Hybrid Method of Local Binary Pattern and Classification Tree for Early Breast Cancer Detection by Mammogram Classification

O.K. Girija, M. Sudheep Elayiodm

Abstract—Breast cancer is an alarming disease due to a mutation in breast cells and it is one type of cancer among women which highly leads to their death. One of the most effective tools for early detection of breast cancer is mammography, which is a screening tool used to examine the human breast by using low-dose amplitude X-rays. Computer-Aided Diagnosis (CAD) is used as an important tool to help the medical professionals for classifying breast tissues into a different class. Computer-Aided Diagnosis (CAD) can be used to reduce human error in reading the mammograms and it shows effective results in the classification of benign and malignant abnormalities. The proposed method presents a new classification approach to detect the abnormalities in mammograms using Local Binary Pattern and Decision Tree Classification. A Uniform Local Binary Pattern(uLBP) is an extension of the original Local Binary Pattern in which only patterns that contain at most two transitions from 0 to 1 (or vice versa) are considered. In uniform Local Binary Pattern (LBP) mapping, there is a separate output label for each uniform pattern and all then on uniform patterns are assigned to a single label. These patterns are utilized to detect breast cancer by classification employing the Decision Tree Classification. Specificity and sensitivity are the two statistical measures used in this proposed method to verify and measure the significance of the test related to abnormalities in the breast tissues. Thus, it can be a measurement of performance tests for classifying the patients who do and do not suffer from cancer. The mini-MIAS mammography database is used for testing the accuracy of the proposed method and the results are promising.

Index Terms—Breast Cancer, Mammography, Classification, Local Binary Pattern, Classification Tree, Early Detection.

I. INTRODUCTION

Breast cancer is an alarming disease due to a mutation in breast cells and it is one type of cancer among women which highly leads to their death. The incidence rate of breast cancer is more in developed countries than in developing countries but the mortality rate is less in developed countries. The reason for the reduction in mortality rate is mainly due to the awareness about screening techniques and treatment facilities.

The analysis of breast image using various modalities such as mammograms along with computer-aided methods assists the radiologist in identifying the presence of abnormalities in its initial stage. The detection of abnormalities in the early stage avoids the spread of malignancy that leads to cancer death [1]. An analysis in India [Agarwal, 2008] reports annual breast cancer diagnosis of about 10,000 new breast cancer cases.

A study conducted by Govt. Medical College, Thiruvananthapuram (under Pain and Palliative Clinic) shows that the most commonly occurring cancer in women in India is breast cancer [Asha, 2014]. India continues to have a low survival rate for breast cancer, with only 66.1% of women diagnosed with the disease between 2010 and 2014 surviving, a Lancet study found. Major breast cancer symptoms are Calcification, Architectural distortion, Mass, Bilateral Asymmetry.

The pertinence of CAD algorithms and the process of screening for the detection of abnormalities in an early stage is very important [2]. Computer-aided techniques help the Radiologists in mammography analysis for early breast cancer detection.

Computer-Aided Detection and Computer-Aided Diagnosis are two types of Computer-Aided Diagnosis systems. When those methods are applied, the region of interest (ROI) extracted using different feature extraction methods and by using these features or patterns, the system automatically classified as normal X abnormal or benign X malignant.

Mammography is one of the techniques for finding early detection of breast cancer and it is a proven method. This technique proposed for Classification of normal tissue, abnormal tissue and then classify it into the benign condition and the malignant condition is a complex procedure because of the variations like tissue characteristics, such as the shape of the tissue, grey level, and size of the tissue, its intensities, and location [3].

With the high sensitivity, the specificity of visual inquisition is low. To improve the interpretation of mammograms became a major issue because it causes accuracy, also can reduce the rate of false negative and false positive.

II. LITERATURE SURVEY

Preceding studies shows, the Computer-Aided Design can be used to help the professionals to simplify the action of interpreting the mammograms and get more accurate results. There are different classes of classification methods proposed to create an efficient and efficient CAD for the detection of cancer in mammograms proposed by researchers in the last two decades.

Neural networks are a good choice for classification of mammograms and they always have a good performance for the classification of normal breast tissues and abnormal breast tissues from mammogram images.
To classify the normal breast tissues and abnormal breast tissues from mammograms using a convolution neural network are proposed by B. Sahiner, et al in 1996. In this experiment, features were computed from sub-regions by using the methods, Gray Level Difference Statistics, and Spatial Gray Level Dependence. Then, these features further used as input to classify the breast issues as normal and abnormal using the Convolution neural network [4]. Sung-Nien Yu and Yu-Kun Huang proposed a wavelet filter method that is used to detect the cancer region by calculating the mean pixel value. They used MRF, fractal models, and statistical texture features to extract features from mammograms. Then they used neural network classifier-Back Propagation neural network (BPNN) and it is used for the classification process using previous features with 3 layers [5].

In 2017, Esa Prakasa used the Local Binary Pattern method to extract texture feature, then he applied pattern recognition algorithm might be applied for further implementations, such as texture segmentation and grading on the regularity of texture patterns [6]. Detection of breast cancer proposed by Amirhossein Ketabdar et al using combined method of wavelet transform and Local Binary Pattern (LBP) algorithm. The combined method of wavelet and LBP algorithm archived an accuracy of 96.2 percent for normal cancer and 94.7 percent for abnormal cancer [7].

Developed a system, an effective technique to classify digital mammograms as a benign stage and malignant stage by IslamM.et.al.in 2010[8]. They have used a Multi-layer Perceptron (MLP) with several layers for classification from seven texture features. Man To Wong et al has proposed a model that comprises Co-Occurrence Matrix by Gray Level (GLCM) and Neural Network (NN).

In this method, GLCM is used for feature extraction and ANN is used to classify the mass region or non-mass region in the breast [9]. A method proposed by Liu et al. for segmenting mass and detection using a rule-based algorithm called MCL and NBAC (narrowband region-based active contour). The Complete Local Binary Pattern-CLBP are extracted from ROIs of segmented suspicious regions and these selected regions are classified using SVM [10].

Choi et al. developed a method to classify mammograms in normal breast tissues and breast masses using time resolution LBP and extracted features classified using Support Vector Machine Recursive Feature Elimination (SVMr-RFE) [11]. A modified algorithm proposed by Liu et al. is the use of the median value of the neighboring pixels as a threshold value instead of the value of the central pixel in the rotation invariant LBP algorithm for mammograms classification (malignant or benign)[12].

For classifying ROIs as either breast cancer mass or normal breast tissue by grouping LBP patterns to better characterize texture micro patterns not losing the spatial structure of masses is proposed by Llad et al[13]. Rinku Rabidas et al proposed a comparative analysis, this study contains different texture features they are based on the local operator. This study produced for the determination of mammographic masses as benign or malignant [14].

Based on the previous works that are described in the previous paragraphs, this paper enumerates the advantages of uniform Local Binary Pattern (LBP) for feature extraction and Classification Tree (CT) is used to classify the extracted features using LBP. Based on related work, we propose an enhanced method that combines both uniform Local Binary Pattern (LBP) and Classification Tree (CT) for the classification of digital mammograms.

III. METHODOLOGY

The suggested method utilizes the advantages of both uniform Local Binary Pattern (LBP) [15] for extracting features and using these extracted features are used as the input of Classification Tree [16] to classify normal and abnormal breast tissues to get the accurate prediction. This method consists of three processes, they are preprocessing, feature extraction and the classifier.

First, pre-process the data for the more accurate result then the texture features are extracted using a uniform Local Binary Pattern (LBP) and get the texture information of the mammogram that contains in a feature vector. Then classify the breast tissues using Classification Tree (CT). The classification process contains two major phases, which are Phase I and Phase II. They are described as follows,

1. Phase I (Normal and Abnormal): To find any abnormalities Phase I can be used as the primary phase. During Phase I, training the classifier using a training set that contains normal and abnormal mammograms. The result will be treated as the input of the next phase.

2. Phase II (Benign and Malignant): In Phase II, input will be taken from the previous phase. By using these features, find the abnormalities type. These abnormalities classified into two classes, they are benign and malignant. Figure 1 shows the methodology of classification.

![Classification Methodology](image-url)
A. Pre-Processing

Data from the real world is not relevant for our application because the data is incomplete because some values may be missing or inconsistent, and/or deficient in certain behaviors or drift, and may contain many errors.

Data preprocessing is a technique used to resolving issues in the data from the real world. Data preprocessing prepare raw data for further processing.

The intention of pre-processing is to enhance the image data that will snuff out unwilling distortions or enhances some image features which are important for further processing, although geometric transformations of images (e.g. rotation, scaling, translation) are classified among pre-processing methods here since similar techniques are used [17].

B. Feature Extraction

Preprocessing is the widespread method to reduce the dimensionality in high dimensional data analysis, visualization and modeling. One of the common and simple methods to reduce the dimensionality is feature extraction.

Feature Extraction is a more general method in which one tries to develop a transformation of the input space onto the low dimensional subspace that preserves most of the relevant information.

To improve performance such as estimated accuracy, visualization, and comprehensibility of learned knowledge, feature extraction, and selection methods are used alone or in combination [18].

A well-known texture analysis method that considers the spatial behavior of intensity values in any given neighborhood is the Local Binary Pattern (LBP) [19] and it is mostly used for recognition techniques by extracting the texture feature.

The local binary pattern operator transforms an image into an array or image of integer labels describing details of textures in the image. The image labels from the LBP can be used directly or their statistics for further analysis. A binary code is generated for each pixel in an image is,

\[ LBP_{p,r}(N_c) = \sum_{p=0}^{P-1} S(N_p - N_c)2^p \]  

Where,
- Neighborhood pixel \( N_p \) in each block of the image
- is threshold by its center pixel value \( N_c \)
- Sampling points= p.
- Radius =r.

\( s(x) \) is the binary threshold function,

\[ s(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \]

The different steps for generating texture features using LBP are described as follows,

1. Divide the image into different blocks.
2. For each pixel (center) in a block, compare the pixel to each of its 8 neighboring pixels and follow this procedure either clockwise or anticlockwise that means along a circle.
3. If the neighbor pixel value is greater than center pixel value, write “1” Otherwise write “0”.
4. All the obtained binary codes are concatenated to form a binary number. For the convenience these binary numbers are converted to decimal values that serve as labels of each pixel. These binary numbers are called LBP codes.
5. Compute the Histogram of the frequency of each number occurring over the block.

\[ h(k) = \sum_{x=0}^{L} \sum_{y=0}^{L} \delta(LBP_{p,r}(x,y),k) \cdot \delta(0,2^p - 1) \]  

Where, \( \delta(u,v) \) is Kronecker delta function and \( \delta(u,v) = \begin{cases} 1, & \text{if } u = v, \\ 0, & \text{otherwise}. \end{cases} \)

Uniform Local Binary Patterns are the important observations of the LBP code in images. Certain binary patterns are the main properties of texture and their frequency exceeds 90%, so these type of patterns are called Uniform.

Uniform patterns are defined uniformity measure corresponding to a spatial transition (at most two circular 0-1 and 1-0 transitions), denoted as \( LBP_{U2}(p,r) \)

The uniformity measurement of LBP is;

\[ U(LBP_{p,r}(N_c)) = \sum_{p=0}^{P} \sum_{r=0}^{R} S(N_{p+1} - N_c) - g(N_p - N_c) \]  

Where,
- Neighborhood pixel \( N_p \) is equivalent to \( N_0 \).
- Uniform LBP is always \( U(LBP_{p,r}) \leq 2 \)

Uniform LBP is an extension of original LBP operator and the examples of original LBP and uniform LBP are 01010010, 10100110 and 11111111, 00000000 respectively. Figures 2 and 3 shows working example of LBP operator and computation of Local Binary Pattern (LBP) [20].

![Figure 2: Example of how LBP operator works](image1)

![Figure 3: Computation Example of LBP](image2)
C. Classification Tree (CT)

A classification tree is a popular classification method. It is a tree-like structure where each tree node denotes a test on an attribute value and each branch represents an outcome of the test. The tree leaves represent the classes.

A classification tree is a model that is both predictive and descriptive and it displays relationships found in the training data using its tree structure. Classification tree consists of two types of nodes,

1. Intermediate nodes - An intermediate node is labeled by a single attribute, and the edges extending from the intermediate node are predicates on that attribute.
2. Leaf nodes – A leaf node is labeled by the class label which contains the values for the prediction.

In many medical image processing systems, the medical professionals may not know what the decision rule is, so he or she would like to discover the decision rule by using data. In such situations, decision trees are often referred to as classification trees. Classification trees apply to data where y-value (outcome) is a classification label, such as the disease status of a patient, and the medical professional would like to construct a decision rule that predicts the outcome using x-variables (dependent variables) available in the data.

Because the data set available is just one sample of the underlying populace, to construct a decision tree rule that is accurate for both data at hand and external data. Assume that each object has several independent attributes and a dependent attribute. Induction is the training process and it generates the tree [21].

D. Correctness or Accuracy

The confusion matrix can be used to evaluate the accuracy of the classifier. The performance for the abnormality detection of each classifier can be compared by using specificity and sensitivity.

Specificity and sensitivity are the two statistical measures used in this proposed method to verify and measure the significance of the test related to abnormalities in the breast tissues.

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \quad (4)
\]

\[
\text{Specificity} = \frac{TN}{TN + FP} \quad (5)
\]

Where,

TP is True Positive.
TN is True Negative.
FP is False Positive.
FN is False Negative.

Based on the above two equations, sensitivity is the count of disease which is predicted correctly using the positive test at the same time specificity is the count of patients without disease whose test result is negative.

Hence, it can be a measurement of performance tests for the purpose of classifying patients who do and do not suffer from cancer.

E. Proposed Method

In the proposed work, we delve into the importance and advantages of Local Binary Pattern (LBP) and Classification Tree (CT) and we analytically discover the best Fine-tuning method to ratify when training a CT model.

Build a CAD system that is based on a Classification Tree (CT) that helps the radiologist classify mammography abnormalities. We used a hybrid version of Local Binary Pattern (LBP) and Classification Tree (CT) to build a CAD system that classifies normal tissues and abnormal tissues in mammograms.

To test the proposed system, we have used the mini-MIAS dataset which contains 322 images including 13 malignant and 12 benign cases [22]. Considering the training set, the ground truth belong to the MIAS database is used.

Figures 4 and 5 show the conceptual level diagrams of the training and test procedures and image before and after preprocessing.

Figure 4: Conceptual block diagram for training and testing

Figure 5: Images before and after the preprocessing using the sample image mdb028.pgm [17]

In this paper, first, feature extracted from the processed images using the Local Binary Pattern (LBP) method. The LBP operator converts an image into an array or image of integer labels describing texture details of the image.

The image labels from the LBP can be used directly or their statistics for further analysis. The texture features extracted is used as the input of the classifier, for more accuracy we used Classification Tree (CT) as the classifier in this method.

The classifier then differentiates the normal and abnormal mammograms. From the abnormal mammograms, the classifier again classifies the mammograms into two classes such as Benign and Malignant mammograms. The algorithm for the proposed method is given below:
ALGORITHM: PROPOSED METHOD

INPUT: Digital Mammograms
OUTPUT: Different classes of input data and final prediction.

1) For feature extraction,
   a) Divide the image into different blocks.
   b) For each pixel (center) in a block, compare the pixel to each of its 8 neighboring pixels and follow this procedure either clockwise or anti-clockwise that means along a circle. If the neighbor pixel value is larger than center pixel value, denote as “1” otherwise denote as “0”. For each pixel of an image, a binary code is produced,
   \[ LBP_{r,r}(N_c) = \sum_{P=0}^{P=\langle r \rangle} s(N_P - N_c)2^P \]
   \[ s(x) \] is the binary threshold function,
   \[ s(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \]
   c) Concatenate the obtained codes to make a binary number.

2) For the classification of extracted features,
   a) The training data be D. The data may contains values are continuous valued, they should be converted into discrete value. If all items in D are in the same class, then stop.
   b) Select an attribute A to split the node from the independent attributes that best divides or splits the object in the node into subsets and create a decision tree node.
   c) According to the value of A, split the node.
   d) Stop the splitting process if either of the following condition is met, otherwise continue with step c:
   e) If the partition divides the data into different sets that belong to a single class no other node needs splitting.
   f) If no other attributes in the sample D may be further divided.
   g) Classifier tree classifies the data items into two classes, which are Normal breast tissue and abnormal breast tissue.
   h) Based on the information from the abnormality class, the classifier again classifies the data in to two classes, which are benign condition and malignant condition.

3) The final prediction is done based on the result from the previous step, as a patient who does have or does not have the disease.

IV. EXPERIMENTAL RESULTS

The propitious experimental results obtained by a combination of Local Binary Pattern (LBP) and Classification Tree (CT) classifier. This experiment used the MIAS database and the dataset was divided into 7 different classes in total. From the 7 different classes, the abnormal classes again divided into 6 classes. The training process and testing were done through the two-phase method, the first phase classifies the data D into Normal class or abnormal class and in the second phase, the data from the abnormal class again classified into abnormality class (Benign/Malignant). In this work, done the preprocessing on the raw dataset for better results. Because of this preprocessing before the feature extraction helped to obtain correct texture patterns using LBP and CT classifier is used for discrimination. The dataset contains total abnormal classes are 270 and normal classes are 52 and benign are 209 and malignant are 61. Based on the experiment conducted, the CT classifier with LBP is more accurate in classifying the breast cancer tissues in mammograms with the rate of sensitivity and specificity is 98.52% and 92.3077% respectively for normal tissue and abnormal tissue classification. It gained 95.6938% sensitivity and 90.1639% specificity for classifying the benign and malignant lesion. Experimental results are tabulated and shown in different tables below. They are:

1. Tables 1 and 2 are shows the confusion matrix obtained for normal - abnormal and benign - malignant in Phase I and Phase II respectively.
2. Table 3 contains the values of overall Accuracy rate, Sensitivity and Specificity values of Proposed Method (LBP+CT).

Table I: Confusion Matrix of Normal and Abnormal classification using Proposed Method

<table>
<thead>
<tr>
<th>Classification</th>
<th>Normal</th>
<th>Abnormal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground Truth</td>
<td>4</td>
<td>48</td>
</tr>
</tbody>
</table>

Table II: Confusion Matrix of Benign class and malignant class using Proposed Method

<table>
<thead>
<tr>
<th>Classification</th>
<th>Benign</th>
<th>Malignant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground Truth</td>
<td>200</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>55</td>
</tr>
</tbody>
</table>

Table III: Confusion Matrix of overall Accuracy rate, Sensitivity and Specificity values of Proposed Method

<table>
<thead>
<tr>
<th></th>
<th>Phase I</th>
<th>Phase II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>97.5155</td>
<td>94.4444</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>98.5185</td>
<td>95.6938</td>
</tr>
<tr>
<td>Specificity</td>
<td>92.3077</td>
<td>90.1639</td>
</tr>
</tbody>
</table>

The overall accuracy of 98% in Phase I and 95% in Phase II was obtained on the MIAS database [22] and by comparing the previous research in mammogram classification [23] a comparative study shown in Figure 6.
By dividing the classification process into two major phases can be improved the accuracy of the method, they are normal tissue class - abnormal tissue class classification and benign condition class - malignant condition class classification.

Figure 6: Comparative graph based on Accuracy of proposed method (CT+LBP) with other methods (CNN+GLCM) and (CT+GLCM)

V. CONCLUSION

Breast cancer is an alarming disease due to a mutation in breast cells and it is one type of cancer among women which highly leads to their death. One of the most effective tools for early detection of breast cancer is mammography, which is a screening tool used to examine the human breast by using low-dose amplitude X-rays.

Computer-Aided Diagnosis (CAD) is used as an important tool to help the medical professionals for classifying breast tissues into a different class. Computer-Aided Diagnosis (CAD) can be used to reduce human error in reading the mammograms and it shows effective results in a classification of benign and malignant abnormalities. We provide a new classification approach to detect the abnormalities in mammograms using the Local Binary Pattern and Classification Tree (CT) classifier.

LBP is commonly used for recognition techniques by extracting the texture feature. The local binary pattern operator transforms an image into an array or image of integer labels describing details of textures in the image. The image labels from the LBP can be used directly or their statistics for further analysis. These extracted texture feature patterns are used to detect breast cancer by employing the Classification Tree (CT) classifier. Results are acceptable and that shows a combination of LBP and CT is better for the classification of mammograms. This technique can be used to classify digital mammograms to detect breast cancer in the future.

ACKNOWLEDGMENT

First of all, we thank The Almighty God for blessing us with his grace and supporting us throughout this endeavor. We extend our sincere and heartfelt thanks to Dr. Joy Mampilly, for providing us with the right guidance and advice at the crucial junctures and for showing us the right way.

REFERENCES

1. ACS, Cancer facts and figures, (2010).
16. Lior Rokach and Oded Maimon, “DECISION TREES”, DATA MINING AND KNOWLEDGE DISCOVERY HANDBOOK.