

**DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED
SCIENCES**

**BAYES GAUSSIAN CLASSIFICATION OF
WISCONSIN BREAST CANCER
DATABASE**

by

Mozhgan MOAZZEN ZADEH

October, 2011

İZMİR

**BAYES GAUSSIAN CLASSIFICATION OF
WISCONSIN BREAST CANCER
DATABASE**

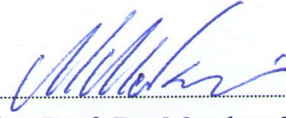
**A Thesis Submitted to The
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Master of Science in
Electrical and Electronics Engineering, Electrical and Electronics
Engineering Program**

**by
Mozhgan MOAZZEN ZADEH**

**October, 2011
İZMİR**

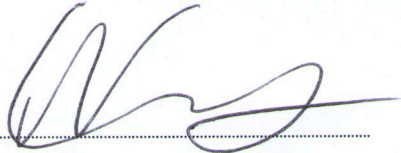
M.Sc. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled “**BAYES GAUSSIAN CLASSIFICATION OF WISCONSIN BREAST CANCER DATABASE**” completed by **MOZHGAN MOAZZEN ZADEH** under supervision of **ASSIST. PROF. DR. METEHAN MAKİNACI** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.



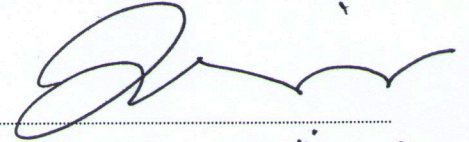
Assist. Prof. Dr. Metehan MAKİNACI

Supervisor



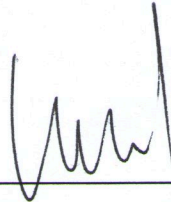
Assist. Prof. Dr. Nalan ÖZKURT

Jury Member



Assist. Prof. Dr. Bonu ÖZGÜR

Jury Member



Prof. Dr. Mustafa SABUNCU

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

I would like to express my profound gratitude and appreciation to my advisor Assist. Prof. Dr. Metehan MAKİNACI, for his consistent help and attention that, he devoted throughout the course of this work. He was always kind, understanding, and sympathetic to me. His valuable suggestions and useful discussions made this work interesting for me. I am deeply grateful to him.

I would like to give thanks to the staff and my all colleagues at Graduate School of Natural and Applied Sciences, in Dokuz Eylül University, for their supports and good wishes.

Finally, I also wish to express my deepest gratitude to my family for their helping, supporting and encouraging me in my whole life.

Mozhgan Moazzen Zadeh

BAYES GAUSSIAN CLASSIFICATION OF WISCONSIN BREAST CANCER DATABASE

ABSTRACT

The correct pattern classification of breast cancer is an important medical problem. Breast cancer etiologies remain unclear and no single dominant cause has emerged. Prevention is still a mystery and the best way to improve patient survival is through early detection. If the cancerous cells are detected before they spread to other organs, the survival rate is greater than 97 percent.

A major class of problems in medical science involves disease diagnosis based on various tests performed on patients. For this reason, the use of classifier systems in medical diagnosis is gradually increasing. There is no doubt that data evaluation taken from patients and experts decisions are the most important factors in diagnosis. Besides, artificial intelligence classification techniques can enhance current research. Classification systems, through minimizing possible errors likely produced due to tiredness or lack of experience, can provide more detailed medical data that can be checked in a shorter period.

In this study, we focused on developing a medical decision-making application using Bayes Gaussian classification method. At the first step, theoretical derivations are adopted into our problem then we used MATLAB to write a computer program to be able to test developed algorithm.

The purposed medical decision making system has been applied on the task of diagnosing breast cancer. Test of the developed classifier is carried out by using Wisconsin Breast Cancer Database. The 10-fold cross validation results show that overall accuracy is 94.38 percent.

Keywords: Bayes gaussian classification, breast cancer.

WISCONSIN GÖĞÜS KANSERİ VERİ TABANININ BAYES GAUSSIAN SINIFLANDIRILMASI

ÖZ

Kanser dünyadaki başlıca ölüm nedenlerinden biridir ve bu yüzden tedavisi bilim dünyası için önemli bir konu haline gelmiştir. Göğüs kanseri nedenleri belirsizliğini korumaktadır ve hiçbir baskın neden ortaya çıkmamıştır. Hastanın yaşam süresini artırmanın en iyi yolu erken tanıdan geçmektedir. Doğru bir tanı sistemi ile durumu erken teşhis edilen kanserli hücrelerin diğer organlara yayılmadan önce tespiti, hastanın yüzde 97 oranında iyileşmesine olanak sağlamaktadır.

Tıp bilimindeki birçok problem, hasta üzerinde yapılan çeşitli testlere dayalı hastalık teşhislerini gerektirir. Bu nedenle, tıbbi tanı sınıflandırma sistemlerinin kullanımı giderek artmaktadır. Hiç şüphe yok ki, uzmanların alınan veriler üzerindeki değerlendirmeleri hastalık tanısında en önemli ve etkin faktörlerdendir. Ancak yapay zekâ sınıflandırma teknikleri mevcut araştırmaları daha da etkinleştirebilir. Sınıflandırma sistemleri, muhtemelen yorgunluk veya tecrübe eksikliği nedeniyle olacak hataları en aza indirerek daha sağlıklı tıbbi kararlar alınmasını sağlayabilir.

Bu çalışmada Bayes-Gauss sınıflandırma yöntemini kullanan bir tıbbi karar alma sistemi geliştirilmesine odaklanılmıştır. İlk aşamada teorik ifadeler probleme adapte edilmiştir, daha sonra, oluşturulan algoritmayı test edebilmek için bir MATLAB programında bilgisayar kodu hazırlanmıştır.

Geliştirilen tıbbi karar verme sistemi göğüs kanseri teşhisinde uygulanmıştır. Geliştirilen sınıflandırıcının testi, Wisconsin Göğüs Kanseri veri tabanı kullanılarak gerçekleştirilmiştir. 10-kat çapraz doğrulama sonuçlarına göre yöntemin doğruluğu yüzde 94,38 olarak elde edilmiştir.

Anahtar Kelimeler: Bayes gauss sınıflandırma, göğüs kanseri.

CONTENTS

	Page
M.Sc. THESIS EXAMINATION RESULT FORM.....	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT.....	iv
ÖZ	v
CHAPTER ONE – INTRODUCTION	1
1.1 Breast Cancer.....	1
1.2 Wisconsin Breast Cancer Database	2
1.3 Literature Review	3
1.4 Outline	4
CHAPTER TWO – INTRODUCTION TO CLASSIFIERS & PATTERN RECOGNITION	5
2.1 Classifier	5
2.2 Pattern Recognition Application.....	5
2.3 Pattern Recognition Approaches	8
2.4 Parametric & Non-parametric Classifiers.....	9
2.4.1 Parametric Classifiers	9
2.4.2 Non-parametric Classifiers	10
2.5 Commonly Used Classifiers	10
2.6 Problems with Classifiers	12
CHAPTER THREE – BAYES GAUSSIAN CLASSIFIER	14
3.1 Bayes Gaussian Classifier.....	14
3.2 Limitations of Bayes Gaussian Classifier	18
CHAPTER FOUR – APPLICATION OF CLASSIFICATION & RESULTS ..	19
4.1 Materials & Methods	19
4.2 Numerical Results.....	19

4.3 Classification Evaluation	21
4.3.1 K-fold Cross-validation	21
4.3.2 Confusion Matrix	22
4.3.3 Comparison with other Classification Methods	24
CHAPTER FIVE – CONCLUSION	27
REFERENCES	29

CHAPTER ONE

INTRODUCTION

1.1 Breast Cancer

Breast cancer is the form of cancer that either originates in the breast or is primarily present in the breast cells. The disease occurs mostly in women but a small population of men is also affected by it. Breast cancer is the most common form of cancer amongst the female population as well as the most common cause of cancer deaths (Sewak, Vaidya, Chan & Duan, 2007). According to the Turkish Health Ministry resources, the number of breast cancer incidents has increased in the last decades. In 2011, the estimated number of breast cancer patients in Turkey is over 50,000. It is estimated that 1 out of every 8 women develop breast cancer at one point in their lives (Ozmen, 2008).

Early detection of breast cancer saves many thousands of lives each year. Besides, there are many similarities between the structures of the malignant and benign tumors. Hence, an extremely difficult and time-consuming task is needed to separate whether the tumor is malignant or benign.

As an illustration, in Figure 1.1, two different biopsy images are given to compare the difference between the structures of the malignant and benign tumors. For an untrained eye, it is hard to classify whether the image corresponds to a malignant or benign tumor. In medical care process, accurate classification is crucial as the effect of the cytotoxic drugs administered during the treatment can be life threatening or may cause another cancer. Nowadays, analyses and biopsies for the structures of the tumor are carried out manually. Hence, the accuracy becomes lower and, the process takes longer time to be completed. An automated system is needed to achieve a faster and more reliable method for predicting the type of the tumor, with avoiding the human nature errors.

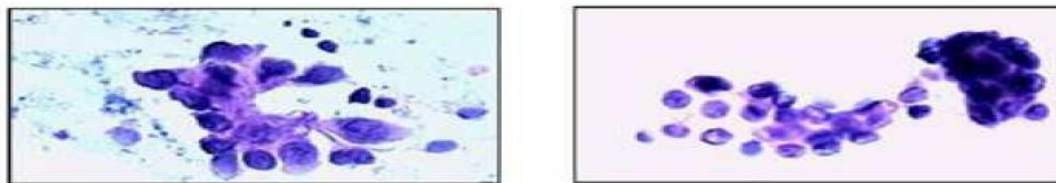


Figure 1.1 Fine needle biopsies of breast. Malignant (left) and Benign (right) (Sewak, Vaidya, Chan & Duan, 2007).

1.2 Wisconsin Breast Cancer Database

The Wisconsin Breast Cancer Database (WBCD) was originally provided by Dr. William H. Wolberg (Mangasarian & Wolberg, 1990). In our experiments, WBCD is used which is taken from the University of California, Irvine (UCI) machine learning repository. This dataset is commonly used among researchers who use machine-learning methods for breast cancer classification, so it provides us to compare the performance of our method with that of others.

The current database consists of 699 instances taken from needle aspirates from patients' breast tissue. While 458 cases belong to the benign class, remaining 241 cases belong to the malignant class. This dataset contains 16 instances with missing attribute values. Because of disregarding the missing 16 data samples, 683 cases are used in the current study. There are nine attributes to classify the samples into two categories: benign or malignant. These nine attributes are listed in Table 1.1. Attributes are graded 1-10, with 10 being the most abnormal state. The class attribute is represented as 2 for benign and 4 for malignant cases.

Table 1.1 Wisconsin breast cancer database: attribute information.

Number	Attribute description	Possible Value
1	Sample code number	id number
2	Clump thickness	1-10
3	Uniformity of cell size	1-10
4	Uniformity of cell shape	1-10
5	Marginal Adhesion	1-10
6	Single Epithelial cell size	1-10
7	Bare Nuclei	1-10
8	Bland Chromatin	1-10
9	Normal Nucleoli	1-10
10	Mitoses	1-10
11	Class	2 for benign, 4 for malignant

$N = 683$ observations, 242 malignant and 441 benign.

1.3 Literature Review

There has been a lot of research on medical diagnosis of breast cancer with WBCD in literature, and most of them reported high classification accuracies. We investigated some thesis and article mentioned below.

In Albrecht, Lappas, Vinterbo, Wong & Ohno-Machado (2002), a learning algorithm that combined logarithmic simulated annealing with the perceptron algorithm was used and the reported accuracy was 98.8%. In Pena-Reyes and Sipper (1999), the classification technique used Fuzzy-GA method reaching a classification accuracy of 97.36%. In Setiono (2000), the classification was based on a feed forward neural-network rule extraction algorithm. The reported accuracy was 98.10%. Quinlan (1996) reached 94.74% classification accuracy using 10-fold cross-validation with C4.5 decision tree method. Hamiton, Shan, & Cercone (1996) obtained 94.99% accuracy with RIAC method, while Ster & Dobnikar (1996) obtained 96.8% with linear discrete analysis method. The accuracy obtained by Nauck & Kruse (1999) was 95.06% with neuro fuzzy techniques. In Goodman, Boggess & Watkins (2002), three different methods, optimized learning vector quantization (LVQ), big LVQ, and artificial immune recognition system (AIRS), were applied and the obtained accuracies were 96.7%, 96.8%, and 97.2%, respectively. In Abonyi & Szeifert (2003), an accuracy of 95.57% was obtained with the application of supervised fuzzy clustering technique. In Polat & Gunes (2007), least square SVM was used and an accuracy of 98.53% was obtained. In Hassanien (2004), the classification technique used rough set method reaching a classification accuracy of 98%. In Sahan & Polat (2007), a new hybrid method based on fuzzy-artificial immune system and K-NN algorithm was used and the obtained accuracy was 99.14%. In Maglogiannis (2009), three different methods, SVM, Bayesian classifiers and Artificial Neural Networks were applied and the obtained accuracies were 97.54%, 92.80% and 97.90%, respectively. Besides, in Karabatak & Ince, (2009), the method combined with association rules, and neural-networks were utilized and accuracy of 95.6% was obtained. Multilayer perceptron neural network, four different methods, combined neural network, probabilistic neural network,

recurrent neural network, and SVM were used respectively; highest classification accuracy of 97.36% was achieved by SVM (Ubeyli, 2007).

1.4 Outline

This thesis consists of four chapters. In the first chapter the breast cancer, WBCD and typical attributes of this database are introduced. A literature review and outline are also presented in this chapter.

The second chapter discusses pattern recognition, commonly used classifiers, and problem with these classifiers. In chapter three an introduction and theoretical background of Bayes Gaussian classifier are given. The chapter four represents applications of classifiers, results, and evaluations. Implementation of Bayes Gaussian classifiers on WBCD is performed by using MATLAB software. Comparison of simulation results with the other classification methods is also presented in chapter four.

CHAPTER TWO

INTRODUCTION TO CLASSIFIERS & PATTERN RECOGNITION

2.1 Classifier

As the number of patients is increasing, a computerized medical system becomes crucial in the decision process. A classifier is required as part of this decision process to assist the analyzers in diagnosing a disease. Computer based medical system can provide more accurate and faster data analyses in comparison with the manual data analyses. As a result, development of computer based classification techniques is very important to assist the analyzers in diagnosis.

Classifier is known as an algorithm which provides an output depending on the features defined as input. In other words, classifier is based on information that is defined into the classifier algorithm and its parameters. Even the output of the classifier is generally a label, it also contains reliable values.

2.2 Pattern Recognition Application

In a pattern recognition system, the given sensor data is segmented and the features are extracted from it. Using these features as an input vector, a classifier is designed. Based on a decision rule, the class of the data is estimated. Actual pattern recognition systems may be more complicated and may have many more elements. A simplified block diagram of a recognition system is presented in Figure 2.1, with all its main functional components.

The most important physical senses of living organisms are vision, hearing, taste, smell and touch. Vision is the ability of the brain and eye to recognize a face, written characters, and color of light. The sense of sound perception is hearing, such as recognizing spoken words or speaker. Taste, smell and touch are the activities of the receptors and/or neurons in the body. All these activities are the complex processes of the pattern recognition.

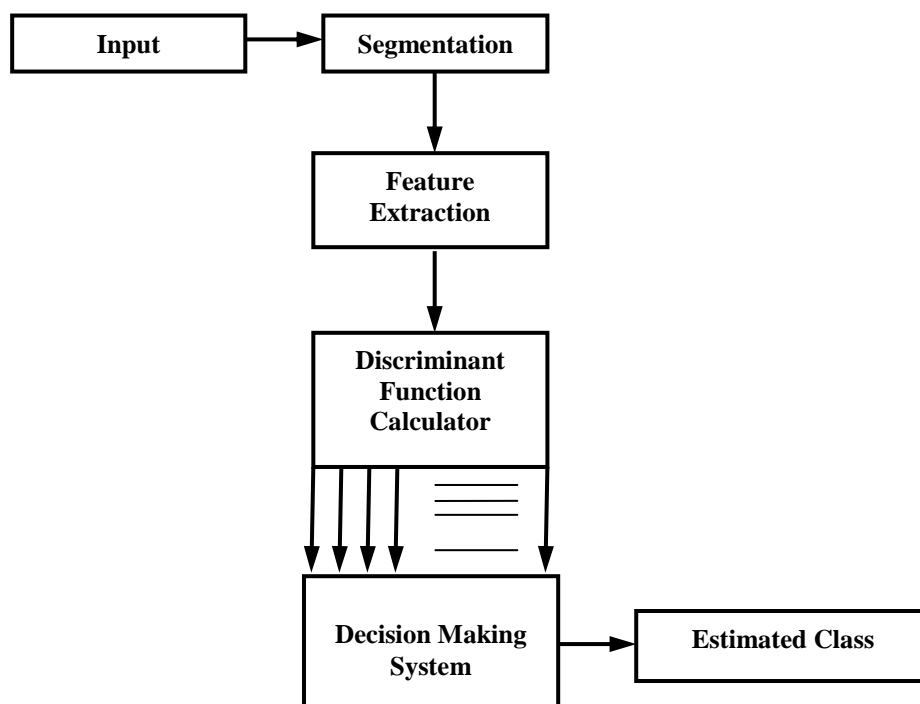


Figure 2.1 Recognition system with classifier.

Pattern recognition is the act of taking in raw data and making an action based on the “category” of the pattern (Duda, Hart & Stork, 2001). In the most instances, humans are the best pattern recognizers. The main goal of pattern recognition is to design systems that can recognize patterns (Jain, Duin & Mao, 2000).

The final decision of the classifier depends on how well certain “features” or “properties” of patterns are extracted and how the classifier is trained. Training a classifier is the process of obtaining design samples or features of the different categories of the data. The greater the training samples the better estimate. In practical applications, it may not possible to collect large number of training samples, and may need to design a classifier with limited number of training signals.

A typical pattern recognition system includes:

- Data collection,
- Feature selection,
- Model selection,
- Training,

- Evaluation.

A transducer, appropriate for particular data, is used for data acquisition. The characteristics of transducer such as bandwidth, gain, distortion, and signal to noise ratio will dictate the quality of data collected. Choosing features that carry the most discriminatory information across the pattern classes and selecting the model (Statistical, neural or structural approaches) are critical design steps. Training a classifier is the process of determining the parameters of the classifier from the feature set. The final step involves an evaluation of the performance of the pattern recognition system using test data. In all practical pattern recognition problems, majority of time has to be spent in learning and it is difficult to predict the classification model without effective learning.

Learning can be divided into two types:

- Supervised learning in which the class of an input pattern is predicted from labeled training samples.
- Unsupervised learning or clustering involving unlabeled training data.

Applications of Pattern recognition is in a variety of engineering, marketing and scientific disciplines. And there are countless pattern recognition applications where classifiers are used. Some examples are given here:

- Automatic form reading,
- Face recognition,
- Fingerprint processing,
- Automatic target recognition,
- Bottle cap recognition,

In automatic form reading, character recognition techniques are used to read the forms and identify their content. For example, in automatic mail processing, all the different regions of interest on the envelope are segmented such as; main address,

return address, barcode etc. In addition to these, the zip code is the main region of interest.

Face recognition is another developing application, which is often used to restrict building access to authorized people. For this application, the selected database contains the faces of those individuals who have access. When a person seeks access, the recognition system extracts features from the face images and decides whether the face is similar to the one of the stored ones.

Fingerprint processing is similar to previous method, but it is more simple and practical, since fingerprints are almost described two-dimensional. Automatic target recognition is a defense related application where we try to locate man-made objects and classify them as friend or foe.

Bottle cap recognition is used by various airlines where they need to sort out different beverages bottles automatically.

2.3 Pattern Recognition Approaches

Three common pattern recognition approaches are listed as followings:

1. Statistical: In statistical approach, pattern classification is based on an underlying statistical model of the features. The statistical model is defined by a family of class-conditional probability density functions $p(x|\omega_c)$ (Probability of feature vector x given class ω_c).

2. Neural: In neural approach, classification is based on the response of a network of processing units (neurons) to input stimuli (pattern). “Knowledge” is stored in the connectivity and strength of the synaptic weights. Neural pattern recognition is trainable and non-algorithmic. Neural pattern recognition is very attractive since it requires minimum prior knowledge with enough number of layers and neurons. An ANN can create any complex decision region.

3. Syntactic (structural): In structural approach, pattern classification is based on measures of structural similarity. “Knowledge” is represented by means of formal grammars or relational descriptions (graphs). Syntactic Pattern Recognition approaches formulate hierarchical descriptions of complex patterns built up from simpler sub patterns.

2.4 Parametric & Non-parametric Classifiers

In statistical pattern recognition, selection of classifier depends on class conditional densities. The optimal Bayes decision rule can be used to design a classifier if all of the class-conditional densities are completely known. However, the class conditional densities must be estimated from available training pattern because they are usually unknown. Depending on the availability of class conditional densities, classifiers can be divided into two types: 1) Parametric classifiers, and 2) Non-Parametric classifiers (Jain, Duin & Mao, 2000).

2.4.1 Parametric Classifiers

Parametric classifiers use a statistical approach to classify patterns. Parametric classifiers can be represented by their discriminant functions. Examples of discriminant functions include:

- Gaussian Discriminant Function:

$$g(\omega_c|x) = -\frac{1}{2}(x - \mu_c)^T \Sigma_c^{-1} (x - \mu_c) - \frac{1}{2} \ln |\Sigma| \ln P(\omega_c) \quad c = 1, 2, \dots, C \quad (2.1)$$

Where x is a D-dimensional column feature vector, μ_c is the D-dimensional mean column vector of class ω_c , Σ is the $(D \times D)$ covariance matrix of class ω_c , and $P(\omega_c)$ is the prior probability of class ω_c .

- Quadratic Discriminant Function:

$$g(\omega_c|x) = -\frac{1}{2}(x - \mu_c)^T \Sigma^{-1}(x - \mu_c) + \ln P(\omega_c) \quad c = 1, 2, \dots, C \quad (2.2)$$

- Mahalanobis Discriminant Function:

$$g(\omega_c|x) = -(x - \mu_c)^T \Sigma^{-1}(x - \mu_c) \quad c = 1, 2, \dots, C \quad (2.3)$$

- Nearest Mean Discriminant Function:

$$g(\omega_c|x) = -\|x - \mu_c\|^2 \quad c = 1, 2, \dots, C \quad (2.4)$$

In practice, the mean vectors and covariance matrices in the discriminant functions are estimated using the feature vectors in the training set.

2.4.2 Non-parametric Classifiers

Supervised learning is conducted with the assumption that the class-conditional densities are known. However, in most practical pattern recognition problems, this assumption may not be true. The non-parametric procedures can be used with arbitrary distribution and without assumption that the forms of class conditional densities are known. Nonparametric classifiers are better alternative to parametric classifiers however; they usually require a large amount of training set. The nearest neighbor (1-NN) and k-nearest neighbor rule classifiers are the most widely used nonparametric classifiers.

2.5 Commonly Used Classifiers

Following classifiers are commonly used in pattern recognition:

- The Bayesian classifiers,
- Bayes Gaussian classifier,
- Nearest neighborhood classifier,
- k-Nearest (k-NNC),

- Neural Networks classifiers.

Bayesian classification and decision making is based on probability theory and the principle of choosing the most probable or the lowest risk (expected cost) option (Theodoridis & Koutroumbas, 2000). The major advantage of the Bayes classifier is its short computational time for training since it requires relatively small amount of training data to estimate the parameters for classification. Bayes classifier is also robust to missing values because these values are simply ignored in computing probabilities and thus have no impact on the final decision.

Bayes Gaussian classifier is a Bayes classifier for data input classes having Gaussian distribution. The classifier learns from training data and estimates the posterior probabilities of the classes given particular instance of the features using Bayes theorem assuming Gaussian pdf for the data features. Prediction of the class is determined by identifying the class with the highest posterior probability.

The nearest neighbor rule for classifying a test pattern is to assign a label of the class of a training sample that is the nearest as measured by a distance metric. This is the most fundamental and simplest supervised classification technique; however, it tends to be computationally intensive. This method would be a first choice when there is no prior knowledge about the distribution of the data. The nearest neighbor method is also very robust to noise. The removal of few random samples or artifacts in training data does not affect the performance of the classifier. Furthermore, the nearest neighbor rule and the optimal Bayes classifier are identical under certain conditions (Duda, Hart & Stork, 2001).

The k -nearest neighbor classifier is an extension of nearest neighbor classifier. A test pattern is assigned to the class that is the most frequent among k -nearest neighbors in the training set, and generally, k is chose to be odd. The k -nearest neighbor rule becomes optimal when k tends to infinity. The nearest neighbor and k -nearest neighbor classifiers have been used in hand written character recognition, to predict the secondary structure of proteins, breast cancer detection, chromosome

image classification, infrared face recognition, EEG classification, and detecting bruise in apples.

Several Artificial Neural Networks (ANN) methods have also been used for classification. ANN can typically undergo supervised learning. In supervised learning, there exist the input feature vector, X_p , and the feature vector's class label, $i_c(p)$. Multi-layer Perceptrons (MLP), radial basis function (RBF) networks and support vector machines (SVM) are trained using supervised learning techniques. ANN classifiers are usually trained to minimize the Mean-Square Error (MSE) over the number of iterations.

2.6 Problems with Classifiers

In Bayes classifiers, the required conditional probability densities are usually not available. Only approximations from parametric and non-parametric modeling approaches are available.

The k-Nearest Neighbor classifier is quite simple, but computationally very intensive to design. Even for simple classification problem, it requires high memory area in calculations, which makes the classification a complex process. Theorems on convergence to Bayes error do exist for nearest neighbor classifiers (NNCs) and k-NNCs (Duda, Hart & Stork, 2000; Fukunaga, 1990), which also have the advantage of being easy to design in a short period. Nevertheless, owing to the time-consuming procedure to apply, the NNC and k-NNC methods are preferred rarely. Also, ANN classifiers have many problems. The most common problems of the back propagation algorithm in MLP training are the possibility of ending up in a local minimum of the error function and the time for convergence. Training time for MLP and RBF classifiers can be long and they may suffer from over fitting (Duda, Hart & Stork2000). SVM classifiers avoid over fitting but usually require several orders of magnitude more hidden units than RBF and MLP networks. In addition, because of the number of patterns provided during training, MLP can suffer from memorization problems. As SVM's require hundreds or thousands of parameters, it takes long time

to apply. A large numbers of support vectors are needed to obtain satisfactory performance for SVM's.

CHAPTER THREE

BAYES GAUSSIAN CLASSIFIER

3.1 Bayes Gaussian Classifier

A classifier calculates discriminant functions for each class and makes the decision according to which class's discriminant is largest or smallest. Bayes Classifier is a simple probabilistic classifier based on the Bayes rule (Bayes, 1763). This classifier can be designed if statistical information of the system including conditional probability density of the feature vectors is available and well defined. Our goal in Bayes classifier design is to develop the discriminant function that minimizes the probability of classification error.

In this study, Bayes Gaussian classifier is used because of some properties of the Gaussian, such as:

- Analytically tractable,
- Completely specified by the 1st and 2nd moments,
- Has the maximum entropy of all distributions with a given mean and variance,
- Many processes are asymptotically Gaussian (because of Central Limit Theorem),
- Linear transformations of a Gaussian are also Gaussian,
- Uncorrelatedness implies independence.

The Bayes Gaussian Classifier (BGC) is a Bayes classifier where the conditional pdf $p(x|\omega_c)$ is assumed Gaussian. With the other sentence a Bayes classifier is a probabilistic classifier that makes decisions by combining two sources of information, i.e., the prior and the likelihood, to form a posterior probability using Bayes' rule (Bayes, 1763). We can say that when the feature vectors are jointly Gaussian, the result is the Bayes Gaussian classifier (BGC). Most of the data available in the real world is Gaussian because of the “*Central Limit Theorem*”

(Papoulis, 2002), so this classifier is applicable in many real world applications.

The classifier uses Bayes theorem with Gaussian distribution for pattern classification (Theodoridis & Koutroumbas, 2009; Conrath, 2004). First we consider the univariate case, with a continuous random variable x , whose pdf, given class, ω_c , is a Gaussian with mean μ and variance σ^2 .

Using Bayes' theorem we can write:

$$p(\omega_c|x) = \frac{p(x|\omega_c)p(\omega_c)}{p(x)} \quad (3.1)$$

Where,

- $p(x|\omega_c)$: Class conditioned probability or likelihood,
- $p(\omega_c)$: A priori or prior probability,
- $p(x)$: Evidence (*usually ignored*),
- $p(\omega_c|x)$: Measurement-conditioned or posterior probability

$$p(\omega_c|x) = \frac{p(x|\omega_c)p(\omega_c)}{\sum_{k=1}^C p(x|\omega_k)p(\omega_k)} \quad (3.2)$$

$$p(\omega_c|x) \propto p(x|\omega_c)p(\omega_c) \quad (3.3)$$

An essential statistical approach to solving the problem of pattern recognition is Bayesian decision theory. It supposes that the decision problem is revealed in probabilistic terms. With choosing the state of nature that maximizes the posterior probability $p(x|\omega_c)$, probability of error in a classification problem can be minimized. Bayes formula allows us to calculate such probabilities given the prior probabilities $p(\omega_c)$ and the conditional densities $p(x|\omega_c)$ for different categories.

To simplify the above equation, we take the logarithm of the equation,

$$\log p(\omega_c|x) \propto \log p(x|\omega_c) + \log p(\omega_c) \quad (3.4)$$

$$LP(\omega_c|x) \propto LL(x|\omega_c) + LP(\omega_c) \quad (3.5)$$

where $LP(\omega_c|x)$ is the log of the posterior probability, $LL(x|\omega_c)$ is the log of the likelihood and $LP(\omega_c)$ is the log of the prior probability. The log of the posterior probability ratio (LPPR) is then defined as,

$$\log \frac{p(\omega_a|x)}{p(\omega_b|x)} = LP(\omega_a|x) - LP(\omega_b|x) \quad (3.6)$$

$$\log \frac{p(\omega_a|x)}{p(\omega_b|x)} = [LL(x|\omega_a) - LL(x|\omega_b)] + [LP(\omega_a) - LP(\omega_b)] \quad (3.7)$$

For the one-dimensional case, the Gaussian probability density function (Univariate Gaussian pdf) is defined as

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (3.8)$$

where, mean (μ), expected value of x ,

$$\mu = E[x] = \int_{-\infty}^{\infty} xp(x) dx \quad (3.9)$$

and variance (σ), expected squared deviation of x ,

$$\sigma^2 = E[(x-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx \quad (3.10)$$

If we assume the probability density function is Gaussian, the log of the likelihood becomes,

$$\begin{aligned} LL(x|\mu,\sigma^2) &= \log p(x|\mu,\sigma^2) \\ &= -\log \left\{ \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \right\} \\ &= -\log(\sqrt{2\pi\sigma^2}) - \frac{(x-\mu)^2}{2\sigma^2} \end{aligned} \quad (3.11)$$

Therefore, the log of the posterior probability is

$$LP(\omega_c|x) \propto \frac{1}{2} \left[-\log(2\pi) - \log \sigma^2 - \frac{(x-\mu)^2}{\sigma^2} \right] + \log p(\omega_c) \quad (3.12)$$

If ω_a and ω_b are modeled by Gaussians with means μ_a and μ_b , variances σ_a^2 , σ_b^2 then we can write the log ratio of posterior probabilities as follows:

$$\begin{aligned} \ln \frac{p(\omega_a|x)}{p(\omega_b|x)} &= \frac{1}{2} \left[-\log(2\pi) - \log(\sigma_a^2) - \frac{(x-\mu_a)^2}{\sigma_a^2} \right] \\ &\quad - \frac{1}{2} \left[-\log(2\pi) - \log(\sigma_b^2) - \frac{(x-\mu_b)^2}{\sigma_b^2} \right] + [\log p(\omega_a) - \log p(\omega_b)] \end{aligned} \quad (3.13)$$

$$\begin{aligned} \ln \frac{p(\omega_a|x)}{p(\omega_b|x)} &= -\frac{1}{2} \left[\frac{(x-\mu_a)^2}{\sigma_a^2} - \frac{(x-\mu_b)^2}{\sigma_b^2} + \log(\sigma_a^2) - \log(\sigma_b^2) \right] \\ &\quad + [\log p(\omega_a) - \log p(\omega_b)] \end{aligned} \quad (3.14)$$

If we consider D -dimensional data x from class C modeled using a multivariate Gaussian,

$$p(x|c) = p(x|\mu, \Sigma) = N(x; \mu, \Sigma) \quad (3.15)$$

$$p(x|c) = \frac{1}{(2\pi)^{\frac{d}{2}} \Sigma^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right] \quad (3.16)$$

$$\Sigma = \text{Cov}(x) = E \left[(x-\mu)(x-\mu)^T \right] = \int_{-\infty}^{\infty} (x-\mu)(x-\mu)^T p(x) dx \quad (3.17)$$

We can take the logarithm to obtain the log of the likelihood,

$$\begin{aligned} LL(x|\mu, \Sigma) &= \log p(x|\mu, \Sigma) = -\frac{d}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| \\ &\quad - \frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \end{aligned} \quad (3.18)$$

We know that, $(x-\mu)^T \Sigma^{-1} (x-\mu)$ is the Mahalanobis distance and we can write the log of the posterior probability,

$$\log p(\omega_c|x) \propto \frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) - \frac{1}{2} \log |\Sigma| + \log P(\omega_c) \quad (3.19)$$

3.2 Limitations of Bayes Gaussian Classifier

The most important limitation of the BGC is that the distribution of the given data should be Gaussian. But, the covariance matrix is generally singular (non-invertible), That causes problems during calculation of the inverse covariance matrix required in the discriminant function. The error curve could not be monotonically non-increasing because the features may not be arranged in order of their importance. Additionally, the weights calculated by the statistical information of the data may not be exact.

CHAPTER FOUR

APPLICATION OF CLASSIFICATION & RESULTS

4.1 Materials & Methods

For having a classification, database has to be separated to two parts. They are train and test datasets. Our method was trained with the same training data set and tested with the same evaluation data set to comparatively evaluate the performance of the various classifiers presented in this study. 480 samples was used in this study, the network was trained with 432 samples, which consist of 216 malignant and 216 benign samples. The testing set consisted of 24 malignant samples and 24 benign samples. Both of these data sets have equal number of instances from each class, *B* or *M*. Figure 4.1 shows this allocation.

The BGC algorithm was developed in MATLAB[®] (version 7.4, R2007a).

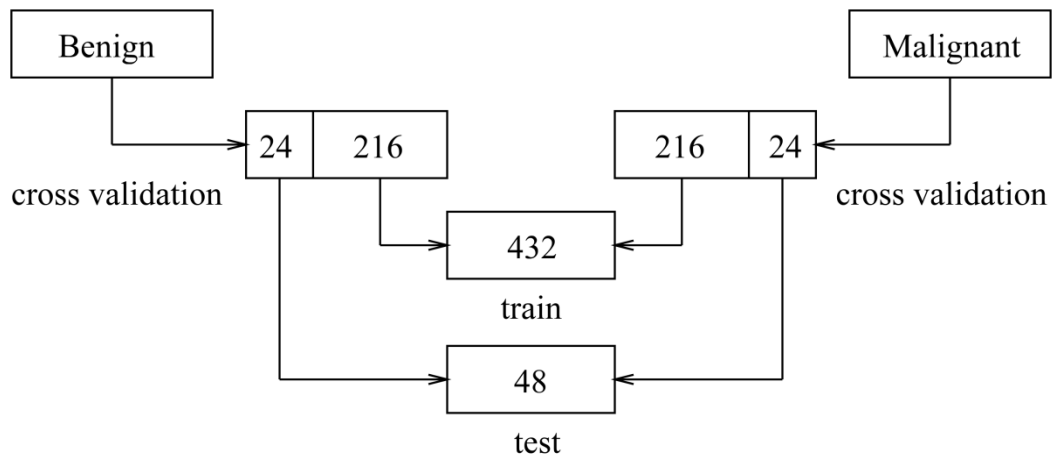


Figure 4.1 Choosing train and test dataset.

4.2 Numerical Results

Table 4.2 and table 4.3 show the accuracy percentages of Bayes Gaussian classification method. As we have 10 set of 24 benign and 24 malignant data, the minimum and maximum accuracy values of the classifier implementation are obtained as 87.50% and 97.92%. The overall accuracy of 94.38% is calculated after

10-fold cross-validation.

Table 4.2 Overall results

Average Classification Result	Maximum Classification Result	Minimum Classification Result
94.38%	97.92%	87.50%

In Table 4.3 it can be seen that, in the minimum accuracy result, true predictions of the malignant instances have higher values than the benign ones, and in the maximum accuracy case predictions are vice versa. In minimum accuracy result, the Bayes Gaussian classifier could predict 80% of true positive (Truly predicted benign in benign category) and 20% of false positive (falsely predicted binge in malignant category). Maximum accuracy result consists of 100% of benign instances in the benign group and 95.66% of malignant instances in the malignant group.

Table 4.3 Minimum, maximum and overall accuracy of Bayes Gaussian classification

Cross-validation	Original	Predicted Group Membership		Total (%)
		Benign (%)	Malignant (%)	
Min. accuracy 87.50%	Benign	80	20	100
	Malignant	9.09	90.91	100
Max. accuracy 97.92%	Benign	100	0	100
	Malignant	95.66	4.34	100
Overall accuracy 94.38%	Benign	90.42	9.58	100
	Malignant	2.56	97.44	100

In overall accuracy result, the Bayes Gaussian classifier could accurately predicts 90.42% of benign instances in their proper category and 97.44 percent of malignant instances in the malignant category. It means that the overall accuracy of the classifier have 2.56% of malignant samples in their proper category and 9.58% of benign in false category.

4.3 Classification Evaluation

4.3.1 K-fold Cross-validation

Cross-validation method is used to increase the accuracy of test results in current study (Duda, Hart & Stork, 2001). It minimizes the bias associated with the random sampling of the training (Sirakaya, Delen & Choi, 2005). In this method, whole data is randomly divided to “ k ” mutually exclusive and approximately equal size subsets. The classification algorithm trained and tested “ k ” times. In each case, one of the folds is taken as test data and the remaining folds are added to form training data. Thus, for each training–test configuration, a total of “ k ” test results are obtained. The average of these results gives the test accuracy of the algorithm (Sirakaya, Delen & Choi, 2005).

Creating a K-fold partition of the dataset is presented in Figure 4.2. For each of K experiments, use K-1 folds for training and a different fold for Testing. This procedure is illustrated in the following figure for K=4.

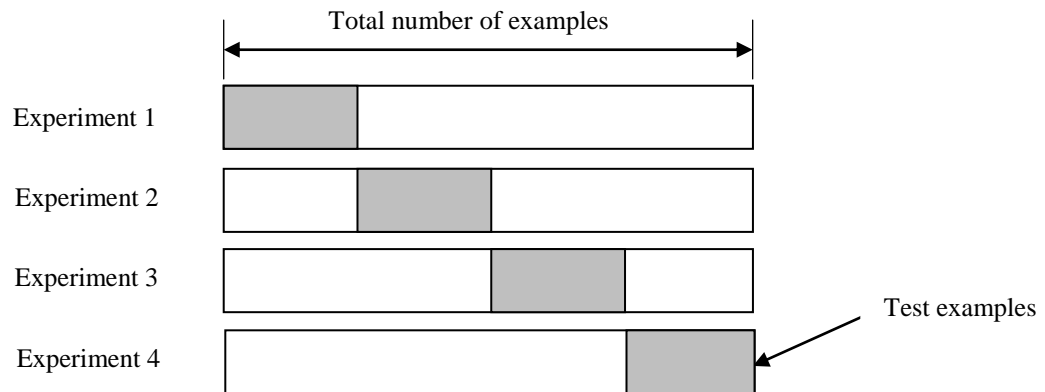


Figure 4.2 create a K-fold partition of the dataset.

The aim of using cross validation is to have an average result for calculation. By using this way realistic values can be obtained and avoid different peak results of our classification. Cross-validation is repeated for choosing test dataset before every classification process.

4.3.2 Confusion Matrix

Confusion matrix is obtained from “Wisconsin Breast Cancer Database”. In order to evaluate the prediction performance of BG classifier, we define and compute the classification accuracy, sensitivity, specificity, and confusion matrix.

Classification accuracy is measured using the following equation,

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.1)$$

where TP , TN , FP , and FN denote true positives, true negatives, false positives, and false negatives, respectively. Explanations for each abbreviation is given below,

- *True positive (TP)*: An input is from a patient with breast cancer (input is abnormal), but diagnosed as breast cancer by the clinic experts,
- *True negative (TN)*: An input is normal and is labeled as a healthy individual by the expert clinicians,
- *False negative (FN)*: An input is from a patient with breast cancer, but is labeled as a healthy person by the expert clinicians,
- *False positive (FP)*: An input is normal but diagnosed as breast cancer.

The following expressions are used for sensitivity and specificity analyses,

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100 \quad (4.2)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \times 100 \quad (4.3)$$

A confusion matrix (Kohavi & Provost, 1998) contains information about actual and predicted classifications performed by a classifier. Performance of the classifier is commonly evaluated using the data in the matrix.

Table 4.4 shows the confusion matrix for a two-class classifier. Classification accuracy, sensitivity, and specificity can be defined by using the elements of the confusion matrix with the formulations, which are defined above.

Table 4.4 Confusion matrix representation

		Predicted	
		<i>Positive(Malignant)</i>	<i>Negative(Benign)</i>
Actual	<i>Positive(Malignant)</i>	TP(MM)	FN(MB)
	<i>Negative(Benign)</i>	FP(BM)	TN(BB)

B: Benign, *M*: Malignant

Table 4.5 shows the classification results obtained in the best simulation for Bayes Gaussian classifier that used in this study in a confusion matrix.

Table 4.5 Confusion matrixes for WBCD

		Predicted	
		<i>Positive(Malignant)</i>	<i>Negative(Benign)</i>
Actual	<i>Positive(Malignant)</i>	234	6
	<i>Negative(Benign)</i>	21	219

The system was able to accurately predict 453 of the 480 samples yielding an accuracy of 94.38%. The classifier was able to classify 219 benign samples in their proper category and 21 benign in false category, whereas on 6 out of 240 times it falsely classified cases malignant into the benign category.

We present values of sensitivity and specificity in Table 4.6. When the performance was considered individually, they carried out the prediction with an average accuracy of 94.38%. However when the majority decision was considered, the prediction accuracy was 94.375%. Thus the Bayes Gaussian classifier performed well to distinguish between benign and malignant characteristics of the WDBC dataset.

Table 4.6 Sensitivity and specificity

Classification Accuracy	$\frac{TP+TN}{TP+TN+FP+FN} \times 100$	94.375%
Specificity	$\frac{TN}{FP+TN} \times 100$	97.33%
Sensitivity	$\frac{TP}{TP+FN} \times 100$	91.76%

A sensitivity of 100% means that the test recognizes all sick people as such. Thus in a high sensitivity test, a negative result is used to rule out the disease. A specificity of 100% means that the test recognizes all healthy people as healthy. Thus, a positive result in a high specificity test is used to confirm the disease.

4.3.3 Comparison with other Classification Methods

Bayes Gaussian classification results compared with the best results obtained by other researchers using the same database. Results are summarized in Table 4.7 and there has been a lot of research on medical diagnosis of breast cancer with WBCD in literature, and most of them reported high classification accuracies.

Table 4.7 Classification accuracies obtained with our method and other classifiers from literature.

Authors	Method	Classification accuracy (%)
Marcano-Cedeño et al.(2010)	AMMLP	99.63
Übeyli (2007)	SVM	99.54
Akay (2009)	SVM-CFS	99.51
Sewak (2007)	SVM	99.29
Albrecht et al. (2002)	LSA MACHINE	98.8
Polat and Güneş (2007)	LS-SVM	98.53
Yang (2000)	NN	98.5
Mu et al. (2007)	SVM	98.4
Setiono(2000)	NEURO-RULE 2a	98.1
Anagnostopoulou et al. (2005)	NN	97.9
Wolberg (2007)	SVM	97.5
Karabatak & Ince (2009)	AR+NN	97.4
Pena-Reyes & Sipper (1999)	FUZZY-GA1	97.36
Güneşer (2009)	SVM	97.3
Güneşer (2009)	NN	97.2
Ster & Dobnikar (1996)	LDA	96.8
Conforti & Guido (2010)	SVM-SDP	96.79
Güneşer (2009)	LDA	96.6
Güneşer (2009)	KNN	96.06
Guijarro et al. (2007)	LLS	96
Li (2007)	SVM	95.6
Abonyi & Szeifert (2003)	SFC	95.57
Nauck & Kruse (1999)	NEFCLASS	95.06
Hamiton et al. (1996)	RAIC	94.99
Quinlan(1996)	C4.5	94.74
Current Study(2011)	BGC	94.38
Elouedi et al. (2010)	NB	94.19
Elouedi et al. (2010)	DT	93.12
Elouedi et al. (2010)	KNN	92.66

In Albrecht et al. (2002), a learning algorithm that combined logarithmic simulated annealing with the perceptron algorithm was used and the reported accuracy was 98.8%. In Pena-Reyes & Sipper (1999), the classification technique used fuzzy-GA method reaching a classification accuracy of 97.36%. In Setiono (2000), the classification was based on a feed forward neural network rule extraction algorithm. The reported accuracy was 98.10%. Quinlan (1996) reached 94.74% classification accuracy using 10-fold cross-validation with C4.5 decision tree method. Hamiton et al. (1996) obtained 94.99% accuracy with RIAC method, while Ster & Dobnikar (1996) obtained 96.8% with linear discreet analysis method. The accuracy obtained by (Nauck & Kruse, 1999) was 95.06% with neuro-fuzzy techniques.

In Abonyi & Szeifert (2003), an accuracy of 95.57% was obtained with the application of supervised fuzzy clustering technique. In Polat & Gunes (2007), least square SVM was used and an accuracy of 98.53% was obtained. Akay (2009) presented an SVM-based model using a grid search to optimize model parameters and an *F*-score calculation to select input features. Akay reached a classification accuracy of 99.51%. Übeyli (2007) used five classifiers SVM, probabilistic neural network, recurrent neural network, combined neural network and multilayer perceptron neural networks reported an accuracy of 99.54%. The BGC results were compared with the recently proposed algorithms applied to the WBCD database by Conforti & Guido (2010) in their report proposed an optimization model-based approach for learning the best kernel function to be embedded into the support vector machine (SVM) classifier. They generated an optimal kernel function by formulating and solving a semi-defined programming (SDP) model. They obtained accuracy results of 96.79%. However, their learning algorithm cannot be completed in a reasonable amount of time because the SDP/SVM model is computationally inefficient in the case of very large-scale sets. Marcano-Cedeño, Quintanilla-Domínguez & Andina (2010) obtained the best result so far with the AMMLP algorithm is 99.63%. Karabatak & Ince (2009) result with an expert system for detection of breast cancer based on association rules and neural network was 97.4%. Guijarro-Berdias, Fontenla-Romero, Perez-Sanchez & Fraguera (2007) reached

96.00% with linear learning method for multilayer perceptrons using least squares. Elouedi, Lefèvre & Mercier (2010) used three of well-known classifiers namely k-nearest neighbors (KNN), Naive Bayes (NB), and decision trees (DT) and their result was 92.66%, 94.19% and 93.12%. Sewak, Vaidya, Chan & Duan (2007) used SVM for WBCD and got 99.20% accuracy result. Anagnostopoulou, Rouskas, Kormentza & Vergados (2005) had an accuracy of 97.90% by using advanced neural network techniques, as well as Yang, Lu, Yu & Yu (2000) had 98.50%. These differences with the same methods on network specifications, importance order of attributes and some other reasons. Li, Mu & Wolberg (2007) made a classification application for WBCD. They all used SVM algorithm and they got 95.50%, 98.40% and 97.50% accuracies, respectively. Günşer's results for (KNN), (NN), (LDA), and (SVM) was 96.06%, 97.20%, 96.60%, 97.30%, respectively (Günşer, 2009). In this study, the accuracy is obtained as 94.38%.

CHAPTER FIVE

CONCLUSION

In this study, we focused on developing a medical decision-making application by using Bayes-Gaussian classification method. Bayes-Gaussian classifier is an effective and fundamental methodology for solving classification problem. The proposed medical decision making system based on Bayes Gaussian classifier has been applied on the task of diagnosing breast cancer. Experiments have been carried out on different portions of the WBCD, which is commonly used among researchers who use machine learning methods to diagnose breast cancer.

At the first step, theoretical derivations are adopted into our problem, then we used MATLAB software to code a computer program to be able to test developed algorithm. The functionality of the decision making system is verified by using 10-fold cross validation. The results of the cross validation method are compared with other studies. It is observed that the proposed method yields one of the highest classification accuracies (94.38%).

Additional performance measures such as sensitivity, specificity, and confusion matrices are also presented for Bayes Gaussian classifier. Considering the results, the developed Bayes Gaussian classifier gives very promising results in classifying the breast cancer.

Owing to the parametric and linear behavior of the Bayes Gaussian classifier, in the current study, lower accuracy values are achieved in comparison with other classifiers. Generally classifiers with high overall accuracy are nonparametric classifiers. Nonparametric classifiers are better alternative to parametric classifiers.

More accurate predictions obtain with using the nonparametric classifiers, but they usually require a large amount of the training set. It is obvious that, for the Bayes Gaussian classifier, higher classification accuracies could be achieved with increasing the number of instances obtained from the database.

As a conclusion, we believe that the proposed system can be very helpful to the physicians for their final decisions on their patients. By using such a tool, they can make very accurate decisions.

Further exploration of the data can yield more interesting results. This will be the focus of our future work. Also, besides of breast cancer problems, other medical diagnosis applications can be conducted by this system.

REFERENCES

- Abonyi, J. & Szeifert, F. (2003). Supervised fuzzy clustering for the identification of fuzzy classifiers. *Pattern Recognition Letters*, 24(14), 2195-2207.
- Akay, M.F. (2009), Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Systems with Applications*, 36 (2), 3240–3247.
- Albrecht, A.A., Lappas, G., Vinterbo, S.A., Wong, C.K. & Ohno-Machado, L. (2002). Two applications of the LSA machine. *Proceedings of the 9th International Conference on Neural Information Processing. (ICONIP)*, 2(1), 184-189. Nanyang Technical University.
- Anagnostopoulou, I., Anagnostopoulou, C., Vergados, D., Rouskas, A. & Kormentzas, G. (2006). The Wisconsin Breast Cancer Problem: Diagnosis and TTR/DFS time prognosis using probabilistic and generalized regression information classifiers. *J Oncol Reports*, 15, 975-81.
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53, 370-418.
- Bors, A.G. & Pitas, I. (1996). Median radial basis functions neural networks. *IEEE Transaction on Neural Networks*, 7(6), 1351-1364.
- Burges, C.J.C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2), 955-974.
- Boser, B. E., Guyon, I. M. & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers, *5th Annual ACM Workshop on COLT*. 144-152, Pittsburgh, PA, ACM Press.

- Casdagli, M. (1989). Nonlinear prediction of chaotic time series, *Physica D* (35), 335-356.
- Conforti, D. & Guido, R. (2010). Kernel based support vector machine via semidefinite programming: Application to medical diagnosis. *Computers & Operations Research*, 37(8), 1389-1394.
- Cristianini, N. & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press.
- Delashmit, W. H. & Manry, M.T. (2005). Recent Developments in Multilayer Perceptron Neural Networks, *Proceedings of the 7th annual Memphis Area Engineering and Science Conference (MAESC)*.
- Duda, R. O., Hart, P. E. & Stork, D. G. (2000). *Pattern Classification*, (2nd ed.). Wiley Interscience.
- Elouedi, Z., Lefèvre, E., & Mercier, D. (2010). Discountings of a Belief Function using a Confusion Matrix. *ICTAI*, 1, 287-294.
- Fukunaga, K. (1990). *Statistical Pattern Recognition*, (2nd ed.). Academic Press, NY.
- Goodman, D., Boggess, L. & Watkins A. (2002). Artificial Immune System Classification of Multiple Class Problems. *Artificial Neural Networks In Engineering, (ANNIE)*.
- Guijarro-berdi, B., Fontenla-romero, O., Perez-Sanchez, B. & Fraguera, A. (2007). A Linear Learning Method for Multilayer Perceptrons Using Least-Squares, *Lecture Notes in Computer Science*, 4881, 365-374.

- Gutierrez-Osuna, R. (n.d.) *validation, L(13)*, Retrieved 16/01/2010, from http://courses.cs.tamu.edu/rgutier/ceg499_s02/l13.pdf.
- Güneşer, C. (2009). *Classification of Wisconsin Breast Cancer Database*. Master's thesis, Dept. of Elektrik-Elektronik Engineering, Univ. of Dokuz Eylül.
- Hamilton, H. J., Shan, N. & Cercone, N. (1996). RIAC: A Rule Induction Algorithm Based on Approximate Classification, *Technical Report, University of Regina*.
- Hassanien, A. E. (2004). Rough set approach for attribute reduction and rule generation: A case of patients with suspected breast cancer. *Journal of the American Society for Information Science and Technology*, 55 (11), 954–962.
- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*, (2nd ed.). New Jersey, U.S.A: Prentice Hall.
- Jain, A.K., Duin, R.P.W. & Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22 (1), 4-37.
- Karabatak, M. & Ince, M. C. (2009). An expert system for detection of breast cancer based on association rules and neural network. *Expert Systems with Applications*, 36(2), 3465-3469.
- Kohonen, T. (1987). *Self-Organization and Associative Memory*, (2nd ed.). Springer-Verlag.
- Kolahdouzan, M. & Shahabi, C.(2004).Voronoi-Based K Nearest Neighbor Search for Spatial Network Databases, *Proceedings of the 30th Very Large Data Bases (VLDB) Conference, Toronto, Canada*.
- Kinto ,E. A., Hernandez, M., Marcano-Cedeño, A. & Pelaez, J. R.(2007).A Preliminary Neural Model for Movement Direction Recognition Based on

- Biologically Plausible Plasticity Rules, *Lecture Notes in Computer Science* , 4528, 628-636.
- Maglogiannis, I., Zafiroopoulos, E. & Anagnostopoulos, I.(2007). An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. *Applied Intelligence*, 30 (1), 24-36.
- Mangasarian, O. L. & Wolberg, W. H.(1990). *Cancer diagnosis via linear programming*, 23 (5), 1 – 18, SIAM News.
- Marcano-Cedeño, A., Quintanilla-Domínguez J. & Andina, D. (2011).WBCD breast cancer database classification applying artificial metaplasticity neural network.*Expert Systems with Applications*, 38(8), 9573-9579.
- Nauck, D. & Kruse, R. (1999).Obtaining interpretable fuzzy classification rules from medical data. *Artificial intelligence in medicine*, 16(2), 149-69.
- Ozmen, V. (2008). Breast cancer in the world and Turkey. *Meme Sağlığı Dergisi*, 4, 7-12.
- Papoulis, A. (2002) *Probability, Random Variables, and Stochastic Processes*, (4th ed.).New York: McGraw- Hill.
- Peña-Reyes, C.A. & Sipper, M. (1999).A fuzzy-genetic approach to breast cancer diagnosis. *Artificial intelligence in medicine*, 17 (2), 131-155.
- Poggio,T& Girosi, F.(1990).Networks for approximation and learning, *Proc. IEEE* 78(9), 1484-1487.
- Polat, K. & Güneş, S., Breast cancer diagnosis using least square support vector machine. (2007). *Digital Signal Processing*, 17(4), 694-701.

- Quinlan, J. R. (2006). Improved Use of Continuous Attributes in C4. 5. *Journal of Artificial Intelligence Research*, 4, 77-909.
- Sahan, S., Polat, K., Kodaz, H. & Güneş S. (2007). A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis. *Computers in biology and medicine*, 37(3), 415-23.
- Sewak, M., Vaidya, P., Chan, C.C. & Duan, Z. H. (2007). SVM Approach to Breast Cancer Classification. *Second International Multi-Symposiums on Computer and Computational Sciences (IMSCCS)*, 2, 32-37.
- Ster, B. & Dobnikar, A. (1996). Neural networks in medical diagnosis: comparison with other methods, *International Conference on Engineering Applications of Neural Networks (EANN)*, London, United Kingdom, 427–430.
- Theodoridis, S. & Koutroumbas, K. (1999). *Pattern Recognition*. Academic Press.
- Ubeyli, E.D. (2007). Implementing automated diagnostic systems for breast cancer detection. *Expert Systems with Applications*, 33(4), 1054-1062.
- UCI, (n.d.). *UCI Machine Learning Repository*, Retrieved 15/01/2010, from <http://archive.ics.uci.edu/ml/>.
- Vapnik, V. N. (1999). An Overview of Statistical Learning Theory, *IEEE Trans. On Neural Networks*, 10(5), 988-999.