THESIS APPROVAL

GRADUATE SCHOOL, KASETSART UNIVERSITY

Doctor of Engineering (Computer Engineering)

DEGREE

Computer Engineering	Computer Engineering
FIELD	**DEPARTMENT**

TITLE: Feature Selection and Dimension Reduction for Medical Image Analysis

NAME: Mrs. Rapepun Piriyakul

THIS THESIS HAS BEEN ACCEPTED BY

_____ **THESIS ADVISOR**

(Associate Professor Punpiti Piamsa-nga , D.Sc.)

_____ **THESIS CO-ADVISOR**

(Assistant Professor Yodyium Tipsuwan, Ph.D.)

_____ **THESIS CO-ADVISOR**

(Assistant Professor Arnon Rungsawang, Docteur de I''ENST)

_____ **DEPARTMENT HEAD**

(Assistant Professor Kemathat Vibhatavanij, Ph.D.)

APPROVED BY THE GRADUATE SCHOOL ON _____

_____ **DEAN**

(Associate Professor Gunjana Theeragool, D.Agr.)

THESIS

FEATURE SELECTION AND DIMENSION REDUCTION FOR MEDICAL IMAGE ANALYSIS

RAPEPUN PIRIYAKUL

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Engineering (Computer Engineering)
Graduate School, Kasetsart University
2008

Rapepun Piriyakul 2008: Feature Selection and Dimension Reduction for Medical Image Analysis. Doctor of Engineering (Computer Engineering), Major Field: Computer Engineering, Department of Computer Engineering. Thesis Advisor: Associate Professor Punpiti Piamsa-nga, D.Sc. 77 pages.

Feature Selection has a crucial role in Medical Image Analysis due to a great deal of image features, higher derivatives of mathematical computation and more of time complexity in detection process. Therefore, only some features that can convey enough information about the image should be concerned.

We construct three novel features selection techniques using collaboration with statistical methods. The experiments conducted on a data set of 113 ROIs from Database of the Mammographic Image Analysis Society (MIAS) of UK .First, we used factor analysis and logistic regression classifier to reduce features with compatibly equal accuracy. Second, feature selection was performed in graph based analysis using analogy via path analysis and Bayes inference. The experimental result shows that our selected 13 features performed as well as original 50 features. Third, are pruned by ANOVA, its quality is similar to SFS method but computation is less required.

_____ _____ __ / __ / __

Student's signature Thesis Advisor's signature

ACKNOWLEDGEMENT

First of all, I would like to gratefully thank to my advisor, Associate Professor Dr. Punpiti Piamsa-nga, for providing the basis and guidance that has enabled this work to emerge. He introduced me to the field of multimedia processing, advised me knowledge in research methodology, and worked hard in reviewing on my journal article. Also, I would like to present deeply thank to Dr. James Edward Brucker of Department of Computer Engineering for language improvement and comments on writing in these research articles.

I am also grateful to the other faculties of Department of Computer Engineering —Assistant Professor Dr. Yodyium Tipsuwan, Assistant Professor Dr. Arnon Rungsawang, Associate Professor Dr. Asanee Kawtrakul, Associate Professor Dr. Kitsana Waiyamai, Assistant Professor Dr. Pirawat Watanapogse, Dr. Thitiwan Srinark, Assistant Professor Dr. Jittat Fakcharoenphol, Assistant Professor Dr. Charay Lerdsudwichai and Associate Professor Pradondet Nilagupta who provided useful comments and gave me the opportunity on studying in the DCPE .

I would like to present deeply thanks to Dr. Nutakran Somsanit, Md., for her good advice, Mammographic Image Analysis Society (MIAS) of UK and the Digital Database for Screening Mammography (DDSM), university of South Florida for the beneficial resource data. Moreover, this research was formally supported by the Kasetsart University Research and Development Institute and the Graduate School of Kasetsart University.

I am heartfully thankful to all my young friends of the MAD Laboratory, which provided a home with a friendly environment.

Finally, most of all, my appreciations devote to my family for their love and encouragements on my success.

Rapepun Piriyakul

October, 2008

TABLE OF CONTENT

Page

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF FIGURES

FEATURE SELECTION AND DIMENSION REDUCTION FOR MEDICAL IMAGE ANALYSIS

INTRODUCTION

Medical Image Analysis is a kind of CAD.CAD is defined as a diagnosis made by physicians who takes into account the computer the diagnostic accuracy, and the consistency of the radiologists' image interpretation (Doi *et al.*,1999). Since the causes of some types of cancer are still unknown, it is difficult to detect whether a tissue is cancer or not. Then, medical image detection is one of promising cancer control strategies for cancer treatment. Recently radiologists can refer to an automated system as their second opinion, as it is often difficult to distinguish malignant from normal healthy tissues. An automated system can detect and diagnose malignant in medical images to suspicious regions for further evaluation. Since various medical images for CAD such as X-ray, CT scan, MRI, and mammogram, consist of considerable amount of image features, easing the detection of the malignant is the issue of interest. According to the highly increase rate of medical image and a large amount of features, the improvement of efficient screening process is challenging for CAD. The system consists of four parts: preprocess, image transformation, feature extraction, feature selection and detection. All four parts are equally important and are significantly determined the system accuracy.

Due to lots of features and newly modified traditional features, there has a serious effect upon the time complexity. From previous studies we found that adding non-significant features increase not only cost of processing but the feature space as well. The larger feature space, the large barrier is, for constructing an efficient kernel or transformation to the detection step. Therefore to explore an efficient management for feature selection is the crucial point for CAD. Based on the desires on minimizing time complexity while still maintain the rate of precision, to construct an feature selection is our research objective.

Research procedure contained two feature management strategies. The first strategy was the dimension reduction while maintain all relevant information from them. The second strategy was the technique filtering selection. The techniques are all used feature selection; only outstanding important features from mammogram are chosen for breast cancer detection.

The research is conducted by using mammograms for experimenting on CAD in addition to the rapid growth rate of breast cancer and now a day there is no preventive way of breast cancer.

Breast cancer is among the most frequent forms of cancers found in women (Lung Cancer, 2007. http://ww.lungcancer.org/patients/fspc_lc_101htm). There are many ways to diagnose the breast cancer such as biopsy, ultrasound and imaging breasts (Almato *et al*.,2005) Ultrasound cans diagnosis simple cysts in the breast with an accuracy of 96-100% (Song *et al*., 2005) ; however, when use for unequivocal differentiation between solid benign and malignant masses was proven to be very difficult. Despite the considerable efforts toward Ultrasound, improving imaging techniques is still necessary. Mammography is now commonly used in combination with computer-aided diagnosis (CAD).

Image features are conceptual descriptions of images which are needed in image processing for analyzing image contents or meanings. Features are usually represented as data structures of directly extractable information from images, such as colors, and grays, and higher derivatives from mathematical computation of the basic information such as its edges, histograms, and Fourier descriptors. There must be a specific algorithm to process each type of features. Therefore, only features that carry enough information about the image should be concerned. Moreover, the feature-extraction technique should be practical and easy to be computed. Many researchers tried to improve their techniques by introducing more new features on the assumption that the better analysis would be ascertained when sufficient information is reached. However, the more features to be used, it definitely introduces the higher costs and

the longer computation efforts both in collecting them and utilizing them, in prediction purposes.

The addition of more features sometimes does not improve the system efficiently since a newly added feature may not contribute more accurate results. There are two different approaches to improve quality of feature extraction process; one is to extract more new features and the other is to investigate the feature pruning techniques (Cosit *et al*., 1997; Chiou and Hwang, 1993; Foggia *et al*.2006; Widodo and Yang, 2007; Majcenic and Loncaric, 2007)

It has been proven successively that using more features did not leave significant improvement of classifier performance. Foggia *et al*.,2006 used graph based method with only 6 features and found the performance was 82.83% true positive rate (TP) and 0.08% false positive per image rate (FP), where as Fu *et al*.,2005, used sequential forward search (SFS) and found that only 25 features with Mean Square Error (MSE) 0.02994 by using General Regression Neural Networks (GRNN). When support vector machine (SVM) was applied, it has shown that there were 11 features left with MSE 0.0283. There are many ways to discard non-significant features such as sequential forward search (SFS), sequential backward search (SBF) and stepwise regression. SFS and SBF are focusing on the reduction of MSE at the detection process while stepwise regression concerns both the interaction of features and the MSE value. Using stepwise logistic regression is costly since this technique is based on the number of experiments over all possible permutations of every feature in the prediction model.

Exploration of feature extraction analysis suggest four crucial problems

1. The multicollinearity problem might exist if high correlated features are used.

2. The cost of image processing may be high if more features are used.

3. The predicting precision is not depend on number of features.

4. How a feature extraction can be constructed in order that speed the classifiers.

To tackle these problems, we proposed two strategies:

1. Keeping all extracted features, by using an efficient management to reduce feature space and constructing a powerful kernel trick to enhance the performance of detection .

2. Improving and exploring feature selection techniques to discard non-significant features.

To support our objective on two strategies, we propose statistical techniques: factor analysis for reducing the feature space, correlation analysis, Bayes inference path analysis and analysis co-variance for pruning less important features.

Due to enormous medical images and lack of expert, to improve CAD is a challenging in this decade. Our work is a sub crucial part to support this objective. Since feature extraction does not warrant a great deal to detection process, adding an efficient feature selection layer to CAD system provide the accuracy and reduce time complexity.

We used data from Database of the Mammographic Image Analysis Society (MIAS) in our experiment.

Contributions

The contribution of research consists of:

1. A technique for Computation Reduction using Factor Analysis

2. A feature selection technique Using Graph Based Analysis

3. A feature selection under Feature Interaction Constraint (using PCA and ANOVA)

4. To construct the feature selection basis for other in future work.

Thesis Organization

The further document is detailed as follows:

- OBJECTIVES describe about background and the objectives of the research.

- LITERATURE REVIEW presents and discusses medical image feature in previous work, the problems and the feature domains in medical image.

- MATERIALS AND METHODS explains about

1. Data
2. Computation Reduction Method using Factor Analysis
3. Feature Reduction Method

   3.1 Graph Based Analysis
   3.2 PCA and ANOVA

- RESULTS AND DISCUSSION

1. Experiment on Factor Analysis
2. Experiment on Graph Based Analysis
3. Experiment on PCA and ANOVA

- CONCLUSION AND RECOMMENDATION provides conclusions and future work about the deep analysis in some interesting topics.

OBJECTIVES

To find methods to select the most appropriate features for Medical Image Analysis in order to maintain the accuracy and reduce computation requirement.

LITERATURE REVIEW

Introduction to Feature Selction and Dimension Reduction for Medical Image Analysis and Related Terms

This section details about medical image feature, feature extraction and feature selection techniques used in previous research. A summarized of some previous studies with different classifiers and its performance has detailed.

1. Medical Image Analysis

The diagnosis of cancerous and pre-cancer malignancies from medical images is significantly improved by the use of computer-aided diagnosis (Zhao *et al*.,2005, Lyin *et al*.,2005), which can be defined as a diagnosis made by a physician taking into account the computer output as a second opinion, to improve the diagnostic accuracy and the consistency of the medical image interpretation.

The incidence and mortality attributed to cancer has been rising steadily, mainly due to environmental factors. Lung cancer has been the leading cause of cancer deaths among men since the early 1950's; in 2005, it surpassed breast cancer to become the leading cause of cancer deaths among women in the United States and other countries (American Cancer Society, Breast Cancer Facts&Figures 2005-2006, http://www.lungcancer.org/patients/fs_pc_lc_101.htm, 2007). Lung cancer causes more deaths than the next three most common cancers combined (colon, breast and prostate). Cancer is difficult to detect in early stages due to the small size of the malignancy or non-palpable microcalcification clusters. To overcome this problem, investigators have developed computer-aided diagnosis (CAD) schemes to identify regions of potential microcalcification clusters in region of interest tissue. CAD is defined as a diagnosis made by a physician who takes into account the computer output as a second opinion, to improve the diagnostic accuracy and the consistency of the radiologists' image interpretation. Currently, medical image analysis is one of the most promising cancer control strategies since the causes of some types of cancer are

still unknown. Radiologists can refer to an automated system as their second opinion, as it is often difficult to distinguish malignant from normal healthy tissues. An automated system can detect and diagnose malignancy in medical images as suspicious regions for further evaluation. There are various medical images for CAD such as X-ray, CT scan, MRI, and mammogram. The image information consists of image features; therefore, detecting malignant tissue depends on quality of the image features (Liang, *et al*., 1992).

2. Image Features

Image features are conceptual descriptions of images that are useful for analyzing image contents or meanings (Cosit and Loncaric, 1997). Features are usually represented as data structures of directly extractable information such as color, gray level, and higher derivatives from mathematical computation of basic information such as edges, histograms, or Fourier descriptors. Consequently, a specific algorithm to process each feature must be developed. Nevertheless, many researchers tried to improve existing techniques by introducing new features based on the assumption that more information would lead to the better analysis. Incorporating more features generally increases cost and requires longer computing times both in collecting and utilizing them in image analysis.

3. Medical Image Feature Survey and Related Works

Medical image detection from mammogram is limited to analysis of the gray-scale level. Identification, between image density of normal tissue and malignant, was nearly impossible because the difference between them is really small (Foggia *et al*., 2006). Thus, most feature extraction methods are extended from the derivation of limited gray scale information (Hiroyuki *et al*., 2003; Cosit and Loncaric, 1997; Guler and Ubeyli, 2007; Zhao *et al*., 2005; Zoran and Soren, 2007). Medical image features can be divided into three domains, spatial, texture, and spectral. Spatial domain refers to the gray-level information in an arbitrary window size. It includes gray levels, background and foreground information, shape features, and other

statistics derived from image information intensity. Texture refers to properties that represent the surface or structure of an object in reflective and transmissive images. Texture analysis is important in many applications of computer image analysis for classification, detection or segmentation of images based on local spatial variations of intensity. Spectral density or spectrum of signal is a positive real value function of a frequency associated with a stationary stochastic process, which has dimensions of power or energy. However, all useful features must be represented in a computable form.

From previous study (Shiraishi *et al*., 2006) we found that most of them were extracted on the assumption that more features would be enhancing the detection system. There are many ways to extract new features such as modify old features, use more knowledge from syntactic image (Pietikainen *et al*., 2004) and sometime use knowledge base (Miller *et a*l., 2007). Adding more features does not promise to improve the performance but it will be increase time consuming and analyzing complexity in detection step. As a result of enormous number of features and feature dependency, the feature selection process is a crucial task that the CAD problem is tackled immediately.

Since quality, not quantity, of feature is the key to be succeeded; to find appropriate number of good features is a necessary condition for CAD. Theoretical speaking, to investigate optimal features is not confront with the extraction problem but also the selection problem. However, from the past through present, many researchers in the field of medical analysis have put lots of efforts in finding only best feature or best combination of features that gives highest classification rate using appropriate classifier. There are many perspectives on the situation of feature extraction and selection listed as follows.

Fu *et al. (*2005) used 61 features to select a best subset of features that produced optimal identification of microcalcification using sequential forward search (SFS) and sequential backward search (SBS) reduction followed by a General Regression Neural Network (GRNN) and Support Vector Machine (SVM). Due to

their feature selection method we found the high inconsistency between two methods i.e.: one feature which was the top-five significant on the SFS but it was discarded on the SBS.

Zhang *et al.* (2004) attempted to develop feature selection based on the neural-genetic algorithm. Each individual in the population represents a candidate solution to the feature subset selection problem. With 14 features on their experiment, there are 2^{14} possible feature subsets. The results showed that a few feature subsets (5 features) have achieved the highest classification rate 85%. In the situation of a huge number of features and mammography, it is very costly to select features using neural-genetic approach.

Information Retrieval in Medical (IRMA) (Lehmann *et al*., 2006) project used global, local, and structure features in their studies in lung cancer by using global and local feature. The global feature consists of the anatomy of object, a local feature is based on local pixel segment, and the structural feature cooperates with medical apriori knowledge on a higher level of semantics. In addition to the constraints of the global feature construction and lack of prior medical semantic knowledge, this procedure was quite difficult and costly.

For researchers' choices on medical image features, it depends on the objectives of each research topics. Cosit *et al.* (1997), Chiou and Hwang (Chiou and Hwang, 1995) and Zoran (Zoran and Soren, 2007) used simple statistical features on gray scale intensity, while Samuel *et al*. (Samuel *et a*l., 2005) used volume, sphericity, mean of gray level, standard deviation of gray level, gray level threshold, radian of mass sphere, maximum eccentricity, maximum circularity, and maximum compactness in their CAD system. Miller *et al.* (2007) used average gray scale, standard deviation, skewness, kurtosis, maximum and minimum of gray scale, and gray level histogram to identify and detect lung cancer. Shiraishi *et al.* (2006) studied 150 images from Japanese Society of Radiological Technology (JSRT) database by using patient age, RMS of power spectrum, background image, degree of irregularity, full width at half maximum for inside of segment region. Dodd *et al.* (2004) studied

on personal profile, region of interest properties, nodule size, and shape. Ping *et al*. (2004) extended the new modified features, number of pixel in ROI, average gray level, energy, modified energy, entropy, modified entropy, standard deviation, modified standard deviation, skewness, modified skewness, contrast, average boundary gray level. On further investigation on using more features besides the medical image analysis, the experiment of Windodo *et al.* (2007) for fault diagnosis of induction motors to improve the feature extraction process by new proposing kernel trick. On his study, 76 features were calculated from 10 categories of time domain. These categories are mean, RMS, shape factor, skewness, kurtosis, crest factor, entropy error, entropy estimation, histogram lower and histogram upper. We cannot find their common way to select features; however, we can conclude that they tried to add more features in order to increase the efficiency of their methods. According to abundant features and medical image data, the main problem in image processing on CAD is high cost of processing.

Based on each research, there are varieties of feature extraction and classification used in Medical Image Analysis. Table 1 shows the list of features and classifer techniques from previous studies.

Table 1 Features selection and classification method from previous work.

Researcher	Domain	Features used (examples)	Classifier
Fu *et al.* (2005)	Texture	Co-occurrence matrix rotation with angle 0°, 45°, 90°, 135°: Difference entropy, Entropy, difference variance, contrast, angular second moment, correlation	GRNN (SFS, SBS)
	Spatial	Mean, area, standard deviation, foreground/ background ratio, area, shape moment intensity variance, energy –variance	
	Spectral	Block activity, Spectral entropy	
Samuel, G. *et al.* (2005)	Spatial	volume, sphericity, mean gray level, gray level standard deviation, gray level threshold, radius of sphere, maximum eccentricity, maximum circularity, maximum compactness	Rule-based, linear discriminant analysis
Dodd. L.E. *et al.* (2004)	Spatial, Patient Profile	Patient profile, nodule size, shape (measured with ordinal scale)	Regression analysis
Shiraishi *et al.* (2006)	Multi Domain	Patient profile, root-mean-square of power spectrum, histograms frequency, full width at half maximum of the histogram for the outside region of the segmented nodule on the background–corrected image, degree of irregularity, full width at half maximum for inside region of segmented nodule on the original image	Linear discriminant analysis

Table 1 (Continued)

Researcher	Domain	Features used (examples)	Classifier
Miller (2007)	Spatial	Average gray level, standard deviation, skew, kurtosis, min-max of the gray Level, gray level histogram	SVM
Zhao et al.(2005)	Spatial	number of pixels, histogram, average gray, boundary gray, contrast, difference, energy, modified energy, entropy, standard deviation, modified standard deviation, skewness, modified skewness	ANN
Ping *et al*. (2004)	Spatial	Number of pixels, average, average gray level, average histogram, energy, modified energy, entropy, modified entropy, standard deviation, modified standard deviation, skew, modified skew, difference, contrast, average boundary gray level	ANN and Statistical classifier
Songyang and Ling, (2000)	Mixed features	Mean, standard deviation, edge, background, foreground-background ratio, foreground-background difference, difference ratio of intensity, compactness, elongation, Shape Moment I-IV, Invariant Moment I-IV, Contrast, area, shape, entropy, angular second moment, inverse different moment, Correlation, Variance, Sum average	Multi-layer Neural Network

4. Feature Domian

This Section has detailed on feature domains that are usually used for medical image classification. Generally, original digital medical image is in the form of gray-scale or multiple spectrum bitmap. Elements of bitmap are integer values corresponding to properties (i.e. brightness, color) of the corresponding pixel of the sampling grid. Image information in the bitmap is accessible through the coordinates of a pixel that corresponds with row and column indices. All features that can be extracted directly using mathematical or statistical models are recognized as low-level features. High-level features are usually summarized from low-level features by machine-learning models. Most research in medical image analysis usually has to deal with low-level features in order to identify high-level features. In this research, we investigate many types of low-level features in order to identify mammograms whether they are benign or malignancy. Our detection results can be valuable high-level features in other applications. In case of low-level features, it can be divided into three types: spatial (or frequency), texture, and spectral domains.

First, spatial domain composes of features that are extracted and summarized directly from grid information. It implicitly contains spatial relations among semantically important parts of the image. Examples of spatial features that has been used for indentify suspected-area of calcification are shapes, edges, foreground information, background information, contrasts and set of intensity statistics, such as mean, median, standard deviation, coefficient of variation, variance, skewness, kurtosis, entropy, and modified moment. In this research, we added radian of mass.

Second, texture features are features of relation among pixels in bitmap. Representation of texture features usually use co-occurrence matrices to describe their properties. The co-occurrence matrix of texture description is based on the repeated occurrence of some pixel configurations in the texture. Suppose that the probability of intensity value of each pixel depends only on a certain spatial relation r between a pixel of brightness z and a pixel of brightness y; then information about the relation r is recorded in the square co-occurrence matrix C_r whose dimension corresponds to

the number of brightness levels of the image. The following algorithm calculates the co-occurrence matrix C_r, from the image $f(i, j)$: 1) assign $C_r(z, y) = 0$ for all $z, y \in [0, L]$, where L is the maximum brightness; and 2) for all pixels (i_1, j_1) in the image, determine (i_2, j_2) which has the relation r with the pixel (i_1, j_1) and perform $C_r[f(i_i, j_1), f(i_2, j_2)] = C_r[f(i_i, j_1), f(i_2, j_2)] + 1$. Co-occurrence matrix of image used $P_{\phi,d}(a,b)$ describes the frequency of two pixels with gray-levels, a and b, that appear in the window separated by a distance d in direction ϕ. In this research, we set ϕ as 0˚, 45˚, 90˚ and 135˚.

The frequencies of co-occurrence as functions of angle and distance can be defined as

$P_{0^\circ, d}(a, b) = | \{ [(k, l), (m, n)] \in D : k-m = 0, | l-n | = d, f(k, l) = a, f(m, n) = b\} |$

$P_{45^\circ, d}(a, b) = | \{ [(k, l), (m, n)] \in D : (k-m = d), l-n = -d \text{ OR } (k-m = -d, l-n = d), \quad f(k, l) = a, f(m, n) = b\} |$

$P_{90^\circ, d}(a, b) = | \{ [(k, l), (m, n)] \in D : | k-m | = d, l-n = 0, f(k, l) = a, f(m, n) = b\} |$

$P_{135^\circ, d}(a, b) = | \{ [(k, l), (m, n)] \in D : (k-m = d, l-n = d) \text{ OR } (k-m = -d, l-n = -d), f(k, l) = a, f(m, n) = b\} |$

where $|\{\ldots\}|$ refers to set cardinality and D = $(M \times N) \times (M \times N)$

Examples of features in texture domain are listed as follows:

Energy or angular second moment (an image homogeneity measure): $\sum_{a,b} P^2{}_{\phi,d}(a,b)$

Entropy: $\sum_{a,b} P_{\phi,d}(a,b) \log_2 P_{\phi,d}(a,b)$

Maximum probability: $\max_{a,b}\{P_{\phi,d}(a,b)\}$

Contrast: $\sum_{a,b} | a-b |^k P_{\phi,d}(a,b)$

Inverse difference moment: $\sum_{a,b,a\neq b} \frac{P^\lambda{}_{\phi,d}}{| a-b |^k}$

Correlation (a measure of image linearity, linear direction structures in direction ϕ

$\sum_{a,b,a\neq b} \frac{[(ab)P_{\phi,d}(a,b)] - \mu_x\mu_y}{\sigma_x\sigma_y}$ where $\mu_x, \mu_y, \sigma_x, \sigma_y$ are means and standard deviations.

Third, spectral features (Gliman and Sizzanme, 2004; Thyagarajan, 2006; Zhaeng *et al*., 1996) are used to describe the frequency characteristics of the input image. The features are based on transformation from spatial domain into frequency domain. Basically, most frequently-used spectral features are based on discrete cosine transform (DCT) (Sorwar and Abraaham, 2006) and Wavelet. Examples of features based on frequency domain are listed as follows:

Spectral entropy: $-\sum_i \sum_j \overline{X}(i,j) h(\overline{X}(i,j))$

Block activity: $A = -\sum_i \sum_j |\overline{X}(i,j)|$ where i, j are windows size and $\overline{X}(i,j) = \frac{|\overline{X}(i,j)|}{A}$

Feature Selection Problems

Explorations of feature extraction analysis suggest four crucial points:

1) The multicollinearity problem might exist if high correlated features are used.
2) The cost of image processing may be high if more features are used
3) The predicting precision is not depended on number of features.

To tackle these problems, we formulate statistical techniques for our three feature selection issues:

1. "The Cost Reduction in Medical Image Analysis Using Factor Analysis" issue is using factor analysis for reducing dimension in detection step. Our goal is to construct an extraction process which satisfies two conditions, keep all features with cost reduction and general practice for Medical Image Analysis. We use the strategy

of keeping all features because of discarding will be dropped some important features. The contribution of the technique has impacted through the improvement of constructing kernel function in prediction step.

2. “Feature Reduction in Graph Based Analysis” issue is constructed on the casuse and effect based. Exploration of features extraction analysis have been founded that there are two effects, *direct and indirect effect* (feature interaction). From this pin point, we use path analysis concept to discard non- direct and non-indirect features.

3. “Feature Selection under Feature Interaction Constraints” issue is constructed on the assumptions of “*if two objects are different then their features are different and if two objects are not different then their features are not different”* by this way to discard any feature must realize this limitation.

Our three propose techniques: the first and the second technique are the classifer independent while as the third is a classifier dependent.

MATERIALS AND METHODS

Materials

1. Computer System

    1.1 Hardware High Performance PC 1 System

    1.2 Software

    - Mat lab
    - C++ Borland
    - VBA
    - STATISTICA
    - Weka

2. Data

    2.1 Determination of Sample Size

Due to a large population, the sample size to test our experiment has calculated from the statistics sample size equation: $n = \frac{Z^2 p * q}{E^2}$ where Z is a confident interval, E is an error, p is probability or chance of a woman to get a disease and $q = 1 - p$ is the probability or chance of a woman who does not get a disease. We use confidence interval at level 95%, then Z = 1.96 and let absolute error E equal to 0.0002. Based on the breast cancer statistics in Thailand (Breast Cancer, Faculty of Medicine, Prince of Songkla University 2001), the incidence is 16.3/100,000 population. Finally, the number of sample size is calculating as the following:

p = 16.3/100000 = 0.000163, q =0.999837

Then $$n = \frac{1.96^2 * 0.000163 * 0.999837}{0.0023^2} = 113 \text{ unit}$$

The experiment is conducted by using mammograms from two MIAS databases which region of interest (ROI) were segment by radiologists .The databases list as below.

Database of the Mammographic Image Analysis Society (MIAS) of UK (113 ROIs)

The sample data from databases are shown as Figures 1, 2, 3 and 4.

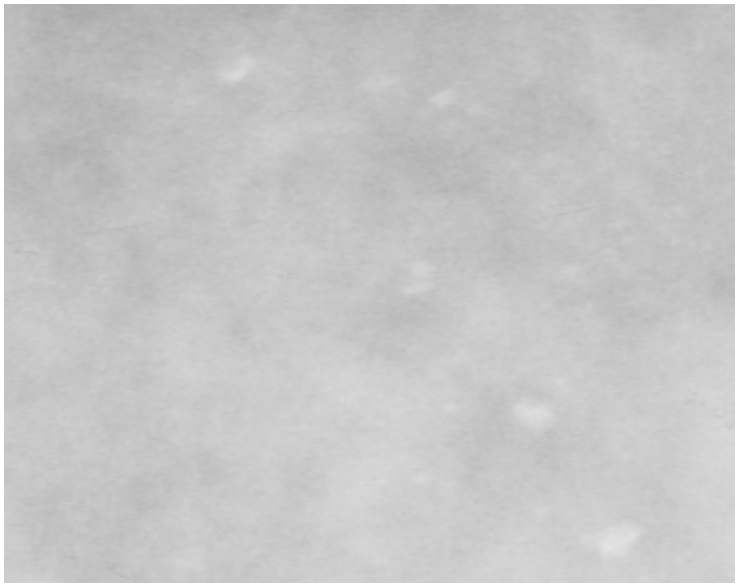

Figure 1 Malignant Mammogram

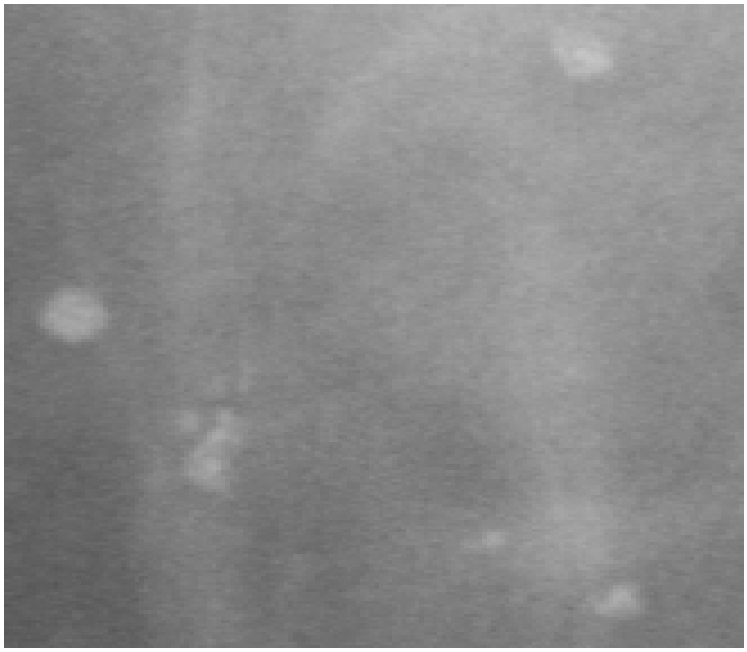

Figure 2 Malignant Mammogram

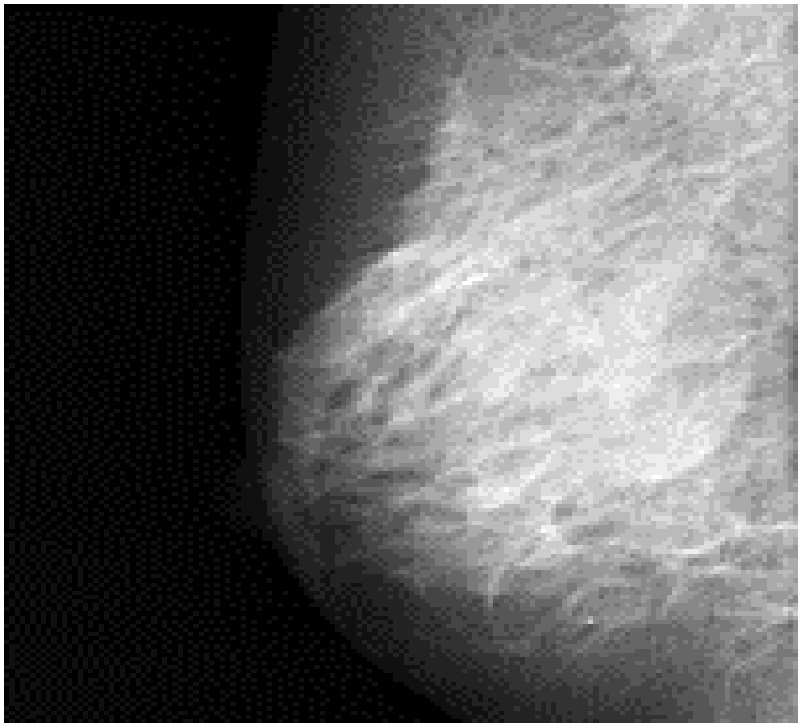

Figure 3 Benign Mammogram

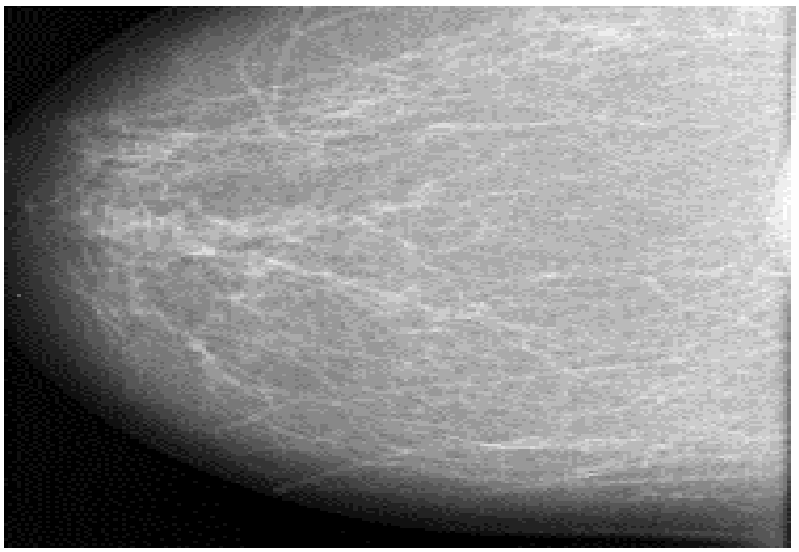

Figure 4 Normal Mammogram

Methods

1. Experiment

The study consists of five phases; the sequence of each process is stated as below:

Phase 1 : Collecting the ROIs from databases (MIST). The first set of ROIs was extracted from MIST

Phase 2 : ROIs are transformed to numeric value using Mat lab software.

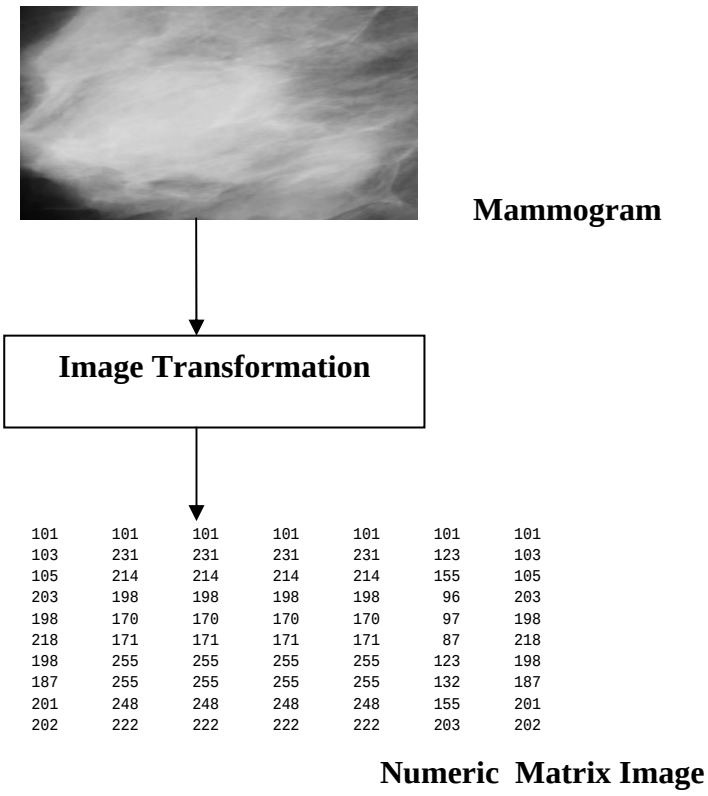

Figure 5 Medical Image Transformation.

Figure 5 shows the process to transform mammogram to numeric value for further extraction process.

Phase 3: Extracting 50 features from (Phase 2) four domains, spatial, texture, frequency and spectral domain as shown in Figure 6.

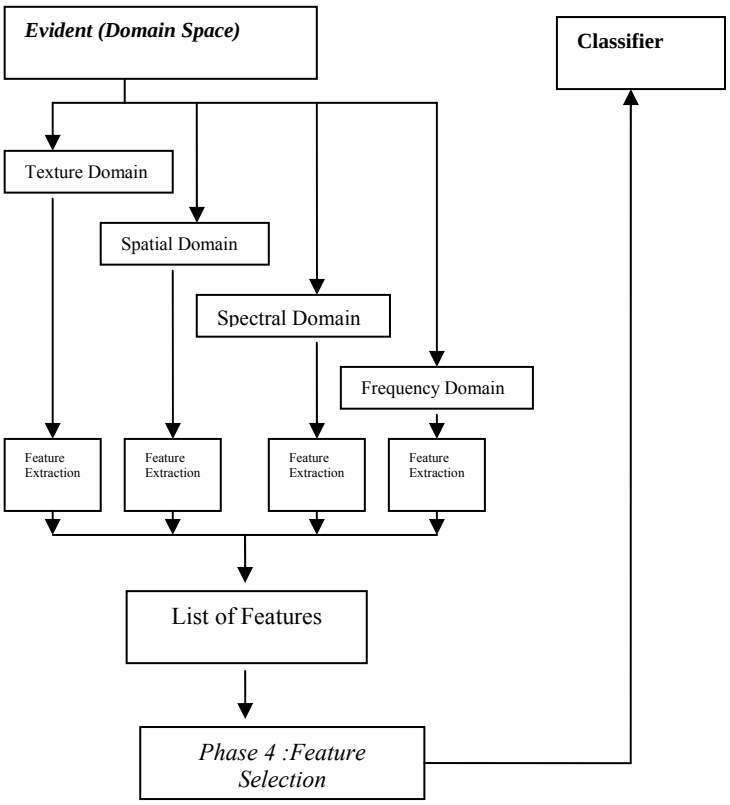

Figure 6 Feature domains.

Phase 4 Experiment on three proposed feature selection techniques:

4.1 Conducting our first feature selection, *"The Cost Reduction in Medical Image Analysis Using Factor Analysis"*, by using factor analysis for feature reduction and then evaluates the result with logistic regression.

4.2 Conducting our third feature selection, *"Feature Reduction in Graph Based Analysis"*, by using correlation analysis, multiple regression and Bayes inference and then evaluates the result with two classifiers, logistic regression and Artificial Neural Network, to check the consistency of this proposed method.

4.3 Conducting our second feature selection, *"Feature Selection under Feature Interaction Constraint"*, by using Principal Component Analysis and Analysis of Variance and then evaluate the result with two classifiers, Binary Logistic Regression and Linear Discriminant Analysis.

Phase 5: Evaluation of three proposed methods.

Our work is proposed three methods Features Reduction using Factor Analysis, Feature Selection in Graph Analysis and Feature Selection using PCA and ANOVA. Figure 7 is our experiment process.

Figure 7: Frame work of sub tasks and three proposed feature selection techniques

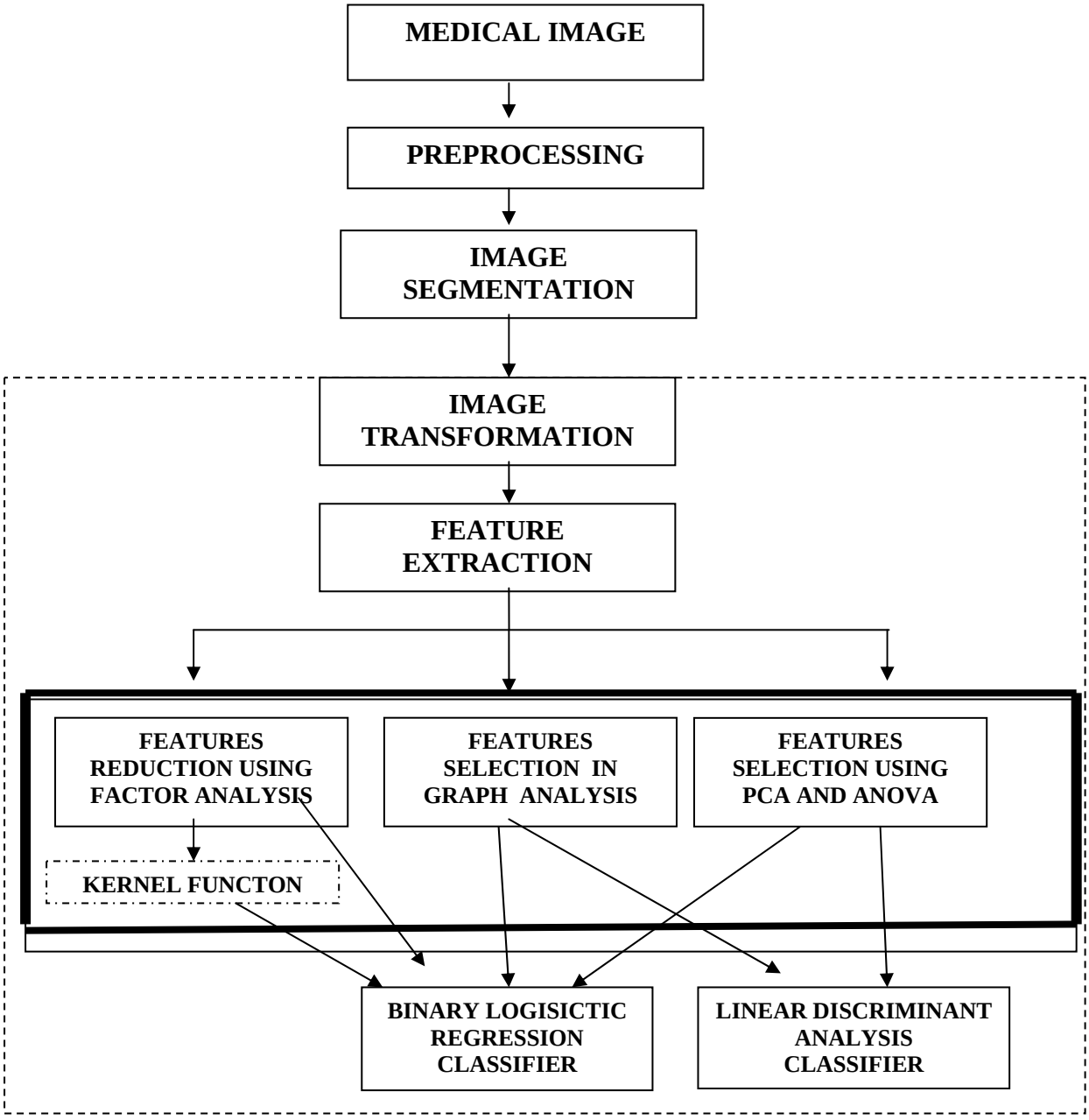

Figure 7 A framework of three feature selection techniques.

In the feature extraction process, the lists of features are shown as Table 2.

Table 2 List of 50 features.

List of Features from 4 Domains (spatial, texture frequency and spectral)
Entropy rotation from 0˚,45˚,90˚,135˚
Energy rotation from 0˚,45˚,90˚,135˚
Inverse difference Moment rotation from 0˚,45˚,90˚,135˚
Mean Co-occurrence rotation from 0˚,45˚,90˚,135
Max Co-occurrence rotation from 0˚,45˚,90˚,135
Contrast rotation from 0˚,45˚,90˚,135˚
Homogeneity rotation from 0˚,45˚,90˚,135˚
Standard deviation on X rotation from 0˚,45˚,90˚,135˚
Standard deviation on Y rotation from 0˚,45˚,90˚,135˚
Modified entropy rotation from 0˚,45˚,90˚,135˚
mean, maximum, median, standard deviation (SD), coefficient of variation (CV), skewness, kurtosis (intensity of gray level)
block activity, spectral entropy, mass radian

Cross-validation

In order to test the stability of feature classification across two classifiers, i.e., Binary Logistic Regression (BLR) and Linear Discriminant Analysis (LDA), we designed the experiments as:

Divide 113 ROIs selected by each 3 different selection algorithm; Factor Analysis, Graph Based Mechanism and Analysis of Variance; into 10 equal subgroups. Therefore we have 60 subgroups available for subsequent test.

In each 10 subgroups, we applied BLR and LDA to them by 10 fold cross-validation mechanism, i.e., one subgroup stands for training group while other subgroups were testing set and average error were obtained.

The data splitting reveals that two classifiers are sufficiently good for further use since there absolute error and other kind of indicator are small. Cross-validation results in more detail are shown in Table3.

Table 3 The 10 folds cross-validation experiment on 50 features, 13 features and 5 factors with Binary Logistic and Linear Discriminant Analysis.

Method	Mean Absolute Error	TP	FP	Precision	Recall	F-Measure	Class
Binary Logistic 50 features	0.1686	0.800	0.118	0.891	0.790	0.838	Benign
		0.882	0.210	0.776	0.882	0.826	Malignant
Binary Logistic 5 factors	0.3854	0.810	0.431	0.690	0.79	0.737	Benign
		0.569	0.21	0.690	0.569	0.624	Malignant
Binary Logistic 13 features	0.1403	0.823	0.098	0.911	0.823	0.864	Benign
		0.902	0.117	0.807	0.902	0.852	Malignant
LDA 50 features	0.115	0.861	0.098	0.915	0.871	0.893	Benign
		0.902	0.129	0.852	0.902	0.876	Malignant
LDA 5 factors	0.3982	0.847	0.763	0.589	0.903	0.713	Benign
		0.835	0.097	0.667	0.235	0.348	Malignant
LDA 13 features	0.0708	0.859	0.059	0.950	0.919	0.934	Benign
		0.841	0.081	0.906	0.941	0.941	Malignant

2. The Methodology to Construct Three Feature Selection Methods

Because the research has proposed three selection methods, the simple way to arrangement the detail of each technique is dividing to three sub parts.

2.1 The Cost Reduction in Medical Image Analysis Using Factor Analysis

Most medical image processing uses not only the image information from the texture, spatial, and spectral domains, but also new extended features that are not independent from each other (Zhou and Gordon, 1989.). To reduce computation required by learning process factor analysis is introduced in this research as a way that correlated features can be grouped into sets (factors) and then a small number of independent factors are selected for image analysis. In other words, factor analysis can perform data reduction by using these interdependent properties, as in Songyang (Songyang and Ling, 2000).

Factor analysis is a statistical technique to reduce the number of observed random variables into fewer unobserved random variables called latents or factors (Mitchell, 1997). The observed or manifest variables are modeled as linear combinations of the factors or latent plus error terms. Expressed mathematically, denote the feature set as $X = \{X_{i,j}\}$ $;i=1,2,...,p_i$ where i is a domain space, j is a feature from domain i and p_j is the number of features in domain i.

To simplify this analysis, let $X_i = (X_{i1}, X_{i2}, ..., X_{ipi})$ represent the feature vector of all domains (spatial, spectral, and texture); the model is then

$$X_i = \mu_i + l_{i1}F_1 + l_{i2}F_2 + ... + l_{im}F_m + \varepsilon_i \qquad (1)$$

where μ_ι is a mean of feature vector X_i , l_{ij}, is a loading of X_i on factor, j, F_j is the j^{th} common factor or latent, j = 1,2,..,m and ε_ι is the error. The matrix representation of (1) is

$$\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ \vdots \\ X_p - \mu_p \end{bmatrix} = \begin{bmatrix} l_{11} & l_{12} \ldots l_{1m} \\ l_{21} & l_{22} \ldots l_{2m} \\ \vdots & \\ \vdots & \\ l_{p1} & l_{p2} \ldots l_{pm} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ \vdots \\ F_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

or $$X - \mu = LF + \varepsilon$$

where X- μ, F, and ε are column vectors and **L** is the loading matrix. Using the maximum likelihood or principal component methods, we can obtain estimators $\mathbf{L}^*$ and F^* satisfying

$$X - \mu = L^* F^* + \varepsilon \qquad (2)$$

F_i is a factor or latent that contains significant features in it with a threshold of α. Factor analysis is a divide–and–conquers technique that fulfills our purpose. The beneficial contributions are time saving and ease of detection If n features grouped into p factors then the time reduction is expected to be $(n\text{-}p)\times c$ where c is the processing time per feature. Figure 8 illustrates the workflow using all features, while Figure 9 uses factor analysis to cluster dependent features together in the detection process. Figure 8 illustrates how factor analysis and kernel or transformation functions are used together.

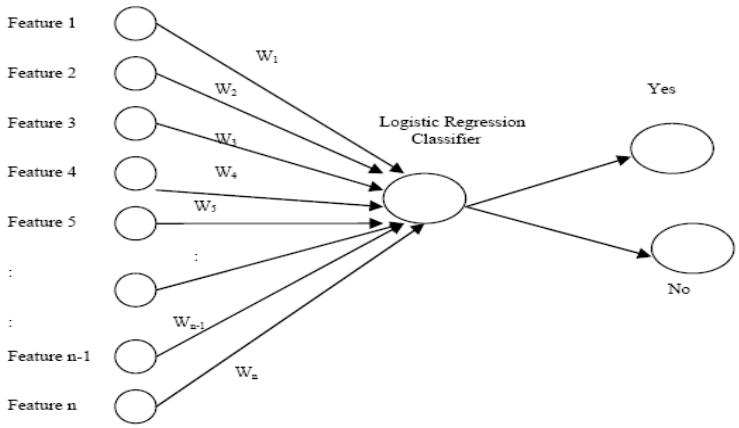

Figure 8 The traditional classification uses all features.

As shows from the Figure 9, a general statistical model is represented by $\beta = [X'X]^{-1}X'Y$ where β is the parameter vector of w_i; i=1,2, …, n; X and Y are the input and output feature vector for a training data set .Computation of the vector β is $O(n^3)$ due to the matrix product of. $[X'X]^{-1}X'Y$

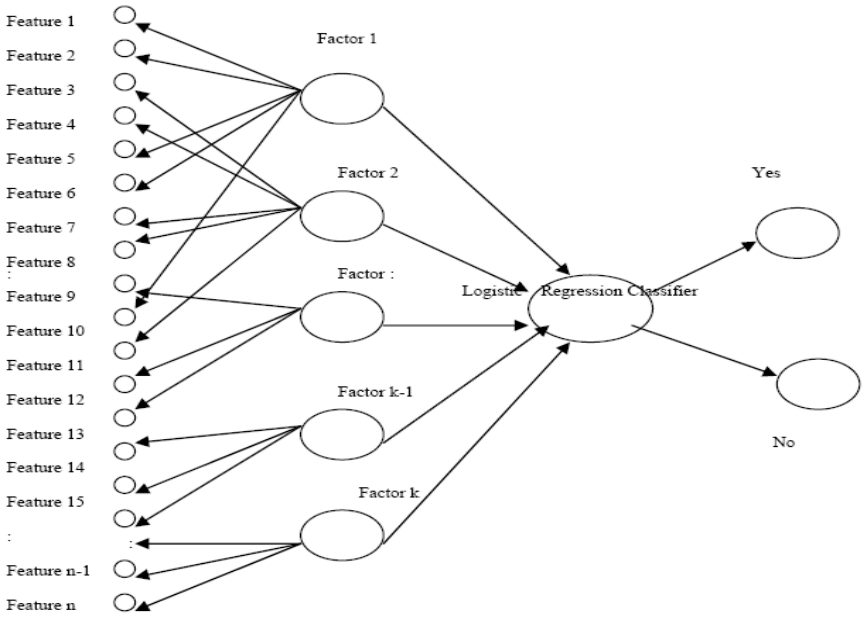

Figure 9 General pattern of feature clustering and detection by using binary logistic regression.

Because the factor analysis method utilizes the correlation matrix of the feature vector which is always symmetric, the computation time is $O(n^2)$ and time to predict output on m factors is $O(m^3)$; $m < n$ (from equation 2).

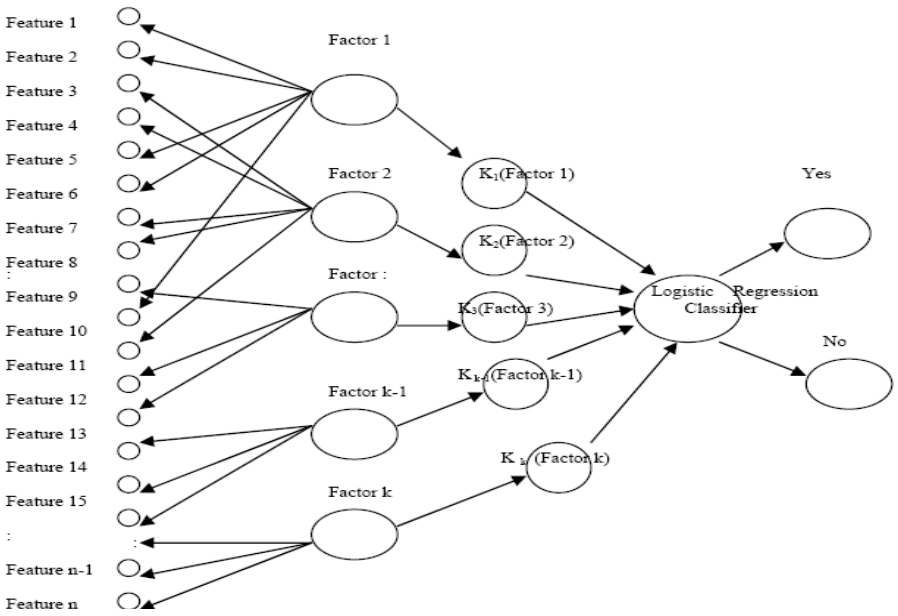

Figure 10 Kernel construction K_i (Factor i) for improving the detection precision.

Figure 10 is designed to improve the accuracy of detection using an appropriated kernel function set. On the assumption that, there are r possible kernel tricks (transformation function) then time consuming in this process is $O(r^k)$ for experiment to select a best kernel for each factor. Exploring a kernel trick for each feature is hard in case of n , a large number of features, with time $O(r^n)$. Finally, the total processing time without clustering features, is

$$O(n^3)+O(r^n) = O(r^n) \; ; r \geq 2$$

Feature extraction is needed to improve speed and correctness of detection. Previously study (Jiang *et al*., 2004), we used some selected kernel, exponential function for example, transform features to new hyper plane (Jiang *et al*., 2004). Since kernel trick is one of the crucial tricks for machine learning and it depends on the feature space then it is necessary to reduce feature space. The problem of large feature space of gray scale will be extended to three times if we process on RGB color image. Fortunately, our challenge technique can overcome such limitation without dropping any chosen features. From Figure 10, if we run an experiment on n features that cluster into m factors and try on r possible kernels, then the total time for

spending at the beginning feature extraction step through the final detection process is $O(n^2)+O(r^m)$.

In addition, we use the binary logistic for classification step; the classifier is a predicted statistical model used to predict the response features to 0 or 1. Rather than classifying an observation into one group or the other, logistic regression can predict the probability of being either group. It predicts the log odds that an observation will have an indicator equal to 1. The odds of an event are defined as the ratio of the probability that an event occurs to the probability that it fail to occurs. Thus, let p be the probability of an event occurs then 1-p is probability of fail. The model is $log\ (p/(1-p))= X\beta + \delta$; β is a parameter vector, δ is a random error vector and X is a feature vector which satisfied independent assumption.

2.2 Feature Selection in Graph Based Analysis

We believed that using only one statistical method for classification will not be successful because of the restriction on measurement values of features and output. As this restriction, we investigate statistical techniques to fulfill the feature selection process. These statistical techniques consist of four parts: 1) feature classification, 2) path analysis, 3) exploration on relations among features and outputs, and 4) hypothesis testing. In the feature classification, we use correlation analysis to transform a number of features into a number of classes. In the path analysis, we construct the conceptual relations among different feature classes. Then, we find the relations between features and outputs using three methods: logistic regression, simple regression, and multiple regression. Finally, hypothesis of feature relationship is tested by a Bayesian technique.

1) Feature classification

Since most low-level features are extracted from spatial and texture based which are highly correlated, the feature selection strategy must realize this limitation. The correlation coefficient is used to analyze this limitation. Correlation coefficient, the most well-known measures of correlation between two random variables, is used

to describe the strength of relationship. Correlation coefficient p between random variable X and Y is defined as $p(x,y) = \frac{\text{cov}(x,y)}{\sqrt{V(x)V(y)}}$ where $\text{cov}(x,y)$ denotes the covariance of X and Y, $V(x)$ and $V(y)$ are variance of X and Y. Correlation matrix is assumed to comply a property of symmetry that correlation of X and Y is the same as correlation of Y and X, and its range is between -1 to 1, where $\rho = 0$ is indicate no linear relation between X and Y.

Correlation coefficients of features can be used to classify many highly-related features into the same groups.

2) Path analysis

From previous phase (feature classification), we can determine many groups of highly-related features. We believe that relationships of features within each group and relationships among groups to final output could be determined by path analysis.

Path analysis exploits benefit of multiple regression analysis. Generally, regression analysis has been used to refer to the analysis of causal models when single indicators are endogenous variables in the model. In path model, there are two types of variable: exogenous and endogenous variables. Exogenous variables may correlate and may leave direct effects as well as indirect effects to the endogenous variables. Causality is the relationship between cause exogenous variable and effect endogenous variable while philosophy causation refers to the set of all particular "causal" relations.

However, path analysis has limitation that it is a regression-based technique which is expected that all variables, especially the dependent variables, are continuous ones. Because our study is based on the continuous cause variables while the endogenous output variable is a dichotomous variable (value is 0 or 1), we cannot use path analysis directly; however, the analysis is still a graph based process. Causal relation analysis can be explained by dependent variables that are measured at interval

or ratio scale (Joreskog and Sorbom, 1989). Thus, in consideration to path analysis with continuous endogenous, the categorical endogenous might leave difficulty both in theoretical background and prediction implication and need some other techniques in turn. Goodman (Goodman, 1971) considered path analysis of binary variable by using logistic regression. Hagenaars (1998) made a general discussion of path analysis of recursive causal systems of categorical variables by using the directed log-linear model approach, which is a combination of Goodman's approach and graphical modeling. Example of the different models of trait effects on output *Y* is illustrated in Figure 11. Figure 11A is a multiple regression model showing that each trait operates simultaneously on fitness *Y*. Figure 11 B is the path analysis model showing four traits at four time periods.

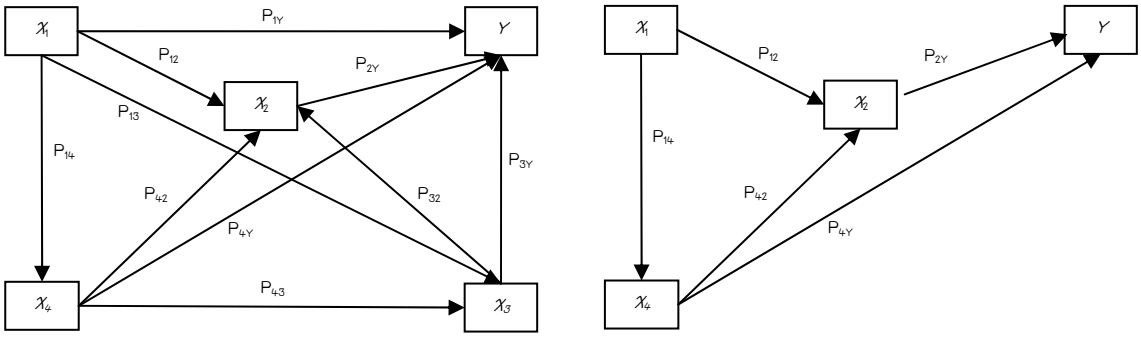

Figure 11 Example of general recursive causal system with four independent features and a dependent output. (A) illustrates possible relations among features and output. (B) is a result from feature selection by analogy with graph base.

A path diagram not only shows the nature and directions of causal relationships but also estimates the strength of relationships. In case of the comparatively weak relationships, those relationships can be discarded; thus some features are eliminated. A path coefficient is the standardized slope of the regression model. This standardized coefficient is a Pearson product – moment correlation. Basically, these relationships are assumed to be unidirectional and linear.

Since our study is based on the binary effect output (cancer or not?) while the cause features are continuous value, we cannot use path analysis directly. To overcome the limitation problem, we use regressions and Bayes inference to construct our graphical model.

3) Relations among features and outputs

From the previous details about features and the path analysis, it is necessary to explore the cause and effect features by regression analysis. In our purpose, we suggest to use logistic, simple, and multiple Regressions.

3.1) Using logistic regression

Logistic regression is a regression model for Bernoulli-distributed dependent variables. It is a linear model that utilizes the logit as its link function. Logistic regression has been used extensively in medical and social sciences (Dodd *et al*., 2004). The logit model takes the form:

$$\log(\frac{p_i}{1-p_i}) = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + e_i; i = 1,2,...,n,$$

where $p_i = \Pr(Y_i = 1)$, β_j ; j =1, 2 … k are parameters (weight) of feature x_i and e_i is a random error (bias) of vector features i (a sample data)

Logistic regression model can be used to predict the response features to be 0 or 1 (0 is benign and 1 is malignant in case of mammogram detection). Rather than classifying an observation into one existing group or the other, logistic regression predicts the probability of being in either group. The model predicts the log odds *(p/(1-p))* that an observation later be transformed to *p* as value of 0 or 1 with an optimal threshold. The odd of an event is defined as ratio of probability that that event occurs to the probability that it fails to occur. Thus, if *p* is the probability of an event occurs then 1-*p* is probability of failure. The general model is $\log(p/(1-p)) = X\beta + \epsilon$ where X is feature vector, β is parameter vector, and ϵ is random error vector.

3.2) Using simple regression and multiple regression

Simple regression has the same basic concepts and assumptions as logistic regression but the dependent variable is on continuous scale and the model has only one single independent variable. The simple regression can be modeled as $Y_i = \beta_0 + \beta_1 X_{1i} + e_i \quad , i = 1,2,..,n$ where Y_i is the dependent variable, β_0, β_1 are parameters (weight), and *n* is the size of training data. X_{1i} is an explained variable of data rerecord i and e_i is random error of X_i . If the results from the analysis has been shown that *p* value of β_1 is significant then we can accept that Y can be determined by X . Multiple regression is the extended simple regression model of multiple variables.

Simple logistic regression and multiple logistic regression were used to explore the cause features to effect output.

4) Hypothesis testing

Although the statistical techniques in previous Section can explore the cause features, it is unable to identify those features whether they are direct or indirect causes. We formulate this issue on the hypothesis testing style.

An appropriate way to test the hypothesis about the direction of causal relationships is easier to illustrate an abstract concept by analogy with Bayesian inference. Bayesian inferences use aspects of the scientific method, which involves collecting evidences that may or may not relevant to a given phenomena. The more evidences accumulated, the more degree of belief in a hypothesis changes. With enough evidence, it will often become very high or very low. Thus, it can in theory be considered a suitable logical basis to discriminate between conflicting hypotheses. Hypotheses with a very high degree of belief should be accepted as true; those with a low degree of belief should be rejected as false. Bayesian inference uses a numerical estimate of the degree of belief in hypothesis *before* evidence has been observed and calculates a numerical estimate of the degree of belief in the hypothesis *after* evidence

has been observed. Bayesian inference usually relies on degrees of belief, or subjective probabilities. Bayes's theorem adjusts probabilities given new evidence as $P(H_0 \mid E) = \frac{P(E \mid H_0)P(H_0)}{P(E)}$, where H_o represents the null hypothesis; that was inferred before new evidence E became available; $P(H_o)$ is the prior probability of H_o; $P(E| H_o)$ is the conditional probability of availability the evidence E given that the hypothesis H_o is true; and $P(E)$ is the marginal probability of E, which is the probability of witnessing the new evidence E under all mutually exclusive hypotheses. $P(E|H_o)$ is the posterior probability of H_o given E.

Using hypothesis testing on the regression, we can use path analysis for the discrete output.

5) Proposed Algorithm

To solve this solution, simple regression, logistic regression and Bayesian inference take into account of causality extraction problem. The algorithm is described as following steps.

Step 1: Partitioning the original feature sets (X_1, X_2... X_n) by using coefficients of correlation matrix to a subset features on the basis of high correlation. Let the number of subset feature be S_i; i=1, 2 ... k and each feature in S_i = (X_{i1}, X_{i2} ... X_{ij}) be related with correlation coefficient p_{ij} (X_i and X_j) .

This step is to partition all features into feature sets S_i, where S_i and S_j ($i \neq j$) are not dependent between each other.

Step 2: Perform the simple logistic regression of each independent feature $X_{ij} \in S_i$ j=1, 2 ... R_i and dependent output Y and then select the possible solution which satisfied the P statistic value (threshold value).

The result from this step shows that the cause features in set S_i is $A_i = (X_{ri}, X_{pi} \ldots X_{ik})$ where each X_{ji} is a direct cause feature of effect output Y.

Step 3: Perform the multiple logistic regression by using all features in set S_i; i=1, 2 ... k into the model and selecting the signify features $B_i = (X_{ti}, X_{li} \ldots X_{zi})$ from the model, where B_j is a set of direct cause features and indirect cause features. Hence in mathematical notation, Step 3 imply that $B_i \subseteq A_i$.

Step 4: Let, $D_i = A_i \ \partial \ B_i$; where ∂ is our *testing hypothesis operator* for exploring the causal relations using Bayes inference conceptual framework.

This step is proven by using Bayesian inference as the following example for two features:

If feature X_i signify to be the cause of Y	$\approx P(Y\|X_i) > C$	(1)
If feature X_t is related (highly correlate) to X_i	$\approx P(X_iX_t) > C$	(2)
If feature X_t is not significant to Y	$\approx P(Y\|X_t) < C$	(3)
If features X_i, X_t are significant to Y	$\approx P(Y\|X_t,X_t) > C$	(4)

where C is a given threshold.

From (1) to (4), we can accept the *hypothesis* that X_i and the interaction of X_i, X_t cause Y. Figure14 illustrates the relation of X_i, X_t to Y and X_i to X_t

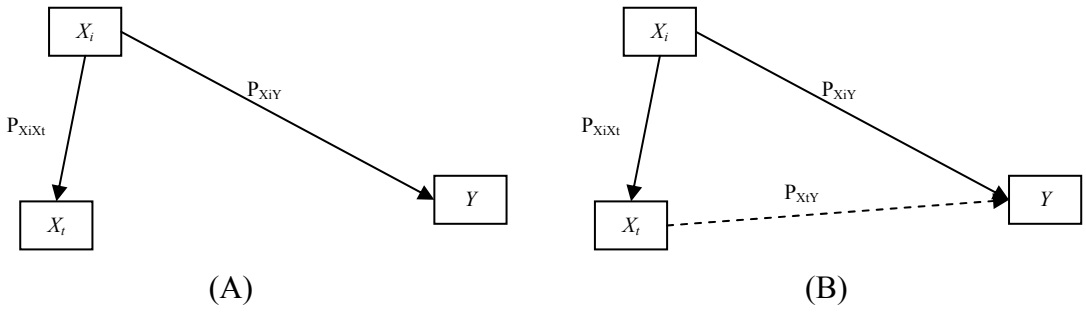

Figure 12 the connected graph on two cause features and effect Y.

There is no direct effect from feature X_t to Y in (A) but, as shows in (B), there is the interaction effect of feature X_t in addition with X_i to Y.

Step 5: Go To Step 2 while $i \leq k$

/* This Step will produce set D_i, where i=1, 2 ... k; and some sets of D_i may be null set.. */

Step 6: Merge sub graph $D_i \rightarrow G(V, E \mid Y) = \cup^{k}_{i=1} D_i$; $V = (v_j)$; $E = (e_j)$; Y is the effect or dependent vertex.

2.3 Feature Selection under Feature Interaction Constraint

Most medical image processing use extended features in addition to texture, spatial, and spectral domains. Since these original and modified features are extracted from the same basis, there are problems of features dependency and interaction. An efficient selection process must concerns about whether the features will be able to provide good information for malignant detection with minimum cost. Our propose method is based on the *assumption* that the object in different class (benign and malignant) must have dissimilarity in same feature. Thus, if any feature has similarity property in two classes (in case of two class labeling), we can discard these features out of the detection. Discarding any feature without concerns about the feature dependency and interaction will caused detection problems, then statistical methodologies i.e. Principal Component Analysis (PCA), in other words, factor analysis can perform data reduction by using these interdependent properties, as in Songyang (Songyang and Ling, 2000) , Analysis of Variance (ANOVA) and Linear Discriminant Analysis (LDA), take into account of our work.

2.3.1 Statistical Methods for Feature Selection

1) Feature Clustering

There are many ways to classify features in to subgroup such as K means, Fuzzy logic on distance base (Veldkamp and Karssmeijer, 1999), Bayes technique on frequency base and Principal Component Analysis (PCA) on correlation base. PCA is a statistical technique to reduce the number of observed random variables into fewer unobserved random variables called latents or factors. The observed or manifest variables are modeled as linear combinations of the factors or latent plus error terms. Mathematically, the feature set is denoted as $X = \{X_{i,j}\}$ $;i=1,2,...,p_i$ where i is a domain space, j is a feature from domain i and p_j is the number of features in domain i.Let $X_i=(x_{i1},x_{i2},...,x_{ipi})$ represent the feature vector of all domains (spatial, spectral, and texture); the model is then

$$X_i = \mu_i + l_{i1} F_1 + l_{i2} F_2 + \ldots l_{im} F_m + \varepsilon_i \qquad (1)$$

where μ_i is a mean of feature vector X_i, l_{ij}, is a loading of X_i on factor j, F_j is the j^{th} common factor or latent, $j=1,2,..,m$ and ε_i. is the error. The matrix representation of (1) is

$$\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ \vdots \\ X_p - \mu_p \end{bmatrix} = \begin{bmatrix} l_{11} & l_{12} \ldots l_{1m} \\ l_{21} & l_{22} \ldots l_{2m} \\ \vdots & \\ \vdots & \\ l_{p1} & l_{p2} \ldots l_{pm} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ \vdots \\ F_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

where X - μ, F, and ε are column vectors and **L** is the loading matrix.

Using the PCA, we can obtain estimators $\mathbf{L}^*$ and F^* satisfying

$$X - \mu = \mathbf{L}^* F^* + \varepsilon \qquad (2)$$

F^*_i is a factor or latent that contains significant features in it with a significant threshold of $\alpha = 0.05$.

Our proposed feature selection involves two steps; the first step is partitioning the original feature into subset of homogeneous subset (clusters) and the second is discarding non-signify subset. Partition of the features is performed by using loading matrix from PCA with unsupervised training data is our first step. Figure 1 as shows below is the conceptual of feature clustering.

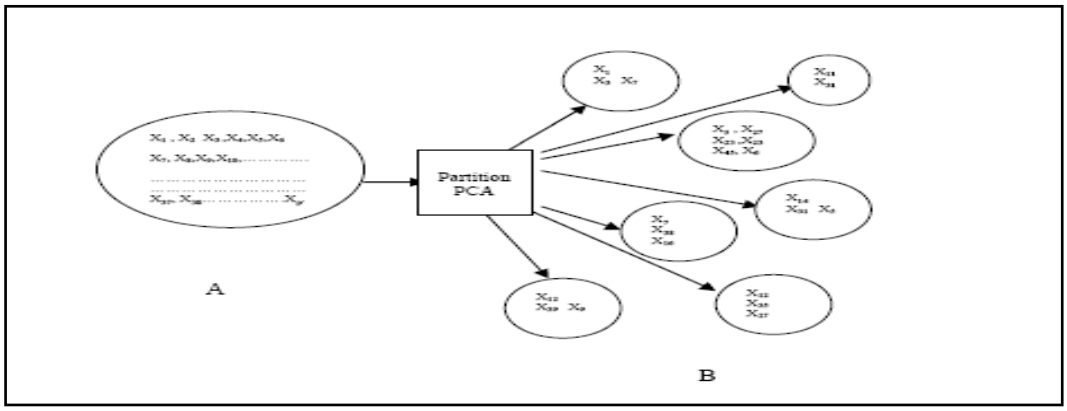

Figure 13 The process of feature clustering into sub group.

Fig 13 illustrates the set of original features extracting from spatial, texture and spectral domain while each independent subset *(B ={S_1,S_2,...,S_k})* are classified by using PCA. Each subset S_i consists of feature vector *(x_{1i},x_{2i},...,x_{ki})* that are high correlate internally with respect to threshold *T*.

In each subset S_i an average value of all features in this set is computed to represent a centroid of each group for further discarding by using ANOVA.

$C_i = \Sigma\ (x_{ij})/m_i$; $i=1,2,..,k$; $j=1,2,..,\ m_i$,where m_j is a number of features in set *i and* C_i *is average mean from all features in set* S_i

2.) Feature Discarding

Feature discarding A data object is describeed by a set of features .A distance similarity on a data *X* is defined to satisfy the follwing condition.

2.1) Symmetry $D(o_i,o_j) = D(o_j,o_i);\ i <> j$

2.2) If $|D(o_i,o_j)| > T$ then o_j and o_i are not similarity Analysis of Variance (ANOVA) uses to test the difference between two or more means, it does this by examining the ratio of variability between two conditions and variability within each condition. A t-test would compare the likelihood of observing the difference in the mean number of words recalled for two group ,while ANOVA test would compare the variability that observe between the two or more conditions to the variability observed within each condition. The measure variability is the sum of the difference of each score from the mean. This technique is used to uncover main and interaction effects of categorical independent variables (features) on an dependent variable . A *main effect* is the direct effect of independent variable on the dependent variable while an *interaction effect* is the joint effect of two or more independent variables on the dependent variable. Since there are interaction among texture features due to extraction process (such as co-occurrence matrix of gray scale , $P_{\square}$, $_d(a, b)$), feature discarding without concern this criterion will be caused detection problem. The following information is the theorectical explanation of ANOVA. Let μ_1, μ_2 represent the mean value of an interesting feature in class 0 and class 1 respectively. The objective is to make an inference about$|\mu_1 - \mu_2|$, the difference between the populations mean of two classes. Suppose that independent sample of *n1, n2* mammograms are randomly selected from each class of supervisory training data (benign and malignant) , and the mean feature values, $\bar{x}_1, \bar{x}_2$ are computed from these two samples. An intuitively appealing estimator for $(\mu_1 - \mu_2)$ is the difference between the sample means $(\bar{x}_1 - \bar{x}_2)$. Statistical assumption on of $(\bar{x}_1 - \bar{x}_2)$ are :

1. The sampling distribution of $(\bar{x}_1 - \bar{x}_2)$ is approximately normal for large samples.
2. The mean of $(\bar{x}_1 - \bar{x}_2)$ is $(\mu_1 - \mu_2)$.
3. If the two samples are independent, the standard deviation of the sampling distribution is

$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$; where σ_1^2, σ_2^2 are the variance of the two populations being sampled, and *n1, n2* are the respective sample size. The hypothesis test for $|\mu_1 - \mu_2|$ (using two sided test) is $H_0 : |\mu_1 - \mu_2| = D_0$ and $H_1 : \neq D_0$ where D_0 is the hypothesis difference between the means. Statistical test for difference is

$$Z = \frac{(\bar{x}_1 - \bar{x}_1) - D_0}{\sigma_{(\bar{x}_1-\bar{x}_2)}}$$

Rejection region $|Z| <$ *Threshold*

Suppose we want to use the two-sample t statistics to compare the mean productivity of two classes, however, we must concern that the assumption of equal variances of two classes may be unrealistic. It would be helpful, to have a statistical procedure to check the validity of this assumption. The common statistical procedure for comparing population variances σ_1^2 and σ_2^2 makes an inference about the ratio σ_1^2 / σ_2^2, based on the ratio of the sample variance (s_1^2 / s_2^2). Thus, we will support the hypothesis that the ratio σ_1^2 / σ_2^2 differs from 1 (i.e., the variances are equal) by testing the null hypothesis that the ratio equal 1 (i.e., the variance are equal).

$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1 (\sigma_1^2 = \sigma_2^2)$ $H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1 (\sigma_1^2 \neq \sigma_2^2)$ and we will test hypothesis by $F - test = \frac{s_1^2}{s_2^2}$. Rejection region: $F > F_{\alpha/2}$ where $s_1^2 > s_2^2$ when $F_{\alpha/2}$ is based on n_1-*1* and n_2-*1* degree of freedom.

3) Classifier Technique

The classification methods use to detect malignant or benign in medical image features are categorized into statistical and machine learning base . Each classier is suitable for its condition such as most of the statistical classifers require the feature independent assupion while Neural Net Work depend on efficient

management of hidden layer and large amount of data (Duda *et al*.,2001). However in case of the dependency problem , some statistical techniques for instant , Linear Discriminant Analysis (LDA), have added the dependency deduction ability. LDA or Mahalanobis distance is widely used like cluster analysis and other clasiification tchniques. In statistics, Mahalanobis distance is a distance intoduced by P.C. Mahalanobis in 1936. It is based on correlation between variables by which different patterns can be idntified and analysed. It is a useful way to classify new sample set to a known one. The technique differs from Euclidean distance in that ,it takes into account the correlations of the data set and is scale-invariant. i.e. not depend on the scale of measurements. In order to use the Mahalanobis distance to classify a test point to one of N classses, one must firstly estimates the covariance matrix of each class, usually based on samples known to belong to each class. Then, given a test sample, computes Mahalanobis distance to each class, and classifies the test pont to that class in which the Mahalanobis distance is minimal .Using the probabilistics interpretation given above, this is equivalent to selectng the class with the highest probability. Also, Mahalanobis distance and leverage are often used to detect outliers espcially in the development of linear regression models. A point that has a greater Mahalanobis distance is said to have higher leverage since it has a greater influence on the slope or coefficients of the regression equation. Mahalanobis distance is closely related to the leverage statistics *h.* The Mahalanobis distance of a data point from the centroid of a multivariate data set is *(N-1)* times the leverage of that data pont, where *N* is the number of data points in the set.

Consider the problem of estimating the probability that a test point in *N*-dimensional Euclidean space belongs to a set, where we are given sample points that definitely belong to that set. The first step would find the average or center of the sample points. Intuiively, the closest the point of question to the center of mass, the more likely belong to that set. However, it is necessary to know how large the set is. The simple approach is to estimate the standard deiation of the distances of the sample points from the center of mass. If the distance between the test point and the center of mass is less than one standard deviation, then we conclude that it is highly probable that test point belongs to the set and vise versa. The assumption behind this

basis is the sample ponits are distributed about the center of mass in sperical manner.Thus ,the probabilty that a test point belong to the set is not only on the distance from the centroid , but also on the elliptical direction . Putting this on a mathematical basis, the ellipsoid that best represent the set's probability distribution can be estimated by building the covariance matrix of the samples . Formally, the Mahalanobis distance *D*, from a group of valus with mean is expressed as

Let vector $\mu = \{\mu_1,\mu_2,...,\mu_p\}^T$ with covariance matrix **P** ;where □□is a mean of feature x_i ;$i=1,2..,p$ and $x = \{x_1, x_2,..., x_p\}^T$ for a multivariate vector is defined as $D_M(x) = \sqrt{(x-\mu)^T P^{-1}(x-\mu)}$

If the covariance matrix is the identity matrix, the Mahalanobis distance reduces to the Eucldean distance. If the covariance matrix is diagonal, the the resulting distance measure is called the normalized Euclidean distance:

$d(\vec{x},\vec{y}) = \sqrt{\sum_{i=1}^{p} \frac{(x_i - y_i)^2}{\sigma^2{}_i}}$; where σ_i is the standard devation of the x_i over the sample set.

4.) Purposed Features Selection Alogoithm

To overcome the detection problem with interraction constraint , less cost of procesing and controllable precision , we collaborate PCA, ANOVA, and LDA (using Mahalanobis distance) to our work. The algorithm is described as the following :

Step1: For all feature ($X_1, X_2,...,X_p$) we formualte an original vector description of each object which represent by ($x_{1i}, x_{2i},\ldots, x_{pi}$) ; where $i = 1,2,..,n$ is numbers of objects , p is number of extraction features and X_i represent the vector of feature i of n objects

./* Input are mammograms and output are set of feature vectors*/

Step 2: Clustering $(X_1, X_2,...,X_p)$ into k subset using PCA , each subset of feature satisfied the conditions that the features within subset are strongly correlated . Let S_i= *{set of features which highly corelation with threshold}; i = 1,2,...,k*

/* Input are set of features of vector images (matrix **X**) and output are subset S_i which consists of subset of correlated features */

Step 3: Divide supervised data set into two class (using class label 0 for benign and 1 fo malignant) .The performing process provides :

S_{ij} *; j=0* , benign classs , *i =1,2,...k*

S_{ij} *; j=1* , malignant classs, *i =1,2,...k*

/* example S_{10} is feature set 1 in class 0 and S_{11} is feature set 1 in class 1 */

Step 4: Compute the avearge value of all features in each subset S_{ij} to used as centroid of each S_{ij} . The result from this step is C_{ij}=*average value of features in* S_{ij}*; i* is subset and *j* is class label (0,1)

/* example C_{10} and C_{11} are average means compute from all features in set S_1 of class 0 and class 1 */

Step 5: *Initial i=1*

Do This step

{Test the differnce of C_{ij} on two calss (j =0,1) using ANOVA with threshold T.

If ($D(C_{i,0}, C_{i,1}, T)$ = 0) ; C_i in class 0 and 1are similar. Then

we can dicard subset feature S_i for further detection step.

Else feature set S_i will remain for further detection step.

i=i+1

} until $i > p$

/* This step is pruning non-significant Set S_i by using ANOVA*/

Step 6: From the discarding process , step 5 , the algorthm provide the leaving set S_i ; $i = 1,..m$. We join all features in S_i together for supporting the next classification step.

$S = \cup_{i=1,..m} S_i$; => $S = (X_1, X_2, ..., X_u)$; $u < p$ (p is the number of original fatures, u is the number of remaining feature subset)

/* Features in S_i ; $i = 1,..m$ combine features for detection step 7 */

Step 7: Use classifier , LDA, for the last step with supervisory training data set.

/* In put is feature selction matrix (supervised training data) and output is a vector of prediction value weight by using LDA */

The system can be visualized in Figure 14.

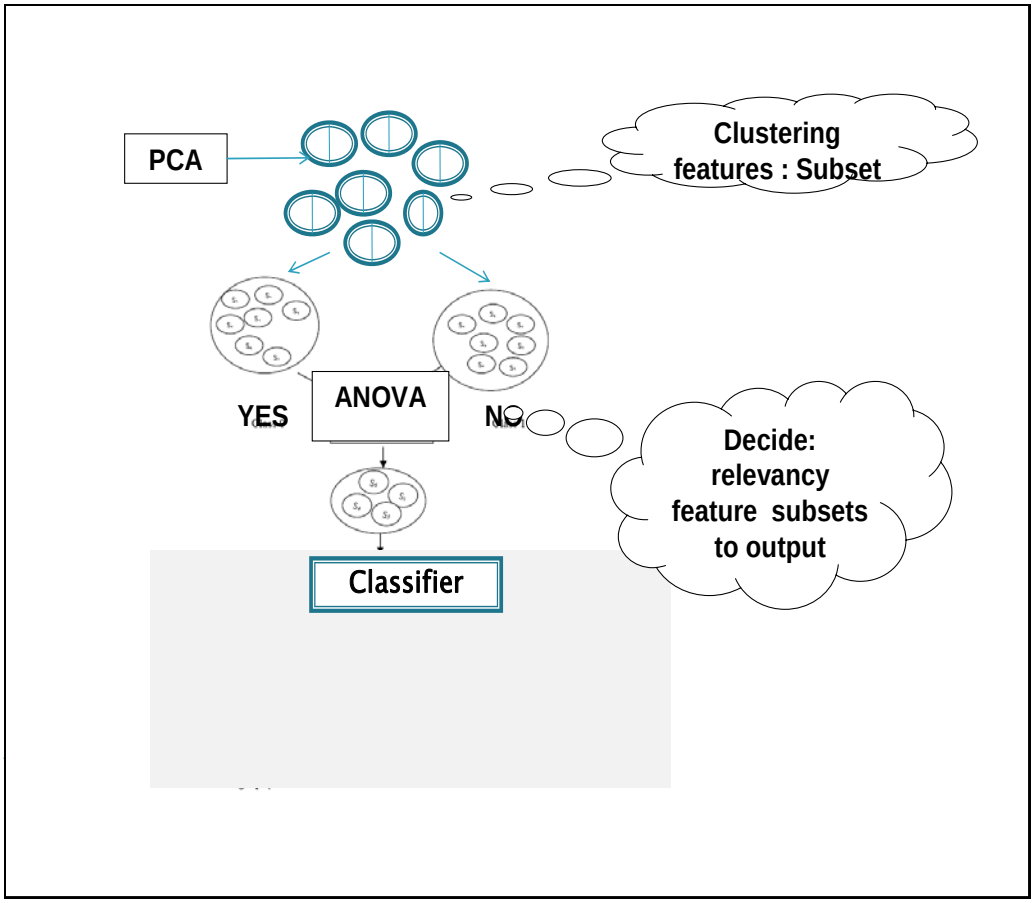

Figure 14 Flow diagram of computer aided detection system based on propose feature selection technique and LDA classifier.

From Figure 14 the experiment features are classified into different set S_i and then S_i of each class (0, 1) of training data is formulated (S_{i0} and S_{i1}) for further step the system in Figure 14. Figure 13 is our purpose algorithms that provide the knowledge based which consists of the significant features and coefficient covariance matrix, **P**, for the mammography CAD. Feature selection is needed to improve precision and reduce time complexity at the detection step. The beneficial contribution for discarding features not only reduces dimensional feature space for gray scale image but also contribute a large amount of profit for RGB in case of color image. Since the problem of large feature space of gray scale will be extended to three times if we process on RGB color image.

RESULTS AND DISCUSSION

Evaluation Methods

The measurements of the proposed methods performance were based on sample ROIs from the mammogram of MIAS. To evaluate each methodology, the inputs were preprocessing already. To evaluate the feature selection from each method, we use standard well-known measures to assess the different approaches i.e. true positive (TP), false positive (FP) and mean square error (MSE), MSE is the average sum of square between predicted value and the actual value of testing data. TP is the ratio of the number of extracted correct results (malignancy) to the total number of extracted results (which are malignant) while FP is the ratio of number of extracted non–correct results to the total of extracted results (which are not malignant)

Table 4 The actual and predicted value to sum up the measurements, TP and FP.

	Actual Malignancy	Actual Non -Malignancy
Predicted Malignancy	a	c
Predicted Non-Malignancy	b	d

$$TP = \frac{a}{a+b} * 100$$

$$FP = \frac{c}{c+d} * 100$$

$$MSE = \frac{\sum_{i=1}^{n} (a_i - p_i)^2}{n}$$; where a_i is actual (supervised data) value and p_i is predicted value.

Results and Discussion

The experimental results of the three selection techniques with different classifiers are summarized according to the resource: MIAS Moreover, the performance and the contribution are discussed at the end of this section. Since, the research has proposed three feature selection techniques to simplify the results we divide this part into three subparts based on each technique and its' experimental data.

1. Computational Reduction Method Using Factor Analysis

1.1 Material and Experimental Results

Our experiment is based on the 113 ROIs from medical images that are segmented by radiologists. After image segmentation, 50 features are extracted from the spatial, texture and spectral domains. The feature set composed of mean, maximum, median, SD (standard deviation), CV (coefficient of variation), skewness, kurtosis from spatial domain; energy, entropy, max-entropy, contrast, inverse different moment, correlation, maximum, SD_x, SD_y (d=1 with angle 0˚, 45˚, 90˚, 135˚) from texture domain; and block activity, spectral entropy from spectral domain. Factor analysis has been trained by varying the number of clusters from 2 to 8. The results appear on Table 5.

Table 5 Feature clustering by varying number of factor from 2 to 8 with the contributed sum of square loading.

No of factors	Sum of square loading	Factor							
		1	2	3	4	5	6	7	8
2	76.84	26	24						
3	83.73	22	24	4					
4	88.28	21	20	4	5				
5	91.71	18	20	4	4	4			
6	93.69	18	20	4	4	3	1		
7	96.51	17	20	4	4	3	1	1	
8	96.64	16	20	4	4	3	2	1	0

Figures in table 5 are sum of square loading and numbers of features tabulated in each cluster.

Table 5 displays the number of features in each group (factor) along with the sum of square loading. For example, in the case of 2 factors, there is 76.84% explained sum of squares with cluster 1 having 26 features and cluster 2 having 24 features. Sum of square loading is the contribution of all features in each factor.

Table 5 also shows the results of the feature clustering. In the table, the optimal solution might be arbitrarily from 5, 6, or 7 factors. We choose the 5 factor case in our experiment because there is no difference in the processing step. The factor composition is shown in Table 6.

Table 6 Factor composition calculates from principal component extraction and varimax rotation with Kaiser Normalization.

Factor Number	List of features in each factor	Spatial	Texture	Spectral
1 (18 features)	Entropy 135˚, Entropy 45˚, Entropy 0˚, Energy 90˚, Energy 135˚, Energy 45˚, Energy 0˚, Max Co-occurence135˚, Max Co-occurrence 45˚, Max Co-occurrence 90˚, Max Co-occurrence 0, Skewness of Region, Kurtosis of Region	2	16	-
2 (20 features)	Inverse difference Moment 45˚, Inverse difference Moment 135˚, Mean of Region, Inverse difference Moment 0˚, Contrast 90˚, Median of Region, Inverse difference Moment 90˚, Block activity, Max of Region, Standard deviation on Y 90˚, Standard deviation on X 90˚, Standard deviation on X	4	15	1

Table 6 (Continued)

Factor Number	List of features in each factor	Spatial	Texture	Spectral
	0˚, Standard deviation on Y 135˚, Standard deviation on X 45˚, Standard deviation on Y 0˚, Standard deviation on Y 45˚, Standard deviation on X 135˚, Contrast 135˚, Contrast 45˚,, mass radian			
3 (4 features)	Mean Co-occurrence 0˚, Mean Co-occurrence 45˚, Mean Co-occurrence 135˚, Mean Co-occurrence 90˚	-	4	-
4 (4 features)	Contrast 0˚, Coefficient of Variation of Region, Correlation 0˚, Standard deviation of Region	2	2	-
5(4 features)	Correlation 135˚, Correlation 45˚, Correlation 90˚, Spectral entropy	-	3	1

From the result in Table 6, the final vectors to be used in the classification step arewhere x_i is the Gaussian transformation of the *i*-th feature.

$$F_1 = \sum_{i=1}^{18} a_i x_i, F_2 = \sum_{i=1}^{19} b x_i, F_3 = \sum_{i=1}^{4} c_i x_i, F_4 = \sum_{i=1}^{4} d_i x_i, F_5 = \sum_{i=1}^{4} e_i x_i$$

From Table 6, we have found that factors 1 to 5 consist of 18, 20, 4, 4, and 4 features, respectively. To enhance the CAD system, reduce cost of processing, and satisfy the requirements of the statistical methods, namely orthogonality and $E(\varepsilon) = 0$,we designed two experiments to compare two prediction methods: one is Logistic regression with 50 features and the other is Logistic regression with 5 factors. Table 7 shows the results of these two alternatives.

Table 7 The logistic regression analysis result computes from two sets of data.

	Method 1 (50 features)	Method 2 (5 factors)
True Positive	82.94	80.46
False Positive	14.51	15.06
MSE	0.52	0.84

The precision of both methods are not significantly different. Furthermore, Method 2 can be enhanced by an appropriated kernel function. Our experiment also shows that as 50 features reduced to 5 factor, time reduction decreased by 50/5=10 times. The new set of factor space is proficient to investigate a suitable kernel function for classifier. The result from the experiment using five linear kernel functions $k_1(F_1)$, $k_2(F_2)$, $k_3(F_3)$, $k_4(F_4)$, $k_5(F_5)$ *is* shown in Table 8.

Table 8 The logistic regression analysis comparison result between method 2 (F_1, F_2, F_3, F_4, F_5) and method 3 $k_1(F_1)$, $k_2(F_2)$, $k_3(F_3)$, $k_4(F_4)$, $k_5(F_5)$.

	Method 2 (F_1,F_2,F_3,F_4,F_5)	Method 3 (K_1,K_2,K_3,K_4,K_5)
True Positive	80.46	84.52
False Positive	15.06	13.15
MSE	0.84	0.69

In our experiment, we explored the power of prediction using adaptive kernel function and finally we used five kernel functions, where $K_i = W_i(Fi)$; i= 1,2,…,5 are our linear kernels

Investigation for appropriated kernel is too difficult for the experiment if there are many features. With our propose method, to construct the kernel function is easy

Table 9 illustrates time in each task of the system, on our experiment the optimal solution is Method 3 with time consume $O(n^2)$ and 84.52 TP rate.

Table 9 The comparative results on time consumption.

Method	Preprocessing Step	Detection Step With kernel	Total Time	TP FP (True Positive) (False Positive)
Method 1 (n features)	$(50)^3$=125,000	-	125,000	82.94. 14.51
Method 1 (n features)	$(50)^3$=125,000	$(4)^{50}=1.26*10^{30}$	≈ $1.26*10^{30}$	No experiment -
Method2 (F_1, F_2, F_3, F_4, F_5)	$(50)^2+(5)^3$ =127,500	-	127,500	80.46 15.06
Method 3 (F_1, F_2, F_3, F_4, F_5) with K_1, K_2, K_3, K_4, K_5)	$(50)^2+(5)^3$ =127,500	$(4)^5$=1024	128,524	84.52 13.15

* Remark: our experiment assumed that there are 4 possible kernel functions for each feature or factor.

1.2 Conclusion of the Experiment

Our research is feature extraction analysis in order to enhance the detection process. We utilize factor analysis to overcome the problem of feature redundancy, improve detection accuracy, and reduce image processing cost. Feature reduction can help to optimize the CAD system. The time reduction by using factor analysis is shown to be on the order of $O(n^3)$ - $O(n^2)$. This is an implementation of feature extraction which is independent of the classifiers. Our work is an intermediate

layer of CAD, for further detection processing. Moreover, this method reduces the features space. We propose factor analysis as a way to extend and enhance traditional extraction processes that that do not incorporate feature selection management

2. Feature Reduction in Graph Based Analysis

2.1 Material and Experimental Results

2.1.1 Training Data and Feature Extraction

Our experiment is based on a training set of 113 ROIs from the Mammographic Image Analysis Society (MIAS) mammogram images that are segmented by radiologists. After image segmentation, 50 features are extracted from the spatial, texture and spectral domains. The feature set composed of mean, maximum, median, standard deviation, skewness, kurtosis of gray level from spatial domain, energy, entropy, modified-entropy, contrast, inverse different moment, correlation, maximum, SD_x (Standard Deviation) and SD_y from the co-occurrence matrix of gray scale used $P_{\phi,d}(a,b)$ with distance d =1 and angle ϕ= 0˚, 45˚, 90˚, 135˚from texture domain and block activity, spectral entropy from spectral domain. *Step*1 of our work is to classify homogeneous feature into 12 feature sets, using the bi-variate correlation coefficient to determine. Table 10 shows list of features in each set.

2.1.2 Feature Selection Experiments

Table 10 12 partitioned feature sets from 50 original features.

Feature set	Number of features	List of Features
#1	4	Entropy rotation from 0˚,45˚,90˚,135˚
#2	4	Energy rotation from 0˚,45˚,90˚,135˚
#3	4	Inverse difference Moment rotation from 0˚,45˚,90˚,135˚
#4	4	Mean Co-occurrence rotation from 0˚,45˚,90˚,135
#5	4	Max Co-occurrence rotation from 0˚,45˚,90˚,135
#6	4	Contrast rotation from 0˚,45˚,90˚,135˚
#7	4	Homogeneity rotation from 0˚,45˚,90˚,135˚
#8	4	Standard deviation on X rotation from 0˚,45˚,90˚,135˚
#9	4	Standard deviation on Y rotation from 0˚,45˚,90˚,135˚
#10	4	Modified entropy rotation from 0˚,45˚,90˚,135˚
#11	7	mean, maximum, median, standard deviation (SD), coefficient of variation (CV), skewness, kurtosis (intensity of gray level)
#12	3	block activity, spectral entropy, mass radian

After the processing of Step 1, the simple and multiple logistic regression analysis in each feature set are manipulated. Tables 11 and 12 illustrate example results from *Step*2 to *Step 4* by using features in feature set #1.

Table 11 The effects among features in feature set #1.

Relations in Feature set #1	Effects to dependent features (using simple linear regression)
Entropy 0˚ to Entropy 45˚	0.000**
Entropy 0˚ to Entropy 90˚	0.004*
Entropy 0˚ to Entropy 135˚	0.000*
Entropy 45˚ to Entropy 90˚	0.000**
Entropy 45˚ to Entropy 135˚	0.022*
Entropy 90˚ to Entropy 135˚	0.000**

* denotes significant with 5% threshold and ** denotes highly significant with 1% threshold.

Significant level , in this table , is number of time out of hundred in believing that relations in feature set #1 is not the determinants of dependent feature when it exactly does.

Table 12 shows probability values, which are computed from training data model under the null hypothesis that the correlation parameter of two testing features are not correlated where as the value which is less than threshold (.05 or 5%) is indicated that the null hypothesis is rejected and hence the alternative hypothesis that there is relationship between two features is accepted.

Table 12 The effects of features in feature set #1 to output

Feature set #1	Effects to output	
	Using simple logistic regression	Using multiple logistic regression
Entropy 0˚	0.034 *	0.026*
Entropy 45˚	0.433	0.031*
Entropy 90˚	0.363	0.241
Entropy 135˚	0.159	0.169

* denotes significant with 5% threshold and ** denotes highly significant with 1% threshold.

Tables 11 and 12 show as the matters of facts that:

- From Table 11: Entropy 0˚ and Entropy 45˚ are highly significantly related.

- From Table 12 (second column): based on the simple logistic model, there is only one evident, Entropy 0˚ causes *Y* (Entropy 0˚ is significant to *Y*).

- From Table 12 (third column): on the multiple logistic regression model, there exist two features, Entropy 0˚ and Entropy 45˚ cause *Y*.

- Finally, with Bayes inference, the direct effect is Entropy 0˚ and the indirect effect is the interaction of Entropy 0˚ and Entropy 45˚ cause *Y*.

Table 12 is the result of Step *4,* $D_i = A_i \ominus Bi$ where i =1. With this analogy on the k iterations of the algorithm, finally our experiment has found that the number features were reduced from original 50 to 13 features. Those features are Entropy 0˚, Entropy 45˚, Max Co-occurrence 45˚, Max Co-occurrence 135˚, Mean Co-occurrence 0˚, Mean Co-occurrence 90˚, Energy 45˚, Homogeneity 0˚, Homogeneity 45˚, Homogeneity 90˚, Homogeneity 135˚, Standard deviation and

Skewness of intensity value. The constructive cause and effect graph, $G(V,E|Y)$, are shown as Figure 15.

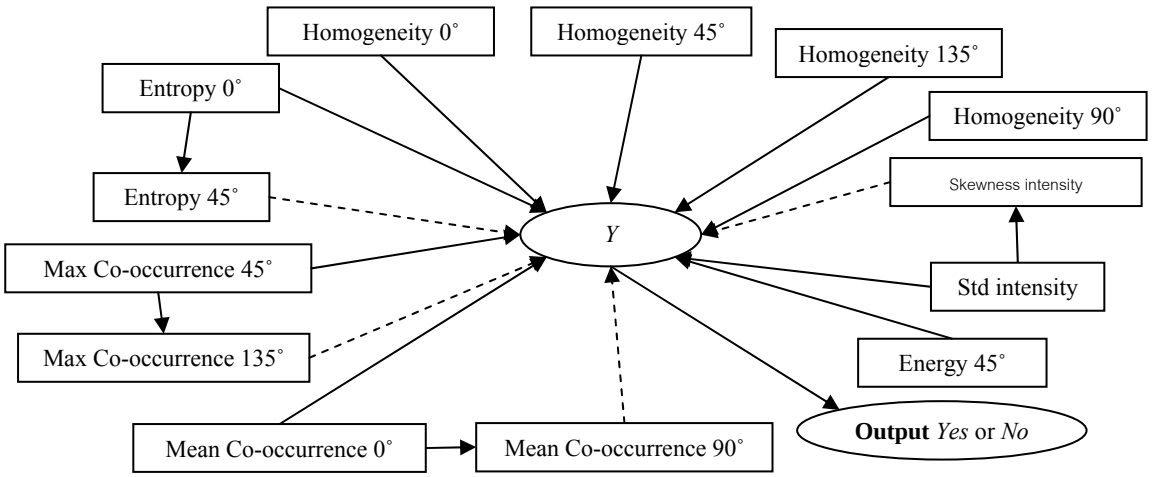

Figure 15 Complete graph on the experiment with direct and indirect effect from retaining process. (Dot line shows indirect effect)

Finally, we evaluated the performance of selected 13 features with original 50 features on two learning models: Linear Discriminant Analysis (LDA) and logistic regression. The results of true positive (TP), false positive (FP) and minimum square error (MSE) are shown in Tables 12 and 13.

Table 13 Performance of logistic regression between using original 50 and selected 13 features.

Logistic regression	TP	FP	MSE
Using original 50 features	82.94	14.51	0.052
Using selected 26 features (SFS-26)	77.41	18.72	0.102
Using selected 13 features	81.64	15.06	0.084

Table 14 Performance of LDA between using original 50 and selected 13 features.

LDA	TP	FP	MSE
Using original 50 features	85.72	12.12	0.025
Using selected 26 features (SFS-26)	80.01	13.32	0.071
Using selected 13 features	84.44	11.02	0.053

The experimental results show that there are no significant difference between both sets of features (50 and 13) when they are applied on both learning model (logistic regression and ANN). Hence our proposed feature selection technique is satisfied for improving CAD.

2.2 Conclusion of the Experiment

In this research, a method to reduce number of features for medical image detection is proposed. We used the mammograms from the Mammographic Image Analysis Society (MIAS).As experiment data and applied our proposed methods to reduce number of features from frequently-used 50 features to 13 features, while accuracies on two learning models are comparably equal. Our method can reduce computation cost of mammogram image analysis and it could be applied to other image-analysis applications.

Graph-based analysis was examined using statistical techniques to identify the crucial direct or indirect features for breast cancer detection in medical images. Our algorithm requires time complexity $O(n^2)$. We can accept the hypothesis that there is no significance between 50 features and 13 features for LDA and logistic regression with threshold 5%. A comparison of the performance between the different configurations of architectures over two set of features (50 and 13 features) with two classifiers (LDA and logistic regression) indicates that the selected 13 features provide the best results in terms of precision with respect to computation time. Using our approach, the detection step improves the temporal ratio of computation by

number of features by 50:13. Moreover, the proposed method demonstrates satisfactory performance and cost compared to SFS.

In our experiment, the 50 features were partitioned into 12 feature sets with S_{11} being the largest set. With this set, the search space for direct cause features (A_7) is (${}^7C_1$) while indirect cause (B_7) exploration was (${}^7Ci$) i=2, 3 … 7. We also found that there were 11 features from the texture domain and two features from the spatial domain that were eliminated from the selection process. The mass radian was not a significant feature because some masses on benign images were larger than on malignant images. Instead of using mass radian (microcalcification), the distribution of micro-calcification is more advantageous.

On the theoretical aspect of finding a best combination feature set, the only way to guarantee the selection of an optimal feature set is an exhaustive search of all possible subsets of features. However, the search space could be very large: 2^N for a set of N features. Our algorithm provides a divide and conquer strategy; with N features (assume that there are r groups with k features each), the number of possible subsets for examining the feature selection is $r\,{}^kC_i$; i=1, 2 ... k.

3. Feature Selection under Feature Interaction Constraint (using PCA and ANOVA)

3.1 Material and Experimental Results

3.1.1 Training Data and Feature Extraction

Our experiment is based on training data that are segmented by radiologists. After image segmentation, 50 features are extracted from the spatial, texture, frequency and spectral domains. The feature set composed of mean, maximum, median, SD (standard deviation), CV (coefficient of variation), skewness, kurtosis from spatial domain; energy, entropy, max-entropy, contrast, inverse different moment, correlation, maximum, SD_x, SD_y (d=1 with angle 0˚, 45˚, 90˚, 135˚)

from texture domain; block activity, spectral entropy from spectral domain and mass radian from ground truth knowledge.

3.1.2 Feature Selection Experiments

Our experiment has formulated in to two sub tasks for comparing the merit of our method.

Sub task 1: Applying ANOVA to prune the original features on the assumption that each feature in two classes are independent and then used the remaining features for the classification.

Sub task 2: Before applying ANOVA to prune the original features, we partition the features in to subset using PCA based on the assumption that some features have *interaction* through dependent feature. Finally, we have discarded feature subset and used the features from remaining sets for further classification.

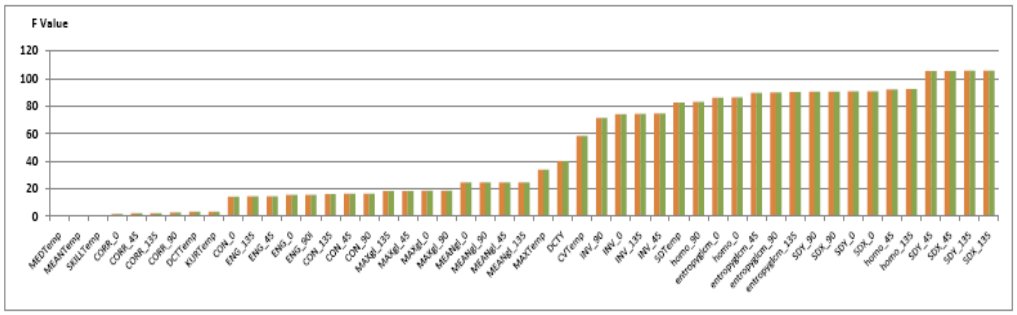

Figure 16 The F test values on each original 50 features.

Figure 16, is the F values which measure the variation ratio of "between" and "within class" variation, therefore the higher value indicate the different feature in two classes. From the experiment in *Sub task 1*, we found that there are 40 features remaining with threshold 5% (significant level) and the 9 discarding features composed of median, mean, kurtosis, skillness from spatial domain, correlation with 0˚, 45˚,90˚,135˚ from texture domain ,and spectral entropy

from spectral domain. In *Sub task 2*, we classify the original features into subsets using loading factor from PCA and the features in each subset have listed below.

Table 15 List of subsets of homogeneous features which partition by PCA (Figure 11 is the visualization of Table 15).

Subset Number	List of features in each Subset
#1 (14 features)	inverse difference moment 0˚, inverse difference moment 45˚, inverse difference moment 90˚, inverse difference moment 135˚, entropy 0˚,entropy 45˚, entropy 90˚, entropy 135˚, inverse difference moment 0˚, inverse difference moment 45˚, inverse difference moment 90˚, inverse difference moment 135˚, coefficient of variation of region intensity, standard deviation of region intensity
#2 (14 features)	mean co-occurrence 0˚, mean co-occurrence 45˚, mean co-occurrence 90˚, mean co-occurrence 135˚, maximum of region intensity, standard deviation on Y 90˚, standard deviation on X 90˚, standard deviation on X 0˚, standard deviation on Y 135˚, standard deviation on X 45˚, standard deviation on Y 0˚, standard deviation on Y 45˚, standard deviation on X 135˚, mean of region intensity
#3 (8 features)	energy 0˚, energy 45˚, energy 90˚, energy 135˚, max co-occurrence 0˚, max Co-occurrence 45˚, max co-occurrence 90˚, max co-occurrence 135˚
#4 (4 features)	homogeneity rotation 0˚, homogeneity rotation 45˚, homogeneity rotation 90˚, homogeneity rotation 135˚
#5 (4 features)	contrast 0˚, contrast 45˚, contrast 90˚, contrast 135˚
#6 (3 features)	median of region , Block activity, mass radian
#7 (1 features)	kurtosis of region intensity

Table 15 (Continued)

Subset Number	List of features in each Subset
#8 (1 features)	skewness of region intensity
#9 (1 features)	Spectral entropy

From Table 15, the average mean of each subset are C_i ; *i=1,2,..9*. The next step in sub task 2 is discarding subset *i* using ANOVA and the results have shown (Figure 16.) that the remaining subsets are S_2, S_4, S_6, S_7 *and* S_8. Consequently, we combine all features from these sets for further classification. By using the same classifier, LDA for two sub tasks with different discarding techniques with different assumptions, the experimental results are illustrated as Table 16 and Table 17.

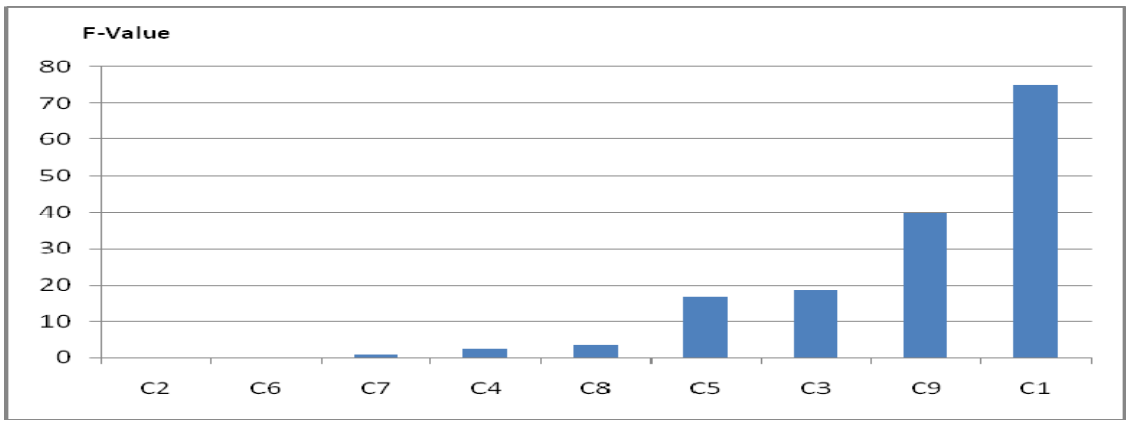

Figure 17 The F-test values of 9 subsets (experiment on Sub task 2).

Based on the F values from the experiment in Figure 17, we have founded that there are 5 subsets S_2, S_4, S_6, S_7 and S_8 were discarding while the remaining subsets are S_1, S_3, S_5 and S_9.

Table 16 List of leaving features on the propose method (experiment of our Sub task2).

Subset Number	List of features in each subset
#1 (14 features)	Inverse difference Moment 0˚, Inverse difference Moment 45˚, Inverse difference Moment 90˚, Inverse difference Moment 135˚, Entropy 0˚, Entropy 45˚, Entropy 90˚, Entropy 135,Homogeneity 0˚, Homogeneity 45˚, Homogeneity 90˚, Homogeneity 180˚, Coefficient of Variation of Region, Standard deviation of Region
#3 (8 features)	Energy 0˚, Energy 45˚, Energy 90˚, Energy 135˚, Max Co-occurrence 0˚, Max Co-occurrence 45˚, Max Co-occurrence 90˚, Max Co-occurrence 135˚
#5 (4 features)	Contrast 0˚, Contrast 45˚, Contrast 90˚, Contrast 135˚
#9(1 features)	Spectral entropy

3.2 Conclusion of the Experiment

In accordance with two discarding methods, we examined the two set of selected features from *sub task 1* (40 features) and *sub task 2* (27 features) with LDA classifier. The results of true positive and false positive are shown in Table 16.

Table 17 Sub-task1 and Sub-task2 with LDA and Logistic classifier are carrying out a comparative evaluation.

LDA	Sub Task 1 (40 features)	Sub Task 2 (27 features)
True Positive (%)	80.17	81.04
False Positive(%)	15.12	14.09

Logistic Regression	Sub Task 1 (40 features)	Sub Task 2 (27 features)
True Positive(%)	78.07	79.45
False Positive(%)	17.22	14.32

The TP and FP of our experiment on two sub tasks with two classifiers are quite different due to different selection technique. The selection technique on sub task1 leaves more features but less precision than sub task 2. From these results, we can accept the hypothesis of sub task 2 of having feature interaction effects through output detection. Therefore, using any features selection method must be concerned with the features interaction effect. *Sub task 1* is a strategy that discards any feature without concern the interaction while *Sub task 2* uses PCA for clustering feature set before discarding.

Our research is to execute feature selection analysis for detection process enhancement. We utilize PCA and ANOVA to overcome the problem of feature interaction, improve detection accuracy, and reduce image processing cost. From the experiment, our proposed technique can reduce features from 50 to 27 features and give satisfying result with respect to the TP and FP .In addition to the acceptance of the hypothesis, the feature interaction gives the most significant effect to the system. For future work, we intend to enhance the CAD using our technique for the first feature selection step and use genetic algorithm for second selection step again to prove consistent of our algorithms.

4. Discussion of Three Feature Selection Techniques

Factor analysis is a practical feature selection technique appropriate for grouping large amounts of relevant features into some feature groups . According to the real problem of cancer detection, all relevant features - such as patient profile, chemical in blood fluid, clinical features of postpartum depression, suspected tissue information and image information are - extracted and grouped for diagnosis purpose. Factor analysis not only the cost reduction technique but also contributes the way to improve kernel function for detection step extraction which is independent from any classifiers.

In discarding methods: we first utilized PCA and ANOVA for pruning the indifferent feature group and then Bayes and Logistic Regression are used to select the significant features. PCA and ANOVA is classifier independent while graph based is classifier dependent. The experiment revealed that the second method gives better satisfactions of TP and FP rate compare to the leaving features. There are 27 leaving features for the first discarding feature method with the same performance with respect to SFS method but our proposed method is less time consuming and ease of use.

CONCLUSION AND RECOMMENDATION

Conclusion

Decision Support System (DSS) for detection and diagnosis (CAD) is increasingly interested due to the high number of patient need appropriated diagnostic while lacking mammogram expert.

Medical Image Analysis used features from mammogram. There are 3 steps in this system, i.e., preprocessing step, feature extraction and detection.

Feature selection aims to extract all significant features for detection purpose. But more features are costly and does not necessary improve system efficiency. There are two feature extraction paradigms; one is keeping all features since they are all useful for patient symptom explanation while the other is selecting only significant features. Each paradigm is applicable on different situation such as to diagnosis of the cancer detection in bone marrow, all features from patient profile; chemical treatment, blood fluid etc are used for detection, where as in malignancy detection only significant feature is selected.

In keeping all features, factor analysis technique is chosen and feature discarding by dropping non-significant is another. Discarding techniques, based on statistics collaborative of correlation analysis, Bayes inference and ANOVA were chosen.

The benefit of feature selection by using Factor Analysis is well performed for case of many features and feature dependency problem. This method is an efficient way to construct small number of factors like a set of hidden layer node in ANN. Factor Analysis reduces feature space and is the classifier independent technique. With small amount of factors, the capability of improving detection proficiency base on exploring new kernel function is quite easy.

Two feature discarding techniques are graph base approach and used ANOVA. The 13 features is leaving by using Feature selection in graph base analysis while using ANOVA, there are 23 leaving features. The performance of using graph base is better than using ANOVA. ANOVA for feature selection gives the contribution same as SFS but less processing cost and the more powerful than SFS.

Recommendation

The limitation of this study is that data used in this analysis are only mammogram. Therefore, the results may not be the same even though we believe that the same process can be applied. Other data that should be considered to add into the analysis are patient profile, laboratory results, symptoms and sickness severity. Factor analysis may reveal semantic explanation. We believe that factor analysis could be a better support medical diagnostic tool for further in-depth study.

Factor analysis (FA) could be thought of as a node construction in a hidden layer of Artificial Neural Network (ANN). Therefore, if FA and ANN are combined by replacing FA to heuristic approach in that construction, better understanding of causal relationships may be attained.

Moreover, researcher should pay more attention to color medical image which provide approximately 3 times of what found in gray scale level. Moreover, other problems in medical image analysis are the lacking of supervised training data should use feature selection methods which are independent from classifiers. In addition to, we can combine our two techniques, pruning by using PCA and ANOVA, and then using Graph Based Analysis for feature selection.

According to our experiment in proposed method, we have found that most of significant features are texture features. In recent literatures, texture analysis is an important generic research area of machine vision. The potential areas of application include biomedical image analysis, industrial inspection, analysis of satellite or aerial imagery, content-based retrieval from image databases, biometric person

authentication, texture synthesis for computer graphics and animation, and image coding. Our research used the most widely, texture measures which derived from gray level co-occurrence matrices or difference histograms, and texture energy measures obtained by local linear transform. Recently, a new interesting approach for texture analysis is nonparametric technique. The approach constructs on the distributions of simple texture measures based on local binary patterns and signed gray level differences are used to provide complementary information about the structural and stochastic properties of image texture. Nonparametric test is a statistical approach which performs well in free condition of assumption about the feature distribution while the other approach such as T-test and ANOVA are based on the assumption of Gaussian distribution. Then, to explore and modify feature selection many new texture analysis is suggested.

For another paradigm where feature selection and feature pruning were employed, we found that graph analysis is an efficient pruning algorithm. The findings should be further investigated in order to study the stability of results if other types of data have been fed into algorithms. Moreover, filtered features should be supplied to another feature screening method such as genetic algorithm.

Beside the deep analysis on texture domain with nonparametric approach, the dispersion of microcalcification is also an interesting theme for our study.

LITERATURE CITED

Almato, G.S. Roy, A.S. MacMahon, H., Li, F., Doi, K., Sone, S., Altman, M.B. 2005. Evaluation of Automated Lung Nodule Detection on Low dose Computed Tomography Scan From a Lung Cancer Screening Program. AUR. **Acad. Radiol.** 12: 337-346.

American Cancer Society, Breast Cancer Facts & Figures. 2005-2006. **National Home Office American Cancer Society, Inc**. Available Source: http://www.lungcancer.org/patients/fs_pc_lc_101.htm, April 15, 2007.

Breast cancer, Maipang, T. Department of Surgery. 2001. **Faculty of Medicine, Prince of Songkla University**. Available Source: http://www.psu. ac.th/smj2/1917.htm, September 1, 2008.

Chiou, G.I. and J.-N. Hwang. 1995. A Neural Network Based Stochastic Active Nodule (NNS-SNAKE) for Contour Finding of Distinct Features. **Image Process. IEEE Trans** 4: 1407-1416.

Cosit, D. and S.L. Loncaric. 1997. Ruled Based Labeling of CT head Image. *In* **Proceeding of 6th Conference on Artificial Intelligence in Medicine**, Europe . 453-456.

Dodd, L.E., R.F. Wagner, S.G. Armato, M.F. McNitt-Gray, S. Beiden, , H.P. Chan, D. Gur, G. McLennan, C.E. Metz, N. Petrick, B. Sahiner, and J. Sayre. 2004. Assessment Methodologies and Statistical Issues for Computer-Aided Diagnosis of Lung Nodules in Computed Tomography. **Acad. Radiol** 11: 462-474.

Doi, K., H. MacMahon, S. Katsuragawa, R.M. Nishikawa, and Y. Jiang. 1999. Computer- aided diagnosis in radiology: Potential and pitfall. Eur. J. **Radiol**. 31:97-109.

Duda. Richard, O. E. Hart.Peter, and G. Stork.David. 2001. **Pattern Classification**, Second Edition,John Wiley & Sons, Inc.

Foggia, P., M. Guerriero, G. Percannella, C.Sansone, F. Tufano, and M.A.Vento. 2006. Graph-Based Method for Detecting and Classifying Clusters in Mammographic Images. **Lect. Notes Comput. Sci., Struct. Syntact. Stat. Patt. Recognition** 4109: 484-493.

Fu, J.C., S.K. Lee, S.T.C. Wong, J.Y. Yeh, A.H. Wang, and H.K. Wu. 2005. Image segmentation feature selection and pattern classification for mammographic microcalcifications. **Comput. Med. Image** 29: 419-429.

Gliman, D.M. and L. Sizzanme. 2005. State of the Art FDG Pet Imaging of Lung Cancer. **Semina. Roentgenol** 40: 143-153.

Goodman L.A. 1971. Exploratory latent structure analysis using both identifiable and unidentifiable models. **Biometrica** 61: 215-231.

Guler, I. and E. D. Ubeyli. 2007. Expert systems for time-varying biomedical signals using eigen-vector methods. **Expert Systems with Applications** 32:1045-1058.

Hagenarrs, J.A. 1998. Categorical causal modeling latent class analysis and discrete log-linear models with latent variables. **Social Methods Research** 26: 436-486.

Hiroyuki, A., M. Herber, S. Junji, L. Qing, E. Roger, and D. Kunio. 2003. Computer – Aided Diagnosis in Chest Radiography: Results of Large -Scale Observer Tests at the 1886-2001 RSNA Scientific Assemblies. **Radiographics** 23: 255-265.

Jiang, W., M. Li, H .Zhang, and J. Gu. 2004. Online Feature Selection Based on Generalized Feature Contrast Model. ***In*** **Proceeding of Multimedia Exp Conference, IEEE** (CME).

Joreskog, K.G. and D. Sorbom. 1989. **LISRL** 7 **User's Reference Guide; SPSS** Inc.: Chicago.

Lehmann, T.M.., M.O. Guld, C. Thres, B. Fischer, and K. Spitzer. 2006. **Content-Based Access to Medical Images.** Available Source: http://phobos.imib.rwth - aachen.de/irma/ps-pdf/MI2006_ Resubmission2.pdf, May 25 ,2007.

Liang, Z., R.J. Jaszczak, and R.E. Coleman. 1992. Parameter Estimation of Finite Mixtures Using the EM Algorithm and Information Criteria with Application to Medical Image Processing. **IEEE T. Nucl. Sci** . 39: 1126 - 1133.

Lung Cancer Home Page. 2007. Available Source: http://www.lungcancer.org/ patients/fs_pc_lc_101.htm, December 25, 2007.

Luyin, Z., L. Boroczky, and K.P. Lee. 2005. False positive reduction for lung nodule CAD using support vector machines and genetic algorithms, **Computer Assisted Radiology and Surgery,** May 2005: 1109-1114.

Majcenic, Z. and S. Loncaric. 2007. CT Image Labeling Using Simulated Annealing Algorithm. Available Source: http://citeseerx.ist.psu.edu, July 12, 2007.

Miller, H., S. Marquis, G. Cohen, P.A. Poletti, C. Lovis, and A. Geissbuhler. 2007. Automatic: Abnormal Region detection in Lung CT images for Visual Retrieval. University and Hospital of Geneva, Service of Medical Informatics, **Department de Radiologie et Informatique Medicale**. Available Source : http://www.simhcuge.ch/medgift ,September 5, 2007.

Mitchell ,Tom, M. 1997. **MACHINE LEARNING**. The McGraw-Hill Companies, Inc.United State.

Pietikainen, M., T. Ojala, and Z. Xu. 2004. **Rotation Invariant Texture Classification Using Feature Distributions**. Available Source: www.mediateam.oulu.fi/publications/pdf/7, March 17 , 2007.

Ping, Z., B. Verma, and K. Kumar. 2004. Neural Vs Statistical Classifier in Conjunction with Genetic Algorithm Feature Selection in Digital Mammography. *In* **Proceedings of 2004, IEEE International Joint Conference on Neural Networks** 3(5-29): 2303 - 2308.

Samuel, G., III. Armato, S. R. Arunabha, H. MacMahon, F. Li, K. Doi, S. Shusuke, and M. B. Atman. 2005. Evaluation of Automated Lung Nodule Detection on Low dose Computed Tomography Scan from a Lung Cancer Screening Program, **AUR**, **Academic Radio 2005** 12 (3): 337-346.

Shiraishi, J., H. Abe, F. Li, R. Engelmann, H. MacMahon, and K. Doi. 2006. Computer-aided Diagnosis for the Detection and Classification of Lung Cancers on Chest Radiographs. **Science Direct. Acad. Radiol** 13: 995-1003.

Song, J.H., S.S.Venkatesh, E.A. Conant, P.H., and C.M. ArgerSehgal. 2005. Comparative Analysis of Logistic Regression and Artificial Neural Network for Computer-Aided Diagnosis of Breast Masses. **Acad. Radiol** 12: 487-495.

Songyang, Y. and G. Ling. 2000. A CAD System for the Automatic Detection of Clustered Microcalcifications in Digitized Mammogram Films. **IEEE T. Med. Imaging** 19: 115-126.

Sorwar, G. and A. Abraaham. 2006. DCT Based Texture Classification sings Soft Computing Approach. **Malaysian Journal of Computer Science** 17 (1): 13-23.

Thyagarajan., K.S. 2006. **Digital Image Processing with Application to Digital Cinema**, Elservier Inc.

Veldkamp,W.J.H. and N. Karssmeijer. 1999. An improved method for detection of microcalcification clusters in digital mamogramsm . *In* **Proceedings of** The **SPIE Conference on Image Processing**, 3661: 512-522.

Widodo, A. and B.-S. Yang. 2007. Application of nonlinear feature extraction and support vector Machines for fault diagnosis of induction motors. **Expert System** 33: 241- 250.

Zhang, P., B. Verma, and K. Kumar. 2004. Neural Vs Statistical Classifier in Conjunction with Genetic Algorithm Feature Selection in Digital Mammography. *In* **Proceedings of IEEE Int. Joint Conf. Neural Networks** 3 (25-29): 2303 - 2308.

Zhao, L., L. Boroczky, and K.P. Lee. 2005. False positive reduction for lung nodule CAD using support vector machines and genetic algorithms. **Computer Assist. Radiol Surgery,** 1281: 1109-1114.

Zhao, W., X. Yu, and F. Li. 2005. Microcalcification Patterns Recognition Based Combination of Auto association and Classifier. **Lect. Notes Computer Science, Computer Intelligent Security** 3801: 1045-1050.

Zheng, B., W. Qian, and L.P. Clarke. 1996. Digital mammography: mixed feature neural network with spectral entropy decision for detection of microcalcifications. **IEEE T. Med. Imaging**. 15: 589-97.

Zhou, X., and R. Gordon. 1989. Detection of early breast cancer an overview and future prospects. **Biomed. Eng**. 17: 203-255.

Zoran M., and L. Soren. 2007. Multiresolution simulated annealing for brain analysis. *In* **Proceedings of SPIE Medical Imaging** 3661: 1139-1146.

CIRRICULUM VITAE

NAME : Mrs Rapepun Piriyakl

BIRTH DATE : October 5, 1949

BIRTH PLACE : Chonburi, Thailand

EDUCATION :

YEAR	INSTITUTE	DEGREE/DIPLOMA
1972	Chulalongkorn Univ.	B.Sc.(Mathematic)
1975	NIDA	M.S.(Applied Statistics)

POSITION/TITLE : Associate Professor

WORK PLACE : Department of Computer Science
Faculty of Science
Ramkhamhaeng University