Malignant Detection of Breast Nodules On BIRADS-Based Ultrasound Images Margin, Orientation, and Posterior

Yuli Triyani, Wahyuni Khabzli, and Wiwin Styorini

ABSTRACT Breast cancer has the largest prevalence in the world in 2020, with 2,261,419 cases or 11.7%. It is also the leading cause of cancer death, accounting for 6.9% of all cancer deaths. Asia and Indonesia have the greatest prevalence and mortality rates. This is an urgent issue that must be addressed. Ultrasonography (USG) is advised for assessing the features of breast nodules. Breast nodules on ultrasound pictures are interpreted using the Breast Imaging, Reporting, and Data System (BIRADS) category, which has five features. Yet, the probability of a False Positive Result (FPR) on ultrasound imaging is relatively high. Computer Aided Diagnosis (CAD) was created to reduce FPR. However, CAD research based on many BIRADS traits is currently margined. As a result, based on three BIRADS characteristics, namely the margin, posterior, and orientation aspects, this study aims to proposed the method for diagnosing breast nodule malignancy. The proposed method consists of 4 stages, namely, pre-processing, automatic segmentation, features extraction, and classification. Pre-processing adaptive median filter maximum window size is 11 pixels, linear histogram normalizing, and Reduction Anisotropic Diffusion (SRAD) filter were used to construct the method. The neutrosophic watershed method was used in the suggested automatic segmentation. Based on the nodule's margin, orientation, and posterior, 10 features were proposed: nodule width, gradient, slenderness, margin sharpness, shadow indicators, skewness, energy, entropy, dispersion, and solidity. MLP is a classification approach. The test used 94 nodule pictures and yielded an accuracy of 88.30%, a sensitivity of 82.35%, a specificity of 91.67%, a Kappa of 0.744, and an AUC of 0.865. As a result, it is feasible to conclude that the proposed method is capable of detecting malignancy in breast nodules in ultrasound images. To make the proposed method more reliable in the future, automatic ROI can be developed.

INDEX TERMS: Nodule, BIRADS, Margin, Orientation, Posterior

I. INTRODUCTION

Cancer is a group of diseases characterized by uncontrolled growth and the spread of abnormal cells [1]. After cardiovascular disease, cancer is the second greatest cause of mortality [2]. FIGURE 1 shows the global cancer prevalence/prevalence in 2020. Overall cancer cases in the world reached 19,292,789 cases, with breast cancer having the highest prevalence at 2,261,419 instances or 11.7%. Breast cancer is also the leading cause of death in the world, accounting for 6.9% of all deaths in 2020. Asia is the continent with the highest frequency, at 45.4%, and the highest mortality rate, at 50.5% [3].

Breast cancer is the most prevalent cancer diagnosed in women in the majority of nations, accounting for 140 of 184 [4]. According to GLOBOCAN, Asian countries account for the greatest proportion of cancer cases worldwide. According to data from the World Health Organization's Global Burden of Cancer Study (Globocan) report, there were 213,546 malignancies affecting Indonesian women, with breast cancer accounting for 30.8%, as shown in FIGURE 1 [5]. This is an urgent issue that must be addressed. Mammmography and ultrasonography are the most commonly used breast cancer radiological imaging (USG). Breast nodules on ultrasound pictures are interpreted by doctors using the Breast Imaging, Reporting & Data System (BIRADS) categorization [6].
The BIRADS characteristics used to detect breast nodule malignancy are shown in TABLE 1. The likelihood of a False Positive Results (misinterpretation) is, however, relatively significant [7]. Computer Assisted Diagnosis (CAD), a system that integrates digital image processing techniques with radiology, was created to reduce FPR [8]. Numerous research [9–12] have created approaches for identifying and classifying breast nodules based on just one BIRADS characteristics, such as shape, texture, margins, or posterior features. However, research on nodule classification using more than one BIRADS trait is still in its early stages. As a result, this study suggests a strategy for diagnosing breast nodule malignancy based on three BIRADS characteristics: margin, postThis algorithm can be part of CAD system development to detect breast nodule malignancy. So that it becomes a tool or as a second opinion provider for radiologists in diagnosing breast nodules based on ultrasound images. This algorithm can also be used as a reference for developing malignancy detection algorithms for other organs that have almost the same characteristics as breast nodules. So that it is expected to produce a breast nodule malignancy detection algorithm on ultrasound images with good performance.

II. METHOD
The proposed method consists of 4 stages, namely, pre-processing, automatic segmentation, features extraction, and classification. Each stage will be tested using the appropriate parameters. The research method to be designed is as shown in FIGURE 3, which consists of several processes including:

1) Pre-processing: aims to improve the quality of the ultrasound image, reduce noise, and remove artifacts or markers so that the optimal image is used at a later stage
2) Segmentation: useful for separating objects to be processed from the background. In this case, of course, the nodule area will be processed
3) Features Extraction: this is a technique for obtaining nodule characteristics that are thought to accurately represent nodular orientation characteristics.
4) Classification: to categorize images in a certain class according to BIRADS, namely malignant or benign.

The dataset used in this study consisted of 94 nodules on ultrasound images and their radiological interpretation results. The classification results were compared to the radiologist's readings as well as the nodule's Anatomical Pathology (PA) test results. Hence, the level of accuracy, specificity, sensitivity, kappa and AUC of the method provided in this study may be seen.
(RGB) is then converted to grayscale, and the markers/labels in the image are removed. Pre-processing employs an adaptive median filter, with the window size continuously increasing until the desired size is attained [13]. This study’s maximum window size is 11 pixels, intending to remove markers/labels from ultrasound pictures.

A linear histogram normalizing approach is also used to boost image contrast [14]. The first stage of histogram normalization is determining the Range using Eq. (1).

\[
\text{Range} = |\text{Imax} - \text{Imin}|
\]  

(1)

Scale the Pixels with equation (2).

\[
l'(x, y) = \left( \frac{I(x,y)}{255 \text{Range}} \right) - \left( \text{Imin} \left( \frac{255}{\text{Range}} \right) \right)
\]  

(2)

If \( l'(x,y) \) is less than 0, then \( l'(x,y) = 0 \). If \( l'(x,y) \) is greater than 255, then \( l'(x,y) = 255 \). One of the best despeckle noise techniques applied to breast ultrasound images is the SRAD filter [15]. Yu, et al [16] use the Anisotropic Diffusion (AD) principle in combination with Adaptive Mean Filters and Anisotropic Diffusion. Both filters are isotropic diffusion from the Partial Differential Equation (PDE) framework. The AD equation that is useful for suppressing noise in ultrasound images is called Reduction Anisotropic Diffusion (SRAD). SRAD technique based on Partial Differential Equation (PDE) and MMSE [17]. SRAD replaces gradient-based edge detection in anisotropic diffusion PDEs with variants of instantaneous coefficients [18]–[20]. The instantaneous coefficients use an edge detector which is a combination of the magnitude of the gradient and the Laplacian operator.

B. SEGMENTATION

Previous research has extensively investigated segmentation methods [21–23]. This study employs an automatic segmentation technique with neutrosophic and watershed methods to improve the consistency and accuracy of segmentation outcomes [10].

The intuitionistic set, classical set, fuzzy set, paraconsistent set, dialetheist set, paradoxist set, and tautological set are all generalizations of neutrosophy. Every element \( x \) (T, I, F) where \( t, i, \) and \( f \) are real values taken from the set T, I, and F with no restrictions on T, I, and F, or their sum is \( n = t+i+f \). Neutrosophic sets generalize:

1. The intuitionistic set supports an incomplete set of theories (0<\(n<100\), 0<\(t,i,f<100\)).
2. Fuzzy logic (n=100, i=0 and 0<\(t,f<100\)).
3. Boolean logic (n=100, i=0, and \(t,f\) between 0 or 100).
4. Multi-valued logic (0<\(t,i,f<100\)).
5. Paraconsistent logic (n>100, dan \(t,f<100\)).

The watershed algorithm makes many contributions to optimizing image segmentation [24]. The gray image is described as a topological surface (landscape) from a location determined based on the \(x,y\) coordinates. The height of the location is related to the intensity of the image. Dam shapeation (watershed line) is the most important thing in the watershed transshapeation process. The shapeation of the dam uses morphological dilation on binary images, which are members of the two-dimensional integer space \(Z^2\).

C. FEATURES EXTRACTION

Features are characteristics that are assigned to objects. The amount of categorization accuracy is determined by the derived features. The diameter of the nodule is calculated using the "Brute force" approach, which involves measuring the largest distance between two spots on the object's edge.

\[\text{Width} = \text{the longest line that connects two pixels on an object that is perpendicular to the maximum length of the object} \] [25]. The width of the object can be obtained from the diameter as shown in Error! Reference source not found. After the two points with the longest distance are obtained, the gradient of the line through the two pixels is calculated using Eq. (3):

\[
\text{grad}1 = \frac{y2 − y1}{x2 − x1}
\]  

(3)

where \(x,y\) pixel intensity position

Furthermore, the line perpendicular to the line with a gradient of \(\text{grad}1\) has a gradient shapulated in Eq. (4):

\[
\text{grad}2 = -\frac{1}{\text{grad}1}
\]  

(4)
The slenderness of the shape is the ratio between the width and length of nodule, which is expressed by the following shpeula (Eq. (5)):

\[
\text{Slenderness} = -\frac{\text{width}}{\text{length}}
\]  

(Eq. (5))

Solidity and dispersion geometry characteristics of the segmented binary images are also analyzed. Solidity is the ratio between the area of the nodule and the convex hull as shpeualated in Eq. (6). Meanwhile, dispersion is the irregularity of the nodule by calculating the ratio between the length of the major axis and the area of the nodule as in Eq. (7) [26].

\[
\text{Solidity} = \frac{\text{Nodule Large}}{\text{Convex Hull}}
\]

(Eq. (6))

\[
\text{Dispersion} = \frac{\text{Length of Axis Major}}{\text{Nodule Large}}
\]

(Eq. (7))

where convex hull is the minimum convex polygon in nodule, nodule large is the area or number of pixels of the segmented nodule, and length of axis major is the farthest distance between 2 pixels around the nodule.

Characteristics of nodule margin were analyzed using margin sharpness [10]. Sharpness is the magnitude of the mean gradient from the nodule margin. If the pixel intensity at position x,y is represented as f(x,y), then the partial derivatives in the x and y directions are expressed in equations (9) and (10), respectively. The magnitude gradient and sharpness features can be obtained based on Eq. (11) and Eq. (12).

\[
G_x = \frac{\delta f(x,y)}{\delta x} = f(x + 1, y) - f(x, y)
\]

(Eq. (9))

\[
G_y = \frac{\delta f(x,y)}{\delta y} = f(x, y + 1) - f(x, y)
\]

(Eq. (10))

\[
G[f(x,y)] = \sqrt{G_x^2 + G_y^2}
\]

(Eq. (11))

\[
\text{Sharpness} = \text{mean} \,(G[f(x,y)])
\]

(Eq. (12))

Where Gx is partial derivatives f(x,y) in the x, Gy is partial derivatives f(x,y) in the y.

Characteristics of the posterior nodule are proposed using the posterior shadow indicator [10]. The posterior shadow indicator detects the lower middle 2/3 of the nodule. Whereas 1/6 of the area on the lower right and lower left of the nodule is ignored to avoid edge shadows as shown in Error! Reference source not found. If the posterior area is segmented as a nodule equal to more than 50%, then the shadow indicator is 1. Conversely, if it is less than 50%, then the shadow indicator is 0.

The margin area is divided into N radial sectors that go through the nodule's center. The average pixel intensity for the inner and outer edge areas is determined in each sector. FIGURE 8 shows that for nodules with a shadow indicator of 1, only the top of the nodule is separated into sectors. Each pair was subjected to a two-tailed Welch's t-test to determine significance. If the p-value is less than 0.001, the sector is classified as having different margins. The equation is used to calculate the sharpness of the margin (11). In this case, n denotes the number of sectors with varying margins, and N is the entire sector (Eq. (13)).

\[
\text{Margin sharpness} = \frac{n}{N} \times 100\%
\]

(Eq. (13))

The proposed algorithm also uses first-order statistical features based on histograms, including skewness, energy, and entropy. Histograms are a simple way to estimate the Probability Density Functions (PDF) of an image.

D. CLASSIFICATION

A classification that aims to distinguish malignant and benign nodules based on previously extracted features. The proposed classification stage uses Multi-Layer Perceptron (MLP). Figure 9 shows the MLP network architecture.
MLP is supervised so it requires a training process to get the optimal weight used in the testing process. The training process is carried out utilizing backpropagation. Optimal weight will produce high accuracy. This weight will continue to be updated during the training process until the desired layer structure has been achieved.

III. RESULT

A. PRE-PROCESSING

In this study, the input image is an ultrasonographic RGB image of the breast with nodules. The image includes the results of the anatomical pathology diagnosis, which included 34 malignant nodules and 60 benign lesions. The RoI procedure is carried out following the radiologist's directions. The RoI method generates a distinct image of the nodule area. The RGB image of the ROI is transformed into a grayscale image.

FIGURE 9. MLP network architecture

FIGURE 10 depicts an example of the RoI procedure results.

Additional signs/artifacts are generally added to ultrasound scans by radiologists to make it easier for doctors to read the images. This, however, may obstruct further image processing. Using a median adaptive filter of size 11 [27], this marker/label is unmarked.

FIGURE 10. RoI Process (a) Ultrasound image (b) RoI results (c) Grayscale image

FIGURE 11 compares the photos before and after the unmarking process. It is possible to delete markers/labels without generating a blur effect.

FIGURE 11. The Image Results of unmarked (a) before unmarking (b) after unmarking

FIGURE 12. Image of pre-processed results: (a) UGS Breast Image (b) Normalized Breast Ultrasound Image (c) SRAD Result

Evaluation is also carried out on normalized breast ultrasound images. The purpose of image normalization is to sharpen the image so that the image margin with blurred nodules can be firmer, which is expected to increase the results during the segmentation process.

FIGURE 12 shows an example of an ultrasound image of the breast, the results of normalization, and the results of the ultrasound image. Visually it can be seen that the SRAD image produces an image that resembles the normalized ultrasound image.

The quantitative evaluation of the SRAD normalized output images is shown in TABLE 2. The variance value is near the variance value of the normalized ultrasound image and can keep the mean value close to the mean value of the normalized image. The skewness value is positive, the kurtosis value is less than three, the image contrast has increased significantly, and the entropy value has increased.

<table>
<thead>
<tr>
<th>Texture Feature</th>
<th>SRAD</th>
<th>Normalized USG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>86.60</td>
<td>86.60</td>
</tr>
<tr>
<td>Variance</td>
<td>3,645.95</td>
<td>3,645.95</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.62</td>
<td>2.62</td>
</tr>
<tr>
<td>Contrast</td>
<td>584.72</td>
<td>584.72</td>
</tr>
</tbody>
</table>
B. SEGMENTATION
Quantitative testing of segmentation results was only carried out on nodules with firm margin because the ground truth results provided by a radiologist on the blurred image were interrupted at the edges of the identified nodule. These sections are difficult to distinguish from the tissue around the nodule and tend to coalesce. A comparison of the application of normalization and without normalization was also carried out.

TABLE 3 shows the quantitative evaluation of the breast ultrasound image segmentation results. In general, quantitative evaluation is accomplished by comparing area- and contour-based differences between segmented binary images and ground truth images.

The dice coefficient, accuracy, and Jaccard parameter values are getting closer to 100% indicating the more accurate the segmentation results are. It can be concluded that the application of normalization can improve segmentation performance. The closer the Hausdorff and MSSD values are to 1, the more accurate the resulting segment results are.

C. FEATURES EXTRACTION
Based on the margin, posterior, and orientation of the features extracted in this study, ten features were hypothesized to indicate the characteristics of nodule malignancy. TABLE 4 shows the statistical feature values of the malignant (M) and benign (B) nodules.

D. CLASSIFICATION
Classification uses the MLP method with a structure as shown in FIGURE 13. MLP consists of 1 hidden layer consisting of 6 nodes. The evaluation uses 3-fold cross-validation so that the shape features are produced as shown in TABLE 4.

It could be concluded that the proposed method, which employs ten features, has an accuracy of 88.3% in distinguishing between malignant and benign nodules on ultrasound images.

IV. DISCUSSION
The proposed method consists of 4 stages, namely, pre-processing, automatic segmentation, features extraction, and classification. The Pre-processing stage is applying manual...
Rol, RGB to greyscale, adaptive median filter, normalization, and SRAD filter. Pre-processing stage succeeded in improving the quality of the ultrasound image, reduce noise, and remove artifacts or markers. The neutrosophic watershed method was used in the proposed automatic segmentation. This segmentation method can separate the nodule area from the background automatically with an accuracy of 90.88%. Based on the nodule's margin, orientation, and posterior, ten features are proposed: nodule width, gradient, slenderness, margin sharpness, shadow indications, skewness, energy, entropy, dispersion, and solidity. Classification uses the MLP method consists of 1 hidden layer consisting of 6 nodes. The proposed method has an accuracy of 88.3% in distinguishing between malignant and benign breast nodules on ultrasound images.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (%)</td>
<td>88.30 %</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>82.35 %</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>91.67 %</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>84.85 %</td>
</tr>
<tr>
<td>NPV (%)</td>
<td>90.16 %</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.7449</td>
</tr>
<tr>
<td>ROC Area</td>
<td>0.865</td>
</tr>
</tbody>
</table>

However, this study still uses manual Rol to determine areas with nodules according to directions from radiologists. Automatic Rol needs to be developed so that this algorithm can be better. Thus this algorithm can work only with ultrasound image input. Several studies [9–12] have developed approaches to classify breast nodules based on only one BIRADS characteristic, such as shape, texture, margin, or posterior features. The results of the classification are the type of each characteristic, neither malignant nor benign. This study developed an algorithm to classify nodules as malignant or benign by analyzing 3 BIRADS characteristics at once, namely margin, orientation, and posterior. So that it can provide information that can be directly used by radiologists as a second opinion in diagnosing nodules on ultrasound images. It can help and increase radiological confidence to diagnose breast nodule on ultrasound images and reduce FPR.

V. CONCLUSION

Based on margin, orientation, and posterior characteristics, this study suggests a method for detecting malignancy in breast nodules using ultrasound imaging. Pre-processing adaptive median filter, normalization, and SRAD filter were used to construct the method. The neutrosophic watershed method was used in the proposed automatic segmentation. Based on the node's margin, orientation, and posterior, ten features are proposed: nodule width, gradient, slenderness, margin sharpness, shadow indications, skewness, energy, entropy, dispersion, and solidity. MLP is a classification method. The test used 94 nodule pictures and yielded an accuracy of 88.30%, a sensitivity of 82.35%, a specificity of 91.67%, a Kappa of 0.7449, and an AUC of 0.865. As a result, it is feasible to conclude that the proposed method is capable of detecting malignancy in breast nodules in ultrasound images. To make the proposed method more reliable in the future, automatic Rol can be developed.

REFERENCES


AUTHORS BIOGRAPHY

**YULI TRIYANI** was born in Pekanbaru, Riau, Indonesia in 1982. She received a associate degree from Politeknik Caltex Riau in 2005, a bachelor's degree from the Sepuluh November Institute of Technology in 2013 and a master's degree from the Universitas Gadah Mada in 2017. From 2006 to 2011, she was an Assistant laboratory at Politeknik Caltex Riau. Since 2009, she has been a lecturer at Politeknik Caltex Riau. Her research interests include medical image processing dan telecommunication systems.

**WAHYUNI KHABZLI** was born in Pekanbaru, Riau, Indonesia in 1984. He received a bachelor's degree from the Sepuluh November Institute of Technology in 2008 and a master's degree from the Bandung Institute of Technology in 2014. Since 2018 the author has begun to focus on research in the field of image processing in the biomedical field. Starting from Texture Analysis and Classification of Echo Pattern Characteristics in Ultrasound Images of the Cervix in 2018, Automatic Detection of Diabetic Retinopathy in Retina Imagery in 2019, Design and Development of a Digital Signal Processing (DSP) Based Cardiac Detection Tool in 2021, and Detection of breast tumor malignancy based on 3 birads characteristics on ultrasound images (USG) in 2022.

**WIWIN STYORINI** was born in Banyuwangi City, East Java province, Indonesia in 1982. She received a bachelor's degree from Brawijaya University in Malang and a master's degree in Engineering from Bandung Institute of Technology, in 2014. Since 2008 until now, she has been a lecturer at Politeknik Caltex Riau. Since 2016, her research focus has been on digital signal processing and image processing. Several research studies have been published in accredited SINTA journals.

Homepage: jeeemi.org