

**ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES**

MSc THESIS

İclal ÇETİN TAŞ

**DEPICTION OF HEMORRHAGIC REGIONS CAUSED BY DIABETIC
RETINOPATHY ON FUNDUS FLOURESCIN ANGIOGRAPHY IMAGES
BY USING IMAGE ENHANCEMENT AND BASIC CLASSIFICATION
METHODS**

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

ADANA, 2012

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

**DEPICTION OF HEMORRHAGIC REGIONS CAUSED BY DIABETIC
RETINOPATHY ON FUNDUS FLOURESAN ANGIOGRAPHY IMAGES BY
USING IMAGE ENHANCEMENT AND BASIC CLASSIFICATION
METHODS**

İclal ÇETİN TAŞ

MSc THESIS

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

We certified that the thesis titled above was reviewed and approved for the award of degree of the Master of Science by the board of jury on 31/10/2012

.....
Asst. Prof. Dr. Turgay İBRİKÇİ
SUPERVISOR

.....
Assoc. Prof. Dr. Zekeriyya TÜFEKÇİ
MEMBER

.....
Asst. Prof. Dr. Murat AKSOY
MEMBER

This MSc Thesis is written at the Department of Institute of Natural And Applied Sciences of Çukurova University.

Registration Number:

Prof. Dr. Selahattin SERİN
Director
Institute of Natural and Applied Sciences

This thesis as financially supported by C.U Research Resource Foundation **MMF2011 YL2**.

Not:The usage of the presented specific declarations, tables, figures, and photographs either in this thesis or in any other reference without citation is subject to “The law of Arts and Intellectual Products” number of 5846 of Turkish Republic.

ABSTRACT

MSc. THESIS

DEPICTION OF HEMORRHAGIC REGIONS CAUSED BY DIABETIC RETINOPATHY ON FUNDUS FLOURESAN ANGIOGRAPHY IMAGES BY USING IMAGE ENHANCEMENT AND BASIC CLASSIFICATION METHODS

İclal ÇETİN TAŞ

**ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING**

Supervisor : Asst. Prof. Dr. Turgay İBRİKÇİ
Year: 2012, Page: 80
Co-Supervisor : Asst.Prof.Dr. Sami ARICA
Jury : Asst. Prof. Dr. Turgay İBRİKÇİ
: Asst. Prof.Dr. Murat AKSOY
: Assoc. Prof.Dr. Zekeriya TÜFEKÇİ

Diabetes can lead very different complications to the human body. The most commonly encountered complications in the human body is eyes. Eye vessels structure corruption, giving rise to a new, abnormal blood vessels and bleeding that occurs as a complication called diabetic retinopathy. Fundus fluorescein angiography (FFA) has been widely used imaging technique for the detection of disease. In recent years, increase in the incidence of this disease has increased the number of studies on the subject.

In this thesis, detection of diabetic retinopathy and to highlight areas that cause hemorrhagic images of FFA on the classification, clustering methods and image analysis methods were used. As supervised learning, back propagation neural networks, statistical methods linear discriminant analysis (LDA), as unsupervised learning techniques, K-means clustering, Gaussian mixture method and the histogram image enhancement techniques transactions were selected for analysis. These approaches have been to emphasize shadows bleeding, and then the results which is obtained by these methods are compared.

Keywords:Classifying, Clustering, Retina, Image Analysis, Medicine

ÖZ

YÜKSEK LİSANS TEZİ

DIYABETİK RETİNOPATİYE YOL AÇAN KANAMALI BÖLGELERİN FUNDUS FLOURESAN ANJİOGRAFİ GÖRÜNTÜLERİ ÜZERİNDE GÖRÜNTÜ İYİLEŞTİRME VE TEMEL SINIFLAMA METOTLARINI KULLANARAK TANIMLANMASI

İclal ÇETİN TAŞ

**ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK ANA BİLİM DALI**

Danışman : Yard. Doç. Dr. Turgay İBRİKÇİ
Yıl:2012, Sayfa: 80
2. Danışman : Yard. Doç. Dr. Sami ARICA
Jüri : Yard. Doç. Dr. Turgay İBRİKÇİ
: Yard. Doç. Dr. Murat AKSOY
: Doç. Dr. Zekeriya TÜFEKÇİ

Şeker hastalığı bilindiği üzere insan vücudunda çok farklı komplikasyonlara yol açmaktadır. Bu komplikasyonların en sık rastlandığı bölgelerden biriside gözlerdir. Göz damarlarının yapısının bozulması, yeni ve anormal damarlar oluşması ve kanamaya sebebiyet veren bu komplikasyonlar diyabetik retinopati olarak adlandırılmaktadır. Hastalığın tespiti için Fundus Floruesan Anjiyografi (FFA) görüntüleme tekniği yaygın olarak kullanılmaktadır. Son yıllarda bu hastalığın görülme sıklığının artması konu üzerinde yapılan çalışmaların sayısını da artırmıştır.

Bu tezde, diyabetik retinopatiye yol açan kanamalı bölgelerin tespiti ve vurgulanması için FFA görüntüleri üzerinde sınıflama, kümeleme metotları ve görüntü analiz yöntemleri kullanılmıştır. Gözetimli öğrenme olarak, yapay sinir ağlarından geri yayılım, istatistiksel metotlardan doğrusaldiskriminant analiz (LDA); gözetimsiz öğrenme olarak kümeleme tekniklerinden K-Means, Gaussian karışım yöntemi ve histogram işlemlerine dayalı görüntü iyileştirme teknikleri analiz için seçilmiştir. Kanamalı gölgelerin vurgulanması bu yaklaşımlarla yapılmış ve sonra elde edilen sonuçlar karşılaştırılmıştır.

Anahtar Kelimeler: Kümeleme, Sınıflama, Retina, Görüntü analizi, Tıp

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Asst. Prof. Dr. Turgay İBRİKÇİ and my co-supervisor Asst. Prof. Dr. Sami ARICA for their guidances and help throughout the thesis work. Many thanks for all of their supports, patiences, co-operations, suggestions on initiating, improving and completing this studyI am grateful to my committe members Assoc. Prof. Dr. Zekeriya TÜFEKÇİ, Asst. Prof. Dr. Murat AKSOYfor his endless support. I would like thank to Sevgi Eye Center’spersonnel especially Dr. Adnan BATMANOĞLU and his assistant for their help was able to take data.

I would also like to acknowledge valuable help from my friends and Çukurova University Department of Electrical and Electronics Engineering staff.

CONTENTS	PAGE
ABSTRACT.....	II
ÖZ.....	III
ACKNOWLEDGEMENTS	III
CONTENTS	IIV
LIST OF FIGURES	VI
LIST OF TABLES.....	VIII
NOTATIONS	IX
1.INTRODUCTION.....	1
2.ARTICLE REVIEWS.....	7
3.MATERIAL AND METHODS	13
3.1. Material	13
3.1.1.Fundus Fluorescein Imaging Technique.....	14
3.2. Methods.....	15
3.2.1. Supervised Learning.....	15
3.2.1.1. Neural Networks.....	18
3.2.1.2. Neural Networks Architectures	20
3.2.1.3. Multi Layer Neural Networks	21
3.2.1.4. Multi Layer Neural Networks Model	21
3.2.1.5. Delta Rule.....	22
3.2.1.6. Linear Discriminant Analysis.....	25
3.2.1.7. Feature Extraction.....	26
3.2.2. Unsupervised Learning.....	30
3.2.2.1. K-Means Clustering	31
3.2.2.2. Gaussian Mixture Model.....	34
3.2.2.3. EM Algorithm	37
3.2.3. ImageSegmentation Using Histogram Processing	39
3.2.3.1. Image Thresholding	41
3.2.3.2. Histogram (Contrast)Stretching.....	42
4.EXPERIMENTS,RESULTS AND DISCUSSIONS	45
4.1. Segmentation With LDA.....	47
4.2. Segmentation With Neural NetworkMethod	50
4.3. K-Means Clustering Results.....	54

4.4. Gaussian Mixture Method Results.....	57
4.5. Segmentation With Image Thresholding Method Results	59
4.6. Image Segmentation With Histogram Stretching Results.....	62
5.CONCLUSION	67
6.FUTURE WORKS	69
REFERENCES.....	71
BIOGRAPHY.....	75
APPENDICES.....	76

LIST OF FIGURES

PAGE

Figure 1.1. Simulation of what someone with diabetic retinopathy might see	1
Figure 1.2. Eye structure and affects of diabetic retinopathy on blood vessel.....	2
Figure 1.3. Comparison of leaking vessel and normal vessel	2
Figure 3.1. Some of data samples.....	13
Figure 3.2. Images taken from the FFA machine	15
Figure 3.3. Comparison of supervised and unsupervised classification	16
Figure 3.4. Biological neuran parts and their task in human body system	19
Figure 3.5. Artificial neuron parts for association these parts with biological neuron	19
Figure 3.6. Artificial neural networks structure	20
Figure 3.7. Linear discriminant analysis classification line for separating data group	25
Figure 3.8. μ_1 and μ_2 is mean value of data class. The distances on the the axes are shown seperability yields difference of between means	26
Figure 3.9. k initial "means" (in this case k=3) are randomly selected from the dataset	32
Figure 3.10. k clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means	33
Figure 3.11. The centroid of each of the k clusters becomes the new means	33
Figure 3.12. Steps 2 and 3 are repeated until convergence has been reached	33
Figure 3.13. Standart Gauss distribution curve	34
Figure 3.14. Clusters μ values.....	36
Figure 3.15. Probability density function for the truncated normal distribution	38
Figure 3.16. Gray scale fundus image histogram.....	41
Figure 3.17. Classes representation on a sample histogram	42
Figure 4.1. Two different form of FFA images.....	45
Figure 4.2. Training part sample for LDA classfier	48
Figure 4.3. Explanation of the LDA classifier results	48

Figure 4.4. Training part sample for backpropagation classifier	50
Figure 4.5. Explanation of the bacpropagation results	52
Figure 4.6. Explanation of GMM result.....	57
Figure 4.7. Comparison of histograms.....	60
Figure 4.8. Piecewise linear transfer function for using contrast stretching method histogram	62
Figure 4.9. Logarithmic transfer function for using contrast stretching method	64
Figure 4.10. Linear transfer function histogram compression algorithm application results histogram	66

LIST OF TABLES	PAGE
Table 4.1. The samples from dataset	45
Table 4.2. Train vector sample	47
Table 4.3. Samples of segmented images obtained by LDA	49
Table 4.4. Train vector sample for neural network classifier	51
Table 4.5. Segmented imaged are result of backpropagation algorithm	53
Table 4.6. The segmented result images obtained by K-means clustering method	56
Table 4.7. Segmented images obtain by GMM.....	58
Table 4.8. Segmented result by using image thresholding method.....	61
Table 4.9. Linear transfer function histogram stretching algorithmapplication results.....	63
Table 4.10.Logarithmic transfer function histogram stretching application results.....	65

NOTATIONS

W	: The synaptic weights
y	: Output matrix
x	: Input matrix
i	: Input node
j	: Output node
α	: Learning rate
$g(x)$	: Activation function
t_j	: Target output
h_j	: The weighted sum of the neuron's inputs
y_j	: Actual output
x_i	: i th input
μ_1	: First group mean
μ_2	: Second group mean
$x_i^{(j)}$	: Data Cluster center
c_j	: An indicator of the distance of the n data points from their respective cluster centre
n_i	: The number of pixel intensity
N	: The total number of pixels in the image
r_k	: The intensity value
n_k	: The number of pixels in the image with intensity r_k
$h(r_k)$	: The histogram of the digital image with gray level r_k
μ	: The mean or expectation (location of the peak)
σ^2	: The variance, the mean of the squared deviation
σ	: The standard deviation
y_j	: The neuron's activation function

1. INTRODUCTION

Diabetic retinopathy (DR) is the most common diabetic eye disease, which occurs when blood vessels in the retina bleed. It causes defecting of vision as from until stages of DR. The simulation of DR patient's vision is shown in Figure 1.1. We can explain DR these symptoms. Sometimes these vessels swell and leak fluid or even close off completely. In other cases, abnormal new blood vessels grow on the surface of the retina. The variation of glucose level in the blood causes the damage to the blood vessels. The thin blood vessels are damaged by the increased glucose level in blood. This can cause blindness in advanced stages of DR.



Figure 1.1. Simulation of what someone with diabetic retinopathy might see

The retina is a thin layer of light-sensitive tissue that lines at the back of the eye. Light rays are focused onto the retina, where they are transmitted to the brain and interpreted as the images. Eye structure and DR regions are illustrated in the Figure 1.2. The macula is a very small area at the center of the retina. It is the macula that is responsible for your pinpoint vision, allowing you to read, see or recognize a face. The surrounding part of the retina, called the peripheral retina, is responsible for your side or peripheral vision.

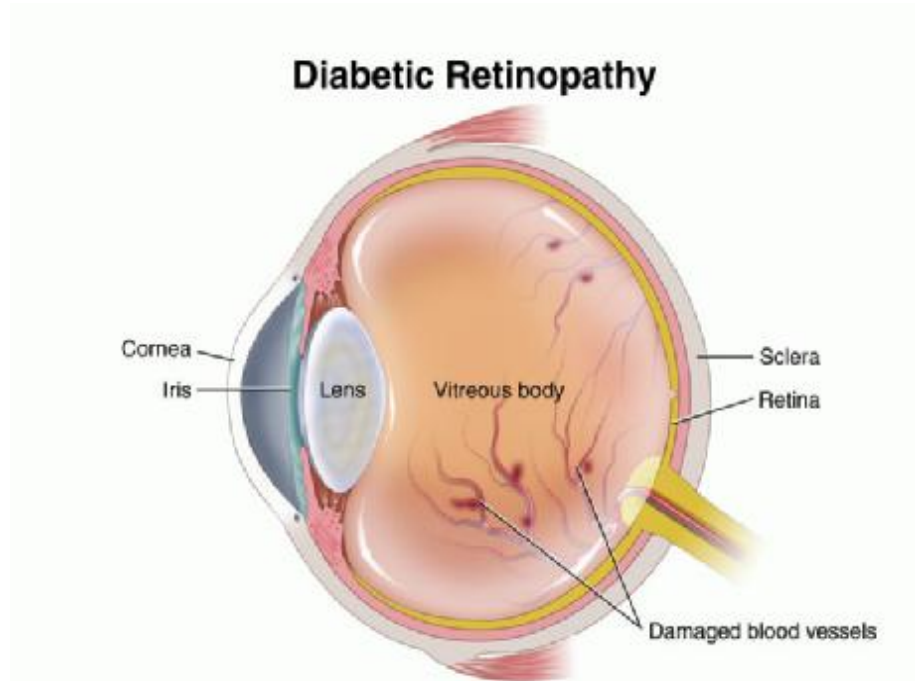


Figure 1.2. Eye structure and affects of diabetic retinopathy on blood vessel

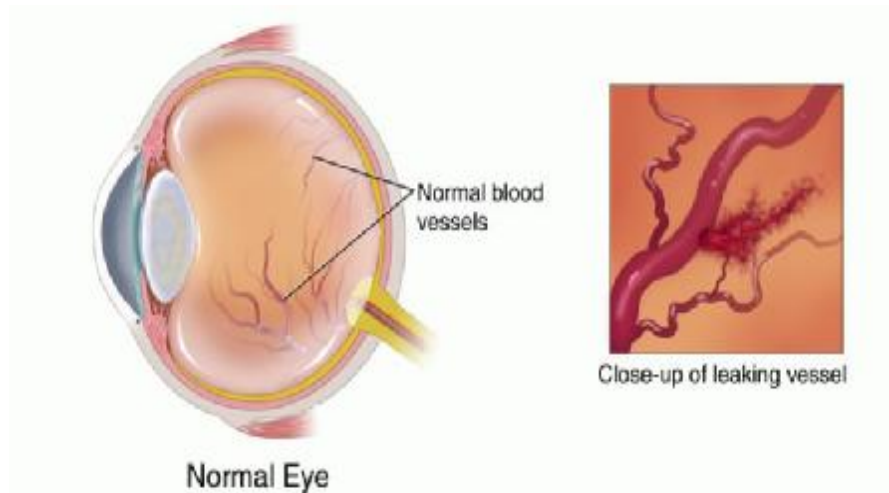


Figure 1.3. Comparison of leaking vessel and normal vessel

You can see normal blood vessels and leaking vessels with DR in Figure 1.3. Diabetic retinopathy affects blood vessels in the retina, leading to new abnormal blood vessel growth, leakage and bleeding. Diabetic retinopathy usually affects both eyes. People who have diabetic retinopathy often don't notice changes in their vision in the disease's early stages. But as it progresses, diabetic retinopathy usually causes vision loss that in many cases cannot be reversed, because of this early diagnosis is very important. For early detection of DR, one of the most important criteria is

screening of disease development. To prevent the progression of the retinopathy or blindness, some grading systems and classifications of diabetic retinopathy have been proposed to assess the severity. Generally, many patients can not notice this disease until its last phase. This situation prevents for effective treatment.

In recent years, researches increase about DR and early detection of this disease. Imaging techniques start to take role about this point like the other medical image analysis. Computer based retinal image analysis system extends doctors capability and productivity during clinical examination. One of the technique outshines the other imaging methods. This is called Fundus Fluorescein Angiography (FFA). Digital fundus imaging technique has been an important tool for detecting DR clinic examines and researches. The images which are taken from FFA machines, process with different methods for emphasizing DR regions. Image processing used for detecting DR clarify fundus images in this study. We can detect DR regions with different image processing and image clustering techniques. First, we explain these methods shortly and touch upon this topic generally.

Image processing technology has many applications in medical area. Image processing is a physical process used to convert an image signal into an image; is the use of computer algorithms to perform processing on digital images. It can perform the visual information test automatically and can offer technique support for establishing objective evaluation system. These techniques are most useful because they use for many different field image in addition to medical images. There is no conscious effort to improve the fidelity of the image with regard to some ideal form of the image. There are a wide variety of techniques for improving image quality. The contrast stretch, density slicing, edge enhancement, and spatial filtering are the more commonly used techniques in this field (Kumar, 2003).

Shome (2011) uses image enhancement techniques to improve the quality of retinal images. Enhancement refers to bring out the finer details of the image under test while emphasizing the features of interest. The goal of image enhancement is to extract the important characteristics from which a detailed description of the target is henceforth possible. Such an image can serve either for improved human perception or as an input to further automated image processing techniques.

The other way, image clustering which is used for this study, is also used in medical image processing field in recent years. There are many different methods. These techniques are useful for detecting diabetic retinopathy. Clustering analysis is based on partitioning a collection of data points into a number of subgroups, where the objects inside a cluster show a certain degree of closeness or similarity. It has been playing an important role in solving many problems in pattern recognition and image processing.

In addition to image enhancement method, unsupervised learning methods are preferred like K-Means and Gaussian Mixture Model (GMM). K-Means is a basic method for clustering application. Because of this we prefer this method for our data. K-Means algorithm classifies the input data points into multiple classes based on their inherent distance from each other. The algorithm assumes that the data features form a vector space and tries to find natural clustering in them. K-Means can be thought of as an algorithm relying on hard assignment of information to a given set of partitions. At every pass of the algorithm, each data value is assigned to the nearest partition based upon some similarity parameter such as Euclidean Distance of intensity. The partitions are then recalculated based on these hard assignments. With each successive pass, a data value can switch partitions, thus altering the values of the partitions at every pass. K-Means algorithms typically converge to a solution very quickly as opposed to other clustering algorithms.

Another unsupervised learning way is GMM for clustering. Clustering can also be made with using gauss distribution. Different GMMs are built with this distribution. GMM classifiers have been used in many fields. Gaussian mixture models are a semi-parametric alternative to non-parametric histograms and it has greater flexibility and precision in modelling the underlying distribution of sub-band coefficients. Gaussian mixture models are often used for data clustering. Clusters are assigned by selecting the component that maximizes the posterior probability. Like K-Means clustering, Gaussian mixture modeling uses an iterative algorithm that converges to a local optimum. Gaussian mixture modeling may be more appropriate than K-Means clustering when clusters have different sizes and correlation within them. Clustering using Gaussian mixture models is sometimes considered a soft

clustering method. The posterior probabilities for each point indicate that each data point has some probability of belonging to each cluster.

The neural networks is another approach used in image classification. The most widely applied neural network algorithm in image classification remains the feedforward backpropagation algorithm. The backpropagation algorithm is a generalization of the least mean square algorithm that modifies network weights to minimize the mean squared error between the desired and actual outputs of the network. Backpropagation is an example of supervised learning method. The network is trained using input train data for which inputs as well as desired outputs are known. After training, the network weights are frozen and can be used to compute output values for new input samples.

In addition to these methods we use the image enhancement methods. There are many image enhancement methods. Histogram processing is one of them. Histogram processing can be made with different methods for example stretching (contrast stretching), compression, equalization etc. Contrast stretching attempts to improve an image by stretching the range of intensity values it contains to make full use of possible values. Unlike histogram equalization, contrast stretching is restricted to a linear mapping of input to output values that it contains to make full use of possible values. The result is less dramatic, but tends to avoid the sometimes artificial appearance of equalized images.

In this thesis, the above classification, clustering and histogram processing methods were applied to enhance diabetic fundus images and tested for the sensitivity and specificity of recognition specificity and sensitivity of these features. The predictive capability of the system to correctly detect DR in fundus images was examined to determine whether a trained classify method could be used in the screening of diabetic patients. In addition to this, histogram processing techniques and Gaussian distribution function were used on fundus image. These methods of image processing were for building segmentation chart.

In the final analysis, computational learning and image enhancement techniques' results were compared to determine the best most valuable method for detecting the diabetic retinopathy disease.

2. ARTICLE REVIEWS

DR is a one of the most important disease by causing of diabetes. Because of this situation, it is a common topic for researchers in recent years. The visual loss is often a late symptom of advanced DR. DR investigated from different angles for researches. Genereally these studies can be about disease stages, automatic detection of DR, which technique is more suitable than the other and prevention of disease for medical researcher.

Screening of DR is an effective way of reducing preventable blindness. With the development of digital image processing techniques, computer-aided fundus image processing becomes useful to clinical work.If the duration of diabetes is longer, the prevalence of the diabetic retinopathy is also greater. Several preventive and therapeutic methods were reviewed to minimize the morbidity rate of diabetic retinopathy. Glycemic control is important at early detection of DR (Chen et al., 2009). Li et al. (2000) were used more than thirty fundus images for their study. Like as our study, their system were detected optic disk, blood vessels and exudates on fundus images. The system were detected most features with high accuracy. The position and size of the optic disk are obtained at the same time.Their algorithm gave the best result on exudates area than optical disk region.The methods which were used at their studies, was different from our methods.Kirsch's method were applied in the red and green components of the colorful FFA. The different parts in the processing resultsof the two components were detected as blood vessels. Morphological operators were applied to do the post processing for. The common parts except optic disk in the processing results by kirsch's method in red component and green component were detected as exudates.The results of the fundus image processing system were promising and could be beneficial to the mass screening of diabetic retinopathy.After they tested more than thirty images, the algorithms can extract most features of the fundus image successfully (Luo et al., 2000).

Considering to the increase in occurrence of diabetes, as well as to the fast advances in computer hardware and software technique, the development of a cost-effective PC-based retinal image analysis system offers a potential of being

very useful in the mass screening of diabetic retinopathy. Their system may not only provide objective diagnosis to the doctor, but also extend his capability and productivity during clinical examination(Liu, 1997).

Acharyal et al. (2009) were tried to detect DR automatically with using morphological image processing and support vector machine (SVM) techniques. They used 331 fundus images for their study. The patient's DR levels were different each other. They could collect under five groups: normal retina, mild non-proliferative diabetic retinopathy, moderate non-proliferative diabetic retinopathy, severe non-proliferative diabetic retinopathy, and proliferative diabetic retinopathy. Four salient features; blood vessels, microaneurysms, exudates, and hemorrhages were extracted from the raw images using image processing techniques and fed to the SVM for classification. Their system has a sensitivity of more than 82% and specificity of 86%.

The other study is prepared by Sanchez et al. (2007). It was about automatic detection of hard exudate. Hard exudates generally associate with DR. This can be seen at early stages of DR. Their algorithm which is based on Fisher's linear discriminant analysis and made using of colour information to perform the classification of retinal exudates. Their dataset has 58 retinal images. Dataset images were colorful. As a result, their algorithm were obtained a sensitivity of 88% with a mean number of 4.83 ± 4.64 false positives per image using the lesion-based performance evaluation criterion, and achieved an image-based classification accuracy of 100% (sensitivity of 100% and specificity of 100%).

Gardner et al.(1996) were used a neural network structure for their study. They used 13 fundus images. The dataset has normal images and diabetic images. They were detected features which were used for training step of classification. A neural network was tested for the sensitivity and specificity of recognition of determined features. The predictive capability of the system to correctly detect retinopathy in fundus images examined to determine whether a trained neural network could be used in the screening of diabetic patients. The detection rate for the classification of vessels from normal retina was 91.7%; detection rates for exudates and hemorrhages from normal retina were 93.1 % and 73.8% respectively with this

algorithm. The dataset classified 13 images were classified as of unsure diagnosis and disagreed with the original diagnosis attributable to the image in four 'normal' images and six diabetic images by ophthalmologist. These images were removed from the comparison with the neural network. For the detection of diabetic retinopathy the network achieved a sensitivity of 88.4% and specificity of 83.5% in comparison with the ophthalmologist.

In the object classification stage a three layer perceptron neural network with 18 input nodes corresponding to the feature set were used for Singh (2010). Experimentation with a hidden layer with a range of 2–35 hidden units to find the optimum architecture (15 hidden units) was done. A single output node was gave the final classification probability. They used a different neural network structure for their study. Fuzzy C-Means (FCM) clustering applied to the 75 color images of the prepared image dataset including 25 normal and 50 abnormal images. The outcome of the segmentation step was 3860 segmented regions consisting of 2366 exudates and 1494 non-exudates, which were considered negative cases. The regions were then labeled as exudates or non-exudates by a clinician to obtain a fully labeled dataset of examples. Two different learning methods, including standard backpropagation (BP) and scaled conjugate gradient (SCG) descent were investigated for. The output threshold gives the balance between sensitivity and specificity. Sensitivity and specificity were adjusted by changing the threshold values (Singh, 2010).

Digital image processing techniques helped to extract the location and size/level of abnormalities, gave an objective grade and compared the changes in objects in sequential images to Sinthanayothin (2002) for his study. High performance pre-processing of the colour images was performed. Automated image analysis systems were used to detect major landmarks of the retinal image (optic disc, blood vessels and fovea). Recursive region growing segmentation algorithms were combined with the use of a new technique, termed a 'Moat Operator', were used to automatically detect features of non-proliferative diabetic retinopathy (NPDR). The features included hemorrhages and microaneurysms (HMA), which were treated as one group, and hard exudates as another group. Sensitivity and

specificity data were calculated by comparison with an experienced fundoscopist. The algorithm for exudate recognition was applied to 30 retinal images of which 21 contained exudates and nine were without pathology. The sensitivity and specificity for exudate detection were 88.5% and 99.7%, respectively, when compared with the diagnosis of ophthalmologist. HMA were present in 14 retinal images. The algorithm achieved a sensitivity of 77.5% and specificity of 88.7% for detection of HMA. Fully automated computer algorithms were able to detect hard exudates and HMA. This paper presents encouraging results in automatic identification of important features of NPDR.

In their study, Saravanan et al. (2008) studied and presented a fast and robust method to correct contrast variation in retinal fundus imagery. Their technique uses a mixture of Gaussians to model the bias of the intensity variation. Typically a three or four component mixture was sufficient to characterize the principal variation due to the spherical geometry of the retina, the high-contrast reflection off the optic nerve and the darker macula. They compared the results with a non-parametric, filtering approach on a standard diabetic retinopathy database of 89 images. Their results indicated that a parametric approach using mixture Gaussian was better at contrast stretching in lesion regions making and was effective as a pre-processing step for manual and computer aided diagnostic techniques.

Yin et al. (2008) also presented an approach for medical image categorization based on Gaussian mixture model in their study. There were distinct differences on texture, shape and intensity characteristics among the images of different parts of body. They extracted the characteristic vectors of the training image set to learn the class model for each class, then categorized the test image using the Bayesian principle. The experimental results indicated that the method performed very well on CT image categorization. They achieved classification accuracy up to 97% in the experiment.

As seen from above mentioned studies, DR is a popular topic for researchers. Different methods are used for detecting this disease. Some of them are the same which are used for our study. All methods provide different results. Their accuracies are various. The previous studies give different ideas to us. They show to us

advantages and disadvantages of the algorithms.Theyhelped us in developingour methods. The previous studies and their results provide different viewpoint to us.

3. MATERIAL AND METHODS

3.1Material

Hemorrhagic eye images are acquired with fundus fluorescein angiography from 30 diabetic retinopathy patients. Some of data samples are shown in Figure 3.1. All data are taken from Sevgi Eye Center;a private hospital specialized for eye diseases in Adana, Turkey. The images were read and stored by the system both in color and gray forms.The mean e age of patients were 40. Although some of patients' laser operated in early stages of theirdiabetic retinopathy, the vessels of these patients continued bleeding.Eye images were also taken from number of normal subjects for classification purposes.

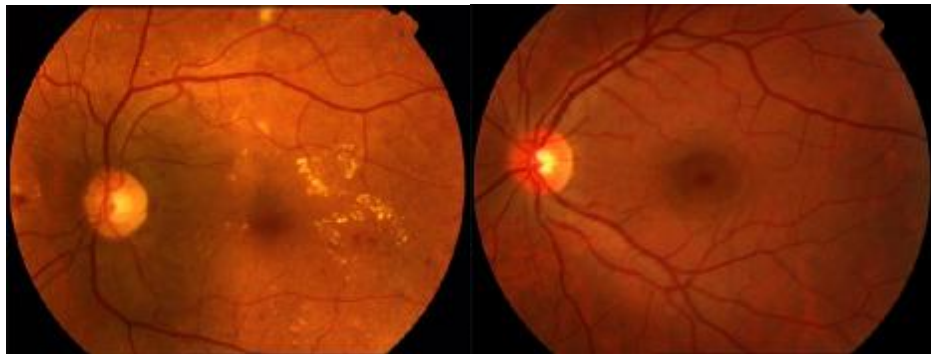


Figure 3.1. The images samples

Hence, we explain some characteristic for understanding DR fundus fluorescein images. Choroidal vessels are visible in the normal eye with near IR imaging and not much well at other wavelengths.The choroidal vessel contrast of the IR is usually dark on a light fundus background. Retinal vessels appear dark or have dark borders at all wavelengths in direct mode imaging. The optic disc rim appears dark at all wavelengths in confocal mode. The optic disc in pathological eyes can appear pale, especially in patients with dark pigmentation or pathology near the disc, such as a retinal detachment or staphyloma.

We defined fundus fluoresan image technique the previous paragraph. Different computational methods are used for this study. These methods are built in Matlab. Matlab computing language is used for processing these images. Matlab is a

high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis, and numerical computation. It allows matrix manipulation, plots the functions and implements the algorithms.

3.1.1 Fundus Fluorecein Imaging

Imaging techniques are very important and effective way for detecting some illness. Various medical devices are designed to human body's different parts. Sometimes, these devices need some medicine or different material for increasing the image and illness detection quality. Fundus fluoresceintechnique is one of them (Figure 3.2.).Fluorescein liquid is used for this imaging. We try to summarize fundus fluorescein angiography details in this section. Interpretation of retinal angiographic studies has been largely qualitative. The temporal properties associated with fluorescein fluorescence of normal and pathologic fundus features are quantified to potentiate pixel assignment and fundus feature quantitation for clinical studies where precise image metrology is vital. Fluorescein angiography images are digitized, and temporally sequential images are spatially registered with polynomial warping algorithms, allowing for the construction of a three-dimensional angiogram vector. Temporal profiles through spatially registered, temporally sequential pixels are computed. Fundus feature fluorescence behavior is quantified. Fundus features may be discriminated on the basis of spatio-temporal fluorescein fluorescence properties. Pixel assignment based on spatio-temporal relationships facilitates fundus are quantition(Berger, 1999).

In summary, we can collect this method function with some step. Fluorescein angiography is a fundal photography, performed in rapid sequence following intravenous injection of fluorescein dye. Sodium fluorescein is an organic dye. It provides following information's:

1. The blood flow characteristics in the vessels as the dye reach and circulate through the retina and choroid.

2. It provides fine details of the retinal circulation and pigment epithelium that may not otherwise be visible.

3. It records functional integrity retinal vessels, so their assessment can be done.

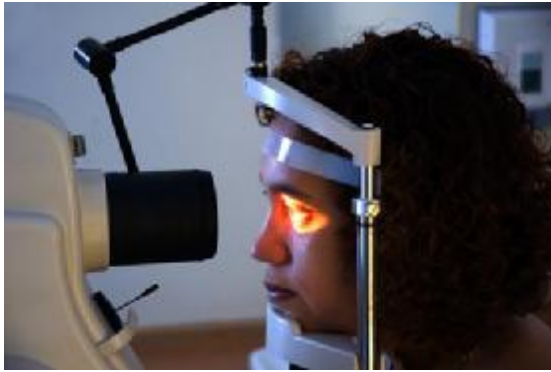


Figure 3.2. Images taken from the fundus fluoresceinangiographi machine

There are two circulations within the fundus. These are choroidal circulation and retinal circulation. The fluorescein freely leaks out of the fenestrated choroidal capillaries, and from there through Bruch's membrane. However, tight junctions between retinal pigment epithelium (RPE) cells prevent dye reaching the retina.

The retinal blood vessel endothelial cells are joined by tight junctions which prevent leakage of fluorescein into the retina. This constitutes the blood retina barrier. Any leakage from the retinal vessels is considered as an abnormal.

3.2. Methods

3.2.1. Supervised Learning

Machine learning algorithms are described into three groups as supervised, unsupervised and reinforcement learning. We emphasize supervised and unsupervised learning types especially. The distinction is drawn from how the learner classifies data. Supervised and unsupervised learning ways are compared in Figure 3.3. In this section we explain supervised learning method.

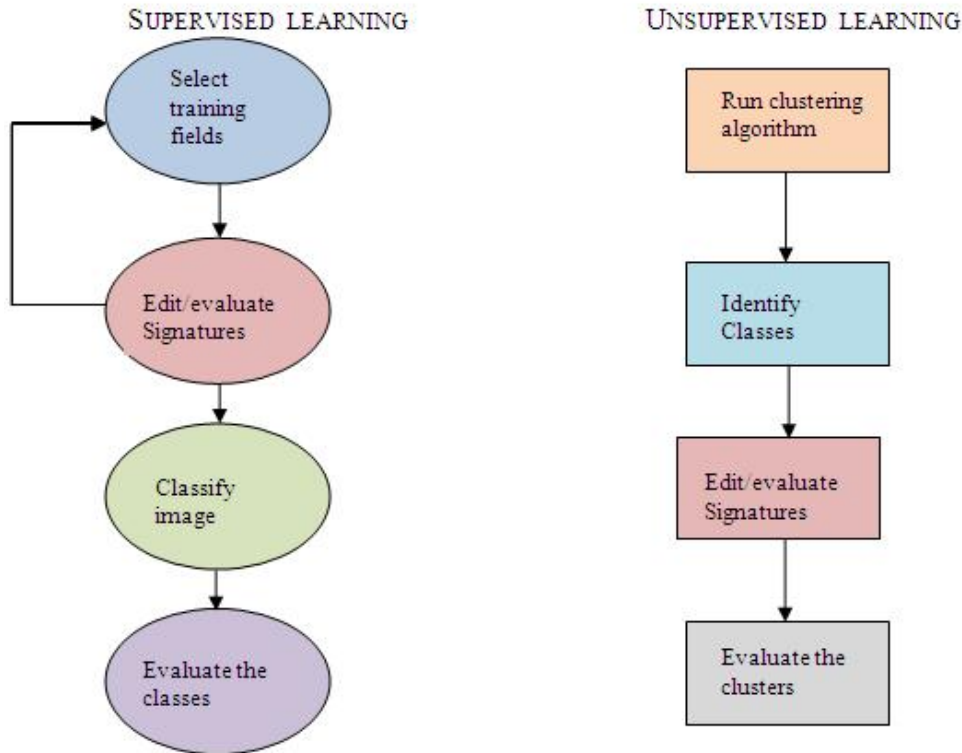


Figure 3.3. Comparison of supervised and unsupervised classification

Supervised learning is the machine learning task of inferring a function from supervised training data. The training data consist of a set of training examples. In supervised learning, each example is a pair consisting of an input object and a desired output value. A supervised learning algorithm analyzes the training data and produces an inferred function, which is called a classifier or a regression function. The inferred function should predict the correct output value for any valid input object. This requires the learning algorithm to generalize from the training data to unseen situations in a reasonable way. In order to solve a given problem of supervised learning, one has to perform the following steps:

1. Determine the type of training examples. Before doing anything else, the user should decide what kind of data is to be used as a training set. In the case of handwriting analysis, for example, this might be a single handwritten character, an entire handwritten word, or an entire line of handwriting.

2. Gather a training set. The training set needs to be representative of the real-world use of the function. Thus, a set of input objects is gathered and corresponding outputs are also gathered, either from human experts or from measurements.

3. Determine the input feature representation of the learned function. The accuracy of the learned function depends strongly on how the input object is represented. Typically, the input object is transformed into a feature vector, which contains a number of features that are descriptive of the object. The number of features should not be too large, because of the curse of dimensionality; but should contain enough information to accurately predict the output.

4. Determine the structure of the learned function and corresponding learning algorithm. For example, the engineer may choose to use support vector machines or decision trees.

5. Complete the design. Run the learning algorithm on the gathered training set. Some supervised learning algorithms require the user to determine certain control parameters. These parameters may be adjusted by optimizing performance on a subset of the training set, or via cross-validation.

6. Evaluate the accuracy of the learned function. After parameter adjustment and learning, the performance of the resulting function should be measured on a test set that is separate from the training set.

In supervised algorithms, the classes are predetermined. These classes can be conceived of as a finite set, previously arrived at by a human. In practice, a certain segment of data will be labelled with these classifications. The machine learner's task is to search for patterns and construct mathematical models. These models then are evaluated on the basis of their predictive capacity in relation to measures of variance in the data itself. Many of the methods referenced in the documentation (decision tree induction, naive Bayes, etc) are examples of supervised learning techniques.

3.2.1.1 Neural Networks

Neural network (artificial neural network) is a mathematical model or computational model that is inspired by the structure and/or functional aspects of biological neural networks. A neural network is configured for a specific application, such as pattern recognition or data classification, through a learning process.

A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. In most cases, an artificial neural network is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. This artificial neuron is a computational model inspired in the natural neurons. Natural neurons receive signals through synapses located on the dendrites or membrane of the neuron. When the signals received are strong enough (surpass a certain threshold), the neuron is activated and emits a signal through the axon. This signal might be sent to another synapse, and might activate other neurons.

The complexity of real neurons is highly abstracted when modelling artificial neurons. These basically consist of inputs (like synapses), which are multiplied by weights (strength of the respective signals), and then computed by a mathematical function which determines the activation of the neuron (Figure 3.4). Another function (which may be the identity) computes the output of the artificial neuron (sometimes in dependence of a certain threshold). ANNs combine artificial neurons in order to process information.

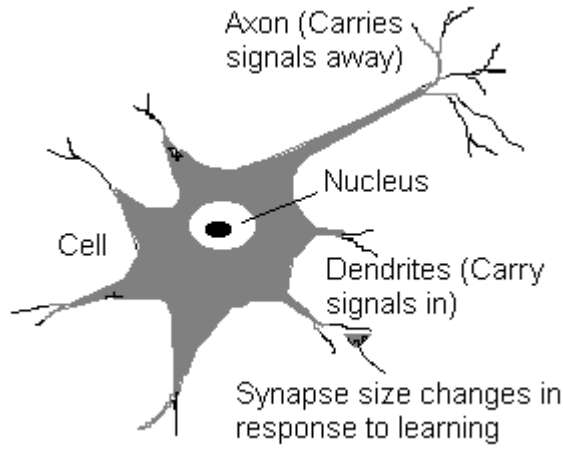


Figure 3.4. Biological neuron parts and their task in human body system

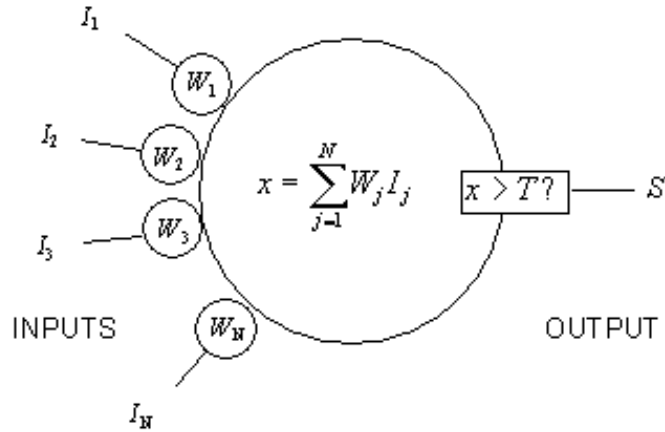


Figure 3.5. Artificial neuron parts for association these parts with biological neuron

ANN analysis may be more able to cope with the variability because this system employs mathematical weights to decide the probability of input data (for example, a feature on a digitised fundus image) belonging to a particular output (Figure3.5). The 'weights' are adjusted by training the network with data of known output. Subsequently, the network can be tested with new data, providing a probability of such data belonging to a particular output.

3.2.1.2. Neural Networks Architectures

The basic architecture consists of three types of neuron layers: input, hidden, and output layers. In feed-forward networks, the signal flow is from input to output units, strictly in a feed-forward direction. The data processing can extend over multiple units, but no feedback connections are present. Recurrent networks contain feedback connections. Contrary to feed-forward networks, the dynamical properties of the network are important. In some cases, the activation values of the units undergo a relaxation process such that the network will evolve to a stable state in which these activations do not change anymore. In other applications, the changes of the activation values of the output neurons are significant, such that the dynamical behavior constitutes the output of the network. There are several other neural network architectures (Elman network, adaptive resonance theory maps, competitive networks, etc.), depending on the properties and requirement of the application. Neural networks are based on three architectures in shown Figure 3.6. These are single layer perceptron, multi layer perceptron and recurrent. All networks type basic charts are shown below.

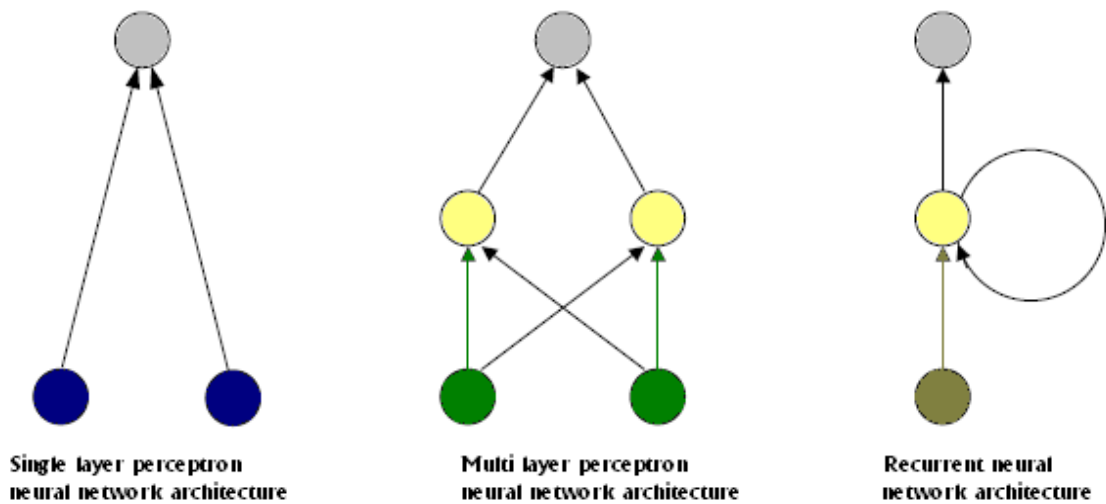


Figure 3.6. Artificial neural networks structure

3.2.1.3 Multi Layer Neural Networks

Multilayer networks solve the classification problems for nonlinear sets by employing hidden layers, whose neurons are not directly connected to the output. The additional hidden layers can be interpreted geometrically as additional hyperplanes, which enhance the separation capacity of the network. The training occurs in a supervised style. The basic idea is to present the input vector to the network; calculate in the forward direction the output of each layer and the final output of the network. For the output layer, the desired values are known and therefore the weights can be adjusted as for a single layer network; in the case of the backpropagation algorithm according to the gradient decent rule. To calculate the weight changes in the hidden layer, the error in the output layer is backpropagated to these layers according to the connecting weights. This process is repeated for each sample in the training set. One cycle through the training set is called an epoch. The number of epoches needed to train the network depends on various parameters, especially on the error calculated in the output layer.

3.2.1.4 Multilayer Neural Network Model

We used backpropagation method for classifying our dataset. Backpropagation is a common method of training artificial neural networks so as to minimize the objective function. Backpropagation can also be considered as a generalization of the delta rule for non-linear activation functions and multilayer networks. We tried to explain supervised and unsupervised learning in previous sections. It is a supervised learning method, and is a generalization of the delta rule. It requires a dataset of the desired output for many inputs, making up the training set. It is most useful for feed-forward networks (networks that have no feedback, or simply, that have no connections that loop). The term is an abbreviation for "backward propagation of errors". Backpropagation requires that the activation function used by the artificial neurons (or nodes) be differentiable.

3.2.1.5. Delta Rule

The delta rule is a gradient descent learning rule for updating the weights of the artificial neurons in a single-layer perceptron. It is a special case of the more general backpropagation algorithm. For a neuron j with activation function $g(x)$ the delta rule for j 's i th weight w_{ji} is given by Equation 3.1.

$$\Delta w_{ji} = a (t_j y_j) g'(h_j) x_i \quad (3.1.)$$

a is a small constant called learning rate

$g(x)$ is the neuron's activation function

t_j is the target output

h_j is the weighted sum of the neuron's inputs

y_j is the actual output

x_i is the i th input.

It holds that Equation 3.2. and 3.3.

$$h_j = \sum x_i w_{ji} \quad (3.2.)$$

$$y_j = g(h_j) \quad (3.3.)$$

The delta rule is commonly stated in simplified form for a perceptron with a linear activation function (Equation 3.4).

$$\Delta w_{ji} = a (t_j y_j) x_i \quad (3.4.)$$

The delta rule is derived by attempting to minimize the error in the output of the perceptron through gradient descent. The error for a perceptron with j outputs can be measured (Equation 3.5).

$$E = \sum_j \frac{1}{2} (t_j - y_j)^2 \quad (3.5.)$$

In this case, we wish to move through weight space of the neuron (the space of all possible values of all of the neuron's weights) in proportion to the gradient of the error function with respect to each weight. In order to do that, we calculate the partial derivative of the error with respect to each weight. Because, we are only concerning ourselves with the j^{th} neuron, we can substitute the error formula above while omitting the summation (Equation 3.6):

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial(\frac{1}{2}(t_j - y_j)^2)}{\partial w_{ji}} \frac{\partial y_i}{\partial w_{ji}} \quad (3.6.)$$

Next, we use the chain rule to split this into two derivatives (Equation 3.7):

$$= \frac{\partial(\frac{1}{2}(t_j - y_j)^2)}{\partial y_i} \frac{\partial y_i}{\partial w_{ji}} \quad (3.7.)$$

To find the left derivative, we simply apply the general power rule (Equation 3.8):

$$= -(t_j - y_j) \frac{\partial y_i}{\partial w_{ji}} \quad (3.8.)$$

To find the right derivative, we again apply the chain rule, this time differentiating with respect to the total input to j, h_j (Equation 3.9):

$$= -(t_j - y_j) \frac{\partial y_i}{\partial w_{ji}} \frac{\partial h_j}{\partial w_{ji}} \quad (3.9.)$$

Note that the output of the neuron y_j is just the neuron's activation function $g()$ applied to the neuron's input h_j . We can therefore write the derivative of y_j with respect to h_j simply as $g'()$'s first derivative (Equation 3.10):

$$= -(t_j - y_j)g'(h_j)\frac{\partial h_j}{\partial w_{ji}} \quad (3.10.)$$

Next we rewrite h_j in the last term as the sum over all k weights of each weight w_{jk} times its corresponding input x_k (Equation 3.11):

$$= -(t_j - y_j)g'(h_j)\frac{\partial(\sum_k x_k w_{jk})}{\partial w_{ji}} \quad (3.11.)$$

Because we are only concerned with the i^{th} weight, the only term of the summation that is relevant is $x_i w_{ji}$ (Equation 3.12). Clearly,

$$\frac{\partial x_i w_{ji}}{\partial w_{ji}} = x_i \quad (3.12.)$$

giving us our final equation for the gradient (Equation 3.13):

$$\frac{\partial E}{\partial w_{ji}} = -(t_j - y_j)g'(h_j)x_i \quad (3.13.)$$

As noted above, gradient descent tells us that our change for each weight should be proportional to the gradient. Choosing proportionality constant a and eliminating the minus sign to enable us to move the weight in the negative direction of the gradient to minimize error, we arrive at Equation 3.14:

$$\Delta w_{ji} = a(t_j - y_j)g'(h_j)x_i \quad (3.14.)$$

3.2.1.6.Linear Discriminat Analysis (LDA)

LDA is method used in statistics, pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of objects or events. We preferred this classifier because our data have 2 different classes. The resulting combination may be used as a linear classifier or more commonly, for dimensionality reduction before later classification. The explanation of linear classifier which is separated two or more data groups with a linear line (Figure 3.7).

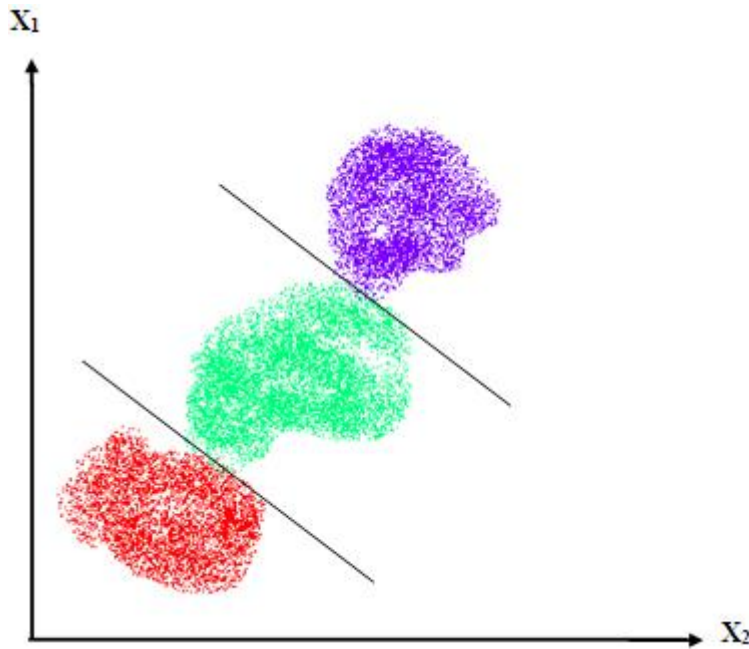


Figure 3.7.Linear discriminant analysis classification line for separating data group

Discriminant analysis has several interrelated objectives that include: identification of how one or more groups of observations differ as described by one or more, usually interrelated variables; assignment or classification of new observations to one or another of the groups based on the values of the variables.

LDA is another technique used for dimension reduction however, it is a supervised learning algorithm. Therefore, it considers the class labels. When reducing the dimensions, LDA seeks to maximize the separation among the different groups by preserving as much as the class discriminatory information as possible. This

means that we must have some measure of the separation between the groups. One possible way is to maximize the distance between the means of the groups when they are projected onto line. However, maximizing the distance between the projection means does not always maximize the distance between the groups because it does not take into account the variability within each group. Instead of this naive approach, LDA maximizes the distance between the projection means while at the same time minimizing the scatter within the group (Figure 3.8).

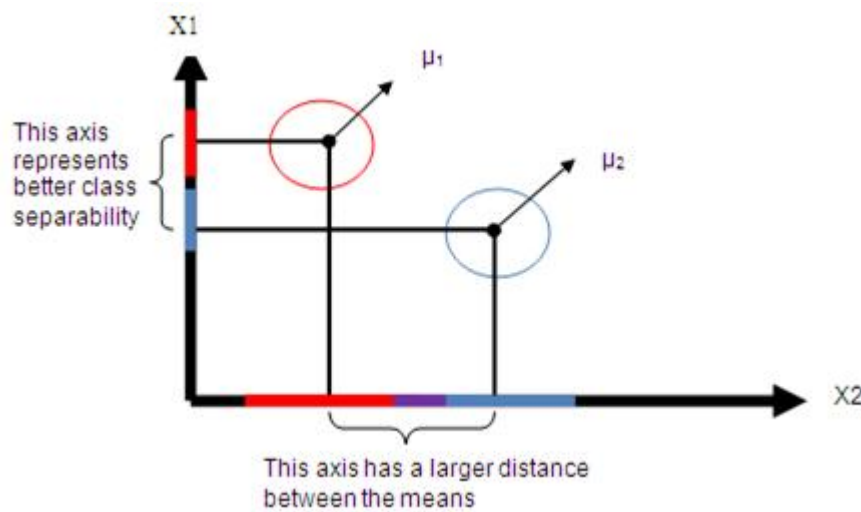


Figure 3.8. μ_1 and μ_2 is mean value of data class

The distances on the the axes are shown seperability yields difference of between means.

3.2.1.7.Feature Extraction

Feature extraction is a special form of dimensionality reduction for pattern recongnition and image processing. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant then the input data will be transformed into a reduced representation set of features (called features vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to

perform the desired task using this reduced representation instead of the full size input.

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which overfits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems.

Four features are selected kurtosis, skewness, median, and variance for our application. Firstly, we can try to explain these methods. Median and variance are main statistical function. We can benefit from lost of mathematical and statistical solutions. First feature is variance. Variance is one of the base point for statistical solutions.

The variance is one of the moments of a distribution. The variance is a parameter describing in part either the actual probability distribution of an observed population of numbers, or the theoretical probability distribution of a sample of numbers. In the latter case a sample of data from such a distribution can be used to construct an estimate of its variance: in the simplest cases this estimate can be the sample variance.

The variance of a random variable or distribution is the expectation, or mean, of the squared deviation of that variable from its expected value or mean. Thus the variance is a measure of the amount of variation of the values of that variable, taking account of all possible values and their probabilities or weightings.

Median is described as the numerical value separating the higher half of a sample, a population, or a probability distribution, from the lower half. The median of a finite list of numbers can be found by arranging all the observations from lowest value to highest value and picking the middle one. If there is an even number of observations, then there is no single middle value; the median is then usually defined to be the mean of the two middle values.

In a sample of data, or a finite population, there may be no member of the sample whose value is identical to the median, if there is such a member, there may be more than one so that the median may not uniquely identify a sample member. Nonetheless, the value of the median is uniquely determined with the usual definition. A related concept, in which the outcome is forced to correspond to a member of the sample, is the medoid.

At most, half the populations have values less than the median, and, at most, half have values greater than the median. If both groups contain less than half the population, then some of the population is exactly equal to the median.

The median can be used as a measure of location when a distribution is skewed, when end-values are not known, or when one requires reduced importance to be attached to outliers, e.g., because they may be measurement errors.

Skewness and kurtosis a fundamental task in many statistical analyses is to characterize the location and variability of a data set. Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable. The skewness value can be positive or negative, or even undefined. Qualitatively, a negative skew indicates that the tail on the left side of the probability density function is longer than the right side and the bulk of the values lie to the right of the mean. A positive skew indicates that the tail on the right side is longer than the left side and the bulk of the values lie to the left of the mean. A zero value indicates that the values are relatively evenly distributed on both sides of the mean, typically but not necessarily implying a symmetric distribution.

We firstly explain skewness feature. Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point. The histogram is an effective graphical technique for showing both the skewness and kurtosis of data set. For univariate data $Y_1, Y_2, \dots, Y_N$, the Equation 3.15 for skewness is:

$$skewness = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^3}{(N-1)s^3} \quad (3.15.)$$

where $\bar{Y}$ is the mean, s is the standard deviation, and N is the number of data points. The skewness for a normal distribution is zero, and any symmetric data should have skewness near zero. Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicate data that are skewed right. By skewed left, we mean that the left tail is long relative to the right tail. Skewed right means that the right tail is long relative to the left tail. Some measurements have a lower bound and are skewed right. For example, in reliability studies, failure times cannot be negative.

Kurtosis is a measure of how outlier-prone a distribution. The kurtosis of the normal distribution is 3. Distributions that are more outlier-prone than the normal distribution have kurtosis greater than 3; distributions that are less outlier-prone have kurtosis less than 3.

Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak. A uniform distribution would be the extreme case.

For univariate data $Y_1, Y_2, \dots, Y_N$, the Equation 3.16 for kurtosis is:

$$kurtosis = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^4}{(N-1)s^4} \quad (3.16.)$$

where $\bar{Y}$ is the mean, s is the standard deviation, and N is the number of data points. From a practical standpoint, it is probably most useful to think of kurtosis as specifically measuring the heaviness of the tail i.e. the extent to which the distribution of the variable falls off relatively slowly near the extremes, or equivalently the propensity by which a random value in the distribution is likely to fall several standard deviations away from the mean. Heavy-tailed distributions are useful in a wide variety of contexts, e.g. for estimation that is robust to small errors in the model assumptions or to occasional erroneous data points. High-kurtosis

distributions do generally have heavy tails, a fact that can be seen directly in the standard formula defining the kurtosis. On the other hand, there are large numbers of exceptional cases where high kurtosis and sharp peaks do not correspond: For example, the pointy-peaked Laplace distribution has an excess kurtosis of 3, whereas the squat, round-peaked Student's t-distribution with degrees of freedom $\nu = 4$ has infinite kurtosis. Furthermore, it is not always intuitively obvious what peakedness means particularly as distinguished from low variance. We tried to explain our feature vector criteria as above.

3.2.2. Unsupervised Learning

Unsupervised learning refers to the problem of trying to find hidden structure in unlabeled data. Since the examples given to the learner are unlabeled, there is no error or reward signal to evaluate a potential solution. This distinguishes unsupervised learning from supervised learning and reinforcement learning.

Unsupervised learning is closely related to the problem of density estimation in statistics. However unsupervised learning also encompasses many other techniques that seek to summarize and explain key features of the data. Many methods employed in unsupervised learning are based on data mining methods used to preprocess data.

Approaches to unsupervised learning include:

- clustering (e.g., k-means, mixture models, hierarchical clustering),
- blind signal separation using feature extraction techniques for dimensionality reduction (e.g., Principal component analysis).

Unsupervised learners are not provided with classifications. In fact, the basic task of unsupervised learning is to develop classification labels automatically. Unsupervised algorithms seek out similarity between pieces of data in order to determine whether they can be characterized as forming a group. These groups are

termed clusters, and there are whole families of clustering machine learning techniques.

In unsupervised classification, often known as 'cluster analysis' the machine is not told how the texts are grouped. Its task is to arrive at some grouping of the data. In a very common of cluster analysis (K-means), the machine is told in advance how many clusters it should form a potentially difficult and arbitrary decision to make.

It is apparent from this minimal account that the machine has much less to go on in unsupervised classification. It has to start somewhere, and its algorithms try in iterative ways to reach a stable configuration that makes sense. The results vary widely and may be completely off if the first steps are wrong.

3.2.2.1. K-Means Clustering

K-Means is the most popular unsupervised learning algorithms that can solve the clustering problems easily. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids for each cluster. These centroids should be placed in a cunning way because of different location causes different result. The better choice is to place them as much as possible far away from each other.

The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to recalculate k new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more.

Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function is shown in Equation 3.17.

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (3.17.)$$

where $\|x_i^{(j)} - c_j\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j , is an indicator of the distance of the n data points from their respective cluster centres. The algorithm is composed of the following steps:

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

We can represent algorithm with a simple drawing. K-Means flow chart can be shown like as Figure 3.9, Figure 3.10, Figure 3.11 and Figure 3.12.

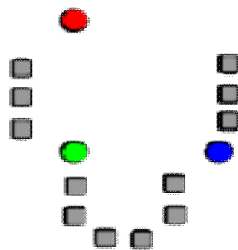


Figure 3.9 k initial "means" (in this case k=3) are randomly selected from the data set

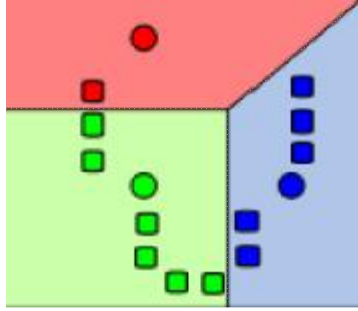


Figure 3.10. k clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means

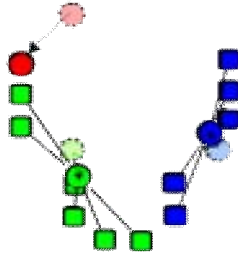


Figure 3.11. The centroid of each of the k clusters becomes the new means

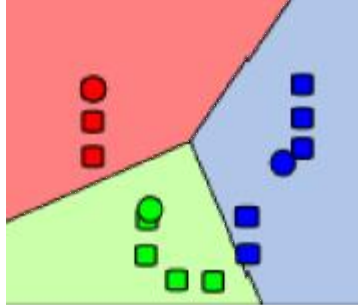


Figure 3.12. Steps 2 and 3 are repeated until convergence has been reached

There is no guarantee that it will converge to the global optimum, and the result may depend on the initial clusters. As the algorithm is usually very fast, it is common to run it multiple times with different starting conditions. However, in the worst case, K-means can be very slow to converge: in particular it has been shown that there exist certain point sets, even in 2 dimensions, on which K-Means takes exponential time, that is $2^{\Omega(n)}$, to converge. These point sets do not seem to arise in practice: this is corroborated by the fact that the smoothed running time of K-Means is polynomial.

3.2.2.2. Gaussian Mixture Model

A mixture distribution is the probability distribution of a random variable whose values can be interpreted as being derived in a simple way from an underlying set of other random variables. In particular, the final outcome value is selected at random from among the underlying values, with a certain probability of selection being associated with each. The underlying random variables may be random vectors, each having the same dimension, in which case the mixture distribution is a multivariate distribution.

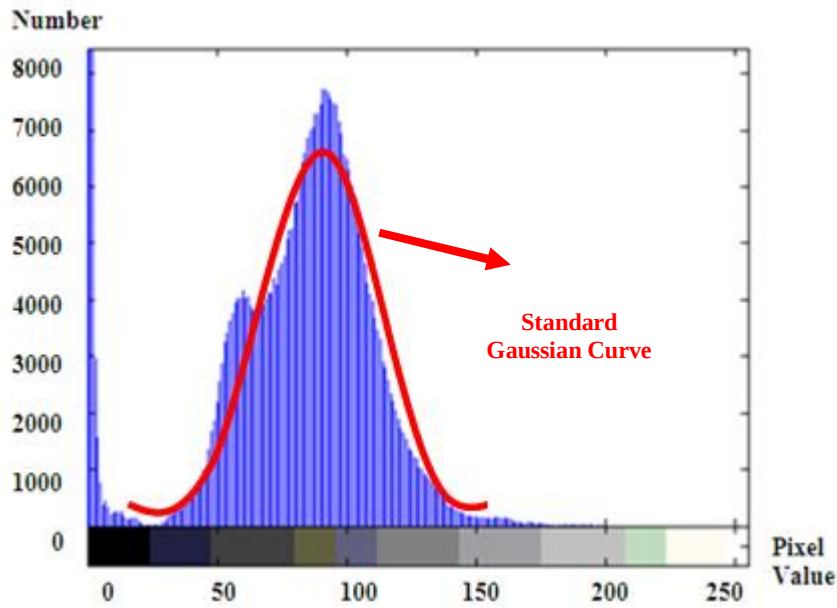


Figure 3.13. Standart Gauss distribution curve

In cases where each of the underlying random variables is continuous, the outcome variable will also be continuous and its probability density function is sometimes referred to as a mixture density (Figure 3.13). The cumulative distribution function (and the probability density function if it exists) can be expressed as a convex combination of other distribution functions and density functions. The individual distributions that are combined to form the mixture distribution are called the mixture components, and the probabilities (or weights) associated with each component are called the mixture weights. The number of components in mixture

distribution is often restricted to being finite, although in some cases the components may be countable. More general cases, as well as the countable case, are treated under the title of compound distributions.

A distinction needs to be made between a random variable whose distribution function or density is the sum of a set of components and a random variable whose value is the sum of the values of two or more underlying random variables, in which case the distribution is given by the convolution operator. As an example, the sum of two normally-distributed random variables, each with different means, will still be a normal distribution. On the other hand, a mixture density created as a mixture of two normal distributions with different means will have two peaks provided that the two means are far enough apart, showing that this distribution is radically different from a normal distribution.

Mixture distributions arise in many contexts in the literature and arise naturally where a statistical population contains two or more sub-populations. They are also sometimes used as a means of representing non-normal distributions. Data analysis concerning statistical models involving mixture distributions is discussed under the title of mixture models, while the present article concentrates on simple probabilistic and statistical properties of mixture distributions and how these relate to properties of the underlying distributions.

The most widely used clustering method of this kind is the one based on learning a mixture of Gaussians. The normal (or Gaussian) distribution is a continuous probability distribution that has a bell-shaped probability density function, known as the Gaussian function in Equation 3.18 or it is called the bell curve.

$$f(x) = \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-m)^2}{2s^2}} \quad (3.18.)$$

where parameter m is the mean or expectation (location of the peak) and s^2 is the variance, the mean of the squared deviation, (a measure of the width of the distribution). s is the standard deviation. The distribution with $m = 0$ and $s^2 = 1$ is

called the standard normal. A normal distribution is often used as a first approximation to describe real-valued random variables that cluster around a single mean value. Each Gaussian has its own mean and covariance matrix that has to be estimated separately. Unlike in the case with just one Gaussian, you do not know which training sample contributes to which Gaussian and hence the existing formulas for mean and covariance matrix are not applicable. Gaussian mixture models are formed by combining multivariate normal density components.

Gaussian mixture models are often used for data clustering. Clusters are assigned by selecting the component that maximizes the posterior probability. Like K-means clustering, Gaussian mixture modeling uses an iterative algorithm that converges to a local optimum. Gaussian mixture modeling may be more appropriate than K-means clustering when clusters have different sizes and correlation within them. Clustering using Gaussian mixture models is sometimes considered a soft clustering method. The posterior probabilities for each point indicate that each data point has some probability of belonging to each cluster.

We can actually consider clusters as Gaussian distributions centred on their barycentres, as we can see in this picture, where the grey circle represents the first variance of the distribution (Figure 3.14):

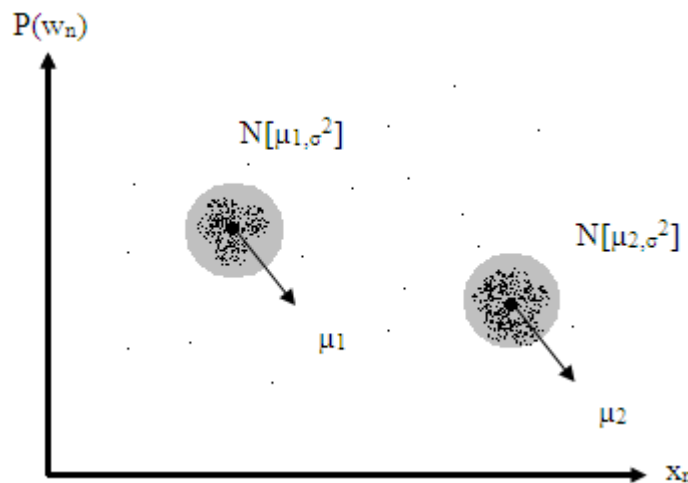


Figure 3.14. Clusters μ values

We used MATLAB gmdistribution class for applying Gauss distribution. It fit data using expectation maximization (EM) algorithm, which assigns posterior

probabilities to each component density with respect to each observation. The fitting method uses an iterative algorithm that converges to a local optimum. Clustering using Gaussian mixture models is sometimes considered a soft clustering method. The posterior probabilities for each point indicate that each data point has some probability of belonging to each cluster. The `gmdistribution` constructors create Gaussian mixture models with specified means, covariances, and mixture proportions. We explain expectation maximization (EM) algorithm in the next section. EM algorithm is related with Gaussian distribution.

3.2.2.3. Expectation Maximization (EM) Algorithm

The linear algorithm which is used in practice to find the mixture of Gaussians that can model the data set is called EM (Expectation-Maximization). EM is a well established clustering algorithm in the statistic community. EM is a distance-based algorithm that assumes the data set can be modeled as a linear combination of multivariate normal distributions and the algorithm finds the distributions parameters that maximize a model quality measure called log likelihood.

EM is chosen to cluster data for following reason among others:

- It has a strong statistical basis.
- It is linear in database size.
- It is robust to noisy data.
- It can except the desired number of clusters as input
- It can handle high dimentionalitiy
- It converges fast given a good initialization.(Abbas, 2008)

The basic approach and logic of this clustering method is as follows. Suppose you measure a single continuous variable in a large sample of observations. Further, suppose that the sample consists of two clusters of observations with different means (and perhaps different standard deviations); within each sample, the distribution of

values for the continuous variable follows the normal distribution. Probability density function is variable for different cases in Figure 3.15.

Probability density function for the truncated normal distribution for different sets of parameters. In all cases $a = -10$ and $b = 10$ for the black: $\mu = -8, \sigma = 2$; blue: $\mu = 0, \sigma = 2$; red: $\mu = 9, \sigma = 10$; orange: $\mu = 0, \sigma = 10$.

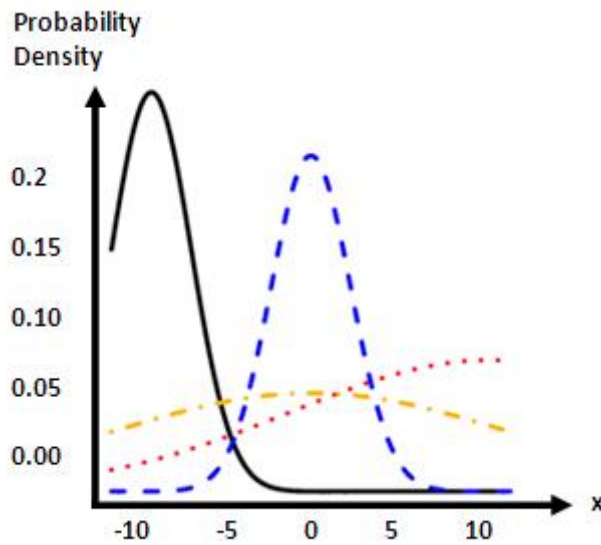


Figure 3.15. Probability density function for the truncated normal distribution

The goal of EM clustering is to estimate the means and standard deviations for each cluster so as to maximize the likelihood of the observed data (distribution). Put another way, the EM algorithm attempts to approximate the observed distributions of values based on mixtures of different distributions in different clusters. With the implementation of the EM algorithm in some computer programs, you may be able to select (for continuous variables) different distributions such as the normal, log-normal, and Poisson distributions.

The EM algorithm can also accommodate categorical variables. The method will at first randomly assign different probabilities (weights, to be precise) to each class or category, for each cluster. In successive iterations, these probabilities are refined (adjusted) to maximize the likelihood of the data given the specified number of clusters.

3.2.3. Image Segmentation Using Histogram Processing

Segmentation is the process of partitioning a digital image into multiple segments. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics.

Image segmentation can be different ways which histogram processing method is one of them. Histogram processing is the act of altering an image by modifying its histogram. Common uses of histogram processing include normalization by which one makes the histogram of an image as flat as possible. This is also known as contrast enhancement. Intensity transformation functions based on information extracted from image such as enhancement, compression, segmentation and description. We use histogram thresholding and contrast (histogram) stretching methods for our image data segmentations.

An image histogram is a type of histogram that acts as a graphical representation of the tonal distribution in a digital image. It plots the number of pixels for each tonal value. By looking at the histogram for a specific image a viewer will be able to judge the entire tonal distribution at a glance.

The horizontal axis of the graph represents the tonal variations, while the vertical axis represents the number of pixels in that particular tone. The left side of the horizontal axis represents the black and dark areas, the middle represents medium grey and the right hand side represents light and pure white areas. The vertical axis represents the size of the area that is captured in each one of these zones. Thus, the histogram for a very dark image will have the majority of its data points on the left side and center of the graph. Conversely, the histogram for a very bright image with few dark areas and/or shadows will have most of its data points on the right side and center of the graph.

The histogram of digital image with the intensity levels in the range $[0, L-1]$ is a discrete function in Equation 3.19.

$$h(r_k) = n_k \quad (3.19.)$$

where

- r_k is the intensity value.
- n_k is the number of pixels in the image with intensity r_k .
- $h(r_k)$ is the histogram of the digital image with gray level r_k

Histograms are frequently normalized by the total number of pixels in the image. Assuming an $M \times N$ image (Equation 3.20), it has a normalized histogram.

$$p(r_k) = \frac{n_k}{MN} \quad (3.20.)$$

$K = 0, 1, 2, 3, \dots, L-1$ is related to probability of occurrence of r_k in the image. where

- $p(r_k)$ gives an estimate of the probability of occurrence of gray level r_k .
- The sum of all components of a normalized histogram is equal to 1.

The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristics. DR regions are suitable for extracting with segmentation. We can hide the blood vessels with image segmentation method. Fundus image segments are blood vessels, DR parts, the remaining healthy parts and iris.

The histogram of gray fundus image is shown each pixel values and number of each pixel in Figure 3.16. All our image segments are represented with different parts of histogram. We used two of histogram processing methods for segmentation.

These are called histogram stretching and histogram thresholding. We firstly explain these methods.

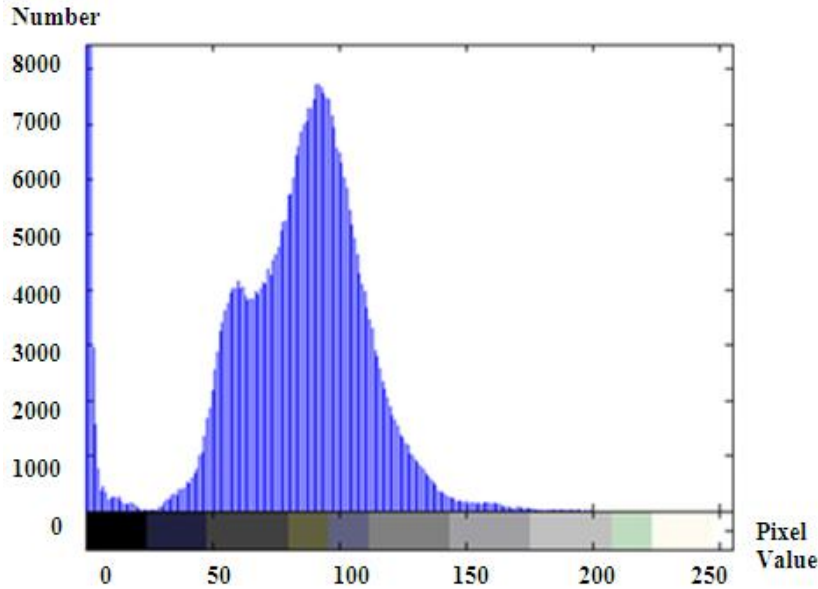


Figure 3.16. Gray scale fundus image histogram

3.2.3.1. Image Thresholding

Thresholding is a basic way for image segmentation. Generally, consists segmented an image with one of the thresholding methods separate 2 region. These are a particle region and a background. Both binary image and color image can be segment with thresholding methods. This process works by setting to white all pixels that belong to a gray-level interval, called the threshold interval, and setting all other pixels in the image to black. Image thresholding can be different methods. These methods could be collected in two basic title, manual techniques and automatic techniques. Manual threshold selection is normally done by trial and error, using a histogram as a guide. Automated thresholding techniques select a threshold which optimizes a specified characteristic of the resulting images. These techniques include clustering, entropy, metric, moments, and interclass variance. Clustering is unique in that it is a multi class thresholding method. In other words, instead of producing only binary images it can specify multiple threshold levels which result in images with three or more gray-level values. The key parameter in the thresholding process is the

choice of the threshold value. Several different methods for choosing a threshold exist; users can manually choose a threshold value, or a thresholding algorithm can compute a value automatically, which is known as automatic thresholding. A simple method would be to choose the mean or median value, the rationale being that if the object pixels are brighter than the background, they should also be brighter than the average. In a noiseless image with uniform background and object values, the mean or median will work well as the threshold, however, this will generally not be the case. A more sophisticated approach might be to create a histogram of the image pixel intensities and use the valley point as the threshold. The histogram approach assumes that there is some average values for both the background and object pixels, but that the actual pixel values have some variation around these average values. However, this may be computationally expensive, and image histograms may not have clearly defined valley points, often making the selection of an accurate threshold difficult. The classes' separation is shown in Figure 3.17.

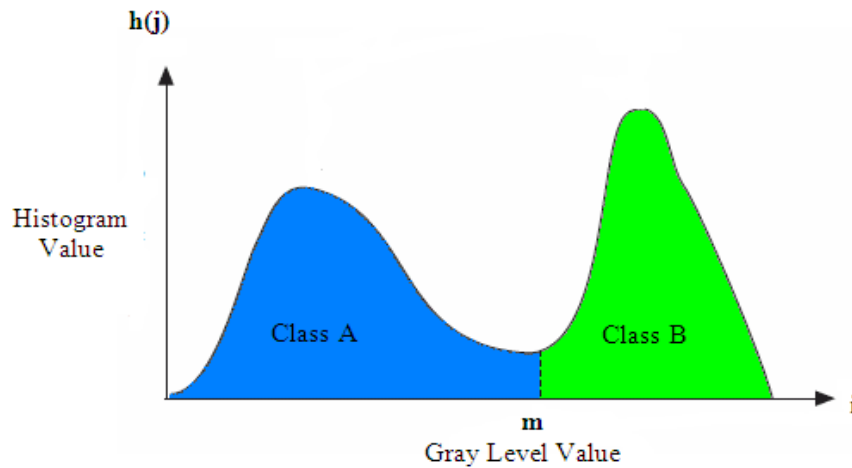


Figure 3.17. Classes representation on a sample histogram

3.2.3.2. Histogram (Contrast) Stretching

Histogram stretching is called contrast stretching or contrast normalization. Contrast stretching attempts to improve an image by stretching the range of intensity values it contains to make full use of possible values. Unlike histogram equalization, contrast stretching is restricted to a linear mapping of input to output values. The

result is less dramatic, but tends to avoid the sometimes artificial appearance of equalized images.

The most straightforward histogram function is referred to as stretching. Many images contain a combination of dark and bright colors, but unless an image takes advantage of the full luminance spectrum, it may appear dim or washed out. It is necessary to adjust the image's colors to better span all possible luminance values for improving an image. Stretching handles this by finding the darkest and lightest values in an image, then performing an automatic conversion between that color range (max – min) and the ideal color range (pure white – pure black).

Contrast is referred as the difference in gray levels or luminance values over some areas of the image. Low contrast images can result from poor illumination, lack of dynamic range in the imaging sensor, or even wrong setting of a lens aperture during image acquisition. Contrast stretching is a simple image enhancement technique that attempts to improve the contrast in an image by “stretching” the range of intensity values it contains to span a desired range of values, e.g. the full range of pixel values that the image type allows.

The general idea in contrast enhancement is to apply a linear or non-linear scaling transform to obtain a new modified image. The first step is to determine the limits over which image intensity values will be extended. These lower and upper limits will be called a and b , respectively. Next, the histogram of the original image is examined to determine the value limits (lower = c , upper = d) in the unmodified picture. If the original range covers the full possible set of values, straightforward contrast stretching will achieve nothing, but even then sometimes most of the image data is contained within a restricted range; this restricted range can be stretched linearly, with original values which lie outside the range being set to the appropriate limit of the extended output range. Then for each pixel, the original value r is mapped to output value s using the function in Equation 3.21:

$$S = (r - c) \frac{(b - a)}{(d - c)} + a \quad (3.21)$$

One problem with this method is that outliers can reduce the effectiveness of the operation. This was already mentioned above when it was suggested that sometimes a restricted range of the input values as determined by inspecting the histogram of the original image might be used. Frequently, it is advantageous to select c and d to be at the 5th and 95th percentiles, respectively, of the input values. Alternatively, one can start at the histogram peak and move up and down the value list until only a small (e.g., 1% to 3%) number of values are rejected and left outside the chosen limits for c and d .

4. EXPERIMENTS, RESULTS AND DISCUSSIONS

Approximately, thirty FFA images are used in this thesis. All of them are shown in Appendix 1. Data group consists of healthy eyes FFA images and DR eyes FFA images. We need 2 types of images for our classification and clustering applications. You can see two type of data samples in Table 4.1. DR stages of the data samples are different. Generally, patients' disease levels are at final stages for our image data set.

The original size of image is 768x576x3. The eye center is visible in the images.. It is a circle on the eye structure and called iris. Iris is intersection point of vessels. Eye center and disease part of retina are illustrated by brighten region of fundus images. Vessel parts can select be on FFA easily. Black region is background of FFA images. The retina is illuminated with an excitation color which fluoresces light of another color where the dye is present. By filtering to exclude the excitation color and pass the fluorescent color, and a very high-contrast image of the vessels is produced.

Angiography machines can give two different images. These are gray form and color form. We can use two types for our different algorithm applications. Color images have three dimensions. This situation sometimes becomes a problem. Use of gray scale images can provide advantages in some methods. We obtain gray and color forms of all images to use with our algorithms. Two different style of a sample are shown in Figure 4.2.

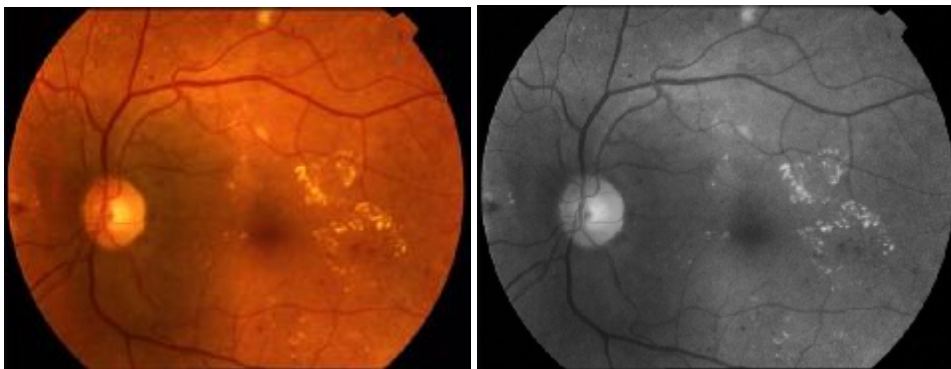
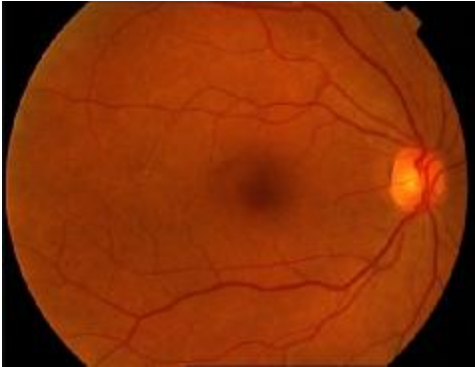
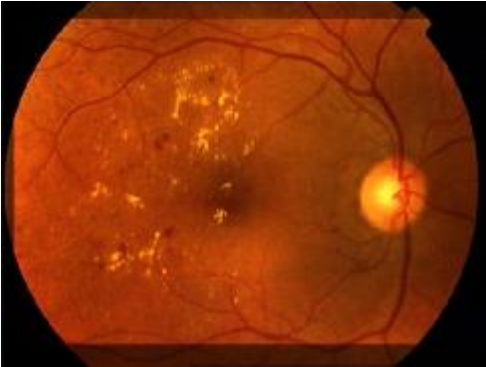


Figure 4.1. Two different form of images

Table 4.1. The samples from dataset.

Healthy Data Samples	DR Data Samples
 Healthy	 Advanced Stage
 Healthy	 Advanced Stage
 Healthy	 Early Stage

4.1 Segmentation With LDA

LDA is a simple classification way. This way is simple but it can gives best results to us. We applied firstly LDA classification method for our dataset. LDA classifier is a classification method for solving big matrices problems. Our classification problem could be handled up because it has just two classes that can desire “eye has a disease region or not? “.For solving all classification problems firstly features should be choosen. Feature extraction is a part of clasification theory. Selected features are important features for classification. Because of these reasons we determined features before application of two classification methods. Our features are minimum value of matrix, maximum value of matrix, median, skewness and kurtosis.

After calculation of min, max, median, skewness, kurtosis features of image matrix then we build a matrix with these values. Train matrix is shown in Table 4.2.

Table 4.2. Train vector sample

Training Vector for LDA Classifier													
0	0	0	0	0	0	0	0	1	1	1	1	1	1
0	0	0	0	0	0	0	0	1	1	1	1	1	1
0	0	0	0	0	0	0	0	1	1	1	1	1	1
0	0	0	0	0	0	0	0	1	1	1	1	1	1
0	0	0	0	0	0	0	0	1	1	1	1	1	1
0	0	0	0	0	0	0	0	1	1	1	1	1	1
0	0	0	0	0	0	0	0	1	1	1	1	1	1

A window can be defined a small pixel group on images. Window size is related with classification result. Ifwe took big windows, calculated valuescan not be represented the big matrix class. Before the detection of window size, we tried different window size. For example 3x3, 8x8, and 20x20 are tried. We considered divisor values of training window. We decided to be 16x16. A sample training image part is shown in Figure 4.2.

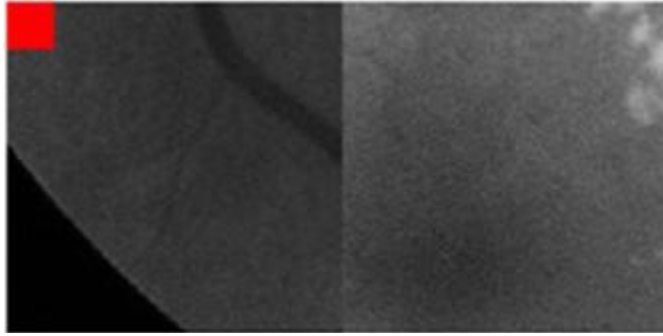


Figure 4.2. Training part sample for LDA classifier

The other parts are not belonged to training window which are used for testing. We determined two groups for LDA classifier. One of them is disease part which is shown in figure with green. Healthy and background parts are represented in figure dark blue. Train data are shown with red windows. Explanation of the image is shown in Figure 4.3. Some of the data samples LDA classification results are shown in Table 4.3.

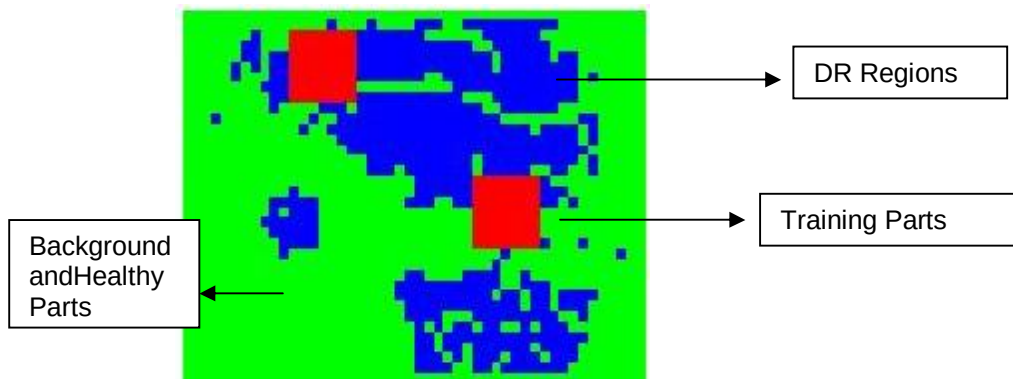
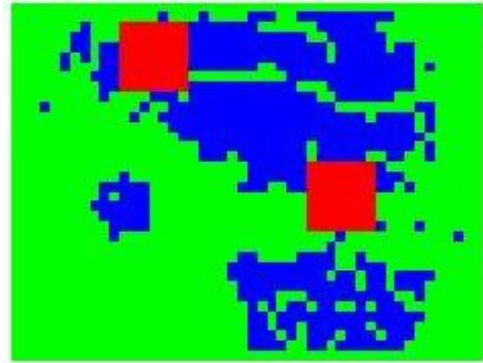


Figure 4.3. Explanation of the LDA classifier results

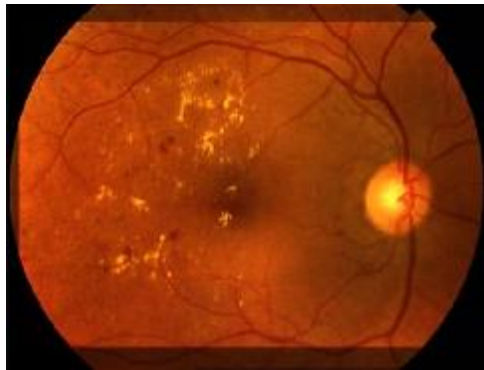
Table 4.3. Samples of segmented images obtained by LDA

LDA Classifier Results

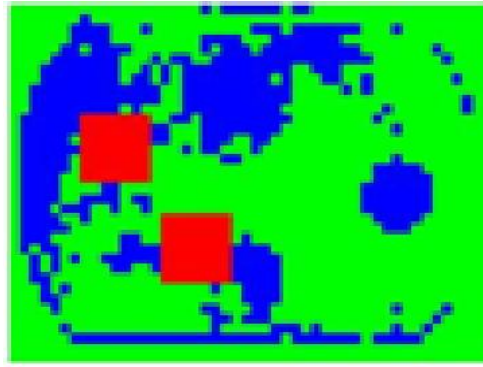
Sample 1



Segmented image of sample 1



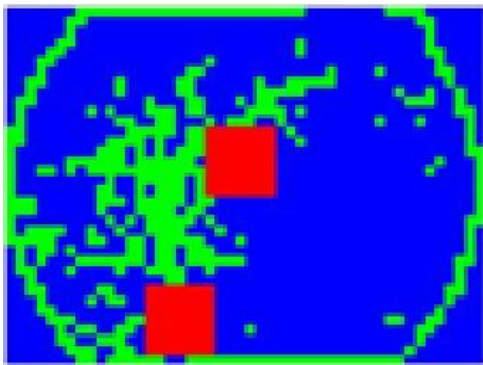
Sample 2



Segmented image of sample 2



Sample 3



Segmented image of sample 3

LDA classifier separates dataset two or more classes with a linear line. We thought that the classifications problem could be solved easily by this way. Datasets have two different classes healthy or disease. Determining the training vector values is very important and related with the result closely. Unfortunately, results aren't at the expected point. It can be caused from various reasons such as fixed window size, distribution of DR variability.

4.2. Segmentation with Neural Network Method

Our second classification method is backpropagation that is a neural network branch. We tried to explain our first method LDA in previous section. Backpropagation is basic method of the artificial neural network classification. This method is a type of supervised learning algorithm. Before application of backpropagation, we designed neural network structure. Design is very important because it is related with algorithm's final accuracy. Some of the base criterias are number of neuron, number of layer, number of epoch etc. Number of layers aren't related with algorithm accuracy. A complex classification problem can solve with using one layer. Classification's bases are training and test data. How can we build training vector? Firstly, we determine training data. After that we match classes and train data in Figure 4.4.

Then some parts of image use for training and then build a training vector like as LDA. But we use three classes for this classification. Healthy parts, disease parts and background are our classes. Healthy parts, disease parts and background are represented 0,1,-1 in order in training matrices.

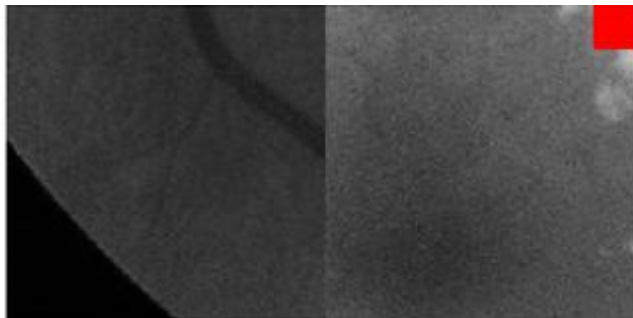


Figure 4.4. Training part sample for backpropagation classifier

We tried to choose training parts different region of our images as LDA classifier. Training vector and training time are very important for classification. This duration spends the biggest time of our program working period. Researchers must be careful during choosing training vector. One of the sample training vectors is given below in Table 4.4.

Table 4.4. Train vector sample for neural network classifier

Train Vector for Neural Network Classifier													
-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1
-1	-1	-1	-1	-1	-1	-1	1	1	1	0	0	1	0
-1	-1	-1	-1	-1	-1	0	0	1	1	1	0	0	0
-1	-1	-1	-1	-1	0	0	0	1	1	1	1	0	1
-1	-1	-1	-1	-1	0	0	0	1	1	1	1	0	1
-1	-1	-1	-1	0	0	0	0	0	0	0	1	1	-1
-1	-1	-1	0	0	0	0	0	0	0	1	1	1	1
-1	-1	-1	0	0	0	0	0	0	1	1	1	1	0

We used two times 112x112 windows at different region for training. Window size is 16x16. We tried different window size as LDA classifier, but results aren't expected level because of disease distribution. The remaining parts of FFA images are used for testing. But this method can't give to us our expected result as LDA classifier. Samples backpropagation results are shown in figure 4.5. To determined three groups for neural network backpropagation classifier. One of them DR parts is represented by green in figure. Healthy parts of eye are represented by dark blue in figure too. Background of image is shown in figure by gray color. Training datais shown by red color windows (in Figure 4.5). This classifier has some problems for FFA image data's classification. This classifier can't separate background and healthy parts clearly. Background parts are black, healthy parts can seem darken underfluorescein. Because o the problems, we decided to this method is not suitable for this dataset. All the results are shown in Table 4.5.

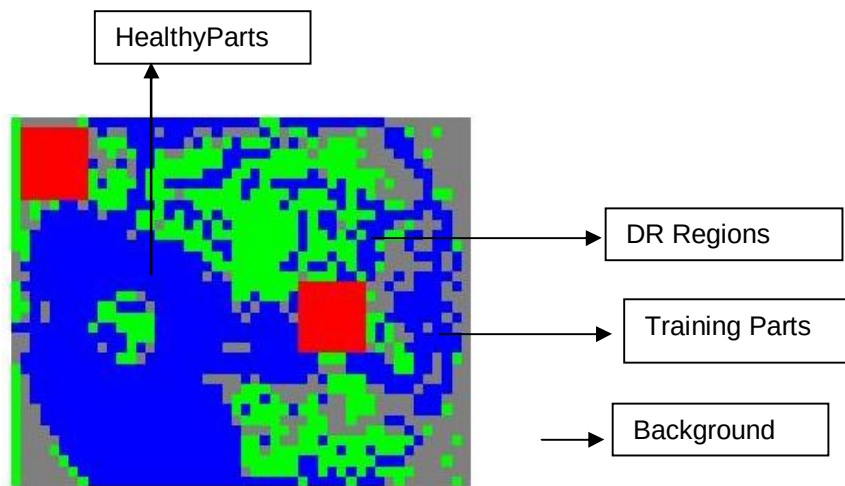
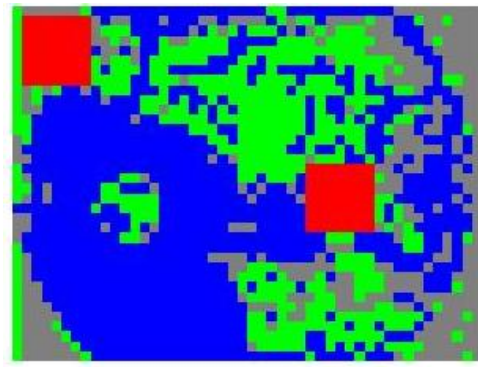


Figure 4.5. Explanation of the backpropagation results

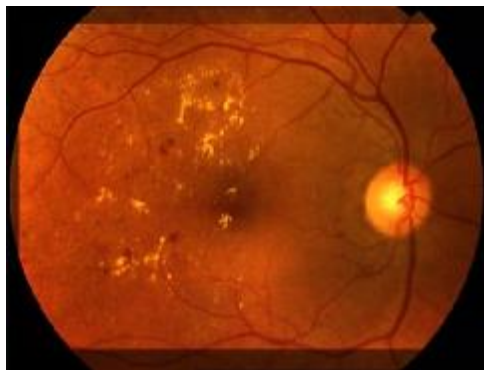
Table 4.5. Segmented imaged are result of backpropagation algorithm

Backpropagation Method Results

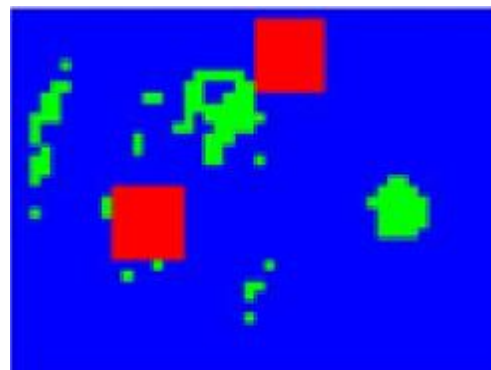
Sample 1



Segmented image of sample 1



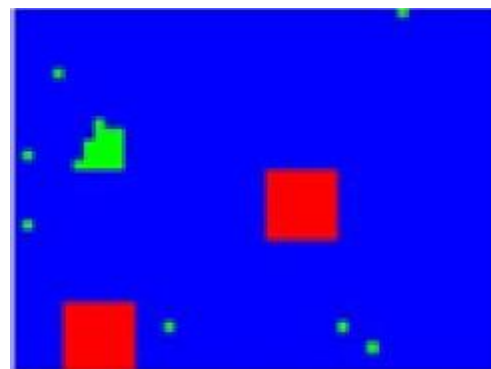
Sample 2



Segmented image of sample 2



Sample 3



Segmented image of sample 3

4.3 K-Means Clustering Results

Clustering is a way to separate two or more groups of data. K-Means clustering treats each object as having a location in space. It finds partitions such that objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible.

We used matlab K-Means algorithm for clustering FFA images. The function K-Means partitions data into k mutually exclusive clusters, and returns the index of the cluster to which it has assigned each observation. Unlike hierarchical clustering, K-Means clustering operates on actual observations (rather than the larger set of dissimilarity measures), and creates a single level of clusters. The distinctions mean that K-means clustering is often more suitable than hierarchical clustering for large amounts of data.

We determined firstly cluster number before construction of algorithm code. The number of cluster is determined as three, which are background, healthy parts and disease parts like background classifier. After determining cluster numbers, algorithm embodies according to this number. K-Means algorithm used Euclidean distance. K-Means returns an index corresponding to a cluster for every object in FFA image data. This method gives to us different accuracy levels for different fundus images.

K-Means clustering requires that you specify the number of clusters to be partitioned and a distance metric to quantify how close two objects are to each other. K-Means algorithm is used widely for clustering. Generally, this method uses for clustering because the algorithm gives high accuracy using with variable data sets. Its usage is easier than the other clustering algorithms.

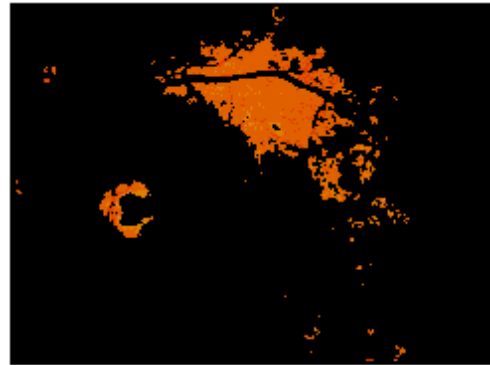
Cluster number is three (DR part, healthy part and background). Background and healthy parts are darkened. They are represented by black. DR parts and iris are shown clearly. Our image data are belonging to patients with different disease levels. DR stages and distributions are very important criteria for K-Means like as the other classification theories. In addition to disease distribution, the area which is enclosed by bleeding is important. The cluster's center changes to relate with disease

distribution. K-means algorithm's performance changes under this condition which is described in the previous paragraph. Finally, the K-Means algorithm can't give the same accuracies for different data samples. If we remark about the method's negative directions, this algorithm can find large DR regions than real dimensions of disease regions. It can find just edges of some disease regions. In spite of all these situations, the algorithm's accuracy is not bad for an unsupervised learning method. We obtained different results for one image each running of our Matlab code. It caused from learning. During application of unsupervised learning method is a normal situation. Unsupervised learning methods don't have any feedback mechanism. Because of this learning can't be consistent everytime. K-Means clustering results are shown in Table 4.6.

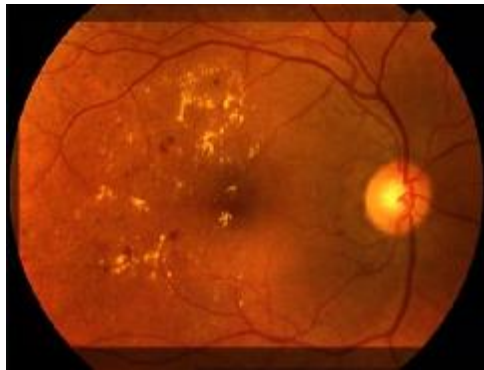
Table 4.6. The segmented result images obtained by K-means clustering method

K-Means Clustering Results

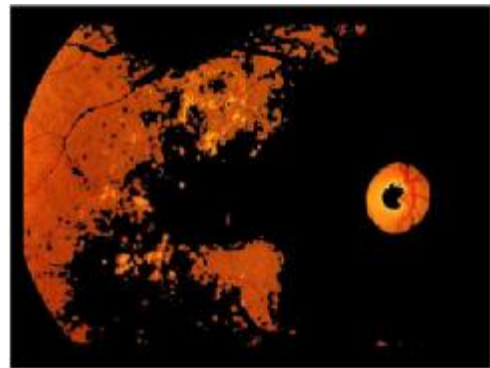
Sample 1



Segmented image of sample 1



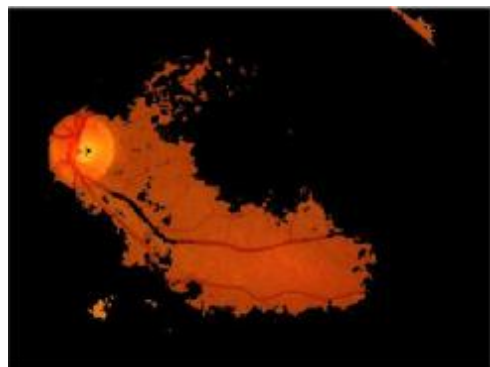
Sample 2



Segmented image of sample 2



Sample 3



Segmented image of sample 3

4.4. Gaussian Mixture Method Results

We tried to cluster our dataset with gaussian mixture distribution according to variance and means of image matrix. Gaussian mixture distributions can be used for clustering data, by realizing that the multivariate normal components of the fitted model can represent clusters. Shown results are belonging to patients being different stage of disease result. Most of the results show eye center and DR parts clearly and suppress the vessels. We tried to show clearly by using contrast color. Explanations of the regions are shown in Figure 4.6.

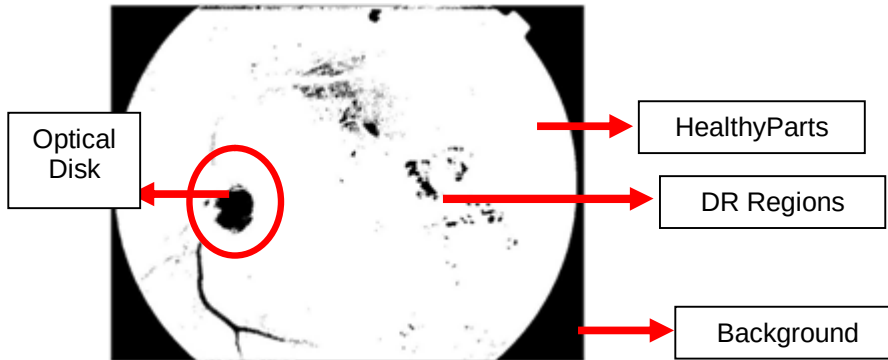


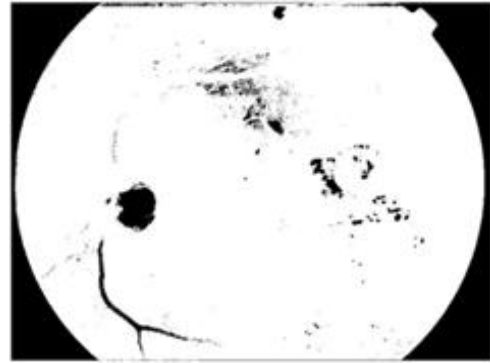
Figure 4.6. Explanation of GMM result

Gaussian distribution model found DR parts clearly for patients having big DR region. The results are very successful because fundus image histogram likes a gauss curve. Gaussian mixture models give good results for Gauss distributions. Gaussian mixture model results accuracy is higher than other explained methods above. Iris and DR parts can be selected with our eyes clearly. This method gives different results than expected result. The difference can be related with some pixel values. These suppressed values are situated without our Gaussian curve and threshold value. For some patient DR pixel value can be different because patients DR level may be at early stage or final stage. If the patients are early stage of DR, generally the pixel values are small vice versa. Pixel value affected our segmentation results. DR part and background of image are shown by black points. You can see Gaussian mixture model results in Table 4.7.

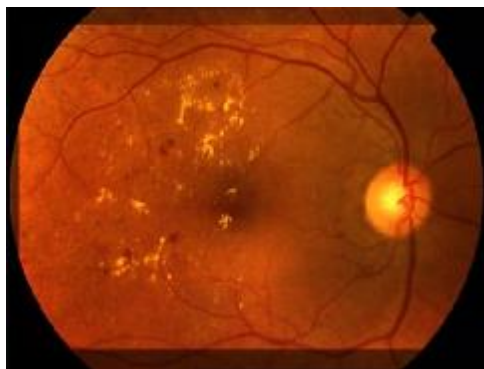
Table 4.7. Segmented images obtain by GMM

GMM Clustering Results

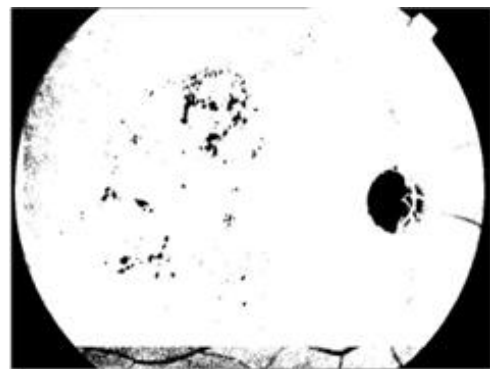
Sample 1



Segmented image of sample 1



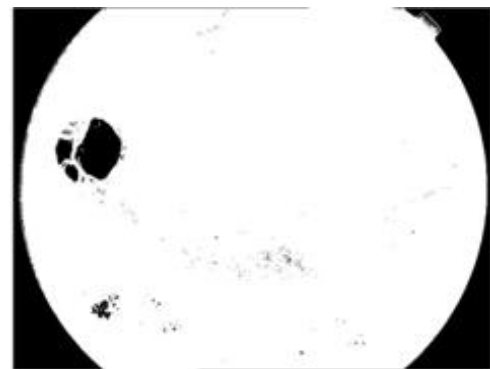
Sample 2



Segmented image of sample 2



Sample 3



Segmented image of sample 3

4.5. Segmentation With Image Thresholding Method Results

Generally, histogram value gives many solutions for image analysis. We use histogram segmentation method. We determine image histogram's parts. Histogram part's of image very important for determining disease. We divided image histogram three parts which are background, healthy region and disease region.

In other words, the histogram stretching is called generally contrast stretching. Contrast stretching is based on the differences between dark and light colors on image. Disease part and optic disc are very light color on the image. Background, some vessels and healthy parts are dark color. Which values on the histogram show which eye parts that is important for histogram processing in Figure 4.7.

Determining the threshold pixel value is the most important criteria for histogram thresholding method. We tried different pixel values that are 120,130,160. This value is determined by investigation of the image histogram and expectation the pixel value range of DR parts. Pixel value is greater than 140 that are light parts of histogram. This part is DR region of eye.

We extracted this part with histogram thresholding. We tried to darken the pixel value bottom of 140, simultaneously closed to white level the pixel value top of 140. With this way, we could try to emphasize DR region. We set the threshold values at 140 for all Gaussian mixture model applications. This value can be enough for DR region extraction. The results are shown in Table 4.8.

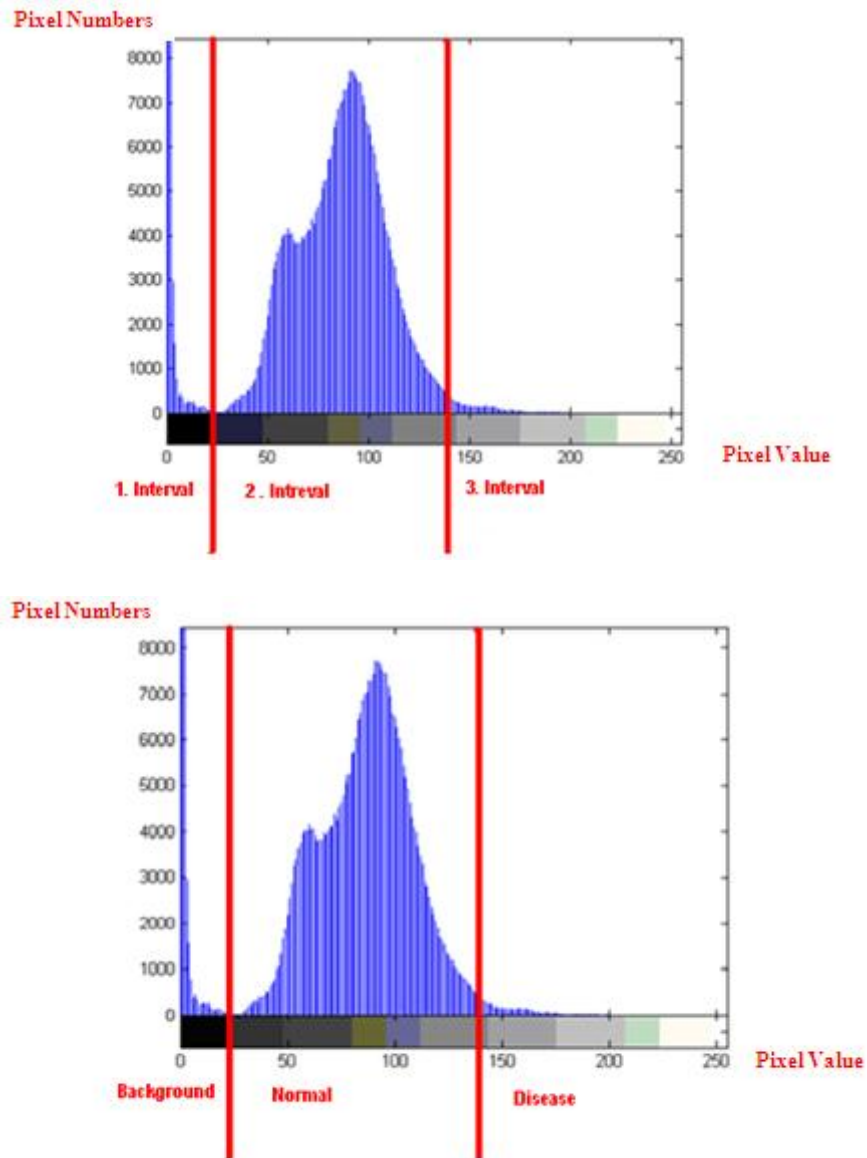


Figure 4.7. Comparison of histograms

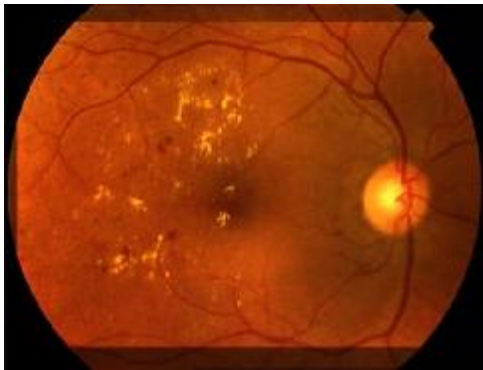
Table 4.8. Segmented result by using image thresholding method

Image Thresholding Results

Sample 1



Segmented image of sample 1



Sample 2



Segmented image of sample 2



Sample 3



Segmented image of sample 3

4.6. Image Segmentation With Histogram Strecthing Results

Image segmentation can be with different methods. We preferred to make with histogram operations. The histogram stretching method is one of the way from processing operations. The most important point is chosen of the transfer functions. The selection of transfer function effects result's accuracy. We obtained different result for different transfer function. The important parameter of this method is transfer function type. We applied our experiment two different transfer functions. This is very important for seeing the differences of transfer functions. The linear transfer function is shown in Figure 4.8 which is used for our application.

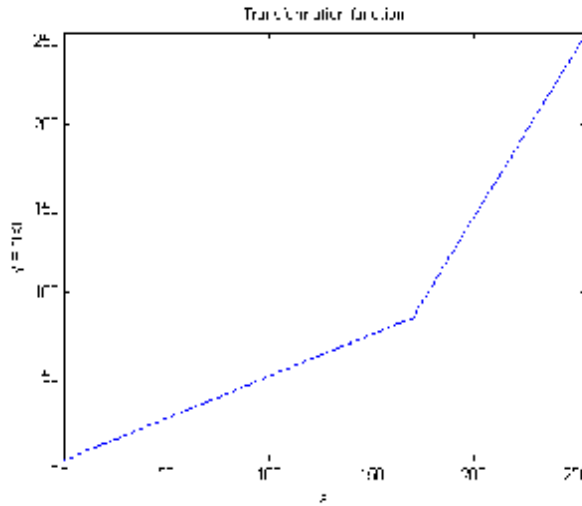


Figure 4.8. Piecewise linear transfer function for using contrast stretching method

All results are obtained using linear transfer function. The results accuracies are related with bleeding level. If bleeding level is high, brightness is increasing with this level. Some of the patients' disease situation is at critical point. These patients fundus image stretching result accuracies are the best than the other patients being at starting level of DR. This method try to get darken healthy part and background. This application results are shown in Table 4.9.

Table 4.9. Linear transfer function histogram stretching algorithm application results

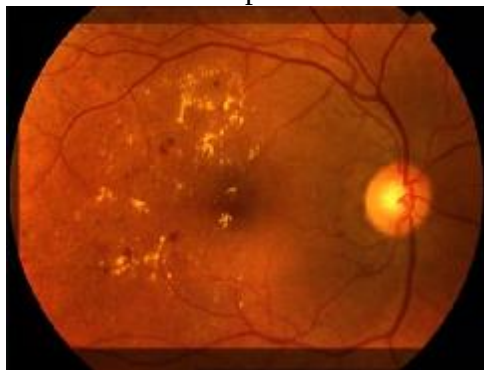
Histogram Stretching with Linear Transfer Function Result



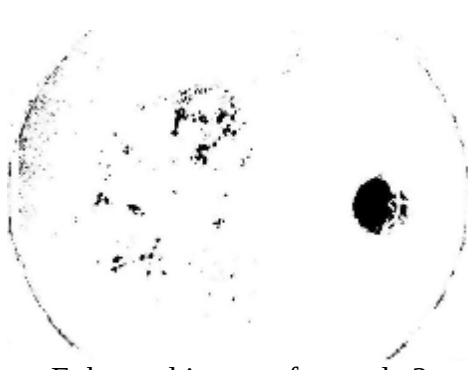
Sample 1



Enhanced image of sample 1



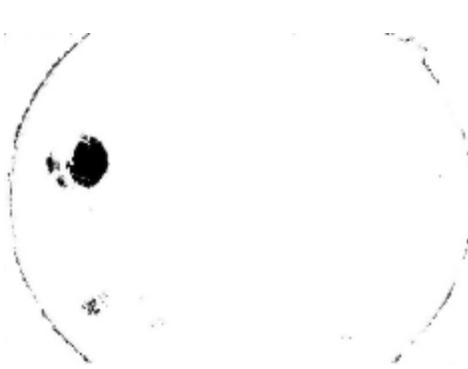
Sample 2



Enhanced image of sample 2



Sample 3



Enhanced image of sample 3

The results obtained with using linear transfer function which histogram values are between small values. DR parts on the image correspond to these small regions. Healthy parts and eye center try to suppress on fundus image. Histogram schedule changes with these contrast stretching methods.

The aim of contrast enhancement techniques expand the range of brightness values in an image so that the image can be efficiently displayed in a manner desired by the analyst. We did contrast enhancement using two different transfer functions.

We tried to explain firstly linear transfer function and its role. The logarithmic transfer function is shown in Figure 4.9 which is used for our application.

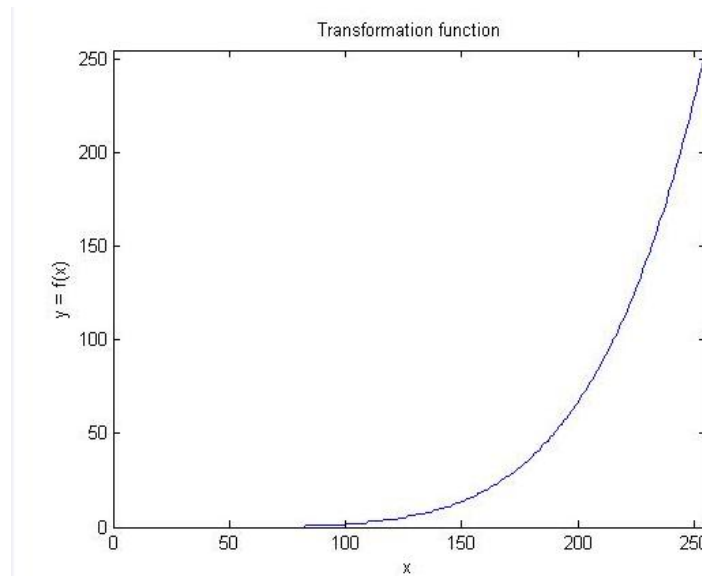


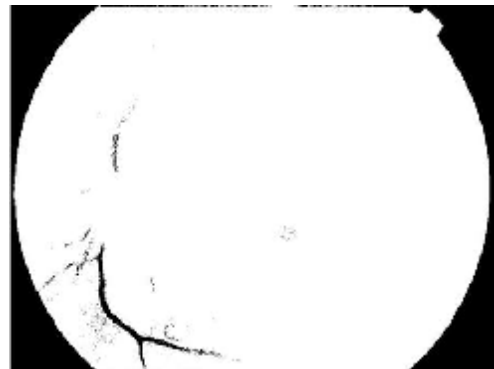
Figure 4.9. Logarithmic transfers function for using contrast stretching method.

Logarithmic function suppressed extremely. Because of this DR parts can not select clearly. The logarithmic function results are shown in Table 4.10.

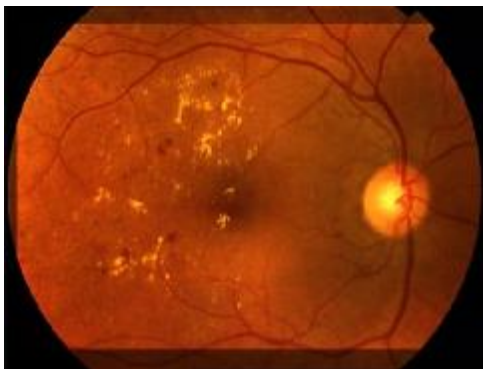
Table 4.10. Logarithmic transfer function histogram stretching application results

Histogram Stretching with Logarithmic Transfer Function Results

Sample 1



Enhanced image of sample 1



Sample 2



Enhanced image of sample 2



Sample 3



Enhanced image of sample 3

The results shown in Figure 4.10 show to us using different transfer function role on the fundus image. Linear transfer function result is better than logarithmic. If we compare linear and logarithmic function results for same image. We can notice difference easily.

Linear transfer function exhibited disease parts clearly and brightly. Logarithmic transfer function result suppressed much for some regions.

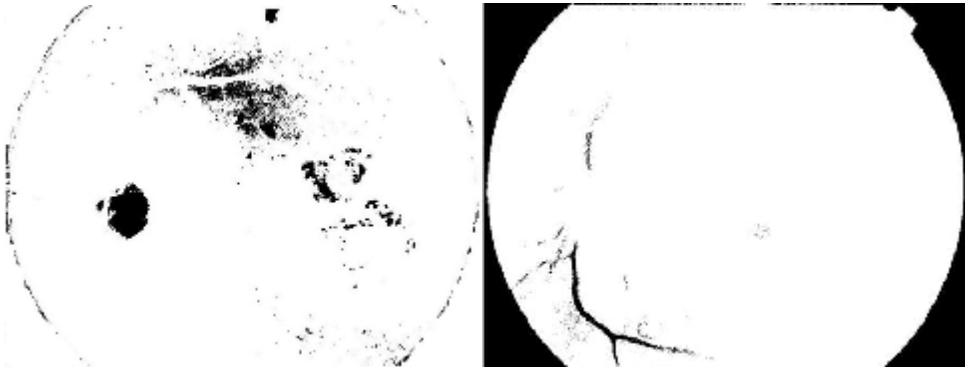


Figure 4.10. Comparison of linear and logarithmic transfer functions results

5. CONCLUSIONS

In this thesis, we used several methods for detecting and highlighting the DR regions of the images to assist doctors and their judgements in the diagnosis. The enhanced images obtained by these methods are compared and interpreted.

AFFA image matrix has 576 lines and 768 columns. We used backpropagation, linear discriminant analysis, K-Means, Gaussian Mixture Model, histogram thresholding and contrast stretching respectively. Before using unsupervised learning methods, we preferred to extract features. The features are selected firstly. The features are some statistical measure of the image matrices. These are minimum value, maximum value, median, kurtosis and skewness. The image is divided into small rectangular pieces. These pieces are called windows. We are determined train parts. We chose train parts which include DR, healthy and background regions. This criteria is important for classification performance. Two sub-images extracted from the whole image are employed to train the classifier. The window size influences the performance. Although we tried to build optimum conditions for classifications methods, both methods couldn't give our expected results. Because of this, the distributions of hemorrhagic regions on the images are not analogous. Some patient hemorrhagic regions are wide but the others are narrow. The hemorrhagic region distribution is not regular and homogeneous. Therefore it is difficult to determine a global optimum window size. We tried different window sizes but the best result took from using the 16x16 window size. After these results, we determined the window size 16x16. The window size causes a decrease for the resolution of the resulted image. We calculated min, max, median, kurtosis and skewness value as feature vector. The vector is the base point of classification. The two classes; hemographi and non-hemographi are used by LDA classification. Three classes are used in the neural network classifier in addition to LDA classes background. The same calculations are done for the two supervised methods. After all segments are labeled, the LDA and backpropagation results are compared.

After application of supervised learning methods, unsupervised learning methods are applied. K-Means algorithm is also used as a supervised learning method to

obtain segment the images. Application of K-Means algorithm is easy and its accuracy is high. These are required properties by researchers. Because of this, K-Means is a preferred method. The main point in the algorithm is the selection of cluster number. We designated 3 classes for K-Means clustering ($k=3$); background, normal of eye and DR.

After K-Means method, we tried GMM for FFA image data. We chose this method because histogram of an FFA image is similar to a Gauss curve. Therefore it seems to be natural to use Gaussian distribution to detect DR. Variance and mean properties are very important parameters for Gaussian distribution. Since they specify the distribution, we decided 2 variance values for 2 groups of data. These groups are healthy part and disease part. Gauss distribution function evaluates these values for clustering.

Subsequently supervised and unsupervised, histogram thresholding method is used on FFA images. Segmentation with histogram specifications is a simple method for image analysis. These methods based on histogram distribution and threshold value. The important point for the method, is to decide thresholds from the histogram. We have seen that the pixel a value greater than 140 corresponds to diabetic retinopathy. We used this threshold value to segment DR sections.

The other way is an image enhancement method. It is called histogram stretching. We used this method with two different approximations, piecewise linear function and logarithmic function. We compared these functions results.

The best results are obtained from the histogram thresholding method. Histogram of our image data are suitable for enhancement methods since pixel distributions are regular. FFA image's histogram is like a 2 different standard Gauss curve. Because of this, the separation of pixels is easy. We can classify the histogram region. And then we can find disease region.

6.FUTURE WORKS

Diabetic retinopathy's incidence is rising all over the world. The importance of early detection is increasing from this cause by day by. Imaging become effective technique for detection of some disease. For example MR (cancer type), ultrasound (using for imaging different disease in human body), retinal imaging techniques (for all retinal disease). For future work of this thesis, wavelet transformation can apply on FFA images. Addition to wavelet transform fuzzy logic controllers and self organizing maps can be suitable for FFA images. We try to explain these methods shortly.

A wavelet is a wave-like oscillation with an amplitude that starts out at zero, increases, and then decreases back to zero. As a mathematical tool, wavelets can be used to extract information from many different kinds of data, including - but certainly not limited to - audio signals and images. Sets of wavelets are generally needed to analyze data fully. A set of "complementary" wavelets will deconstruct data without gaps or overlap so that the deconstruction process is mathematically reversible. Thus, sets of complementary wavelets are useful in wavelet based compression/decompression algorithms where it is desirable to recover the original information with minimal loss.

Fuzzy logic controller technique generally is used for image analysis. It has many different methods. Fuzzy logic is a form of many-valued logic; it deals with reasoning that is approximate rather than fixed and exact. In contrast with traditional logic theory, where binary sets have two-valued logic: true or false, fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Fuzzy image processing is the collection of all approaches that understand, represent and process the images, their segments and features as fuzzy sets. The important point for fuzzy image processing is determination of fuzzy method. For this, firstly image data is analysed ideally.

A self-organizing map (SOM) is a type of artificial neural network that is trained using unsupervised learning to produce a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples,

and called a map. Self-organizing maps are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space. A self-organizing map consists of components called nodes or neurons. Associated with each node is a weight vector of the same dimension as the input data vectors and a position in the map space. The usual arrangement of nodes is a regular spacing in a hexagonal or rectangular grid. The self-organizing map describes a mapping from a higher dimensional input space to a lower dimensional map space. The procedure for placing a vector from data space onto the map is to first find the node with the closest weight vector to the vector taken from data space. Once the closest node is located it is assigned the values from the vector taken from the data space. While it is typical to consider this type of network structure as related to feedforward networks where the nodes are visualized as being attached, this type of architecture is fundamentally different in arrangement and motivation.

Finally, all these methods are used for solving large matrices signals and images problems. Explained techniques are main statistical analysis way for clustering. Data clustering and classification are very important for analysing data. Data type can change for application or research. Especially, these methods are used for signal and image processing. Image processing can separate into two groups. These are medical image processing and image processing for visual qualifications. Medical image processing has become important way for detection of different disease in recent years. Medical image processing techniques provide early and easy detections of disease for doctor. For all important disease, early detection is very important criteria for example cancer, eye disease, tumor etc.

Our data set is very suitable for development. Many problems can be seen at human eye. With these methods and other methods (maybe will find in future) can develop the best software. Different methods can give better results than our application results.

REFERENCES

- ABBAS, O.A., 2008. Comparison Between Data Clustering Algorithms. The International Arab Journal Of Information Technology, vol:5, no:3.
- ARCHARYA, U.R., LIM, C.M., et al., 2009. Computer based detection of diabetes retinopathy stages using digital fundus images. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, vol. 223, 7:545-553.
- BERGER, JW., 1999. Quantitative Image Sequence Analysis Of Fundus Fluorescein Angiography. Scheie Eye Institute, University of Pennsylvania, Philadelphia, USA.
- CHEN, K.Y, FUH, C.S., 2009. High Dynamic Range Image Processing To Screen Diabetic Retinopathy. Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University.
- DESHPANDE, P.D., 2006, Survey of Satellite Image Segmentation and Classification, NCETE2006, MAE, Alandi, Pune.
- GARDNER, G.G., KEATING, D., 1996. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. British journal ophthalmology. Tennent Institute of Ophthalmology, Scotland, 80:940-944.
- GIANCARDIO, L., 2008. Quality Analysis of Retina Images For The Diagnosis Of Diabetic Retinopathy. A Thesis Degree Of Msc Erasmus Mundus In Vision and Robotics.
- GOLDBAUM, M., MOEZRI, S. and GOLDBAUM, T.A.M., 1996. Automated diagnosis and image understanding with object extraction, object classification, and inference in retinal images. International conference on image processing, vol. 3, 695-698.
- GONZALES, R.C., WOODS, R.C., 2009. Digital Image Processing. Second edition. Prentice Hall, New Jersey, chap:3, 88-103.
- IQBAL, M.I., AIBINU, A.M., 2006. Automatic Diagnosis Of Diabetic Retinopathy Using Fundus Images. Blekinge Institute Of Technology, Master Thesis.

- KUMAR, M., 2003. Satellite Remote Sensing and GIS Applications in Agricultural Meteorology, 81-102.
- KUMARI, V.V., NARAYANAN, N.S., 2010. Diabetic Retinopathy-Early Detection Using Image Processing Techniques, International Journal on Computer Science and Engineering, vol. 02, 357-361.
- LUO, G., CHUTATAPE, O., LI, G., KRISHNAN, S.M., 2001, Abnormality detection in automated mass screening system of diabetic retinopathy, Computer-Based Medical Systems. Proceedings 14th IEEE Symposium On, 132-137.
- NEERA SINGH, N., TRIPATHI, R.C., 2010. Automated Early Detection of Diabetic Retinopathy Using Image Analysis Techniques. International Journal of Computer Applications (0975-8887) vol:8, no:2.
- SARAVANAN, P., 2008. Retinal fundus image contrast normalization using mixture of Gaussians. Signals. Systems and Computers, 42nd Asilomar Conference on, 647-650.
- SANCHEZ, C.I., HORNERO, R., 2007. A Novel Automatic Image Processing Algorithm For Detection Of Hard Exudates Based On Retinal Image Analysis. Medical Engineering & Physics, 350-357.
- SHOME, S.K., VADALI, S.K.R., 2011. Enhancement of Diabetic Retinopathy Imagery Using Contrast Limited Adaptive Histogram Equalization. International Journal of Computer Science and Information Technologies, vol:2 (6) , 2694-2699.
- SINGH, N., TRIPATHI, C.R., 2010, International Journal of Computer Applications BIJCA Journal, 1186-1648.
- SINTHANAYOTHIN, C., BOYCE, J.F., 2002. Automated detection of diabetic retinopathy on digital fundus images. Diabetic Medicine, vol:19(2), 105-112.
- STAURENGHI, G., LEVI, G., 2007. New Developments in cSLO Fundus Imaging. chapter:2, 23.
- VIJ, K., SINGH, Y. Enhancement of Images Using Histogram Processing Techniques Int. J. Comp. Tech. Appl., vol 2 (2), 309-313.

- YIN, D., PAN,J., 2008. Medical Image Categorization Based on Gaussian Mixture Mode, BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on , 128 - 131.
- ZHENG, L., OPAS, C., KRISHNAN and S.M., 1997. Automatic image analysis of fundus photograph. Proceedings of the 19th Annual International Conference of the IEEE, Nanyang Technology University, vol.2, 524 – 525.

BIOGRAPHY

İclal ÇETİN TAŞ was born on March 11, 1985 in Adana, TURKEY. She graduated from Adana Anatolian High School in 2004. She received her B.Sc. degree from Electrical and Electronics Engineering Department of Çukurova University in 2009.

After she graduated from University at same year, she started as a master student at the Electrical and Electronics Engineering Department of Çukurova University.

APPENDICES

APPENDIX 1

FFA IMAGES	
