Abstract- The most common type of cancer seen in women is breast cancer. To improve from this severe disease, monitoring and early detection must be provided and related safety measures must be taken as the first step. During diagnosis, some of the cases may be overlooked depending on fatigue and eyestrain, because determination of abnormalities is a repetitive procedure. In this study, a Computer Aided Diagnosis (CAD) system, using wave atom transform (WAT) algorithm and Support Vector Machine (SVM), is proposed to evaluate mammography images. During the process, the region of interest is defined before applying the method. The system includes a feature extraction approach based on WAT algorithm. In terms of classification, the process has two main stages as classification of normal-abnormal regions and malignant-benign ones.

Keywords: Mammography, CAD system, Wave atom transform, SVM

1. INTRODUCTION
Today breast cancer is the most frequent form of cancer in women above 40. The World Health Organization’s international agency for research on cancer estimates that more than 150,000 women worldwide die of breast cancer each year. Wavelet decomposition and SVM is investigated for classifying the masses as benign or malignant is explained in [1]. Feature extraction and classification are the two stages which are performed by the decision making. Two novel texture features which provides smoothness and the architectural distortion is used to classify the abnormalities in the mammogram is explained in [2]. Three layer back propagation network is used to classify the masses.

A computer aided diagnostic system based on new texture shape feature coding is used to classify the masses is explained in [3]. The mass classification on mammogram presents a great challenge for design of computer aided diagnostic systems due to the complexity of mass background and its mammographic characteristics. SVM is also used for the classification.

Manly transformation and Principal Component Analysis is also used to reduce the high dimensional data in [4]. The classifiers like SVM and Bayesian Networks are trained and tested using a k-fold cross-validation test method. Ranklet is used as a classification features for classifying the masses in [5]. The dimensionality of the classification problem is reduced by using a recursive feature elimination technique without affecting the classification performances.

2. MATERIALS AND METHODS
2.1. Wave atom Transform
Wave atom transform is presented by Demanet in 2007. The transformation, obeying the parabolic scaling law, can be considered a variant of 2D wavelet packets. Wave atom transform have two very significant properties. First one is the ability to adapt to arbitrary local directions of a pattern. The second one is the ability to sparsely represent anisotropic patterns aligned with the axes. Wave atoms offer sharp frequency localization than other wave packets. It also has significant sparse expansion for oscillatory functions when compared with wavelets, curvelets and Gabor atoms.

The forms of wave packets, known as wavelets, Gabor, ridgelets, curvelets and wave atoms, are created using two parameters, which are $\alpha$ and $\beta$. These variables symbolize decomposition and directional ability for all wave forms. $\alpha$ and $\beta$ values are 1/2 for wave atoms and Figure 1 shows wave packet’s support in space and in frequency plane. Here, $\alpha$ corresponds to the multiscale
structure of the transform and $\beta$ corresponds to directional selectivity.

Actually, wave atoms are built from tensor products of 1D wave packets. One-dimensional wave packets can be represented as $\psi_{n,m}(x)$, where $j, m \geq 0$, and $n \in Z$. frequency restrictions are $\pm \omega_{j,m} = \pm \pi 2^j m$ with $C_1 2^j \leq m \leq C_2 2^j$. space restrictions is defined as $X_{j,n} = 2^j n$ (1)

Two-dimensional wave atoms $\varphi_{\mu}(X_1,X_2)$ are constructed with subscript $\mu = (j,m,n)$, where $m=(m_1, m_2), n=(n_1, n_2)$ 2D orthonormal basis is written as follows:

$$\varphi_{\mu}^+(X_1, X_2) = \psi_{m_1}^-(X_1-2^{-j} n_1) \psi_{m_2}^-(X_2-2^{-j} n_2)$$ (2)

$$\varphi_{\mu}^-(X_1, X_2) = \psi_{m_1}^+(X_1-2^{-j} n_1) \psi_{m_2}^+(X_2-2^{-j} n_2)$$ (3)

Where, $H$ is Hilbert transform. The wave atom tight frame is formed by combination of (2) and (3).

$$\varphi_{\mu}^{(1)} = \frac{\varphi_{\mu}^+ + \varphi_{\mu}^-}{2}, \varphi_{\mu}^{(2)} = \frac{\varphi_{\mu}^+ - \varphi_{\mu}^-}{2}.\text{ (4)}$$

2.2 Support Vector Machine

Support vector machines (SVMs) are a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis (Rejani and Selvi, 2009). The standard SVM is a non-probabilistic binary linear classifier, i.e. it predicts, for each given input, which of two possible classes the input is a member of. A classification task usually involves with training and testing data which consists of some data instances. Each instance in the training set contains one “target value” (class labels) and several “attributes” (features) (Gorgel et al., 2009). SVM has an extra advantage of automatic model selection in the sense that both the optimal number and locations of the basic functions are automatically obtained during training. The performance of SVM largely depends on the kernel.

SVM is essentially a linear learning machine. For the input training sample set

$$(x_i, y_i), i = 1...n, x \in R^n, y \in \{-1,+1\}$$

Let the classification hyper plane equation is to be

$$\omega \cdot x + b = 0$$

Thus the classification margin is $2/|\omega|$. To maximize the margin, that is to minimize $|\omega|$, the optimal hyper plane problem is transformed to quadratic programming problem as follows,

$$\begin{align}
\min \phi(\omega) &= 1/2(\omega \cdot \omega) \\
\text{s.t.} \ y_i ((\omega \cdot x_i) + b) \geq 1, i = 1,2...J
\end{align}\text{ (5)}$$

After introduction of Lagrange multiplier, the dual problem is given by,

$$\begin{align}
\max Q(\alpha) &= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \\
\text{s.t.} \ y_i \alpha_i = 0, \alpha_i \geq 0, i = 1,2...n
\end{align}\text{ (7)}$$

According to Kuhn-Tucker rules, the optimal solution must satisfy

$$\alpha_i (y_i((\omega \cdot x_i) + b) - 1) = 0, i = 1,2...n$$

That is to say if the option solution is

$$\alpha^* = (\alpha_1^*, \alpha_2^*, ..., \alpha_n^*), i = 1,2...n$$

Then

$$w^* = \sum_{i=1}^{n} \alpha_i^* y_i x_i$$

(8)

For every training sample point $x_i$, there is a corresponding multiplier. And the sample points that are corresponding to $\alpha_i = 0$ don’t contribute to solve the classification hyper plane while the other points that are corresponding to $\alpha_i > 0$ do, so it is called support vectors. Hence the optimal hyper plane equation is given by,

$$\sum_{x_i \in SV} \alpha_i y_i (x_i, x_j) + b = 0$$ (9)

The hard classifier is then,

$$y = \text{sgn}\left(\sum_{x_i \in SV} \alpha_i y_i (x_i, x_j) + b\right)$$ (10)

For nonlinear situation, SVM constructs an optimal separating hyper plane in the high dimensional space by introducing kernel function $K(x,y) = \phi(x) \cdot \phi(y)$, hence the nonlinear SVM is given by,

$$\begin{align}
\min \phi(\alpha) &= 1/2(\alpha \cdot \alpha) \\
\text{s.t.} \ t_i (\langle \alpha \phi(x_i) \rangle + b) \geq 1, i = 1,2...J
\end{align}\text{ (11)}$$

And its dual problem is given by,

$$\begin{align}
\max L(\alpha) &= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \\
\text{s.t.} \ y_i \alpha_i = 0, 0 \leq \alpha_i \leq C, i = 1,2...,J
\end{align}\text{ (12)}$$

Thus the optimal hyper plane equation is determined by the solution to the optimal problem.

3. Experimental setup

The proposed method consists of two stages namely feature extraction stage and classification stage. The two stages are described below:

3.1 Feature Extraction stage:

Feature extraction is an essential pre-processing step to pattern recognition and machine learning problems. The main objective of this study is to distinguish between the types of mass as benign or malignant. MIAS database, a benchmark database used by many researchers is taken to evaluate this study. The original mammograms in MIAS are very big size (1024×1024 pixels). The whole
image consists of 50% of noise. In order to remove this unwanted noise, a cropping operation is done manually before the feature extraction by choosing the given center of abnormality as the center of the ROI. The size of ROI image is 256*256. In the proposed system, the wave atom coefficients are used as features to classify the images. The original image is decomposed by using the WAT. All coefficients are considered as features for the classification process. This method is applied to all the training images and the feature vectors are stored in the database called as Feature DataBase. The energy of each sub-band of the image I is calculated by using the formula in

\[ \text{Energy}_e = \frac{1}{RC} \sum_{i=1}^{R} \sum_{j=1}^{C} |I_e(i, j)| \]

where \( I_e(i, j) \) is the pixel value of the eth sub-band and \( R, C \) is width and height of the sub-band respectively. Figure 1 shows the original mammogram image (mdb134) and Fig. 1b shows the cropped image. The block diagram of the proposed method is shown in figure 2.

### 3.2 Classification stage:

Classification phase executes two phases. In the first one, the classifier is applied to classify mammograms into normal and abnormal cases. Then the mammogram is considered as abnormal if it contains tumor (mass). Finally, the abnormal mammogram is classified into malignant or benign in the final stage. In this classification stage, SVM classifier in every phase is trained at specific number of training set in each category. Table 1 shows the number of training and testing images.

<table>
<thead>
<tr>
<th>Images</th>
<th>Normal</th>
<th>Abnormal</th>
<th>Benign</th>
<th>Malignant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>47</td>
<td>37</td>
<td>25</td>
<td>13</td>
</tr>
<tr>
<td>Testing</td>
<td>70</td>
<td>56</td>
<td>37</td>
<td>19</td>
</tr>
</tbody>
</table>

### 4. EXPERIMENTAL RESULTS

The assessment is carried out in relation with two stages of mass classification system namely (i) normal/abnormal classification (ii) benign/malignant classification. The performance of the proposed system is carried on 70 normal images and 37 mass images. Among the 37 abnormal images, there are 25 benign and 13 malignant images available. All the images are considered for the classification test. Table 2 shows the success rates of SVM method for the classification of images as normal and abnormal. Table 3 shows the success rates of SVM method for the classification of images as normal and abnormal.

At first, the dataset is classified as normal and abnormal. Then the accuracy is calculated for both normal and abnormal images. After that the same process is done for both benign and malignant images.

### 5. CONCLUSION

This paper describes a CAD system for recognizing breast cancer in ROIs of digital mammograms. The study also investigates the presentation of the system with wave atom transform, and SVM method. The best success rates in this work were obtained with the coefficients at scale of 1, 2 and 3 by using SVM. These results demonstrate that wave atom transform and SVM are useful and prevailing methods to distinguish the mammographic images as normal, benign and malignant.
REFERENCES:


