effect of principal component analysis selected features is very limited on the performance of the probabilistic neural networks. Therefore, we may conclude, that principal component analysis reduced features set does not improve the performance of the classifiers in general, and may even worsen it. In other words, to classify normal versus abnormal retina digital images it is more suitable to employ an EMD-SVM system with twelve features (six from intrinsic mode function one and six from intrinsic mode function two) without using principal component analysis at all. On average, this approach leads to the highest detection rate of pathologies in retina photographs.

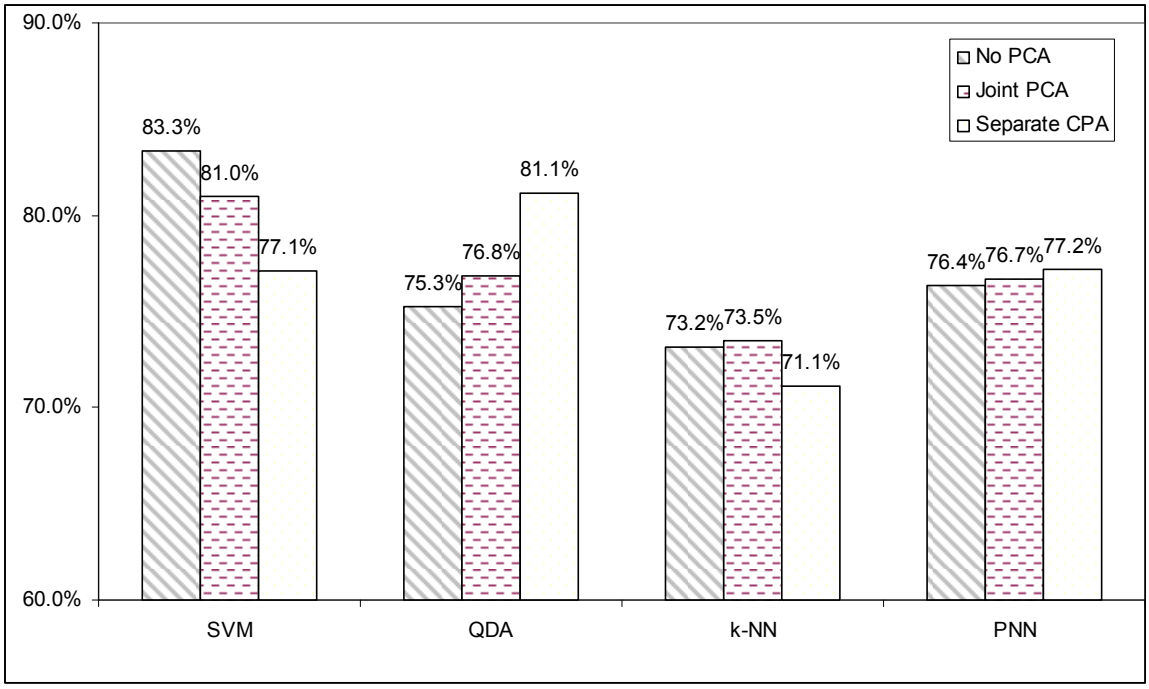

Figure 4.12 Average correct classification rate over all pathologies for each classifier

Table 4.11 SVM classification results

SVM	No PCA	Joint PCA	Separate CPA
	Artery	**Artery**	**Artery**
C.C.R	0.7333	0.7837	0.7471
Sensitivity	0.7616	0.92	0.7616
Specificity	0.707	0.6382	0.7364
	Blot	**Blot**	**Blot**
C.C.R	0.7345	0.6966	0.5628
Sensitivity	0.781	0.7384	0.4538
Specificity	0.6815	0.6548	0.6803
	Circinate	**Circinate**	**Circinate**
DR	0.9511	0.9136	0.935
Sensitivity	0.9731	0.8824	1
Specificity	0.9256	0.9481	0.8604
	Drusen	**Drusen**	**Drusen**
C.C.R	0.7869	0.7494	0.7312
Sensitivity	0.8461	0.872	0.7705
Specificity	0.7217	0.6098	0.6853
	MA	**MA**	**MA**
C.C.R	0.9611	0.9076	0.8803
Sensitivity	0.9797	0.8143	0.7796
Specificity	0.9424	1	1

Table 4.12 QDA classification results

QDA	No PCA	Joint PCA	Separate CPA
	Artery	**Artery**	**Artery**
C.C.R	**0.6624**	**0.7547**	**0.6619**
Sensitivity	0.5625	0.7888	0.6003
Specificity	0.7849	0.7168	0.7282
	Blot	**Blot**	**Blot**
C.C.R	**0.6902**	**0.7011**	**0.7623**
Sensitivity	0.6717	0.8054	0.8135
Specificity	0.708	0.5847	0.7133
	Circinate	**Circinate**	**Circinate**
C.C.R	**0.9384**	**0.8703**	**1**
Sensitivity	0.9209	0.8356	1
Specificity	0.961	0.9051	1
	Drusen	**Drusen**	**Drusen**
C.C.R	**0.5439**	**0.5826**	**0.7331**
Sensitivity	0.5678	0.6676	0.6629
Specificity	0.5235	0.4987	0.8062
	MA	**MA**	**MA**
C.C.R	**0.9284**	**0.9336**	**0.9001**
Sensitivity	1	0.9524	0.9381
Specificity	0.8559	0.9106	0.862

Table 4.13 k-NN classification results

kNN	No PCA	Joint PCA	Separate CPA
	Artery	**Artery**	**Artery**
C.C.R	**0.5901**	**0.701**	**0.7176**
Sensitivity	0.5974	0.6819	0.7009
Specificity	0.5746	0.7264	0.7425
	Blot	**Blot**	**Blot**
C.C.R	**0.5061**	**0.5212**	**0.46384**
Sensitivity	0.5875	0.34262	0.5039
Specificity	0.4398	0.70483	0.43548
	Circinate	**Circinate**	**Circinate**
C.C.R	**0.9508**	**0.8201**	**0.9182**
Sensitivity	1	0.8762	0.9798
Specificity	0.8961	0.7446	0.8548
	Drusen	**Drusen**	**Drusen**
C.C.R	**0.7005**	**0.725**	**0.6176**
Sensitivity	0.7214	0.8299	0.5745
Specificity	0.6807	0.6042	0.6599
	MA	**MA**	**MA**
C.C.R	**0.9113**	**0.9063**	**0.8388**
Sensitivity	0.9269	0.8845	0.8159
Specificity	0.8956	0.928	0.864

Table 4.14 PNN classification results

PNN	No PCA	Joint PCA	Separate CPA
	Artery	**Artery**	**Artery**
C.C.R	**0.6218**	**0.7679**	**0.7765**
Sensitivity	0.6306	0.802	0.7132
Specificity	0.6074	0.7297	0.843
	Blot	**Blot**	**Blot**
C.C.R	**0.653**	**0.7133**	**0.6964**
Sensitivity	0.7137	0.8482	0.662
Specificity	0.5904	0.5719	0.7351
	Circinate	**Circinate**	**Circinate**
C.C.R	**0.936**	**0.9143**	**0.9196**
Sensitivity	1	0.8858	0.9784
Specificity	0.881	0.9463	0.86
	Drusen	**Drusen**	**Drusen**
C.C.R	**0.7675**	**0.6484**	**0.6489**
Sensitivity	0.9056	0.9632	0.9687
Specificity	0.6178	0.305	0.2709
	MA	**MA**	**MA**
C.C.R	**0.841**	**0.7903**	**0.8177**
Sensitivity	0.8968	0.7265	0.8109
Specificity	0.7844	0.872	0.821

4.2.4 Comparison of simulation results

This section compares the experimental results obtained with DWT and EMD. The average correct classification rate, average sensitivity, and average specificity of DWT and EMD based features are given respectively in Figure 4.13, 4.14, and 4.15.

We found that the experimental results show strong evidence of the superiority of EMD-based features over DWT extracted features in terms of correct classification rate (Figure 4.13). For instance, the average recognition (classification or detection) rate (Figure 4.6) obtained with EMD-based features is 83.34% with SVM, 75.27% with QDA, 73.18% with *k*-NN, and 76.39% with PNN. On the other hand, the average recognition (classification or detection) rate obtained with DWT-based features is 71.95% with SVM, 63.32% with QDA, 62.07% with *k*-NN, and 54.15% with PNN. In addition, the simulations demonstrate the high effectiveness of SVM over the other classifiers. Indeed, whether using EMD-based features or DWT-based features; the SVM outperforms QDA, *k*-NN, and PNN. In terms of sensitivity, the EMD-based features outperform features extracted from DWT processed images when the classifiers SVM, *k*-NN, and QDA are employed; except for PNN (Figure 4.14). For instance, the maximum average value over all experiments is obtained with the DWT-based features using PNN classifier (99.08%).

Finally, in terms of specificity, the simulations show evidence that EMD-based features achieve higher average specificity over all experiments than DWT-based features (Figure 4.15). For instance, overall average specificity obtained with EMD features is 79.57%, with SVM, 76.67% with QDA, 69.74% with *k*-NN, and 69.62% with PNN. On the other hand, overall average specificity with extracted features from DWT processed images is 74.28% using SVM, 57.65% using QDA, 60.73% using *k*-NN, and 1.57% using PNN.

In summary, EMD-based features are strongly superior to DWT-based features in the diagnosis of retina digital images. In addition, we have found that the SVM outperforms QDA, *k*-NN, and PNN for both EMD-based features and DWT-based features.

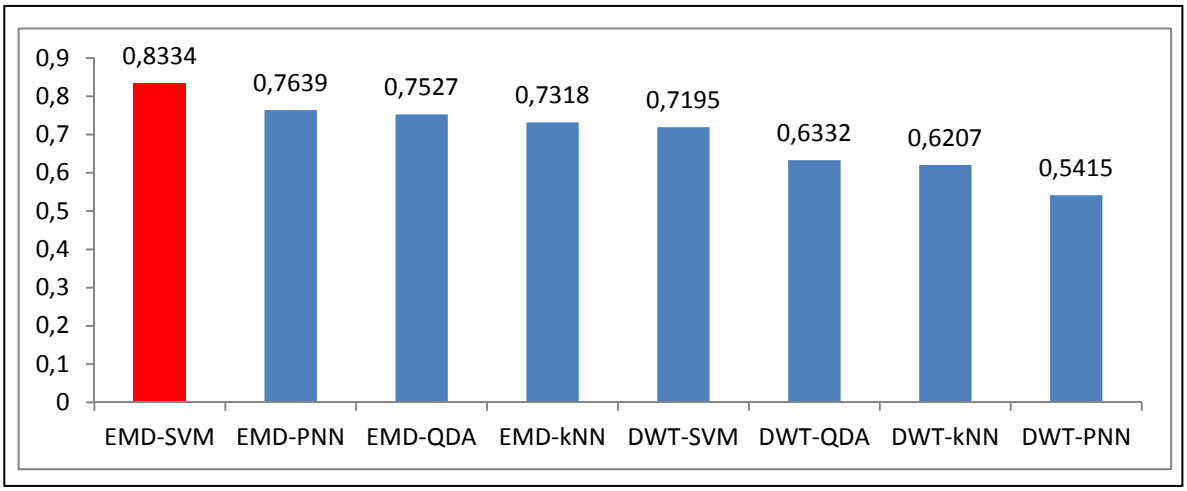

Figure 4.13 Average correct classification rate over all experiments

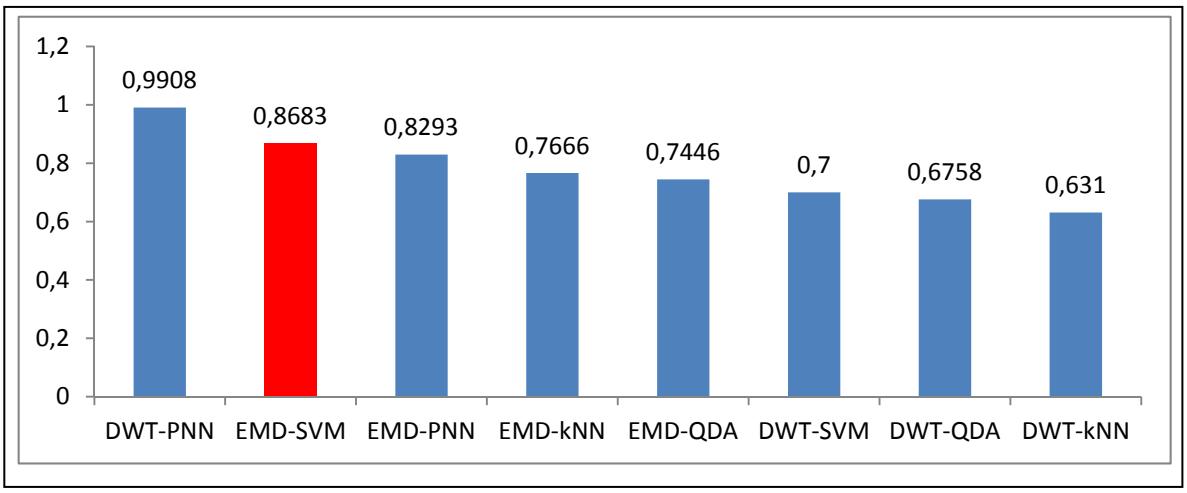

Figure 4.14 Average sensitivity rate over all experiments

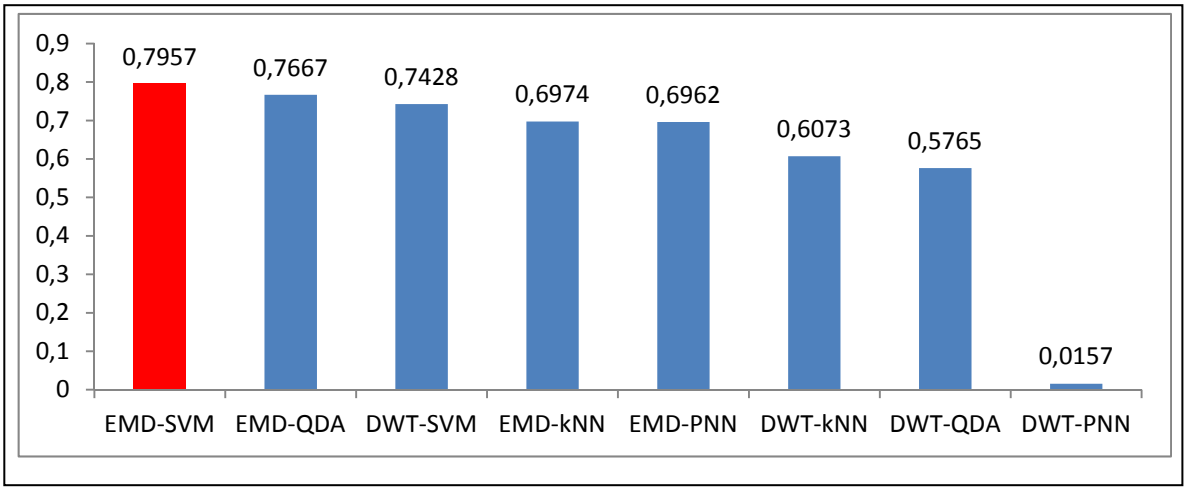

Figure 4.15 Average specificity rate over all experiments

CHAPTER 5

DISCUSSION OF THE OBTAINED RESULTS

Our main results could be summarized as follows:

a) empirical mode decomposition based features are strongly superior to DWT extracted features for all classifiers employed in our study;

b) support vector machines outperform quadratic discriminant analysis, *k*-NN, and probabilistic neural networks. Thus, SVM confirms its superiority mentioned in previous papers;

c) all classifiers obtain their highest accuracy to classify normal-versus-MA or normal-versus-circinate images using EMD-based features;

d) principal component analysis reduced features set does not improve the performance of the classifiers in general, and may even worsen it.

The EMD-based approach suggested in our study appears to compares favorably with previous results in terms of correct recognition rate. For instance, Yagmur et al., (2008) achieved 85% as maximum detection rate with and DWT-based processing approach and back-propagation neural networks (BPNN) as main classifier. Table 5.1 situates our study with previously reported researches. It is a non exhaustive summary of previous works.

Table 5.1 Some results reported in the literature

Authors	Image processing	Classifier	Performance
Lamard et al., (2007)	Haar wavelet	Template matching	Specificity: 96.18% Sensitivity: 87.94%
Khademi and Krishnan (2007)	Belkyns's shift-invariant DWT (SIDWT) + Gray level co-occurrence matrices (GLCM)	LDA Leave One Out Method (LOOM)	Accuracy: 82.2%
Quellec et al., (2008)	DWT	Template matching	Sensitivity: 93.74%
Yagmur et al. (2008)	DWT	Back-propagation neural networks (BPNN)	Accuracy: 50% -95%
Xu and Luo (2009)	Stationary discrete wavelet transform (SWT) and GLCM	SVM	Accuracy: 84% Sensitivity:84% Specificity: 80%
Baroni et al., (2007)	GLCM	BPNN	Accuracy: 79% Sensitivity:71% specificity: 87%
Meier et al., (2007)	Gabor filter Fourier Transform Histogram of the intensity Pixel intensities.	SVM	Accuracy Gabor: 80% Fourier Transform: 76% Histogram: 73% Pixel intensities: 83%
Anmar et al., (2008)	Fourier transform	Regression analysis	*R-squared* = 0.59
Lee et al., (2008)	Morphological scale space analysis + GLCM	Probabilistic boosting algorithm	Accuracy Normal images (81.3% - 92.2%) Abnormal images (71.7% - 85.2%)
Jagadish Nayak et al., (2009)	Morphological measures	BPNN	Accuracy: 90%
Our study (2011)	EMD	SVM + 10 folds cross validation	Accuracy: 73% - 96% Sensitivity: 86.83%-74.46% Specificity: 79.57%-69.62%

The performance of the EMD over the DWT may be explained as follows. The EMD is adaptive time-frequency decomposition (Oonincx , 2002; Flandrin et al., 2004; Demir and Ertürk, 2010). In other words, the EMD is like an adaptive filter banks. Therefore, it contains relevant information after the end of the sifting process.

In particular, the results suggest that high frequency signals (first and second intrinsic mode functions) characterize much better normal and abnormal retina digital images than DWT. On the other hand, the DWT requires a predetermined wavelet function and level of decomposition (Dunn et al., 1994; Livens and Van de Wouwer, 1997; Janusauskas et al., 2005). It would be more effective if the choice of the wavelet mother and the decomposition level was optimal.

If we hypothesize that high frequency components are able to detect pathologies in retina, then DWT could perform better if HH sub-band features are kept and LH and HL sub-bands features are omitted.

The extraction of features is based on statistical measures; including the mean, standard deviation, third moment, smoothness, uniformity and entropy. Based on these statistics, we made an assumption that they are sufficient to characterize the distribution of texture in retina digital images. They were considered because they are widely employed in biomedical image classification. However, other statistics were omitted such as the energy, and kurtosis.

The high performance of SVM may be associated with its theoretical foundations. Indeed, SVM are based on the theory of Structural Risk Minimization and usage of kernel function for data mapping (Vapnik, 1995). In addition, the SVM was originally designed for binary classification; which is the case of our problem (normal versus one abnormal category). SVM confirms its superiority in the classification of one dimensional and two dimensional biomedical signals (Gil-Pita et al., 2007; Melgani and Bazi, 2008; Ai et al., 2011; Salas-Gonzalez et al., 2011).

It is important to notice that we have used support vector machines with polynomial kernel of order two. A Gaussian kernel could be also employed and may give better results. However, it requires finding the optimal values of the Gaussian distribution used to model the data; for example the mean and the standard deviation. This is a difficult task and many researches were written on this subject (Friedrichs and Igel, 2005; Sullivan and Luke, 2007; Diosan et al., 1011). We have also used polynomial kernels with orders three and four.

The results are not reported here since in many case the SVM take a long time to converge or do not converge because of limited memory of the computer. For example, it is required to create an optimization options structure using Matlab Optimization Toolbox © to solve the problem of limited memory for each classification problem. Thus, the polynomial kernel of order three and four are omitted in our study.

It should be noted that both QDA and PNN are based on strong statistical assumptions. Indeed, the QDA vector of features is supposed to be multivariate normally distributed; which is not always true in real life problems. Besides, PNN is based on Bayesian decision and the probability density function (PDF) of each group of patterns is assumed to follow a Gaussian distribution; which is also not always true in real life problems. Moreover, PNN requires a large training set to obtain appropriate estimates of the mean and variance that characterize its Gaussian function (PDF) used to perform the classification task. Finally, the performance of *k*-NN depends on the number of the nearest neighbor *k* which was not optimized. For instance, a larger *k* makes *k*-NN less sensitive to noise; whilst a small *k* allows capturing finer structure of space. In our study, we made the assumption that there is no potential noise in the data; thus the parameter *k* was set arbitrarily to 1 to guarantee a better characterization of the structure space.

Furthermore, the performance of *k*-NN depends also on the choice of the distance used for learning and classification; which is often a difficult decision to make. In our study, the Euclidean distance was chosen because it is the most commonly employed type of distance, and is less sensitive to outliers.

In particular, the distance between two objects is not affected by the addition of new objects to the learning database. However, it is very sensitive to scale.

Metric distances other than the Euclidean distance do exist and may be considered for future researches. For instance, we may mention *cityblock* (Sum of absolute differences), *cosine* (one minus the cosine of the included angle between points), `correlation` (one minus the sample correlation between points), and `hamming` (percentage of bits that differ).

In an attempt to improve the accuracy of the classifiers, principal component analysis (PCA) was used to select the features that capture more of the variability in the initial features space. The most representative features selected by PCA are uniformity and entropy for all pathologies (See Tables 4.9 and 4.10), except the mean and entropy for circinate. However, we found that in general the reduced features set do not improve significantly the detection rate. The main problem of PCA is that it is a transformation which performs a linear projection; thus it fails to capture nonlinear relationships in the original features space. In our experimental results, we found that in general PCA does not improve the accuracy of the classifiers. This could be explained by the fact that feature extraction by PCA does not consider the class information of the data (Park and Choi, 2009). The PCA technique could be replaced by artificial neural networks that can perform a kind of non-linear PCA (Markey et al., 2003; Faro et al., 2005; Maiorana, 2008). The ANN are very strong to capture nonlinear dynamics in the data and very robust to noise. However, the obtained results from ANN could not be explained in a statistical sense. For instance, ANN acts as a black-box.

In our study, we have employed a feature transformation method to transform data from the original high-dimensional feature space to a new space with small dimension using principal component analysis. This approach does not necessarily consider relevant features that may help improving the accuracy of the detection system. Then, statistical feature selection algorithms that select a subset of features from the original feature set may be considered; namely filter methods, wrapper methods, and ranking feature approach (Saeys et al., 2007).

Both filter and wrapper methods use characteristics of the data to evaluate and to select the feature subsets. However, filter approach does not involve the chosen learning algorithm; whilst the wrapper approach does. Unlike filter and wrapper methods, the ranking features procedure considers interaction between features.

The main problems with our methodology can be summarized as follows.

First, our database is small. A larger database would be preferable to make strong conclusions. However, 10 folds cross validation approach was adopted in our study to encounter the problem of the limited database. Moreover, five pathologies were considered to make robust conclusion. The second problem is related to the image time processing using the empirical mode decomposition (EMD).

Indeed, EMD algorithm of Rato et al. (2008) that we employed is highly time consuming (340 seconds to 5100 seconds), versus a very fast processing of the images by the discrete wavelet transform (DWT). Therefore, it could not be appropriate to use this approach for real time applications unless a solution to reduce time processing of the EMD is found.

CONCLUSION

Conventional retinal disease identification techniques based on manual observation by physicians for retina diagnosis is time consuming. Therefore, the automatic analysis of retina images has received a large scientific attention with the purpose of providing computational and intelligent tools to detect pathologies in the texture of digitized medical images. The main purpose of our study is to design an automatic screening system which aims to detect Artery sheath (Coat's disease), blot hemorrhage, retinal degeneration (circinates), age-related macular degeneration (drusens), and diabetic retinopathy (microaneurysms and exudates). The system is based on features extractions from retina digital images my means of empirical mode decomposition (EMD) technique in introduced by Huang et al., (1998). Unlike the discrete wavelet transform (DWT), this technique is an adaptive time-frequency analysis approach that requires no assumption about the underlying data and requires no choice about the wavelet mother and level of decomposition. The extraction of features is based on statistical measures; including the mean, standard deviation, third moment, smoothness, uniformity and entropy. Based on these statistics, we made an assumption that they are sufficient to characterize the distribution of texture in retina digital images. They were considered because they are widely employed in biomedical image classification. These statistical features were extracted from the first and the second intrinsic mode functions for the EMD approach, and from LH, LH, and HL for the DWT approach at levels of decomposition one, two and three as done in the literature. In order to classify (detect) normal images versus abnormal ones, four classifiers were considered; including support vector machines, quadratic discriminant analysis (QDA), *k*-nearest neighbor algorithm (*k*-NN), and probabilistic neural networks (PNN). These classifiers were chosen because they are widely used in pattern recognition.

The performance measures were the correct classification rate, sensitivity and specificity. All the experiments were performed using ten folds cross-validation technique.

Then, average values of correct classification rate, sensitivity and specificity are computed to make performance comparisons.

There experimental results show strong evidence of the superiority of EMD-based features over DWT extracted features for all classifiers employed in our study. In addition, the simulations show that support vector machines with a polynomial of order two as kernel function outperform quadratic discriminant analysis, *k*-NN, and probabilistic neural networks. Thus, SVM confirms its superiority mentioned in previous papers. The high performance of SVM may be associated with its theoretical foundations. Indeed, SVM are based on the theory of Structural Risk Minimization and usage of kernel function for data mapping. On the other hand, both QDA and PNN are based on strong statistical assumptions. Finally, the performance of *k*-NN depends on the number of the nearest neighbor *k* which was not optimized.

Furthermore, the experimental results show that all classifiers obtain their highest accuracy to classify normal-versus-MA or normal-versus-circinate images using EMD-based features.
In subsequent experiments, principal component analysis was performed to select most representative features. They were uniformity and entropy for all pathologies, except the mean and entropy for circinate. We found that although principal component analysis reduced features set it does not improve the performance of the classifiers in general, and may even worsen it. Indeed, the main problem of PCA is it is a transformation that performs a linear projection; thus it fails to capture nonlinear relationships in the original features space.

In comparison with the literature, the EMDA-based approach suggested in our study achieved higher accuracy – in terms of recognition rate – than what have been obtained with previous papers. For example, Yagmur et al., (2008) achieved 85% as maximum detection rate with and DWT-based processing approach and back-propagation neural networks (BPNN) as main classifier.

Our approach based on EMD features and support vector machines achieved as much as 96%. Finally, our study confirms the superiority of the EMD extracted features over the DWT method.

For instance, in the problem of context of one dimensional biomedical signal processing and classification; Ai et al., (2011) found that the EMD outperform the DWT using SVM as classifier to distinguish essential tremor (ET) and Parkinson's disease (PD).

For future work, several routes can be considered. First, it is important to gather a larger database of retina images. The only way to do so is to employ private databases which are usually very expensive. Second, the empirical mode decomposition should be applied to other types of biomedical images such as brain magnetic resonance images, mammograms, skin, and prostate digital images in order to check the superiority of the EMD. Third, original and fast EMD algorithm should be considered for retina image processing and high frequency signal extraction. Thus, we may reduce the processing time while probably extracting very accurate high frequency signal. Fourth, features may be obtained from HH sub-bands of the DWT. Fifth, one may investigate classification power of each feature using statistical feature selection approaches such filter, wrapper, and rank method. The purpose of the search algorithm is to find a subset of predictors that optimally model measured responses, subject to constraints such as required or excluded features and the size of the subset. The artificial neural networks may be also considered since they are very strong to capture nonlinear dynamics in the data and very robust to noise. Sixth, other statistical features could be taken into account such as energy of the signal and its kurtosis (fourth moment).

In sum, our study has shown strong evidence; as mentioned previously; of the superiority of the empirical mode decomposition over the discrete wavelet transform in the automatic recognition of certain pathologies in retina digital images.

For future work, many avenues may be explored to improve the performance of the EMD. Finally, other types of biomedical images should be investigated to validate the effectiveness of this very promising adaptive time-frequency analysis method.

BIBLIOGRAPHY

Ai, L., J. Wang, R. Yao. 2011. « Classification of parkinsonian and essential tremor using empirical mode decomposition and support vector machine ». Digital Signal Processing, vol. 21, p. 543-550.

Amara's Wavelet Page. Online. http://www.amara.com/IEEEwave/IW_see_wave.html>. Consulted on May 22th, 2011.

Anzalone, A., F. Bizzari, M. Parodi, and M. Storace. 2008. « A modular supervised algorithm for vessel segmentation in red-free retinal images ». Computers in Biology and Medicine, vol. 38, p. 913-922.

Ayres, F.J., and R.M. Rangayyan. 2005. « Characterization of Architectural Distortion in Mammograms ». IEEE Eng Med Biol Mag, vol. 24, no. 1, p. 59-67.

Baroni, M., S. Diciotti, A. Evangelisti, P. Fortunato, and A. La Torre. 2007. « Texture Classification of Retinal Layers in Optical Coherence Tomography ». Proceedings of Medicon IFMBE 2007, p. 847-850.

Brandon, L., and A. Hoover. 2003. « Drusen detection in a retinal image using multi-level analysis ». Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science, vol. 2878, p. 618-625.

Buciu, I., and A. Gacsadi. 2009. « Gabor Wavelet Based Features for Medical Image Analysis and Classification ». IEEE 2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL), November 24-27, Bratislava, Slovak Republic, p. 1-4.

Chen Y., X. Zhang, M.Q. Yang, and J.Y. Yang. 2007. « Ensemble of Probabilistic Neural Networks for Protein Fold Recognition ». Proceedings of the 7th IEEE International Conference on Bioinfomatics and Bioengineering, p. 66-70.

Chen, Y., Y. Jiang, C. Wang, D. Wang, W. Li, and G. Zhai. 2009. « A Novel Multi-focus Images Fusion Method Based on Bidimensional Empirical Mode Decomposition».IEEE 2nd International Congress On Image and Signal Processing, p. 1-4, 2009.

Coat's disease. Online. <http://www.coatsdisease.org/index.html>. Consulted on May 22th, 2011.

Cover, T., and P. Hart. 1967. « Nearest neighbor pattern classification ». IEEE Transactions On Information Theory, p. 21-27.

Cristianini, N., and J.S. Taylor. 2000. « An Introduction to Support Vector Machines ». Cambridge University Press, Cambridge, MA.

Daubechies, Ingrid. 1988. « Orthonormal bases of compactly supported wavelets». Communications on Pure and Applied Mathematics, vol. 41, nº 7, p. 909-996.

Demir, B., and S. Ertürk. 2009. « An empirical mode decomposition and composite kernel approach to increase hyperspectral image classification accuracy ». IEEE International Geoscience and Remote Sensing Symposium (IGARSS), vol. 2, p. 855-858.

Demir, B., and S. Ertürk. 2010. « Empirical Mode Decomposition of Hyperspectral Images for Support Vector Machine Classification ». IEEE Transactions on Geoscience and Remote Sensing, vol. 48, nº 11. p. 4071-4084.

Digital Reference of Ophthalmology. Online. http://dro.hs.columbia.edu/dotblot.htm. Consulted on May 22th, 2011.

Diosan, L., Rogozan, A., Pecuchet, J.-P. 2011. « Improving classification performance of Support Vector Machine by genetically optimising kernel shape and hyper-parameters ». Advances in Intelligent and Soft Computing, vol. 95, p. 99-108.

Duda, R., P.Hart, and D.Stork. 2000. « Pattern classification ». 2nd ed. New York: Wiley/Interscience.

Dunn, D., Higgins, W.E., and Wakeley, J. 1994. « Texture segmentation using 2-D Gabor elementary functions ». IEEE Trans. Pattern Anal. Mach. Intell., vol. 16, no. 2, p. 130-149.

El-Dahshan, E-S.A., T. Hosny, and A.-B. M. Salem. 2010. « Hybrid intelligent techniques for MRI brain images classification ». Digital Signal Processing, vol. 20, p. 433-441.

Faro, A., Giordano, D., Maiorana, F. 2005. « Discovering complex regularities by adaptive Self Organizing classification». Proceedings of WASET, vol. 4, p. 27-30.

Feng, P., Y.-J. Pan, B. Wei, W. Jin, and D.-L. Mi. 2007. « Enhancing retinal image by the contourlet transform ». Pattern Recognition Letters, Vol. 28, p. 516-522.

Flandrin, P., Rilling, F., and Goncaleves, P. 2004. « Empirical mode decomposition as a filter bank ». IEEE Signal Process. Lett, vol. 11, p. 112-114.

Foracchia, M., E. Grison, and A. Ruggeri. 2005. « Luminosity and contrast normalization in retinal images ». Medical Image Analysis, vol. 9, p. 179-190.

Franzco, A.M.A.-R., T. Molteno, and A.C.B.M. Franzco. 2008. « Fourier analysis of digital retinal images in estimation of cataract severity ». Clinical and Experimental Ophthalmology, vol. 36, p. 637-645.

Friedrichs, F., Igel, C. 2005. « Evolutionary tuning of multiple SVM parameters ». Neurocomputing, vol. 64, p. 107-117.

Fu, J.C., S.K. Lee, S.T.C. Wong, J.Y. Yeh, A.H. Wang, and H.K. Wu. 2005. « Image segmentation feature selection and pattern classification for mammographic microcalcifications ». Computerized Medical Imaging and Graphics, vol. 29, p. 419-429.

Gil-Pita, R., P. Jarabo-Amores, M. Rosa-Zurera, F. López-Ferreras, and M. Utrilla-Manso. 2007. « Divide and Conquer Approach to Improve Performance on ATR Systems ». Pattern Recognition and Image Analysis, vol. 17, nº2, p. 284-291.

Gonzalo, P., and J. M. De La Cruz. 2004. « A wavelet-based image fusion tutorial ». Pattern Recognition, vol. 37, p. 1855-1872.

Guangtao, G., S. Enfang , L. Zhuofu, and Z. Beibei. 2008. « Bidimensional EMD Infinite Sifting ». IEEE International Conference Neural Networks & Signal Processing, Zhenjiang, China, June 8-10, p. 348-352.

Hastie T., R. Tibshirani , and J. Friedman. 2001. «The Elements of Statistical Learning: Data Mining, Inference and Prediction ». Springer, New York.

Haykin, Simon. 2008. « Neural Networks and Learning Machines ». Publisher: Prentice Hall; third edition.

Huang N.E., Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-Ch. Yen, C.C. Tung and H.H. Liu. 1998. « The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis ». Proc. R. Soc., Lond. A 454, p. 903-995.

Jai-Andaloussi, S., M. Lamard, G. Cazuguel, H. Tairi, M. Meknassi, B. Cochener, and C. Roux. 2010. « Content Based Medical Image Retrieval based on BEMD: optimization of a similarity metric ». 32nd Annual International Conference of the IEEE EMBS, Buenos Aires, Argentina, August 31 - September 4, p. 3069-3072, 2010.

Janusauskas, A., Jurkonis, R., Lukosevicius, A., Kurapkiene, S., and Paunksnis, A. 2005. « The empirical mode decomposition and the discrete wavelet transform for detection of human cataract in ultrasound signals ». Inf. Lith. Acad. Sci., vol. 16, no. 4, p. 541-556.

Jolliffe, I. 2002. « Principal component analysis ». Springer. Second edition.

Khademi, A., and S. Krishnan. 2007. « Shift-invariant discrete wavelet transform analysis for retinal image classification ». Med Bio Eng Comput, vol. 45, p. 1211-1222.

Lamard, M., G. Quellec, P. M. Josselin, G. Cazuguel, B. Cochener, and C. Roux. 2007. « Detection of lesions in retina photographs based on the wavelet transform ». World Congress on Medical Physics and Biomedical Engineering, vol. 14, Part 15, p. 2435-2438.

Le, D.-D., and S. Satoh. 2005. « An Efficient Feature Selection Method for Object Detection ». ICAPR, Lecture Notes in Computer Science 3686, p. 461-468.

Lee, N., A.F. Laine, and T.R. Smith. 2008. « Learning non-homogenous textures and the unlearning problem with application to drusen detection in retinal images ». The 5th IEEE international symposium on biomedical imaging, May 4-17, Paris, France, p. 1215-1218.

Liu, W., W. Xu, and L. Li. 2007. « Medical Image Retrieval Based on Bidimensional Empirical Mode Decomposition ». Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering (BIBE), Boston, MA, 14-17 October, p. 641-646.

Livens, S., Van de Wouwer, G. 1997. « Wavelets for Texture Analysis: An Overview ». In Proceedings of the Sixth International Conference on Image Processing and its Applications (IPA'97), Dublin, Ireland, p. 581-585.

Lorena, A.C., and A.P.L.F. de Carvalho. 2009. « Evaluation functions for the evolutionary design of multiclass support vector machines ». International Journal of Computational Intelligence and Applications, vol. 8, n° 1, p. 53-68.

Ma, H., C. Jia, and S. Liu. 2005. « Multisource Image Fusion Based on Wavelet Transform ». Int. Journal of Information Technology, vol. 11, nº. 7, p. 81-91.

Maiorana, F. 2008. « Feature Selection with Kohonen Self Organizing Classification Algorithm ». World Academy of Science, Engineering and Technology, vol. 45, p. 47-52.

Malhi, A., and R.X. Gao. 2004. « PCA-Based Feature Selection Scheme for Machine Defect Classification ». IEEE Transactions on Instruments and Measurement, vol. 53, nº 6, p. 1517-1525.

Mallat, Stéphane. 1999. « A Wavelet Tour of Signal Processing ». Academic Press.

Markey, M.K., Lo, J.Y., Tourassi, G.D., and Floyd, C.E. 2003. « Self-organizing map for cluster analysis of a breast cancer database ». Artificial Intelligence in Medicine, vol. 27 p. 113-127.

Martins, C.I.O., F.N.S. Medeiros, R.M.S. Veras, F.N. Bezerra, and R.M. Cesar Jr. 2009. « Evaluation of Retinal Vessel Segmentation Methods for Microaneurysms Detection ». IEEE ICIP, p. 3365-3368.

Martins, C.I.O., R.M.S. Veras, G.L.B. Ramalho, F.N.S. Medeiros, and D.M. Ushizima. 2008. « Automatic Microaneurysm Detection and Characterization Through Digital Color Fundus Images ». Proceedings of The Second Workshop on Computational Intelligence (WCI'08), Salvador, Brazil, October, p. 45–50.

Matlab. Online. < http://www.mathworks.com>. Consulted on May 22th, 2011.

McGonigle, J., M. Mirmehdi, and A.L. Malizia. 2010. « Empirical Mode Decomposition in data-driven fMRI analysis ». IEEE Workshop on Brain Decoding: Pattern Recognition Challenges in Neuroimaging, Istanbul, Turkey, 22-22 Aug, p. 25-28.

Medical Dictionary. Online. <http;//medicaldictionary.thefreedictionary.com/circinate>. Consulted on May 22th, 2011.

Meier, J., R. Bock, L.G. Nyúl, and G. Michelson. 2007. « Eye Fundus Image Processing System for Automated Glaucoma Classification ». 52nd Internationales Wissenschaftliches Kolloquium Technische Universität Ilmenau, 10-13 September, 2007.

Melgani, F., and Y. Bazi. 2008. « Classification of electrocardiogram signals with support vector machines and particle swarm optimization ». IEEE Transactions on Information Technology in Biomedicine, vol. 12, nº5, p. 667-677.

Miri, M.S., and A. Mahloojifar. 2009. « A Comparison Study to Evaluate Retinal Image Enhancement Techniques ». IEEE International Conference on Signal and Image Processing Applications, p. 90-94.

Nayak, J.,R.U. Acharya, P.S Bhat, N. Shetty, and T.-C. Lim. 2009. « Automated Diagnosis of Glaucoma Using Digital Fundus Images ». J Med Syst, vol. 33, p. 337-346.

Niemeijer, M., B. Van Ginneken, S.R. Russel, M.S.A. Suttorp-Schulten, and M.D. Abramoff. 2007. « Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis ». Investigative Ophthalmology & Visual Science, vol. 48, p. 2260-2267.

Nunes, J.C., Y. Bouaoune, E. Delechelle, O. Niang, and Ph. Bunel. 2003. « Image analysis by bidimensional empirical mode decomposition ». Image and Vision Computing, vol. 21, p. 1019-1026.

Oonincx, P.J. 2002. « Empirical mode decomposition: a new tool for S-wave detection ». CWI Rep. Probab., Networks Algorithms (PNA) PNAR0203, (2002) ISSN 1386-3711.

Pan, N., Mang V.I., M. Un, and P. Hang. 2007. « Accurate removal of baseline wander in ECG using empirical mode decomposition ». The Joint Meeting NFSI & ICFBI, p. 177-180.

Park M.S., J.Y. Choi. 2009. « Theoretical analysis on feature extraction capability of class-augmented PCA ». Pattern Recognition, vol. 42, p. 2353-2362.

Parzen, E. 1962. « On Estimation of a Probability Density Function and Mode ». Annals of Mathematical Statistics, vol. 33, p. 1065-1076.

Pinheiro, E., O. Postolache, and P. Girão. 2011. « Empirical Mode Decomposition and Principal Component Analysis implementation in processing non-invasive cardiovascular signals ». Article in press in Measurement. doi:10.1016/j.measurement.2011.03.022

Pizer, S.M., E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. Ter Haar Romeny, B. Zimmerman, and K. Zuiderveld. 1987. « Adaptive histogram equalization and its variations ». Compurer Vision, Graph. Image Processing, vol. 39, p. 355-368.

Poggio, T., and F. Girossi. 1990. « Networks for approximation and learning ». Proc. of The IEEE, vol. 78, p. 1481-1497.

Qin, X., S. Liu, W. Zhengqiang, and J. Han. 2008. « Medical Image Enhancement Method Based on 2D Empirical Mode Decomposition ». The 2nd International Conference on Bioinformatics and Biomedical Engineering (ICBBE 2008), 16-18 May, Shanghai, China, p. 2533-2536.

Quellec, G., M. Lamard, P.M. Josselin, G. Cazuguel, B. Cochener and C. Roux. 2008. « Optimal wavelet transform for the detection of microaneurysms in retina photographs ». IEEE Transactions on Medical Imaging, vol. 27, Nº 9, p. 1230-1241.

Rato, R.T., M.D. Ortigueira, and A.G. Batista. 2008. « On the HHT, its problems, and some solutions ». Mechanical Systems and Signal Processing, vol. 22, p. 1374-1394.

Saeys, Y., Inza, I., Larranaga, P. 2007. « A review of feature selection techniques in bioinformatics ». Bioinformatics, vol. 23, no. 19, p. 2507-2517.

Salas-Gonzalez, D., J.M. Górriz, J. Ramírez, I. Álvarez, M. López, F. Segovia, and C.G. Puntonet. 2011. « Two approaches to selecting set of voxels for the diagnosis of Alzheimer's disease using brain SPECT images ». Digital Signal Processing. In press. doi:10.1016/j.dsp.2011.01.006

Sert, S.B., and Ö. Kalenderli. 2009. « Determination of source voltage from audible corona noise by neural networks ». IEEE Transactions on Dielectrics and Electrical Insulation, vol. 16, nº. 1, p. 224-231.

Sheshadri, H.S., and A. Kandaswamy. 2005. « Detection of breast cancer by mammogram image segmentation ». Journal of Cancer Research and Therapeutics, vol.1, p. 232-234.

Soliz, P., S.R. Russell, M.D. Abramoff, S. Murillo, M. Pattichis, and H. Davis. 2008. « Independent Component Analysis for Vision-inspired Classification of Retinal Images with Age-related Macular Degeneration ». IEEE SSIAI, p. 65-68.

Sopharak, A., B. Uyyanonvara, S. Barman, and T.H Williamson. 2008. « Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods ». Computerized Medical Imaging and Graphics, vol. 32, p. 720-727.

Specht, D. 1988. « Probabilistic Neural Networks for Classification, Mapping, or Associative Memory ». IEEE International Conference on Neural Networks, p. 525-532.

Specht, D. 1990. « Probabilistic Neural Networks ». Neural Networks, vol. 3, p. 109-118.

STARE database. Online. < http://www.ces.clemson.edu/~ahoover/stare/>. Consulted on May 22th, 2011.

Sullivan, K., Luke, S. 2007. « Evolving kernels for SVM classification ». Genetic and Evolutionary Computing conference (GECCO-ACM), p. 1702-1707.

Sun, J., D. Wen, and G. Li. 2009. « An efficient guide stars classification algorithm via support vector machines ». IEEE Second International Conference on Intelligent Computation Technology and Automation, p. 148-152.

Tafreshi, A.K., A.M. Nasrabadi, and A.H. Omidvarnia. 2008. « Epileptic seizure detection using empirical mode decomposition ». 8th IEEE ISSPIT, p. 238-242.

Vapnik, V.N. 1995. « The Nature of Statistical Learning Theory ». Springer-Verlag.

Wang, L., and D. Wen. 2008. « Recognition of Arterial Pulse Wave of Pregnant Woman Based on Probabilistic Neural Network ». Proceedings of the 8th IEEE International Conference on Bioinfomatics and Bioengineering, p. 1955-1958.

Xu, G.L., X.T. Wang, and X.G. Xu. 2011. « Improved bi-dimensional empirical mode decomposition based on 2D-assisted signals: analysis and application ». IET Image Processing, vol. 5, nº 3, p. 205-221.

Xu, L., and S. Luo. 2009. « Support vector machine based method for identifying hard exudates in retinal images ». IEEE Youth Conference on Information, Computing and Telecommunication (YC-ICT), Beijing, China, September 20-21, p. 138-141.

Yagmur, F.D., B. Karlik, and A. Okatan. 2008. « Automatic Recognition of Retinopathy Diseases by Using Wavelet Based Neural Network ». IEEE International Conference on The Applications of Digital Information and Web Technologies (ICADIWT), Ostrava, Czech Republic, August 4-6. p. 454-457.

Yuping, Z. 2006. « Hilbert–Huang transform andmarginal spectrum for detection of bearing localized defects ». Proc. 6th World Congr. Intell. Control Autom. (WCICA 2006), Dalian, China, p. 5457-5461.

Zhao, Q., and H. Lu. 2007. « PCA-based web pagewatermarking ». Pattern Recognition, vol. 40, p. 1334-1341.

Zhiming, W., and T. Jianhua. 2006. « A Fast Implementation of Adaptive Histogram Equalization ». Proc. of the IEEE International conference on signal Processing.

Zimmerman, J.B., S.M Pizer, E.V. Staab, J.R. Perry, W. McCartney, and B.C. Brenton. 1988. « An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement ». IEEE Transactions on Medical Imaging, p. 304-312.