Abstract— Breast cancer is one of the fastest growing cancer and Architectural distortion is one of the important finding in the detection of cancer. The objective of this work is to develop a method for detecting architectural distorted area. In order to identify Architectural Distorted area this paper presents a novel method to distinguish Architectural Distortion (AD) and Non Architectural Distortion (Non AD) of Mammogram images, for the development of an efficient system. Among various breast abnormalities, architectural distortion is the most difficult type of tumour to detect. The proposed method could be applied as a solution to reduce the chances of human errors. The algorithm begins with a modified Gabor filters and orientation filter for identifying tissue patterns. Each detected tissue pattern is characterized into AD and NON AD based on mesh plot and contour plot which clearly states the variation in texture. Further the obtained result of modified gabor filter is tested with control chart to examine the effect of the algorithm. Statistical Parameter result obtained from control chart, for pre and post application of algorithm have showed a significance demarcation between AD and NON AD and also has proven the efficiency of the algorithm. 28 images from MIAS database were tested, with an efficiency of 89.4 percent to distinguish between the AD and Non AD And thus results obtained are promising in comparison with the previously published work.

Index Terms— Architectural Distortion (AD), Non Architectural Distortion (Non AD), Mammogram, Orientation, Andrew plot, Mesh Plot, Contour Plot and Control Chart

I. INTRODUCTION

Breast Cancer is a type of cancer which generally originates from breast tissues, usually in the ducts (tubes that carry milk to the nipple) and lobules (glands that make milk). It occurs in men and women although it has a high probability of occurrence in women and can prove to be fatal. Mammograms is a test performed [1] to carry out early onset of breast cancer and can prove to be a life saving and could reduce the mortality rate. But architectural distortion is one of the commonly missed signs which could be very fatal [2]. Each year almost 24,000 women under the age 45 are diagnosed with cancer in the United States [3]. Over 100,000 new breast cancer patients are estimated to be diagnosed annually in India [4]. Breast cancer cases are expected to increase by 26% by 2020 and most of these will be seen in developing countries [5]. Breast cancer incidence and death rates generally increase with age with Ninety-five percent of breast cancer deaths occurring in women with 40 years of age and older in United States [6]. Diagnosing error which could be a result in human error can lead to abnormality being undetected. Around 50,000 to 90,000 American Women are being Incorrectly Diagnosed with Breast Cancer Annually [7]. Good skill and experience is needed in understanding the complex problem which is generally encountered in diagnosing cancer, so developing new ways in terms Computer aided diagnosis to limit the errors is a priority. In Fig. 1, we can see mammogram images varying from normal, benign to malignant respectively.

(a): Normal  
(b): Benign  
(c): Malignant

Fig. 1: Mammogram Images

Initial screening is always carried out by Mammography, although there are many others. All women who are have a suspect of cancer go through mammography screening procedures. Early confirmation leads to an effective and better plan of action for treatment. Various algorithms are developed to aid radiologist to provide a second opinion and the final decision is left to the radiologists. A mammogram image typically shows variation in gray levels of contrast inside the breast which would characterize normal tissue or the presence of calcification and masses. Oriented texture pattern which typically converges toward the nipple in case of normal images, is distorted in the presence of architectural distortion.
[8]. No definite mass is visible in case of architectural
distortion. Microcalcifications, circumscribed masses,
spiculated masses and architectural distortions are some of the
signs always a radiologist look for [9] which would play a
significant role in deciding the condition of the patient. Fig. 2
describes different variants like Architectural Distortion,
Spiculated Mass, Micro-calcification& Circumscribed
mammogram images are shown in Fig. 2(a) to Fig. 2(d)
respectively.

(1): Architectural distortion  (b): Spiculated Mass

(c): Micro-calcification (d): Circumscribed

Fig. 2: Mamogram Region of Interest

Several algorithm for Discrimination and classification of
mammograms with definite mass for benign, malignant and
normal tissues [10] have been carried out. An architectural
distortion is the one with no definite mass visible [11] hence
making it the most difficult type of tumour to detect and it is a
third most-common appearance of breast cancer [12].
Computer aided diagnosis method using Gabor filters [13],
[14], phase portrait [14] analysis, fractal analysis [15], and
texture analysis [16], [17], [18] are mainly used applied for
mammogram images. Gabor filters and phase portrait maps
[14], [19], [20], [21] to characterize oriented texture patterns
in mammograms to detect architectural distortion have been
discussed vastly.

A method based on statistical analysis of pixel orientation
[22], where increase in pixels pointing to a region is found
based on which the pixels are marked suspicious. Also a
technique for the enhancement of spiculations in which a linear
filter is applied to the Radon transform of the image. The
enhanced image is filtered with radial spiculation filters to
detect speculated masses and architectural distortion [23].
Though these approaches were effective for the estimation of
architectural distortion, the complexity observed is higher in
those algorithms. A need is felt to develop low complex
algorithm for the estimation of architectural distortion. So the
objective was to develop a method which would clearly state
the difference between AD and NON AD region graphically
and statistically. We believe that AD region is made up of a
cluster of structure resembling like a line with different
orientations. So the method discussed can be summarized
under four stages one with extraction of ROI with potential
Ads and NON AD’s , the second stages is implementation of
Gabor filters the next stage is recognition of potential ADs
using spectral variation , contour plot and Andrew plot with
graphical result and also statistical result by the use of Contour
plot. The final fourth stage would comprise of providing a
comparative result between AD and NON AD region and to
check the efficiency of algorithm results were shown prior and
post application of algorithm. Our work shows a significance
variation of spectral variation as observed through mesh plot
and also a volume of distribution using contour plot as
compared to the previously published result [11]. A new
approach using Andrew plot and control chart is being
introduced which provides a significant influence to the
objective of the work which has been discussed in the result
section.

II. METHODOLOGY

A. Preprocessing

The region of interest is defined by x and y co-ordinate with
r the radius by radiologist in the MIAS database, which
resembles a circular region of interest. For implementing
the objectives of the work, a circular co ordinate needs to be
converted into a rectangular coordinate which can be
extracted. Refer to (1) which will approximately cover the area
of interest as defined by the radiologist. This co-ordinate area
is used to extract Architectural distorted area (I_{AD}) using the
inbuilt function of Matlab® where as for extracting Non
Architectural distorted area (I_{Non \text{ AD}}) refer (2).

\[ I_{AD} = \left[ x - r, 1000 - y, 2r, 2r \right] \] (1)

\[ I_{Non \text{ AD}} = \left[ x + 50 - r, 900 - y, 2r, 2r \right] \] (2)

The image can then be resized to 128 x 128, 512 x 512, and
256 x 256. In this work the ROI were resized to 128 X 128.

B. Final Stage

Entropy is a statistical measure of the degree of randomness
that can be used to characterize the texture of the input image.
During Architecture Distortion the oriented features changes
the normal pattern. There is a possibility that entropy might
also change. If the pattern is that of a normal ROI then it does
not have much impact on the outcome. In short if there is
uncertainty in the images then it can be found. Equation (3)
defines Entropy of 1D data where \( P(X_i) \) is the Probability mass distribution.

\[
E(x) = -\sum_{j=1}^{k} P(X_j) \log P(X_j)
\]  
(3)

For a 2D array refer (4), here \( P \) is the intensity in

\[
E(I) = -\sum_{j=1}^{m} \sum_{j=1}^{n} \text{hist}_N(P) \log(\text{hist}_N(P))
\]  
(4)

\( m \times n \) image \( (i, j) \) and \( \text{hist}_x(P) \) is the histogram properly normalized. Literature surveys also reveal that entropy has a significant role in affecting the signs of neoplasm in breast [24-26]. So the entropy was removed from all the ROI to prepare the data sets for further processing.

C. Gabor filtering for Orientation field

The AD lines are straight lines hence are not oriented towards the nipple and do not have angular edges. Because of the orientation of normal tissue structures, radiologists generally conclude that AD regions are not clear. Finding the exact position of AD is important as this lesion is always identified as BIRADS category 4 or category 5 [27]. This has led to a challenge of determining oriented structure in image. Gabor filter is very significant in determining the oriented texture. Numerous works have been carried out and shown the benefits of Gabor Filter [28 - 30]. To extract the texture in a mammogram, the filtered image of the previous stage is allowed to run over a bank of Gabor filters for specific orientations [28]. Gabor filter is justified due its good detection with an accuracy of linear patterns. The Gabor filter kernel is given by eqn (6). The \( S_x \) and \( S_y \) are variance along \( x \) and \( y \) axis, theta represent orientation and \( f \) is the frequency. The values of \( S_x \), \( S_y \) and \( f \) were chosen empirically. Gaussian kernel is used generally for weighting neighbors by their proximity to the pixel of interest and Fourier kernel is Fourier kernel provides us information about patterns in the image, when these two kernels are taken together we obtain the Gabor filter:

\[
\text{GaborFit}(\sigma, \omega, \theta) = \text{GaussKernel}(\sigma) \ast \text{FourierKernel}(\omega, \theta)
\]  
(5)

\[
E_i(x,y) = \left[ \frac{1}{2 \pi S_x S_y} \times \exp \left[ -\frac{1}{2} \left( \frac{x^2}{S_x^2} + \frac{y^2}{S_y^2} \right) \right] \right] \ast M_i(x,y,f)
\]

\( i = 1, 2; \)  
(6)

\[
M_1(x,y,f) = \cos \left( 2 \ast \pi \ast f \ast \frac{\sqrt{x^2 + y^2}}{\sqrt{x^2 - y^2}} \right)
\]  
(7)

\[
M_2(x,y,f) = \cos(2 \ast \pi \ast f \ast (x \ast \cos \theta + y \ast \sin \theta))
\]  
(8)

The values of \( S_x \), \( S_y \) and theta justified for this work were finalized on the basis of trial and error. For determining the orientation in all direction, angles were changed in a set of six. Orientations are obtained by varying angle theta from 0, 30, 45, 60, 90 to 180 degrees. The output obtained at different angles corresponds to a particular image. Keeping this fact in mind that Gabor filter acts in different manner over the normal portions and over the abnormal patterns the results were computed for both the normal ROI and abnormal ROIs. The results were analyzed between before applying Gabor filter and after Gabor filter as the impact of Gabor filter has to be studied on both types of regions. The output of Gabor filter that is output of \( G_1 \) and \( G_2 \) were modified empirically to obtain the result as desired for AD and Non AD region refer (9) to refer (12). The output of the image obtained from equation (12) is similar to the size of the original extracted region. \( N_{Ga} \) and \( N_{Gam} \) are the number of samples from \( G_a \) and \( G_{am} \) respectively.

\[
G_a = \frac{G_1}{G_2}
\]  
(9)

\[
G_{am} = \frac{1}{N_{Ga}} \sum G_a
\]  
(10)

\[
V = \frac{1}{N_{Gam}} \sum G_{am}
\]  
(11)

\[
I = \frac{G_a}{V}
\]  
(12)

D. Orientation Field and its importance

The orientation Field can be extracted from the Gabor filter but to find the variation in this output, to distinguish between distorted and non-distorted region, directional filters are used. Directional filters are nothing but banks of filters with different orientation. This filter processes the data as per the direction coefficient specified. We can easily find a distinction in such a case where we have an architectural variation from the obtained output. Using Gabor wavelets at different orientation and frequencies, Loan Buciu Alexandru Gascai [31] have described an approach where mammogram sample is filtered and directional features are extracted. Roberto H.Bamberger and Mark J.T.Smith [32] were the pioneers in directional filter estimation (DFE) for image processing which estimates directional flow and the filter has overcome the limited directional selectivity of separable wavelets. It finds its application in various imaging applications such as texture classification [33], image denoising [34], fingerprint image enhancement and recognition [35].

III. PROPOSED NEW ALGORITHM

Following the previous available algorithm, a new efficient and an efficient decision making algorithm is proposed. Matlab® is utilized to implement. The following steps were followed to obtain the desired objective:

Step 1: Load the image.
Step 2: Refer (1) to extract the AD region of the image.

Step 3: Refer (2) to extract the NON AD Region of the image for optimizing our method by comparing AD & NON AD region.

Step 4: Pre-processing of Image - Resizing, Guassian filter implementation (to remove Low component frequency) and Entropy Filtration refer (4).

Step 5: Equation (5) to Equation (8) for Gabor Filter Implementation

Step 6: Equation (9) to Equation (12) is completely a novel approach derived, scope of the equations is to visualize variation in the desired pattern.

Step 7: Visualize the output using Mesh plot, Andrew plot and contour plot for both AD and NON AD.

Step 8: The parameters obtained as an output of Control Chart is used to distinguish AD and NON AD numerically

IV. SIMULATION AND EVALUATION

This study used here discusses a novel approach for the classification of AD and Non AD regions. For AD and NON AD, two methods were used to show the results. One is a graphical output and other is a theoretical output. In this study, all ROIs within Mammogram images were chosen from MIAS database, and features for ROIs were analyzed and calculated. Images were analyzed with AD and NON AD. Results of the most efficient Gabor filter parameters chosen as lambda = 8, theta = π/2, gamma = 1.5, band width = 1 & Neighborhood = 8 is shown in Fig. 7. The output was then visualised using Andrew plot, Mesh plot & Contour plot and the plots were verified before applying algorithm for AD and NON AD. On analyzing the results there is a significant influence of the proposed algorithm on AD images. Andrew plot which observes the variation as a defined for the i-th observation of the ROI we could observe a subtle variation in the AD region due to the presence of change in orientation as shown in Fig. 3. When results obtained from Control Chart were analyzed, a range could be observed for Mu and Sigma, which are the constant parameters obtained from the control chart. Table 1 and Table 3 shows the result for mu and sigma on AD and NON AD images before and after applying algorithm. The average statistical response for MU on applying algorithm will lie in the range of 10.264±3 for AD and for NON AD it is 3.197±3. For Sigma the statistical data response was 5.81±3 for AD image and same for NON AD was 0.03215±1.

Contour Plot as shown in Fig. 4 shows a significant distribution after applying the algorithm, due to variation in orientation less contour mass is observed. Similarly for mesh plot as in Fig. 5 and Fig. 6 we could observe a very less spectral density with spikes observed irregularly where as for the Non Ad region spectral variation could be seen evenly distributed. Y axis and X axis represents the image size from 1: rows to 1: columns of Image respectively and the Z axis represents single-valued function defined over a rectangular grid. Color represented is proportional to surface height.

For evaluating the efficiency of algorithm two parameters were looked upon one being the computation time and the other was the efficiency to detect false images, from a set of 128 images, The average computation time observed was 8.34
seconds. For the second parameter the test was performed on AD region where in 16 data sets were used out of which 14 data were recognized as AD images and a accuracy of 89.4 percent accuracy was observed for a data set of 19 non AD images, then finally the test data of 19 non AD images were mixed with 12 new AD images, all 12 were identified as AD and 17 were identified as Non AD with all the three plots and control chart parameters Mu and Sigma providing a clear result. Remaining two provided a satisfactory result for Mu, Sigma, and Andrew plot whereas clear differentiation could not be observed for mesh and Contour plot and the classification were manually observed for all the three plots.

**TABLE I: CONTROL CHART PARAMETERS MU AND SIGMA BEFORE AND AFTER APPLYING ALGORITHM FOR “AD IMAGES”**

<table>
<thead>
<tr>
<th>Image Sets</th>
<th>Mu Before</th>
<th>Mu After</th>
<th>Sigma Before</th>
<th>Sigma After</th>
</tr>
</thead>
<tbody>
<tr>
<td>006</td>
<td>53.525</td>
<td>8.219</td>
<td>29.659</td>
<td>6.2186</td>
</tr>
<tr>
<td>013</td>
<td>163.827</td>
<td>7.648</td>
<td>63.738</td>
<td>5.8456</td>
</tr>
<tr>
<td>032</td>
<td>153.525</td>
<td>7.791</td>
<td>87.659</td>
<td>5.8942</td>
</tr>
<tr>
<td>039</td>
<td>132.642</td>
<td>7.981</td>
<td>43.789</td>
<td>4.8316</td>
</tr>
<tr>
<td>045</td>
<td>133.428</td>
<td>7.119</td>
<td>21.993</td>
<td>6.2913</td>
</tr>
<tr>
<td>074</td>
<td>169.731</td>
<td>15.603</td>
<td>61.191</td>
<td>3.1296</td>
</tr>
<tr>
<td>065</td>
<td>38.320</td>
<td>24.191</td>
<td>79.112</td>
<td>7.2542</td>
</tr>
<tr>
<td>103</td>
<td>112.691</td>
<td>6.981</td>
<td>91.321</td>
<td>7.1386</td>
</tr>
<tr>
<td>106</td>
<td>93.511</td>
<td>8.193</td>
<td>27.051</td>
<td>9.1662</td>
</tr>
<tr>
<td>113</td>
<td>213.021</td>
<td>9.149</td>
<td>41.104</td>
<td>4.2193</td>
</tr>
<tr>
<td>117</td>
<td>113.005</td>
<td>7.101</td>
<td>89.329</td>
<td>4.8149</td>
</tr>
<tr>
<td>123</td>
<td>196.249</td>
<td>13.201</td>
<td>57.711</td>
<td>4.9913</td>
</tr>
</tbody>
</table>
TABLE II: CONTROL CHART PARAMETERS MU AND SIGMA BEFORE AND AFTER APPLYING ALGORITHM FOR “NON AD IMAGES”

<table>
<thead>
<tr>
<th>Image Sets</th>
<th>Mu Before</th>
<th>Mu After</th>
<th>Sigma Before</th>
<th>Sigma After</th>
</tr>
</thead>
<tbody>
<tr>
<td>002</td>
<td>49.228</td>
<td>9.1217</td>
<td>80.768</td>
<td>0.0186</td>
</tr>
<tr>
<td>007</td>
<td>154.702</td>
<td>0.1216</td>
<td>74.116</td>
<td>0.0296</td>
</tr>
<tr>
<td>011</td>
<td>150.818</td>
<td>0.1284</td>
<td>67.188</td>
<td>0.0459</td>
</tr>
<tr>
<td>020</td>
<td>161.228</td>
<td>0.1252</td>
<td>56.802</td>
<td>0.0346</td>
</tr>
<tr>
<td>031</td>
<td>61.121</td>
<td>9.0007</td>
<td>68.148</td>
<td>0.0186</td>
</tr>
<tr>
<td>049</td>
<td>157.872</td>
<td>0.1969</td>
<td>84.196</td>
<td>0.0296</td>
</tr>
<tr>
<td>062</td>
<td>170.911</td>
<td>0.1940</td>
<td>77.189</td>
<td>0.0459</td>
</tr>
<tr>
<td>088</td>
<td>139.288</td>
<td>0.1121</td>
<td>41.931</td>
<td>0.0346</td>
</tr>
<tr>
<td>091</td>
<td>63.110</td>
<td>8.9218</td>
<td>40.168</td>
<td>0.0186</td>
</tr>
<tr>
<td>100</td>
<td>54.793</td>
<td>10.126</td>
<td>71.116</td>
<td>0.0296</td>
</tr>
<tr>
<td>119</td>
<td>9.7129</td>
<td>0.1916</td>
<td>56.190</td>
<td>0.0459</td>
</tr>
<tr>
<td>127</td>
<td>161.018</td>
<td>0.1249</td>
<td>60.179</td>
<td>0.0346</td>
</tr>
</tbody>
</table>

Fig. 5: Y-Z Axis View of Spectral Plot

(a): Y-Z axis of Surface plot for Mammogram ROI before applying algorithm

(b): Y-Z axis of Surface plot as observed for AD region

(c): Y-Z axis of Surface plot as observed for NON AD region
V. CONCLUSION

This paper presents a novel approach for the classification of AD and Non AD regions through graphs, image plots and numerical values by drawing Andrew plot, Mesh Plot, Contour Plot and by Control chart. Since architectural distortion is the most difficult type of tumour to detect, A new novel method of MU, Sigma and Andrew plot will definitely help radiologists to identify and distinguish AD from Non AD region with confidence and hence reduce the human errors substantially. The potential algorithm has to be studied on other available database. It would also be interesting to combine this algorithm on already available Computer diagnosis systems.

REFERENCES


