Normal Mammogram Detection Based on Local Probability Difference Transforms and Support Vector Machines

Werapon CHIRACHARIT, Student Member, Yajie SUN, Pinit KUMHOM, Nonmembers, Kosin CHAMNONGTHAI, Member, Charles F. BABBS,†††, and Edward J. DELP,†††, Nonmembers

SUMMARY Automatic detection of normal mammograms, as a “first look” for breast cancer, is a new approach to computer-aided diagnosis. This approach may be limited, however, by two main causes. The first problem is the presence of poorly separable “crossed-distributions” in which the correct classification depends upon the value of each feature. The second problem is overlap of the feature distributions that are extracted from digitized mammograms of normal and abnormal patients. Here we introduce a new Support Vector Machine (SVM) based method utilizing with the proposed uncrossing mapping and Local Probability Difference (LPD). Crossed-distribution feature pairs are identified and mapped into a new features that can be separated by a zero-hyperplane of the new axis. The probability density functions of the features of normal and abnormal mammograms are then sampled and the local probability difference functions are estimated to enhance the features. From 1,000 ground-truth-known mammograms, 250 normal and 250 abnormal cases, including spiculated lesions, circumscribed masses or microcalcifications, are used for training a support vector machine. The classification results tested with another 250 normal and 250 abnormal sets show improved testing performances with 90% sensitivity and 89% specificity.

key words: breast cancer, mammogram, computer-aided diagnosis (CAD), second opinion, automated radiographic reading

1. Introduction

Breast cancer is the second major cause of death from cancer among women aged 15 to 54 [1]. It is a serious health problem world-wide. More than 20% of women having breast cancer die from it [2]. Screening x-ray mammography is currently the most effective way to detect the cancer in an early stage and reduce mortality.

There have been several approaches for the use of computer-aided diagnosis (CAD) for abnormalities as a second opinion to help radiologists in reading mammograms. Shen et al. [3] proposed a set of shape features to measure the roughness of a segmented individual calcification. El-Naqa et al. [4] investigated support vector machine based approach for detection of microcalcifications. Comer et al. [5] developed a statistical segmentation method based on Markov random field model to detect mammographic masses. Liu et al. [6] proposed a multiresolution scheme for detection of spiculated lesions by using a linear phase non-separable two-dimensional wavelet transform. Such CAD systems can increase reader sensitivity or true positive fraction (TPF) by 5% to 15%. TPF means “abnormal images” which are classified as “abnormal”. However, among positive readings, only 15% to 34% are histologically proven true positive malignancies [7]–[9]. Because normal mammograms constitute the majority of screening mammograms and all such abnormalities are just the minority part. It makes a high risk of false positive diagnosed by finding abnormalities. Improvement is needed to increase reader specificity or to reduce the false positive fraction (FPF) or number of “normal images” misclassified as “abnormal”.

One approach is to better distinguish normal mammograms automatically. Accurate normal mammogram detection helps to minimize false negative fraction (FNF), or number of “abnormal images” which are classified as “normal”. The true negative fraction (TNF), number of “normal images” misclassified as “normal”, is maximized in order to miss the fewest abnormalities [10], [11], since

$$TPF + FNF = 1.$$  \hspace{1cm} (1)

Accurate normal detection also helps to reduce false alarms, one must increase the true negative fraction (TNF), since

$$FPF + TNF = 1.$$  \hspace{1cm} (2)

Conceptually, one could divide a CAD system into two processes. The first process is normal detection known as “first look” or “pre-screening” to find true normals. The second process is cancer detection which could be used as a “second look” mode to find true abnormalities. An overall CAD system might use a combination of first look and second look methods as shown in Fig. 1 (left).

- A first look method provides normal mammogram detection with minimal false negatives. The objective is 100% sensitivity, so as not to miss any abnormalities, but highest specificity. It is a classifier trained

Manuscript received April 10, 2006.
Manuscript revised July 17, 2006.
†The authors are with the Department of Electronics and Telecommunication Engineering, King Mongkut’s University of Technology Thonburi, Prachu-ubthit Road, Bangkok 10140, Thailand.
††The authors are with the School of Electrical and Computer Engineering, Purdue University, 465 Northwestern Avenue, West Lafayette, IN 47907–2035, USA.
†††The authors are with the Department of Basic Medical Sciences, School of Veterinary Medicine, Purdue University, West Lafayette, IN 47907–1285, USA.
*This work was supported in part by a grant from the Ministry University Affair, Thailand.
a) E-mail: s4510108@st.kmutt.ac.th
DOI: 10.1093/ietisy/e90–d.1.258

Copyright © 2007 The Institute of Electronics, Information and Communication Engineers
by features based on normal mammogram characteristics. If the detection result is negative, the patient can go home. The risk of unnecessary biopsies is also reduced. If the result is positive, the images will be sent to radiologists and possibly to a second look system.

- A second look method or local cancer detection would be performed after human reading.

First look and second look strategies are also shown in Fig. 1 (right). It shows how a receiver operating characteristic (ROC) curve may be generated from the distribution of normal and abnormal mammograms and shows the relationship of true negative ($TN$), false positive ($FP$), true positive ($TP$) and false negative ($FN$). The classification threshold differs for “first look” with 100% sensitivity and “second look” methods.

There are little works on normal mammogram. Sahiner et al. classified mass from normal tissue by a convolution neural network that operates directly on mass and normal images [12]. Heine et al. identified normal mammogram using multiresolution statistical analysis [13]. Mini and Thomas also used wavelet transform and a simple linear marking to subtract normal tissue [14]. Liu and Sun et al. characterized normal mammogram features and proposed a full-field mammogram detection [15]–[17]. In this paper we present an improved classification method based on normal features from Liu and Sun et al. We focus on “hard-to-classify” cases, which may in part be hard to classify owing to pairs of crossed-distribution features and noisy features. Using such crossed features, normal mammograms and abnormal mammograms are not separable in Euclidian space. We introduce a change of variables for the crossed feature pairs to create new hybrid features space, for which an optimized decision hyperplane can be defined. We also introduce a local probability difference (LPD) transformation to create new and less noisy features. These approaches are based on a support vector machine (SVM) classifier.

2. Normal Mammogram Analysis

Normal breast tissue consists of fat, glandular elements, blood vessels and fibrous tissues, appearing as diffuse amorphous clouds of density with indistinct borders [18], as shown in Fig. 2 (DDSM [19]). Normal areas are usually observed to contain linear structures that raise the local brightness in the mammogram image. However, contrast of these lines is depended on breast tissue density. No single feature allows separation of normals from abnormals. In this paper, we will use the feature set described in [16], [17]. A complete description of the feature set is provided in the Appendix.

2.1 Normal Mammogram Features

First, mammograms were enhanced based on $h_{int}$ representation that estimates the height of non-fatty or interesting tissue in the breast for each pixel computed from an x-ray energy and attenuation coefficients of the tissues [20]. Each full-field mammogram, approximately $5,000 \times 3,000$ pixels, is analyzed and covered by overlapped moving blocks with size $512 \times 512$ pixels. Then four sets of 86 features are extracted from the mammograms, including curvilinear features, gray level co-occurrence matrix texture features, Gabor features and multi-resolution statistical features [16], [17]. The description of normal mammogram feature set is in the Appendix.

2.1.1 Curvilinear Features

Because normal mammograms are likely line patterns of ductal structures of the breast tissue, we used our line detection algorithm based to extract curvilinear features [15]. Curvilinear structures are not exactly straight lines and suitable to breast structures. The curvilinear markings tend to radiate from the nipple toward the chest wall. The quasi-linear curves in different widths and angles are extracted. There are 18 curvilinear features extracted for each mammogram’s region, capturing the statistical nature of line pixels, including Line Pixel Count, Diagonal Half Line Pixel.
Fig. 3 Examples of crossed distributions of two features in “×”-shaped distribution (top) and uncrossing crossed-features (bottom). Normal and abnormal mapped features can be separated by a hyperplane near $\phi = 0$.

Counts, Half Ratios, means, standard deviations and entropy of Angles, Local Lines, Local Angles and Local Binary Patterns [17].

2.1.2 Gray Level Co-occurrence Matrix Texture Features

Texture is usually used to represent the statistical spatial arrangement of the pixels in the image [21]. From two-dimensional spatial dependence matrix (GLCM: Gray Level Co-occurrence Matrix), we extracted 16 features, including Energy, Entropy, and their sums and differences, Maximum Probability, Correlation, Diagonal Correlation, Inertia, Homogeneity, $H_{xy}$, sums of Variance, Shade and Prominence [17].

2.1.3 Gabor Features

Gabor filters have been successfully used to provide simultaneous localization in both the space and frequency domain [22], [23]. We used four orientations and four scales for our Gabor filter-bank and extracted 32 Gabor features, including means and standard deviations of Energy for each region images [17].

2.1.4 Multiresolution Statistical Features

The last set of our features was computed from nonlinear wavelet images, decomposed by Quincunx wavelet transform [24], [25]. There are five features obtained from each subimage and 20 features were extracted, including Mean, Variance, Compactness, Fractal Dimension and Entropy of each subband [17].

2.2 Useless Crossed-Features Problem

From the above 86 features, there is the presence of crossed distributions that are hard to classify in any one dimension of feature space. Examples of crossed distribution feature pairs are shown in Figs. 3 and 4 (80 normal and 80 abnormal samples). For example, in Fig. 3, the shapes of feature distributions of the eleventh feature (Standard Deviation of local line) versus the fifty-fifth feature (Standard Deviation of Gabor filtered image) in the top-left figure and the distribution of the third feature (Diagonal Half Line Count) versus the eighty-third feature (Mean of wavelet transformed image) in the top-right figure, extracted from normal and abnormal mammograms, are “×”-like.
While, in Fig. 4, the distribution of the fourteenth feature (Standard Deviation of local angle) versus the eighty-second feature (Entropy of wavelet transformed image) in the top-left figure and the distribution of the seventh feature (Half Ratio) versus the twelfth feature (Entropy of local line) in the top-right figure are similar to Fig. 3, but more likely to be “+” shape.

### 2.3 Feature Distributions Problem

Another problem in our set of 86 features is confounding distributions of the features for both normal and abnormal training cases. For example, Fig. 5 shows examples of probability density functions or histograms of features extracted from 80 normal mammograms and 80 abnormal mammograms, represented by “●” and “×” respectively. A different sampling grid for two distributions is used, because of different range of raw data. The same sampling is performed later after the distribution functions are obtained. The top-left figure shows samples of the third feature (Diagonal Half Line Pixel Count), showing obviously a good separation of normal class and abnormal class. However, in the top-right
figure that is samples of the eighty-second feature (Entropy of wavelet transformed image), it shows much confusing of normal and abnormal features. These noisy features, however, contain useful information revealed by the local probability difference transformation, described in Sect. 3.

3. Methods

As we have described the two main problems of classification of normal mammograms with “crossed distribution feature pairs” and “noisy features”, we now introduce a new method to fix “crossed distribution” and “noise” problems with an uncrossing method and local probability difference transformation based support vector machine classifier (LPD-uncrossed-feature based SVM), compared with a conventional SVM learning method and other classifier such as a neural network. The conventional SVM learning is explained first.

3.1 A Support Vector Machine

Support vector machine is a learning algorithm that is opti-
mized or to find the largest separation margin between two classes [26], [27]. Let \( \mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in})^T \) is the \( N \)-dimensional input feature vector for \( i \)th training data and the class label \( y_i = \{+1, -1\} \) for the abnormal and normal case, respectively. The ground-truth-known data set of mammograms with \( l \)-samples is

\[
(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l) \in \mathbb{R}^N \times \{\pm 1\},
\]

where \( i = 1, 2, \ldots, l \). A support vector machine is obtained by optimization of the linear functional decision hyperplane. In a dual form, the learning rule or support vector machine relaxed constraint is;

maximize

\[
W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j)
\]

subject to

\[
0 \leq \alpha_i \leq C
\]

and

\[
\sum_{i=1}^{l} \alpha_i y_i = 0,
\]

where \( \alpha_i \) is the Lagrange multipliers, \( K() \) is kernel function, and \( C \) is a constant for soft margin technique. After training, the decision function is in the form of

\[
f(\mathbf{x}) = \text{sign} \left( \sum_{i=1}^{SV} \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b \right),
\]

where \( b \) can be derived from the primal form and \( SV \) is a number of “support vectors” such that \( \alpha_i \neq 0 \) and \( SV \leq l \). Typically, only a few training samples will be support vectors and therefore its computation is reduced.

For nonlinear separation, the training input becomes a nonlinear-kernel mapped feature vector in higher-dimensional feature space.

Kernel functions are based on transformation or mapping \( \phi() \) of the input features into higher-dimensional feature space [4]. The kernel functions are defined by the transformation of two vectors as \( K(\mathbf{x}_1, \mathbf{x}_2) = \phi(\mathbf{x}_1) \cdot \phi(\mathbf{x}_2) \). In our experiments, we used three standard kernel functions for training and testing the features.

- **Linear kernel**
  
  \[
  K_1(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^T \cdot \mathbf{x}_2
  \]

- **Polynomial kernel**
  
  \[
  K_2(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_1^T \cdot \mathbf{x}_2 + c)^d
  \]

- **Gaussian radial basis function kernel**
  
  \[
  K_3(\mathbf{x}_1, \mathbf{x}_2) = e^{-\frac{||\mathbf{x}_1 - \mathbf{x}_2||^2}{2\sigma^2}},
  \]

where \( c \) is a constant, \( d > 0 \) is a constant that defines polynomial kernel order, and \( \sigma > 0 \) is a constant that defines the Gaussian kernel width.

3.2 An Uncrossing Method for Crossed Features

Normal and abnormal distributions of many pairs of features \( \mathbf{x}^{(m)} \) and \( \mathbf{x}^{(n)} \), where \( m, n = 1, 2, \ldots, N \) and \( m \neq n \), are poorly separable in Euclidian space if they have a crossed pattern. The new uncrossed-feature based SVM learning system is designed to separate these crossed features.

3.2.1 Identification of Crossed Features

To identify the feature pairs of crossed distributions in a set of \( N \) input features, we examined all \( N(N - 1)/2 \) pairs for all normal and abnormal training data. For each pair including \( m \)th feature and \( n \)th feature, we computed the correlation coefficients \( R(m, n) \) between features \( \mathbf{x}^{(m)} \) and \( \mathbf{x}^{(n)} \) for all normal training data and for all abnormal training data. Then we can identify the confounding features by noting that one, but not both, correlation coefficients is negative. In particular, feature pair \((\mathbf{x}^{(m)}, \mathbf{x}^{(n)})\) is crossed if multiplied correlation coefficient

\[
R_{normal}(m, n) \times R_{abnormal}(m, n) \leq -\epsilon,
\]

where \( 0 < \epsilon < 1 \). A greater value of \( \epsilon \) indicates a more distinct crossed pattern.

Some distributions are crossed in the shape of a “+” sign, rather than an “x”. These feature pairs can be found by the product of correlation coefficients after rotation of the axes through an angle \( \theta = 45^\circ \). After rotation, \((\mathbf{x}^{(m)}, \mathbf{x}^{(n)}) = (\mathbf{x}^{(m)} \cos \theta - \mathbf{x}^{(n)} \sin \theta, \mathbf{x}^{(m)} \sin \theta + \mathbf{x}^{(n)} \cos \theta)\) [28]. The new rotated distributions become “x”-like and are used for subsequent testing using (9).

3.2.2 Uncrossing Mapping

We have designed a new transformation for feature pairs that can separate crossed distribution. The normal and abnormal can be separated linearly by a zero hyperplane perpendicular to a new axis. The uncrossing transform is designed by a solution of the zero-plane separating six crossed-distribution sample points from the feature coordination, \((0, 0), (\frac{1}{2}, \frac{1}{2})\) and \((1, 1)\) for the first class and \((0, 1), (\frac{1}{2}, \frac{1}{2})\) and \((1, 0)\) for the second class, shown in (10),

\[
\phi(\mathbf{x}^{(m)}, \mathbf{x}^{(n)}) = -1 + 2\mathbf{x}^{(m)} + 2\mathbf{x}^{(n)} - 4\mathbf{x}^{(m)}\mathbf{x}^{(n)},
\]

where all values of each feature \( \mathbf{x} \) are normalized and ranged from 0 to 1 before mapping.

In Fig. 3, the left figures are examples of “x”-shaped distribution with multiplied correlation coefficient \(-0.11\) and \(-0.14\) for the right figures. The below figure shows separation of mapped normal and abnormal features in three dimensional space. They can be separated by nearly a zero hyperplane for exactly “x”-shaped distribution and by a hyperplane near \( \phi = 0 \) for likely “x”-shaped distribution.

Figure 4 are examples of a “+”-shaped distribution. The left figures are examples with multiplied correlation coefficient after rotation \(-0.21\) and \(-0.42\) for the right figures.
3.3 Local Probability Difference Features

A new LPD-feature based SVM learning is a support vector machine that uses proposed local probability difference transformed features as input features. The overlapped input features are enhanced or preprocessed before training the classifier, based on estimating the probability density distribution of each feature for known normals and known abnormalities.

3.3.1 Local Probability Difference Transformation

The probability density function \( (pdf) \) is estimated from the ground-truth-known training feature set. For samples of normal \((-1)\) and abnormal \((+1)\) features, we define the local probability for any \(n\)th feature of all abnormal samples:

\[
p^+(x^{(n)}) = \frac{pdf^+(x^{(n)})}{pdf^+(x^{(n)}) + pdf^-(x^{(n)})}
\]

and for all normal samples:

\[
p^-(x^{(n)}) = \frac{pdf^-(x^{(n)})}{pdf^-(x^{(n)}) + pdf^+(x^{(n)})},
\]

or for a two class discrimination, \(p^-(x^{(n)}) = 1 - p^+(x^{(n)})\). These local probabilities tend to capture the information about the relative position of normal and abnormal points of the input vector \(x\) in the feature space. Therefore we can obtain local probability difference (LPD) as in (11),

\[
\Delta p(x^{(n)}) = \frac{pdf^+(x^{(n)}) - pdf^-(x^{(n)})}{pdf^+(x^{(n)}) + pdf^-(x^{(n)})}.
\]

The local probability difference has value 1.0, -1.0, and 0 when the abnormal class, normal class, and two-class equal likely are almost certain, respectively. Therefore, each of our features can be transformed into a new set of local probability difference transformed features, \(\phi(x) = (\Delta p(x^{(1)}), \Delta p(x^{(2)}), \ldots, \Delta p(x^{(N)}))\), used as an input vector for training and testing SVM.

A smooth polynomial of the third order \((ax^3 + bx^2 + cx + d\) where \(a, b, c\) and \(d\) are constants) is fit to the raw probability difference values to estimate each new LPD feature. This process is shown in Fig. 5. The left figures are examples of the feature with ten sampled points and the right figures are examples of the feature with fifteen sampled points. The local probability difference function higher than around zero is assumed to be abnormal class and the value lower than zero is assumed to be normal class.

3.4 A New LPD-Uncrossed-Feature Based SVM Learning System

The approach for enhanced SVM performance is a new LPD-uncrossed-feature based SVM shown in Fig. 6. It is a combination of the uncrossing mapping for any crossed feature pairs and the local probability difference transformation for any features. The features are selected by adaptive floating search (AFS) strategy [29]. Unlike other methods for reducing the dimensionality, e.g. principal component analysis (PCA), linear discriminant analysis (LDA), independent component analysis (ICA), a small feature set is selected without any transformation to preserve all useful information of normal mammograms. The significant features are tested for crossed distribution patterns using \(\epsilon = 0.1\). The crossed pairs are mapped and transformed into the local probability difference space. For each pair of crossed distribution features one new transformed feature, \(\phi\), is generated. The local probability difference features are classified by a support vector machine.

4. Experimental Results

All digital mammograms were taken from the Digital Database for Screening Mammography (DDSM) have known ground-truth [19]. The ground-truth known 1,000 mammograms are from 319 patients aged 31 years to 89 years old. The mammograms include both medio-lateral oblique (MLO) views and cranio-caudal (CC) views for both left and right breast images. There are 250 normal and 250 abnormal mammograms for training. A subsequent testing scheme also utilized 250 normal and 250 abnormal mammograms.

The breast density of normal mammograms is rated from 1 (fatty) to 4 (dense) by an experienced mammographer, as shown in Fig. 2. Abnormal mammograms are included with type 1 (spiculated lesions), type 2 (cirscurbed masses), type 3 (microcalcifications) and type 4 (calcifications and masses). We trained and tested the system covering all normal densities and all abnormal types.
### Table 1  
Results of training and testing LPD-uncrossed-features based SVM learning system and comparison of area under ROC curves.

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Abnormal</th>
<th>Correct Classification Rate</th>
<th>$A_Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TN</td>
<td>FP</td>
<td>TP</td>
<td>FN</td>
</tr>
<tr>
<td>Kernel#1 Training</td>
<td>200</td>
<td>40</td>
<td>250</td>
<td>0</td>
</tr>
<tr>
<td>Testing</td>
<td>196</td>
<td>54</td>
<td>202</td>
<td>48</td>
</tr>
<tr>
<td>Kernel#2 Training</td>
<td>205</td>
<td>40</td>
<td>250</td>
<td>0</td>
</tr>
<tr>
<td>Testing</td>
<td>196</td>
<td>52</td>
<td>221</td>
<td>29</td>
</tr>
<tr>
<td>Kernel#3 Training</td>
<td>215</td>
<td>27</td>
<td>250</td>
<td>0</td>
</tr>
<tr>
<td>Testing</td>
<td>199</td>
<td>51</td>
<td>213</td>
<td>37</td>
</tr>
<tr>
<td>Neural Network</td>
<td>216</td>
<td>34</td>
<td>250</td>
<td>0</td>
</tr>
<tr>
<td>Testing</td>
<td>199</td>
<td>51</td>
<td>213</td>
<td>37</td>
</tr>
<tr>
<td>Binary Tree</td>
<td>221</td>
<td>21</td>
<td>250</td>
<td>0</td>
</tr>
<tr>
<td>Testing</td>
<td>212</td>
<td>38</td>
<td>204</td>
<td>46</td>
</tr>
<tr>
<td>LPD Uncrossed</td>
<td>230</td>
<td>18</td>
<td>250</td>
<td>0</td>
</tr>
<tr>
<td>Testing</td>
<td>222</td>
<td>28</td>
<td>224</td>
<td>26</td>
</tr>
</tbody>
</table>

Examples of uncrossing crossed distribution feature pairs have been shown in Figs. 3 and 4. For all our 86 features, we have 3,655 feature pairs. We used the small value of $\epsilon = 0.1$ to identify all possible crossed distribution feature pairs. After exhaustive searching for crossed distributions in 250 training normal mammograms and 250 training abnormal mammograms, there were total 172 crossed-distribution feature pairs involving 77 features. There were 87 “x” crossed distribution feature pairs and 85 “+” crossed distribution feature pairs. From 20 AFS best selected features (190 feature pairs), there were 12 crossed distribution feature pairs to be uncrossed and their local probability difference functions were obtained.

Examples of the results of estimating the local probability difference function are shown in Fig. 5. The functions are obtained with cubic fitting with 10 sampled points and 15 sampled points, respectively. The purpose of this method is to enhance separability of normal versus abnormal samples.

The proposed method tries to improve a specificity from the previous methods, while keeping a high sensitivity at the same rate. Performing 20 selected features, the results of training and testing proposed method are shown in Table 1. The results are compared to standard support vector machine with kernel#1 (linear), kernel#2 (polynomial), kernel#3 (Gaussian radial basis function), and using neural network classification [12]–[14] and binary tree classification [15]–[17]. The receiver operating characteristic (ROC) curves are plotted in Figs. 7 and 8. We have trained the SVM system to be a “first look” system based on 100% sensitivity ($FNF = 0$) to adjust the system for cover all abnormalities. The conventional SVM results, compared in Fig. 7, show the highest testing classification rate of 84% of kernel#2 methods. It has 85% sensitivity and 81% specificity. In Fig. 8, binary tree decision result shows a classification rate of 83% with 82% sensitivity and 85% specificity. While three-layer feedforward backpropagation neural network result shows a classification rate of 82% with 85% sensitivity and 80% specificity.

The results of training and testing the proposed sys-
tem, a new LPD-uncrossed-feature based SVM learning system that is the combination of uncrossing mapping and local probability difference transformed features, show significant improvement with highest specificity of 89%, while keeping high sensitivity of 90%. And the overall testing classification rate is 89%.

Areas under ROC curves ($A_z$) are estimated and compared. The performances are improved and the results of the proposed method, a new LPD-uncrossed-features based SVM learning system, show the area of 0.88 compared with average area of 0.83 for other conventional methods.

Performed with 2 GHz CPU Pentium 4, 768 MB RAM, run with MS Visual C++, the average processing time for 86-feature extraction is 10.146 minutes-per-image. The time for training the SVM with the features extracted from 250 normal mammograms and 250 abnormal mammmograms is 28.237 minutes.

5. Conclusion

In this paper, we focused on crossed distributions of the feature pairs and the features that have overlapped distributions. A new support vector machine based method is then proposed. This method can be utilized with normal mammogram features more effectively. In particular, there is much hidden information in pairs of crossed features. We anticipate that when useful information from crossed distribution feature pairs is thus extracted and when the features are thus enhanced, improved automated classification of normal mammograms can be obtained.

References

[29] P. Somol, P. Pudil, and J. Kittler, “Adaptive float-reference transformed features, show significant improvement with highest specificity of 89%, while keeping high sensitivity of 90%. And the overall testing classification rate is 89%.

Areas under ROC curves ($A_z$) are estimated and compared. The performances are improved and the results of the proposed method, a new LPD-uncrossed-features based SVM learning system, show the area of 0.88 compared with average area of 0.83 for other conventional methods.

Performed with 2 GHz CPU Pentium 4, 768 MB RAM, run with MS Visual C++, the average processing time for 86-feature extraction is 10.146 minutes-per-image. The time for training the SVM with the features extracted from 250 normal mammograms and 250 abnormal mammmograms is 28.237 minutes.

5. Conclusion

In this paper, we focused on crossed distributions of the feature pairs and the features that have overlapped distributions. A new support vector machine based method is then proposed. This method can be utilized with normal mammogram features more effectively. In particular, there is much hidden information in pairs of crossed features. We anticipate that when useful information from crossed distribution feature pairs is thus extracted and when the features are thus enhanced, improved automated classification of normal mammograms can be obtained.

References

Appendix A: Curvilinear Features

Let \( f(i, j) \) be the pixel graylevel at spatial location \((i, j)\) and \( L(\theta, l) \) be a string of pixels in the direction \( \theta \) and of length \( l \), then the standard deviation of the pixel in \( L(\theta, l) \) is

\[
\sigma(\theta, l) = \sqrt{\frac{\sum_{m,n} (f(m,n) - \overline{f(L(\theta,l))})^2}{N_{L(\theta,l)} - 1}}, \tag{A\-1}
\]

where \( N_{L(\theta,l)} \) is the number of pixels and \( \overline{f(L(\theta,l))} \) is the average gray level within \( L(\theta, l) \) [15]–[17].

Let \( \sigma_{i,j}(\theta, l) = \min_{i,j \in L(\theta,l)} \sigma(\theta, l) \) and \( \sigma_{i,j}(l) = \min \sigma_{i,j}(\theta, l) \). The measure of surrounding pixel difference can be obtained from the standard deviation of \( \sigma_{i,j}(\theta, l) \) with regard to \( \theta \)

\[
\sigma_{\sigma_{i,j}(l)}(l) = \sqrt{\frac{\sum_{i}(\sigma_{i,j}(\theta, l) - \overline{\sigma_{i,j}(\theta, l)})^2}{N_\theta - 1}}, \tag{A\-2}
\]

where \( N_\theta \) is the total number of directions and \( \overline{\sigma_{i,j}(\theta, l)} \) is the average value of \( \sigma_{i,j}(\theta, l) \).

Finally, each pixel \((i, j)\) is determined to be as a line pixel or not according to the following rule for binary curvilinear lines:

\[
CL_{bin}(i, j) = \begin{cases} 
1 \text{ (line) if } \sigma_{i,j}(l) < T_{\sigma}, \\
\sigma_{\sigma_{i,j}(l)}(l) > T_{\sigma_{i,j}}, \\
0 \text{ (not) otherwise}
\end{cases} \tag{A\-3}
\]

where \( T_\sigma \) and \( T_{\sigma_{i,j}} \) are thresholds determined experimentally. We used the curvilinear structure of a region size \( 512 \times 512 \). For localized features, the regions were dissected into \( 8 \times 8 \) disjoint sub-blocks. We define \( l_i \) as the line pixel count and \( a_i \) as the average angle of each sub-block, where \( i = 1, 2, \ldots, 4,096 \). Histogram of 12 bins are then obtained for \( l_i \) and \( a_i \) with the relative frequency \( p_i^l \) and bin value \( x_i^l \) at bin \( j \). For the Local Binary Pattern (LBP) features [30],[31], we obtained LBP values for disjoint \( 3 \times 3 \) neighborhoods and generated a histogram of 12 bins with relative frequency \( p_j^{lbp} \) of bin value \( x_j^{lbp} \) at bin \( j \). A total of 18 features was extracted:

A.1 Line Pixel Count

the total number of curvilinear pixels in the region,

A.2 Upper Right Half Line Pixel Count

\[
A = \sum_{i<j} CL_{bin}(i, j),
\]

A.3 Lower Left Half Line Pixel Count

\[
B = \sum_{i>j} CL_{bin}(i, j),
\]

A.4 Upper Left Half Line Pixel Count

\[
C = \sum_{i+j<512} CL_{bin}(i, j),
\]

A.5 Lower Right Half Line Pixel Count

\[
D = \sum_{i+j\geq512} CL_{bin}(i, j),
\]

A.6 Half Ratio

\[
HalfRatio = A/B,
\]

A.7 Half Ratio 2

\[
HalfRatio2 = C/D,
\]

A.8 Angle Mean

the average line angle of the curvilinear pixels in the region,

A.9 Angle Standard Deviation

the standard deviation of the angle of the curvilinear pixels in the region,

A.10 Local Line Mean

\[
LLM = \sum_{j=1}^{12} p_j^l x_j^l,
\]

A.11 Local Line Standard Deviation

\[
LLS = \sum_{j=1}^{12} p_j^l (x_j^l - LLM)^2,
\]

A.12 Local Line Entropy

\[
LLE = \sum_{j=1}^{12} -p_j^l \log p_j^l,
\]

A.13 Local Angle Mean

\[
LAM = \sum_{j=1}^{12} p_j^a x_j^a,
\]

A.14 Local Angle Standard Deviation

\[
LAS = \sum_{j=1}^{12} p_j^a (x_j^a - LAM)^2,
\]

A.15 Local Angle Entropy

\[
LAE = \sum_{j=1}^{12} -p_j^a \log p_j^a,
\]

A.16 Local Binary Pattern Mean

\[
LBPM = \sum_{j=1}^{12} p_j^{lbp} x_j^{lbp},
\]

A.17 Local Binary Pattern Standard Deviation

\[ LBPS = \sum_{j=1}^{12} p_{lj} \left( x_j^{lj} - LBPM \right)^2, \]

A.18 Local Binary Pattern Entropy

\[ LBPE = \sum_{j=1}^{12} -p_{lj} \log p_{lj}. \]

Appendix B: Gray Level Co-occurrence Matrix Texture Features

Let \( P(i, j, d, \alpha) \) be the number of occurrence of pairwise gray levels \( i, j \) separated by a distance \( d \) and at direction \( \alpha \), then the relative frequency corresponding to \( P(i, j, d, \alpha) \) is

\[ p(i, j, d, \alpha) = \frac{P(i, j, d, \alpha)}{N_{d, \alpha}}, \]

where \( N_{d, \alpha} = \sum_i \sum_j P(i, j, d, \alpha) \) [21]. An isotropic GLCM with \( d = 1 \) obtained from four matrices at \( \alpha = 0^\circ, 45^\circ, 90^\circ, \) and \( 135^\circ \), is

\[ p(i, j) = \frac{1}{4} (p(i, j, 1, 0^\circ) + p(i, j, 1, 45^\circ) + p(i, j, 1, 90^\circ) + p(i, j, 1, 135^\circ)). \]

We can then obtain the estimated marginal probability from \( p(i, j) \) as \( p_x(i) = \sum_{j=0}^{N-1} p(i, j) \) and \( p_y(j) = \sum_{i=0}^{N-1} p(i, j) \), where \( N \) is the number of district gray levels; and their means and standard deviations as \( \mu_x, \mu_y, \sigma_x, \sigma_y \), respectively. We defined the gray level difference histogram (GLDH) as \( D(k) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p(i, j), \) where \( k = 0, 1, \ldots, N-1; \) and the gray level sum histogram (GLSH) as \( S(k) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} i \), where \( k = 0, 1, \ldots, 2(N-1) \) [32]. We extracted 16 features:

B.1 Energy

\[ Energy = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p(i, j)^2, \]

B.2 Entropy

\[ Entropy = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} -p(i, j) \log p(i, j), \]

B.3 Max Probability

\[ MaxProb = \max p(i, j), \]

B.4 Correlation

\[ \rho = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{p(i, j)(i-j)(j-\mu_x - \mu_y)}{\sigma_x \sigma_y}, \]

B.5 Diagonal Correlation

\[ DiagCorr = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p(i, j) |i-j|(i+j-\mu_x - \mu_y), \]

B.6 \( H_o \)

\[ H_o = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} -p(i, j) \log p(i, j), \]

B.7 \( H_y \)

\[ H_y = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} -p(i, j) \log p(i, j), \]

B.8 Difference Energy

\[ DEnergy = \sum_{k=0}^{N-1} D(k)^2, \]

B.9 Difference Entropy

\[ DEnergy = \sum_{k=0}^{N-1} -D(k) \log D(k), \]

B.10 Inertia

\[ Inertia = \sum_{k=0}^{N-1} k^2 D(k), \]

B.11 Homogeneity

\[ Homogeneity = \sum_{k=0}^{N-1} \frac{D(k)}{k^2}, \]

B.12 Sum Energy

\[ SEnergy = \sum_{k=0}^{2(N-1)} S(k)^2, \]

B.13 Sum Entropy

\[ SEnergy = \sum_{k=0}^{2(N-1)} -S(k) \log S(k), \]

B.14 Sum Variance

\[ SVar = \sum_{k=0}^{2(N-1)} (k-\mu)^2 S(k), \] where \( \mu = \sum_{k=0}^{2(N-1)} k S(k), \)

B.15 Sum Shade

\[ SSShade = \sum_{k=0}^{2(N-1)} \frac{(k-\mu_x-\mu_y)^2 S(k)}{\sigma_x^2 + \sigma_y^2 + 2\mu_x \sigma_y + 2\mu_y \sigma_x + \sigma_x \sigma_y}, \] and

B.16 Sum Prominence

\[ SProm = \sum_{k=0}^{2(N-1)} \frac{(k-\mu_x-\mu_y)^3 S(k)}{(\sigma_x^2 + \sigma_y^2 + 2\mu_x \sigma_y + 2\mu_y \sigma_x + \sigma_x \sigma_y)^2}. \]

Appendix C: Gabor Features

Let an impulse response of Gabor filter be

\[ g(x, y) = \frac{\exp \left[ -\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi j W x \right]}{2\pi \sigma_x \sigma_y}, \]

where \( W \) is the modulation frequency, \( x, y \) are coordinates.
in the spatial domain, and \( \sigma_x \) and \( \sigma_y \) are the standard deviations in the \( x \) and \( y \) direction [22], [23]. A Gabor filter-bank consists of Gabor filters with Gaussians of different sizes modulated by sinusoidal plane wave of different orientation from the same mother Gabor filter as

\[
g_{m,n}(x, y) = a^{-m}g(x, y), \quad a > 1, \tag{A-7}
\]

where \( x = a^{-m}(x \cos \theta + y \sin \theta), \) \( y = a^{-m}(-x \sin \theta + y \cos \theta), \) \( \theta = \frac{n\pi}{K}, K \) is number of orientation, \( n = 0, 1, \ldots, K - 1, \) and \( a = \left( \frac{m}{L} \right)^{1/S}, \) where \( S \) is number of scales, \( m = 0, 1, \ldots, S - 1; \) and \( U_h, U_l \) are the upper and lower center frequencies. The discrete Gabor filtered output of an image \( I_g(r, c) \) of size \( H \times W \) is given by a 2D convolution as

\[
I_{mn}(r, c) = \sum_s \sum_t I_g(r - s, c - t)g_{mn}(s, r), \tag{A-8}
\]

where \( * \) indicates the complex conjugate. We obtain the mean and standard deviation of the energy of the filtered image,

\[
\mu_{mn} = \frac{\sum_s \sum_t |I_{mn}(r, c)|}{H \times W}, \tag{A-9}
\]

\[
\sigma_{mn} = \frac{\sqrt{\sum_s \sum_t \left(|I_{mn}(r, c)| - \mu_{mn} \right)^2}}{H \times W}. \tag{A-10}
\]

We chose 4 orientations and 4 scales for our Gabor filter-bank. Then we have a Gabor feature vector of 32 features as

\[
\text{GaborFeatures} = [\mu_{00} \sigma_{00} \ldots \mu_{33} \sigma_{33}]_{1 \times 32}. \tag{A-11}
\]

**Appendix D: Multiresolution Statistical Features**

A nonseparable wavelet transform, the Quincunx Wavelet transform [24], [25], is used. Only the first four even-level low-pass decomposition images, i.e. images of spatial resolution \( 256 \times 256 \) (\( X_1 \)), \( 128 \times 128 \) (\( X_2 \)), \( 64 \times 64 \) (\( X_3 \)) and \( 32 \times 32 \) (\( X_4 \)), are retained for feature extraction. Mean, Variance, Skewness, Kurtosis and Entropy Features extracted from each images are defined respectively as

\[
M_k = \frac{1}{M^2} \sum_{i=1}^{M} \sum_{j=1}^{M} X_L(i, j), \tag{A-12}
\]

where \( M = 512/k \) is the size (height or width) of the decomposition image at even level \( L \) and \( k = 2^L, L = 1, 2, 3, 4 \),

\[
V_k = \sqrt{\frac{1}{M^2 - 1} \sum_{i=1}^{M} \sum_{j=1}^{M} (X_L(i, j) - M_k)^2}, \tag{A-13}
\]

\[
S_k = \frac{1}{M^2 - 1} \frac{\sum_{i=1}^{M} \sum_{j=1}^{M} (X_L(i, j) - M_k)^3}{V_k^3}, \tag{A-14}
\]

\[
K_k = \frac{1}{M^2 - 1} \frac{\sum_{i=1}^{M} \sum_{j=1}^{M} (X_L(i, j) - M_k)^4}{V_k^4}. \tag{A-15}
\]

\[
E_k = -\sum_{i=1}^{12} p_i^L \log p_i^L. \tag{A-16}
\]

We have 20 multiresolution statistical features:

\[
\text{MultFeatures} = [M_2 V_2 S_2 K_2 E_2 \ldots M_{16} V_{16} S_{16} K_{16} E_{16}]_{1 \times 20}. \tag{A-17}
\]

**Werapon Chiracharit** received B.Eng. in Electronics and Telecommunication Engineering in 1999 and M.Eng. in Electrical Engineering in 2001 from King Mongkut’s University of Technology Thonburi, Bangkok, Thailand. He is currently working toward Ph.D. in Electrical and Computer Engineering. His research includes image processing, pattern recognition and computer vision.

**Yajie Sun** received the B.S. degree from Shanghai Jiao Tong University, in 1995, and the M.S. degree from Zhejiang University, in 1998, both in Biomedical Engineering. He received the Ph.D. degree in Electrical Engineering from Purdue University, West Lafayette, Indiana, in 2004. His research interests include image processing, medical imaging, pattern recognition, and algorithm development. Dr. Sun is a member of Eta Kappa Nu and Sigma Xi. He is a member of IEEE and AAPM.

**Pinit Kumhom** received the B.Eng. from King Mongkut’s University of Technology Thonburi (KMUTT), Bangkok, Thailand, in 1988 and Ph.D. in Electrical Engineering from Drexel University, PA, USA, in 2001. He had worked as an instructor in the Electrical Engineering Department of KMUTT from 1988-1992. Currently, he is an assistant professor in the Department of Electronics and Telecommunication Engineering of the same university. His research interests include hardware implementation of DSP algorithms, FPGA applications, system-on-chip (SOC) design, design automation and methodologies, performance models in hardware design, hardware/software co-design.
Kosin Chamnongthai currently works as associate professor at Electronic and Telecommunication Engineering Department, Faculty of Engineering, King Mongkut’s University of Technology (KMU), and also serves as associate editor of ECTI-ECTI Trans and chairman of IEEE COMSOC Thailand. He has received B.Eng. in Applied Electronic Engineering from the University of Electro-communication (UEC), Tokyo in 1985, M.Eng. in Electrical Engineering from Nippon Institute of Technology (NIT), Saitama in 1987 and D.Eng. in Electrical Engineering from Keio University, Tokyo, Japan in 1991. His research interests include image processing, computer vision, robot vision, and natural language processing. He is a member of IEEE, IPS, TRS, TESA and ECTI.

Charles F. Babbs received the M.D. (with honors) and M.S. (anatomy) degrees from Baylor College of Medicine, Houston, TX, the Ph.D. degree in pharmacology from Purdue University, West Lafayette, IN, and the B.S. in experimental psychology from Yale University, New Haven, CT. He is an Associate Research Scholar with the Department of Basic Medical Sciences and an Instructor in family medicine at Indiana University School of Medicine, Indianapolis. He has received an NIH Research Career Development Award, and has written more than 150 refereed articles and ten chapters in scholarly journals and textbooks. His research is in the areas of biomedical engineering, cardiopulmonary resuscitation, tumor blood flow, heat therapy for cancer, digital mammography, the biochemistry of cellular damage after cardiac arrest, including the roles of free radicals and iron in reperfusion injury and the biomechanics of closed head injury.

Edward J. Delp was born in Cincinnati, OH. He received the B.S.E.E. (cum laude) and M.S. degrees from the University of Cincinnati and the Ph.D. degree from Purdue University, West Lafayette, IN. In May 2002, he received an Honorary Doctor of Technology from Tampere University of Technology, Tampere, Finland. From 1980 to 1984, he was with the Department of Electrical and Computer Engineering, The University of Michigan, Ann Arbor. Since August 1984, he has been with the School of Electrical and Computer Engineering and the School of Biomedical Engineering, Purdue University. In 2002, he received a Chaired Professorship and currently is The Silicon Valley Professor of Electrical and Computer Engineering and Professor of Biomedical Engineering. His research interests include image and video compression, multimedia security, medical imaging, multimedia systems, communication, and information theory. Dr. Delp is a Fellow of the IEEE, a Fellow of the SPIE, a Fellow of the Society for Imaging Science and Technology (IS&T), and a Fellow of the American Institute of Medical and Biological Engineering. He is Co-Chair of the SPIE/IS&T Conference on Security, Steganography, and Watermarking of Multimedia Contents that has been held since January 1999. He was the Program Co-Chair of the IEEE International Conference on Image Processing that was held in Barcelona in 2003. In 2000, he was selected a Distinguished Lecturer of the IEEE Signal Processing Society. He received the Honeywell Award in 1990, the D. D. Ewing Award in 1992 and the Wilfred Hesselberth Award in 2004 all for excellence in teaching. In 2001 he received the Raymond C. Bowman Award for fostering education in imaging science from the Society for Imaging Science and Technology (IS&T). In 2004 he received the Technical Achievement Award from the IEEE Signal Processing Society. In 2002 and 2006, he was awarded Nokia Fellowships for his work in video processing and multimedia security.