A Computer Aided Diagnosis System for Lung Cancer based on Statistical and Machine Learning Techniques

Hamada R. H. Al-Absi¹, Brahim Belhaouari Samir²*, Suziah Sulaiman¹
¹Department of Computer & Information Sciences, Faculty of Science and Information Technology
Universiti Teknologi PETRONAS, 31750 Tronoh, Perak, Malaysia
²College of Science, ALFAISAL University, P.O.Box 50927, Riyadh 11533, Kingdom of Saudi Arabia
Email: hamadarasheed@ieee.org, sbelhaouari@alfaisal.edu, suziah@petronas.com.my

Abstract — Lung Cancer is believed to be among the primary factors for death across the world. Within this paper, statistical and machine learning techniques are employed to build a computer aided diagnosis system for the purpose of classifying lung cancer. The system includes preprocessing phase, feature extraction phase, feature selection phase and classification phase. For feature extraction, wavelet transform is used and for feature selection, two-step statistical techniques are applied. Clustering-K-nearest-neighbor classifier is employed for classification. The Japanese Society of Radiological Technology’s standard dataset of lung cancer has been utilized to evaluate the system. The dataset has 154 nodule regions (abnormal) - where 100 are malignant and 54 are benign - and 92 non-nodule regions (normal). An Accuracy of 99.15% and 98.70% for classification have been achieved for normal versus abnormal and benign versus malignant respectively, this substantiate the capabilities of the approach presented in this paper.

Indexed Terms — Computer Aided Diagnosis, Lung Cancer, Statistical Feature Selection, Cluster k-Nearest Neighbor term

I. INTRODUCTION

Cancer is a disease that is referred to as the number one cause of death worldwide. According to the World Health Organization (WHO) [1], cancer was the reason for 7.4 million deaths (13% of all deaths) occurred in 2008. More than 70% of all cancer deaths occur in middle income nations. It is projected that deaths caused by cancer will be growing to reach 13.1 million by the year 2030 [1].

Lung cancer is a form of cancer that has become a significant cause of death globally [2]. According to GLOBOCAN Project – which is carried out by the International Agency for Research in Cancer (An agency of the WHO) – lung cancer has accounted for 12.7% of the total cancer incidents and 18.2% of the total mortality caused by cancer, both rates are more than any other form of cancer. Figure 1 shows a summary for the cancer incident and mortality rates in both sexes worldwide as it was published by GLOBOCAN project [3].

![Cancer incidence and mortality rates](image)

Figure 1. Cancer incidence and mortality rates [3]

Given that the cause of cancer remains unknown, early detection and treatment of cancer is the most promising ways to reduce the number of deaths. In order to diagnose cancer, medical imaging modalities such as Mammography, Computed Tomography and Magnetic Resonance Imaging, etc. have been developed to produce images of body organs that help to identify abnormalities. Radiologists and physicians depend on these images to diagnose diseases; yet, radiologists are incapable of indentifying subtle regions. For that, Computer Aided Diagnosis (CAD) systems have been designed to assist radiologist to recognize subtle regions and discover malignant (cancerous) cells [4].
CAD methods are generally designed to guide radiologists in diagnosing pulmonary nodules which necessarily suggest the presence of lung cancer. CAD systems typically incorporate four stages; preprocessing, feature extraction, feature selection and classification. Figure 2 shows a block diagram of the primary computer aided diagnosis system's elements.

![General block diagram for CAD system](image)

As this is a significant area of research, many CAD systems have been developed for lung cancer and for many other types of cancer. However, High false positives and detecting subtle regions are still among the issues that researchers are working to tackle. CAD systems with high false positive decrease the efficiency of the diagnosis [5], therefore, the number of false negatives should be decreased.

II. RELATED WORK

Large numbers of Computer Aided Diagnosis systems have previously been designed in the area of cancer diagnosis. For lung cancer, Hardie et al. [6] designed a CAD technique for lung nodule detection in chest radiography. The system had been evaluated with a data set which contains 167 chest radiography containing 181 lung nodules; the technique employed adaptive distance-based threshold algorithm for nodule segmentation. Thereafter, features were calculated from each nodule set which contains 167 chest radiography containing 181 nodules. The system attained an A value of 0.915. Lee et al. [8] presented a lung nodule detection employing an ensemble classifier assisted by clustering.

Lung images of 32 individuals which enclosed 5721 images have been used to examine the technique. The system attained sensitivity of 98.33% and specificity of 97.11%. Tan et al [9] developed a system for the detection and classification of lung nodules. Vessel enhancement filters and a computed divergence feature to locate the centers of the nodule clusters. After that, the detected nodules are subjected to a classification stage where invariant features are classified through a feature selective classifier based on Genetic Algorithm and Artificial Neural Network. The classifier is compared with fixed-topology neural network and support vector machine classifiers. Results show that the fixed-topology neural network performs better than the other classifiers where a detection rate of 87.5% is achieved. A system has been presented by Dehmeshki et al. [10] in order to detect lung nodules making use of shape-based genetic algorithm template matching (GATM). In this system, a preprocessing enhancement step had been conducted making use of spherical-oriented convolution-based filtering scheme, and a 3D geometric shape feature has been utilized to determine the fitness function. The system had been examined with a dataset of 70 CT images comprising 178 nodules. The system attained an accuracy rate of 90%. Another system has been proposed by Pereira et al [11] which presented a technique for lung nodule classification with multiclasses. The technique begin by filtering images utilizing a multi-scale filter bank which comprise of 36 filters at various scales and orientation, isotropic filters, gaussian and laplacian of gaussian. A multiple classifiers based on different multiple-layer perceptor (MLP) were utilized to classify the images. The system was examined with JSRT dataset, with 19 classifier combination; the Borda count combination attained 97% sensitivity with 43% error. A computerized method for lung cancer detection has also been designed by Sousa et al. [12]. The system has six stages, namely, thorax extraction, lung extraction, lung reconstruction, structure's extraction, tubular structures elimination, and false-positive reduction. Every single stage carries out a particular activity which leads to detecting lung nodules. The sensitivity attained was 84.84%, and specificity attained was 96.15%.

Although many CAD systems have been developed / proposed, common problems with present systems include the high number of false positives and false negatives. Thus, there exists a need to build up computer-aided diagnosis systems to assist in lung cancer diagnosis with better performance. This work is aimed at designing a system with sophisticated and efficient capabilities in classifying lung nodules and decreasing the number of false positives and false negatives. For that, the proposed system presents two stages of feature selections which results in selecting only relevant features that deliver much better performance. Furthermore, the Cluster k-nearest neighbor algorithm which merges two algorithms, namely, K means clustering and K-nearest neighbor is employed in this work, this classifier has been reported to deliver high accuracy [13].
III MATERIALS AND METHODS

A. Dataset

In this paper, the Japanese Society of Radiological Technology (JSRT) standard dataset of chest radiographs [14-15] is utilized to evaluate the method. The dataset includes 247 chest radiographs. Table I shows the distribution of the dataset images.

<table>
<thead>
<tr>
<th>JSRT DATASET IMAGES DISTRIBUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign</td>
</tr>
<tr>
<td>Abnormal (Nodule)</td>
</tr>
<tr>
<td>Normal (Non-nodule)</td>
</tr>
<tr>
<td><strong>Total Images</strong></td>
</tr>
</tbody>
</table>

Regions of 128x128 were extracted from the original 2048x2048 images. For the abnormal (nodule) regions, the regions have been extracted based on the provided coordinates of all the 154 nodules images. Figure 3 shows an example of one chest radiography.

B. Methods

This section presents a computer aided system for lung cancer diagnosis. The section is divided into parts; each one explains a processes. Figure 4 shows the components of the proposed system.

- Preprocessing
  Two preprocessing techniques are employed in this CAD system. The first technique is histogram equalization to enhance the contrast of the image so the intensity levels (which will be in the range [0 1]) would generate a flat histogram. Let $p(r_j), j = 1,2,\ldots,L$ represent a particular image’s histogram intensity level. The histogram equalization transformation shall be:

$$s_k = \frac{T(r_k)}{k} = \sum_{j=1}^{k} p(r_j) = \sum_{j=1}^{n} \frac{n_j}{n} \quad (1)$$

for $k = 1,2,\ldots,L$, where $r_k$ is the intensity value of the input image and $s_k$ is the intensity value of the output image [16]. Figure 5 demonstrates a lung sub-image before and after implementing histogram equalization.

- Feature Extraction using Wavelet Transform
  The transform of a signal is regarded as an alternative way of presenting the signal. And it also saves the important information contained in the signal. Wavelets employ various sets of fundamental functions to allow for the decomposition of continuous and discrete signals. Wavelet Transform provides a time-frequency description of the signal.

  The Continuous Wavelet Transform (CWT) is given in which $x(t)$ is the signal to be assessed.

  \[ \Psi(t) \text{ is the mother wavelet or the fundamental function. The mother wavelet is regarded as the origin of all the wavelet functions} \]
utilized in the transformation through translation (shifting) and scaling.

\[ X_{WT}(\tau,s) = \frac{1}{\sqrt{P}} \int x(t) \Psi^* \left( \frac{t-\tau}{s} \right) \, dt \]  

(2)

The Wavelet Transform is a multiresolution method that has been widely employed in feature extraction because of the excellent performance in contrast to other feature extraction techniques. An in-depth description of wavelets and its concept is available at mallat [17].

- Feature Selection

It is essential to lower the number of the coefficients generated by the wavelet [18]. This can be done by determining those coefficients which contain relevant information only.

The selected features are computed in two stages, which are calculating a statistical energy as well as a statistical metric. In both stages, only features less than a specific threshold are selected. Details of the two stages are as follows:

**Statistical Energy:** This can be calculated as follows:

\[ \text{Energy}_{\text{metric}}(k) = \frac{\sum |n_i^2(k)|}{\sum n_i(k)} \]  

(3)

where \( n_i^2(k) \) is feature \( k \) of image \( j \) in class \( i \).

After calculating the statistical energy, all energy values less than a certain threshold are removed.

**Statistical Metric:** This can be calculated as follows: Assume \( m_1, m_2 \) and \( m_3 \) are the mean of class 1, class 2, and class 3, respectively and \( m_r \) is the mean of all classes.

\[ \sum_{i=1}^{n} m_i(k) \]

Let \( m_r(k) = \frac{1}{n} \sum_{i=1}^{n} m_i(k) \)

\[ \text{var}_{\text{mod}}(k) = \frac{1}{n} \sum_{i=1}^{n} (m_i(k) - m_r(k))^2 \]  

(4)

where \( \sum (m_i(k) - m_r(k))^2 \) is not adequate to measure the classification contribution of the coefficients because it may provide similar values in the two cases. For that reason, there exists a need to present a different metric to measure the coefficients’ contribution. The metric is as follows:

\[ \text{Var}_{\text{mod}}(k) = \min_{i \neq j} \left\{ \frac{m_i(k) - m_j(k)}{\sqrt{S_i^2(k) + S_j^2(k)}} \right\} \]  

(5)

where \( S_i \) is the statistical metric of class \( i \), \( m_i \) is the mean of class \( i \), and \( n_i \) is the number of classes. \( S_i \) can be calculated as follows:

\[ S_i^2 = \sum_{k=1}^{n_i} (x_{ij}(k) - m_i(k))^2 \]  

(6)

where \( i = 1, 2, 3, \ldots, n, \) and \( n_i \) is the number of the features in class \( i \).

- Classification

A classifier which is a combination of K-means modified algorithm and K-Near Neighbor (K-NN) is used within this study. This classifier was introduced by Samir Brahim [13]. The classifier is called Cluster K-Nearest Neighbor (C-K-NN). The algorithm is explained further below:

Every single class, \( C_i \) should become a cluster to various subclasses, \( C_{ij} \), with \( 1 \leq j \leq 1 \), and each subclass will be represented by its mean, \( \mu_{ij} \). Hence, the cluster analysis determines a set of groups, which will decrease the within-group variation and increase the between-group variation. K-means cluster algorithm is applied for every class for clustering purposes then both the number of subclasses for every class and the initial k-vectors to initialize the K-means cluster algorithm are defined to find the ideal number of the subclasses. The number of subclasses is iterated beginning with 1 and the iteration process should stop by the following two conditions:

- All the representatives, \( \mu_{ij} \), should be close with respect to the metric \( d \) of their classes, \( C_i \), (i.e., if we classify all the representatives, \( \mu_{ij} \), we have found 100% accuracy). If there are some misclassifications of \( \mu_{ij} \), we have to decrease the parameter \( \alpha \) by multiplying it with another factor, \( \alpha' \), which is less than 1.

- The statistical metric of each class \( C_p \) var\_i does not decrease significantly in comparison to the previous iteration.

We may use \( \frac{\Delta \text{Var}}{\text{var}_{\text{ini}}} \leq \alpha \) as a criterion to quantify if there is a decrease or it still approximately remains constant. In certain cases, it is better to stop the iteration if the condition \( \frac{\Delta \text{Var}}{\text{var}_{\text{ini}}} \leq \alpha \) has been checked twice or more (i.e., after which the statistical metric will be smoothened).

For initialization of the K-means cluster algorithm in general, we choose k-vectors, which belong to our classes’ data. This will therefore make the algorithm unstable in the sense of the final variance:

\[ \text{var} \ C_i = \sum_{j=1}^{m_i} \text{var} \ C_{ij} \]

which depends on the initial vectors.

From here, the question "How to choose the initial vectors in order to find a minimal variance?" arises. To answer this, in this paper, we have developed two algorithms: the Hierarchical Near-to-Near and Hierarchical Near-to-Mean algorithms, which might require some modification for different applications.

**Hierarchical Near-to-Near Algorithm**

This algorithm consists of calculating the distance \( d(x_{in}, x_{jm}) \) for all \( x_{in} \in C_i \) and starting to cluster our classes to \( N_i - 1 \) subclasses where \( \text{card}(C_i) = N_i \). We put the two closest data in the same subclasses \( C_{i,1} = \{ x_{in}, x_{in,mo} \} \).
where \( n, m \neq m \) and we put each of the other data in separate subclasses, 
\[ c_{ij} = \{ x_{ij}, \forall j \in \{ 1, \ldots, N \} - \{ n, m \} \}. \]

In the next step, the following index \( n_1 \) and \( m_1 \) are considered for which

\[ \min_{n \neq m} d(x_{in}, x_{im}) = d(x_{i,n1}, x_{i,m1}), \]

\[ (n, m) \neq (n_0, m_0) \]

If \( x_{i,n1} \) and \( x_{i,m1} \) belong to the same subclass, \( c_{ir} \), we shall split this subclass into two other subclasses:

\[ c_{ir + 1} = c_{ir} - \{ x_{i,n1}, x_{i,m1} \} \]  \( (7) \)

\[ c_{ir} = \{ x_{i,n1}, x_{i,m1} \} \]  \( (8) \)

In the case where \( x_{i,n1} \) and \( x_{i,m1} \) belong to two different subclasses, \( c_{ir1} \) and \( c_{ir2} \), respectively, we shall put \( x_{i,n1} \) in the subclass, \( c_{ir2} \), if \( \text{card}(c_{ir2}) > \text{card}(c_{ir1}) \) and we may put \( x_{i,m1} \) in the subclass, \( c_{ir1} \), if \( \text{card}(c_{ir2}) \leq \text{card}(c_{ir1}) \). Indeed, to get the cardinality of the set we may use the distance between the vectors to the set as \( d(\text{vector,mean of seat}) \). When we obtain \( k \) subclasses, we stop the iteration, and our initial \( k \)-vectors will be the mean of each subclass.

**Hierarchical Near-to-Mean Algorithm**

This algorithm is almost the same as the Hierarchical Near-to-Near algorithm, except we will deal with the mean of subclass \( \hat{c}_{ir} \) in the processing. We start by splitting our class \( C_i \) into two subclasses:

\[ c_{i,1} = \{ x_{i,n0}, x_{i,m0} \} \]  \( (9) \)

and

\[ c_{i,2} = \{ x_{ij} \mid j \notin \{ n_0, m_0 \} \}, \]  \( (10) \)

where

\[ d(x_{i,n0}, x_{i,m0}) = \min_{n \neq m} d(x_{i,n}, x_{i,m}) \]

We then updated our classes, \( C_i \), by replacing \( x_{i,n0} \) and \( x_{i,m0} \) with their average,

\[ c_{i,1} = \{ \ldots, x_{i,n0} - 1 \text{SD} x_{i,n0} + 1, \ldots, x_{i,m0} - 1 \text{SD} x_{i,m0} + 1, \ldots \} \]

Where

\[ \text{SO} = (x_{i,n0} + x_{i,m0})/2 \]

Next, we consider \( x_{i,n1} \) and \( x_{i,m1} \) as:

\[ d(x_{i,n1}, x_{i,m1}) = \min\{ d(x_{i,n1}, x_{i,m1}) \mid d(x_{i,n}, x_{i,m}) \neq 0 \} \]

We then replace \( c_{i,1} \) and all the data in \( c_{i,1} \) that are equal to \( x_{i,n1} \) or \( x_{i,m1} \) by \( S^1 \), which is the mean of the union of the two subclasses where \( x_{i,n1} \) and \( x_{i,m1} \) belong to:

\[ S_i = \frac{c_{n1} x_{i,n1} + c_{m1} x_{i,m1}}{c_{n1} + c_{m1}} \]  \( (11) \)

Where \( c_{n1} \) is the number of the repetitions of \( x_{i,n1} \) inside of \( c_{i,1} \), and \( c_{m1} \) is the number of the repetitions of \( x_{i,m1} \) inside of \( c_{i,1} \). Our algorithm stops once the number of distinct vectors inside of \( c_{i,1} \) is equal to \( k \). Our classification algorithm does not need to keep all the data, only the average of each subclass. This is the outstanding feature of this new clustering. To classify a new data or vector \( x \), we use k-NN algorithm, i.e., we assign \( x \) to the class \( C_i \) for which:

\[ \hat{i} = \text{arg} \min_{i,j} d(x, \mu_{ij}) \]  \( (12) \)

Where \( \text{arg} d(x, \mu_{ij}) = i_0 \).

**IV RESULTS AND DISCUSSION**

In this paper, two experiments have been carried out. The first experiment is to classify normal from abnormal images and the second experiment to classify benign from malignant images. The following sub-sections report the results obtained for both experiments.

**A. Classification Performance**

In this section, the performance of the methods on classifying normal from abnormal images as well as benign from malignant images is demonstrated. The accuracy, false negatives and false positives are shown in table II when Haar Wavelet was utilized for the Normal vs. Abnormal experiment. The performance reached 99.15% accuracy rate with zero false positives and zero false negatives with Haar Wavelet levels 3, 4, 5 and 6.

**TABLE II**

<table>
<thead>
<tr>
<th>Performance</th>
<th>Haar wavelet function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>Performance</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.9829</td>
</tr>
<tr>
<td>False Negative</td>
<td>1</td>
</tr>
<tr>
<td>False Positive</td>
<td>0</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.9915</td>
</tr>
<tr>
<td>False Negative</td>
<td>3</td>
</tr>
<tr>
<td>False Positive</td>
<td>0</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.9915</td>
</tr>
<tr>
<td>False Negative</td>
<td>4</td>
</tr>
<tr>
<td>False Positive</td>
<td>0</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.9915</td>
</tr>
<tr>
<td>False Negative</td>
<td>5</td>
</tr>
<tr>
<td>False Positive</td>
<td>0</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.9915</td>
</tr>
<tr>
<td>False Negative</td>
<td>6</td>
</tr>
<tr>
<td>False Positive</td>
<td>0</td>
</tr>
</tbody>
</table>

Furthermore, when the Haar Wavelet function was applied to the benign against malignant experiment, the performance reached 98.70% accuracy with zero false negatives and false positives at the function levels 1, 2, 4 and 6. Table III reports the results obtained at different function level.

**TABLE III**

<table>
<thead>
<tr>
<th>Performance</th>
<th>Haar wavelet function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>Performance</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.9870</td>
</tr>
</tbody>
</table>
B. Usefulness of Feature Selection

Feature selection has an influence on the performance of our computer aided diagnosis systems. It works to eliminate repetitive and needless information which was extracted during the feature extraction stage. Table IV shows how our two-step feature selection methods have reduced the number of coefficients extracted by the haar wavelet.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Normal Vs. Abnormal</th>
<th>Benign Vs. Malignant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haar Wavelet Function Level</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Initial coefficients</td>
<td>16384</td>
<td>16384</td>
</tr>
<tr>
<td>After Energy calculation</td>
<td>6562</td>
<td>11470</td>
</tr>
<tr>
<td>After Statistical Metric calculation</td>
<td>291</td>
<td>291</td>
</tr>
</tbody>
</table>

To show the effectiveness of the feature selection step, figure 6 shows an illustration when the haar wavelet with level 3 was used on the Normal Vs. Abnormal experiment. In this figure, the influence of the number of coefficients is noticed where the highest accuracy with minimal false negatives and false positives occurs at the ideal number of features (coefficients).

C. Cross Validation

Cross validation is a statistical approach to assess learning algorithms. It operates by splitting the data into two parts, one is used to train and the other to test. Cross validation is available in many strategies; in this paper, the k-fold strategy is utilized to validate the system. A 10-fold cross validation was performed on the Haar Wavelet with the highest performance for each experiment. Table V reports the results for the cross validation. The averaged accuracy obtained after the cross validation was 98.75 % for the Normal Vs. Abnormal experiment; while it was 98.70 for the Benign Vs. Malignant experiment. This shows the effectiveness of our computer aided diagnosis system and its ability to diagnose lung nodules.

<table>
<thead>
<tr>
<th>k-folds</th>
<th>Normal Vs. Abnormal</th>
<th>Benign Vs. Malignant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fold 1</td>
<td>100</td>
<td>98.70</td>
</tr>
<tr>
<td>Fold 2</td>
<td>95.83</td>
<td>97.40</td>
</tr>
<tr>
<td>Fold 3</td>
<td>95.83</td>
<td>100</td>
</tr>
<tr>
<td>Fold 4</td>
<td>100</td>
<td>98.70</td>
</tr>
<tr>
<td>Fold 5</td>
<td>100</td>
<td>98.70</td>
</tr>
<tr>
<td>Fold 6</td>
<td>100</td>
<td>98.70</td>
</tr>
<tr>
<td>Fold 7</td>
<td>100</td>
<td>98.70</td>
</tr>
<tr>
<td>Fold 8</td>
<td>100</td>
<td>98.70</td>
</tr>
<tr>
<td>Fold 9</td>
<td>100</td>
<td>98.70</td>
</tr>
<tr>
<td>Fold 10</td>
<td>95.83</td>
<td>98.70</td>
</tr>
<tr>
<td>Average</td>
<td>98.75</td>
<td>98.70</td>
</tr>
</tbody>
</table>

V CONCLUSION AND FUTURE WORK

The paper introduced a computer aided diagnosis system for lung nodules based on cluster k nearest neighbor algorithm and statistical feature selection methods. The results shown in the previous section proves the capabilities of the CAD system in lung cancer classification (Normal Vs. Abnormal and Benign Vs. Malignant). A 99.15 % and 98.70 % accuracy rates have been achieved for the normal vs. abnormal and benign vs. malignant experiments respectively with zero false positives and false negatives. The methods were verified with Cross Validation and the results obtained prove the advantage of the proposed system. Further experiments will be performed to examine the methods in other datasets with different modalities (i.e. CT scan) as well as other forms of cancer.

ACKNOWLEDGMENT

The Authors would like to thank Alfaisal University, Saudi Arabia, for their support with the Internal Research Grant (IRG 2012) Grant Number: IRG 419191811126. The authors would also like to thank Universiti Teknologi
PETRONAS, Malaysia, for supporting this work with the Graduation Assistantship Scheme.

REFERENCES


Hamada R. H. Al-Absi has obtained his Bachelor of Technology in Information & Communication Technology and Master of Science in Information Technology in 2009 and 2010 respectively from Universiti Teknologi PETRONAS, Perak, Malaysia. Currently he is pursuing his PhD at the department of Computer and Information Sciences at Universiti Teknologi PETRONAS, Malaysia. His research interests are in pattern recognition and machine learning.

Brahim Belhaouari Samir received his Master degree from INP/EENSET Toulouse, France 2000 and Ph.D. in Stochastic Processes from EPFL, Lausanne Switzerland in 2006. After completing the Ph.D. degree, he held a Postdoctoral at Ecole Polytechnique Federale De Lausanne, Lausanne, Switzerland. Now, he is an assistant professor at the College of Science and General Studies, Alfaisal University, Kingdom of Saudi Arabia. His main research interests are in data analysis, statistical signal processing, and pattern recognition.

Suziah Bt Sulaiman is a senior lecturer at Computer and Information Sciences Department at University Teknologi PETRONAS.