IT Journal Research and Development (ITJRD)
Vol.7, No.2, March 2023, E-ISSN : 2528-4053 | P-ISSN : 2528-4061
DOI : 10.25299/itjrd.2022.11563

Toward Better Analysis of Breast Cancer Diagnosis: Interpretable AI for Breast Cancer Classification

Alifia Revan Prananda[1], Eka Legya Frannita[2]
Department of Information Technology, Faculty of Engineering, Universitas Tidar[1]
Department of Leather Product Processing Technology, Politeknik ATK Yogyakarta [2]
revan@untidar.ac.id[1], eka.legya@atk.ac.id[2]

Article Info

Article history:

Received Des 30, 2022
Revised Jan 12, 2022
Accepted Feb 9, 2023

Keyword:

Breast cancer
Classification
Deep learning
Interpretable method

ABSTRACT

Recently, some countries have been distressing with the increasing number of breast cancer cases. Those cases were extremely increased in every year. Practicaly, the increasing number of patients was caused by the manual examination. Recently, some researchers have been done in the development of AI method for solving this problem. However, AI itself still has limitation since it worked in the black-box approach which was difficult to be trusted. Thus, to overcome those problems, we proposed a method that was able to classify breast ultrasound images into two classes (benign and malignant) and able to explain how the prediction was made. Our proposed method consisted of four processes i.e., pre-processing step, development of CNN model, interpretable step and evaluation. In this research work, our proposed method performed into 780 breast ultrasound images divided into three classes (133 normal, 210 malignant, and 437 benign). In the training process, our proposed method obtained training accuracy of 0.9795, training loss of 0.0675. The validation process obtained validation accuracy of 0.8000 and validation loss of 0.5096. While, in the testing process, our proposed method achieved accuracy of 0.7923. In the interpretable process using LIME, the LIME result is covered by doctor visualization. It was indicated that LIME was suitable enough in visualizing the important features of breast cancer severity. Regarding to the results, our proposed method has a potensial to be implemented as an early detection method for classifying malignancy of breast cancer in order to help the doctor in the screening process.

Corresponding Author:

Alifia Revan Prananda
Department of Information Technology,
Faculty of Engineering, Universitas Tidar
Jl. Kapten Suparman No.39, Tuguran, Potrobangsan, Kec. Magelang Utara, Kota Magelang, 56116
Email: revan@untidar.ac.id

1. INTRODUCTION

Breast cancer became one of the most shocking issues in the last decade since it has been extremely increased in every year and it was expected to be the leading cause of death in women [1]. The recent increasing number of breast cancer cases reported by SEEM (Surveillance, Epidemiology, and End Results) in the end of 2019 has shown new breast cancer cases around of 268,000 and found 41,760 deaths [2]. This number has been increased from the previous report in which in 2018 American Cancer Society estimated new cases of breast cancer around 266,000 causing 63,960 deaths [3]. According to those facts, giving more awareness to cancer becomes very important in order to reduce the increasing number of patients.

Regarding the large number of breast cancer cases, one of the problems that cause the increasing number of it was that the clinical problem in breast cancer examination. The clinical problem was occured since breast cancer was examined manually using medical instrumentation such as ultrasound, MRI, mammogram, etc. This manual procedure commonly caused some difficulties such as very tedious, time-consuming, need more thoroughness, and highly depends on the doctor's skills and experiences [4]–[6] Thus, an alternative way was needed to overcome those problems.

Recently, some medical studies in breast cancer have been successfully conducted and performed outstanding results. Research that was conducted by Zeimarani et at. [7] proposed a deep convolutional neural network for classifying breast lesion in ultrasound images. This study proposed a modification of CNN by adding a few hidden layer followed by applying reqularization techniques due to small dataset problem and to enhance the classification performance. Those modified CNN was then applied using 5-fold cross-validation divided in 80:20 proporsion of data in which the dataset consisted of 641 ultrasound images (413 benign lesion and 228 malignant lesion). Their proposed method achieved accuracy of 92.05% and AUC of 0.97.

Zheng et al. [8] mathematically modified Deep Learning assisted Efficient Adaboost Algorithm (DLA-EABA) with better computation. Practically, the proposed method comprised of several convolutional layers, LSTM and max-pooling layer used for extracting the features. While, the classification process was conducted in fully connected layer followed by softmax layer for estimating the error. Their proposed method successfully achieved accuracy of 97.2%, sensitivity of 98.3% and specificity of 96.5%. Latif et al. [9] proposed CNN model completed with despeckling process to reduce speckle noises and to increase the performance. Their proposed method was began by applying CNN as a denoising process, afterwards they applied another CNN as a classification process. This method successfully achieved accuracy of 99.89%. Gong et al. [10] proposed multi-view deep neural network support vector machine (MDNNSVM) for breast cancer classification using bi-modal ultrasound. This research successfully achieved accuracy of 86.36% and AUC of 0.9079. Hijab et al. [11] proposed deep convolutional neural network for classifying breast cancer which consists of three training models which are (i) baseline of CNN model trained from scratch, (ii) transfer-learning using VGG16, and (iii) fine-turning learning for handling the overfitting problem. Their proposed method successfully achieved accuracy of 97% and AUC of 0.98.

Wei et al. [12] proposed a machine learning model for classifying breast tumor. Their proposed method performed morphological and texture feature extraction followed by training the data using SVM to analyse the breast ultrasound images. This research work was applied in 1061 ultrasound images (589 malignant lesion and 472 benign lesion) and achieved accuracy of 87.32%. González-Luna et al. [13] proposed a comparison study in several machine learning approach for classifying breast cancer. They performed 2032 ultrasound images (1341 benign and 691 malignant) in seven machine learning methods which are RBNF, SVM, k-NN, LDA, random forest, multinomial logistic regression (MLR), AdaBoost. Their experiment concluded that LDA successfully outperformed other methods with accuracy of 89%, sensitivity of 82%, specificity of 93% and AUC of 0.95.

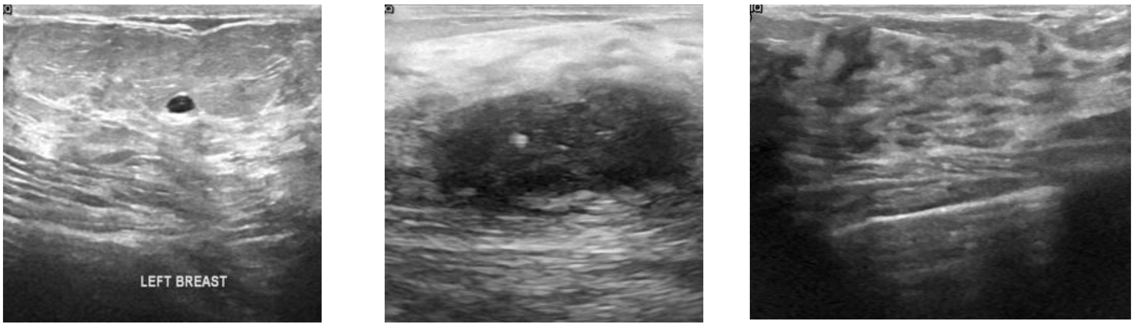
Regardless the outstanding performance of previous research works using both deep learning and machine learning, AI (machine learning and deep learning) itself worked in black-box approach in which it was difficult to understand how the method works and how they make a prediction. Consequently, the doctor was difficult to be trust to the method results. To overcome those problems, an explaination method was needed to complete the existing method with some justification of how the prediction can be made and to produce more valid result. Thus the doctor can be trust to the prediction results [14][15].

In this research work, we focused on the development of deep learning method which was able to classify breast ultrasound images into two classes (benign and malignant) and able to explain how the prediction was made. To address the research problem, we proposed the contributions by developing a CNN model followed by LIME (local interpretable model-agnostic explanations) [16] as an interpretable method. The detail of our proposed method was described as follow: data and methodology were explained in section two, result and discussion were shown in section three, and in the end of paper we described a limitation and future work as a conclusion.

2. RESEARCH METHOD

2.1. Data

The dataset was provided by Baheya Hospital for Early Detection & Treatment of Women's Cancer, Cairo, Egypt [17]. The dataset was collected from 600 female patients aged between 25 and 75 years old. The dataset was a cropped imaged consisted of 780 breast ultrasound images divided into three classes (133 normal, 210 malignant, and 437 benign) and completed with segmentation ground truth.



(a) (b) (c)

Figure 1. Example of dataset: (a) benign lesion, (b) malignant lesion, and (c) normal breast ultrasound.

2.2. Proposed Method

In this study, the proposed method for classifying breast cancer consists of four processes which are pre-processing, train and test model using CNN, explain the result using LIME, evaluation. Figure 2 shows flowchart of the proposed method.

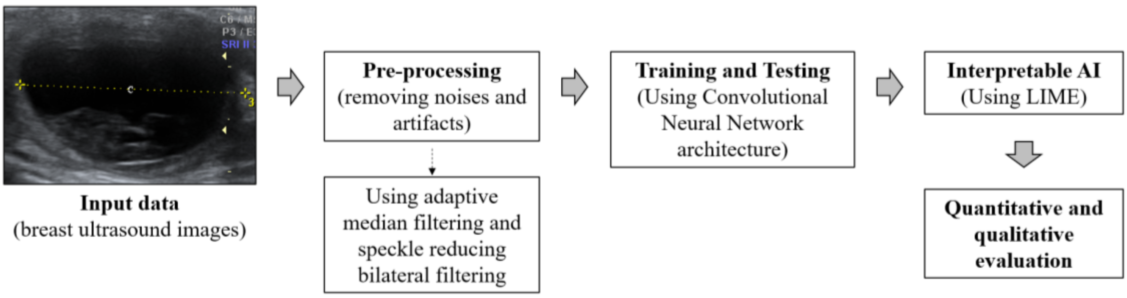


Figure 2. Block diagram of the proposed method

2.2.1 Pre-Processing

Pre-processing step was used to enhance the quality of input image. As depicted in Figure 1, the dataset consists of some noises for example is label written in the image (see a label in Figure 1(a) written in “left breast”). Actually, most of images have some labels in which it can be a distruction for the model to analyse the image. Hence, we conducted pre-processing at the first step to remove those problems. In this case, we applied adaptive median filtering [18]. Basically, adaptive median filtering is new version of median filtering [19] that was adaptively modified to increase the filtering performance. The modification was placed in the window size that can be changed adaptively. Thus, it was powerful to reduce non-impulse noises. Figure 3 shows how adaptive median filtering works. Beside labels, the dataset also has another noises like speckle noise. This noise was occured in the dataset since the dataset was collected using ultrasound modality. To reduce speckle noise, we performed speckle reducing bilatering filtering (SRBF) [18]. Figure 4 shows how SRBF works.

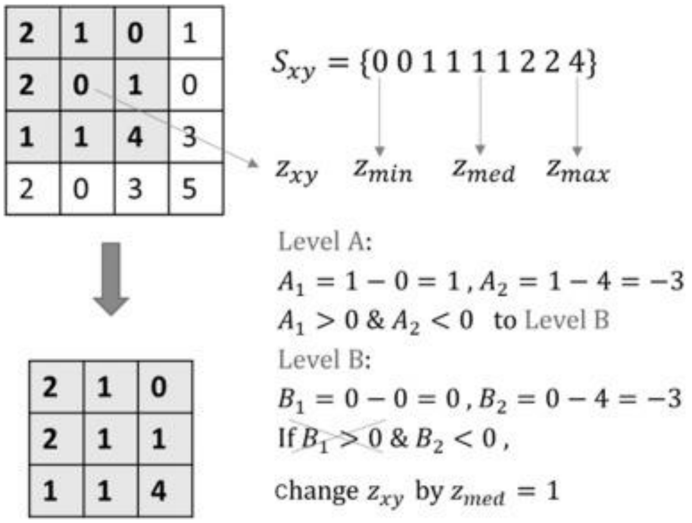


Figure 3. The workflow of adaptive median filtering [18].

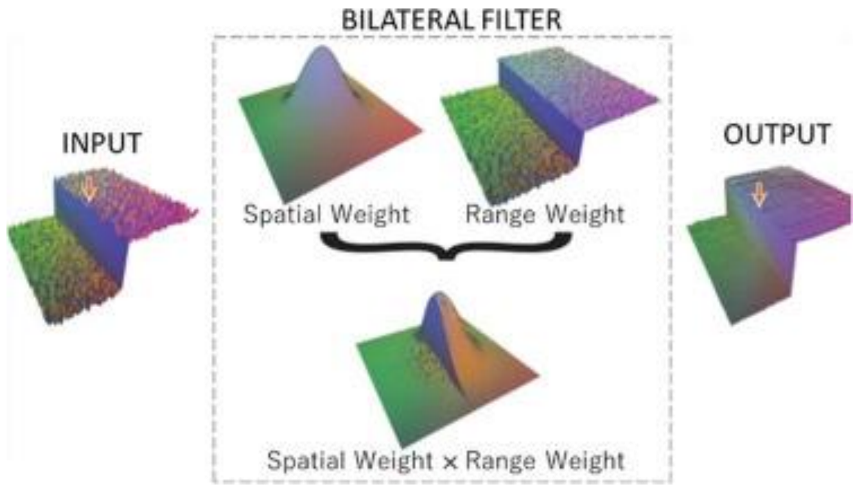


Figure 4. The workflow of SRBF [18].

2.2.2 Training and Testing Data Using CNN

In this step, we trained and tested the dataset using convolutional neural network proposed. The CNN model consisted of 12 layers. The first layer was convolutional layer with 16 filters and filter size of 3x3. In this layer, we performed ReLU activation function. The second layer was max-pooling layer with filter size of 2x2. Then, we continued with dropout layer with dropout rate of 0.25. The architecture was then continued with the second convolutional layer in which we used 32 filters size 3x3 followed by ReLU activation function. Same with the first convolutional network, the second convolutional network was followed by max-pooling with filter size of 2x2 and dropout layer with dropout rate of 0.25. Then, the architecture was continued with flatten layer used to flatten the 2D matrix into single column data. The next layer was two fully connected layers. The first fully connected layer was designed to produce 128 output in which used ReLU activation function. This fully connected layer was then followed with dropout layer to reduce unnecessary weight. The last layer was second fully connected layer with sigmoid activation function. The summary of the architecture is summarized in Table 1.

Table 1. Summary of CNN Architecture

Layer	Component	Kernel size	Activation
Layer 1	Layer input	224x224	-
Layer 2	Convolution	3x3	ReLU
Layer 3	Max pooling	2x2	-
Layer 4	Dropout	0.25	-
Layer 5	Convolution	3x3	ReLU
Layer 6	Max pooling	2x2	-
Layer 7	Dropout	0.25	-
Layer 8	Flatten	-	-
Layer 9	Dense	128	ReLU
Layer 10	Dropout	0.5	-
Layer 11	Dense	1	Sigmoid
Layer 12	Output	1	-

2.2.3 Interpretable AI Using LIME

This step was used to justify the results, hence the doctor can be trust to the prediction results. In this study, we used LIME (local interpretable model-agnostic explanations) [16] as an interpretable method. LIME worked by creating new instance permutated from the model prediction considering mean and standart deviation of the original data as mathematically formulated in Eq.(1), in which argmin L indicates that it would like to find the minimum losses considering the distance between original prediction model (f) and interpretable model (g). While, Ω(g) is used to keep the complexity of model (g) that becomes still be low. The workflow of LIME is illustrated in the Algorithm 1.

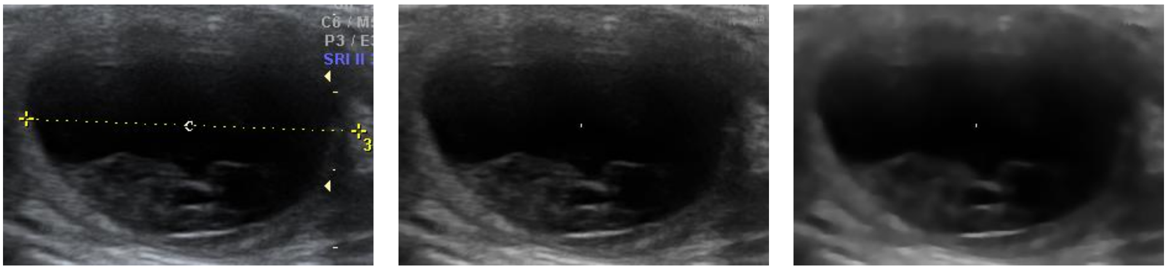
$$Explaination\ (x) = \underset{g \epsilon G}{\operatorname{argmin}}\, L(f, g, \pi_x) + \Omega(g) \tag{1}$$

Algorithm 1. The workflow of LIME [16]

1 : **input**: original trained model f, N samples of data, instance x, interpretable model x', sameness kernel π_x and explaination length K
2 : Z ← {} *Z is a dataset*
3 : **for** a € {1, 2, 3, ..., N} **do**
4 : z_i' ← sample_near_of(x') *using euclidean distance*
5 : Z ← Z ∪ $(z_i', f(z_i), \pi_x(z_i))$
6 : end for
7 : Ω ← K-Lasso (Z, K) *with z_i' as features, f (z) as the prediction target*
8 : **return** Ω

3. RESULTS AND ANALYSIS

As explained in the previous section, our proposed method consisted of four steps. The first step was pre-processing step used to enhance the quality of image. In this study, we performed adaptive median filtering followed with speckle reducing bilateral filtering. Result of this step can be seen in Figure 5.



(a) (b) (c)

Figure 5. Pre-processing result: (a) original image, (b) adaptive median result and (c) speckle reducing bilateral filtering result.

Result of pre-processing step was then resize in size of 224x224. The resized images were then divided into three part which are training data, validation data and testing data with proportion of 80:10:10. The training and validation data were then trained with 70 epoch and batch size of 128. The training process obtained training accuracy of 0.9795, training loss of 0.0675. The validation process obtained validation accuracy of 0.8000 and validation loss of 0.5096. While, in the testing process, our proposed method achieved accuracy of 0.7923. According to those results, we found that the performance decreases in the testing process with with the decreased accuracy of 0.01. The summary of training and testing performance is depicted in Figure 6.

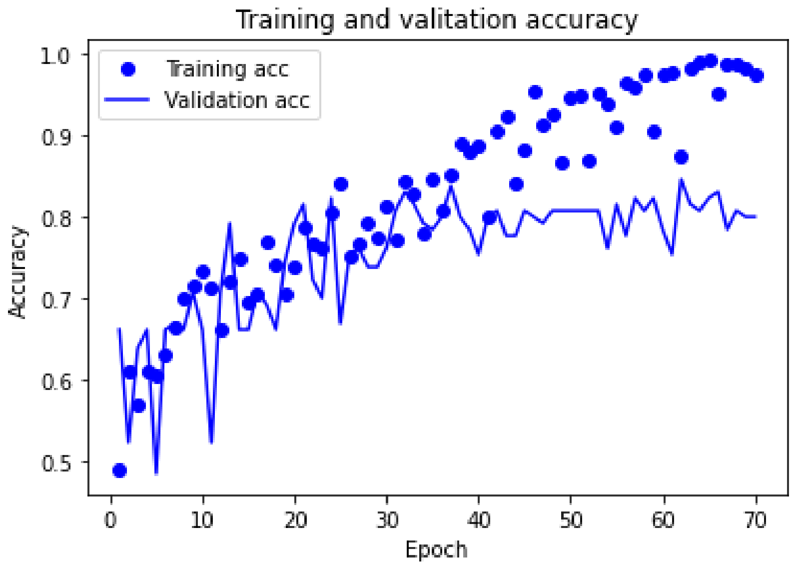
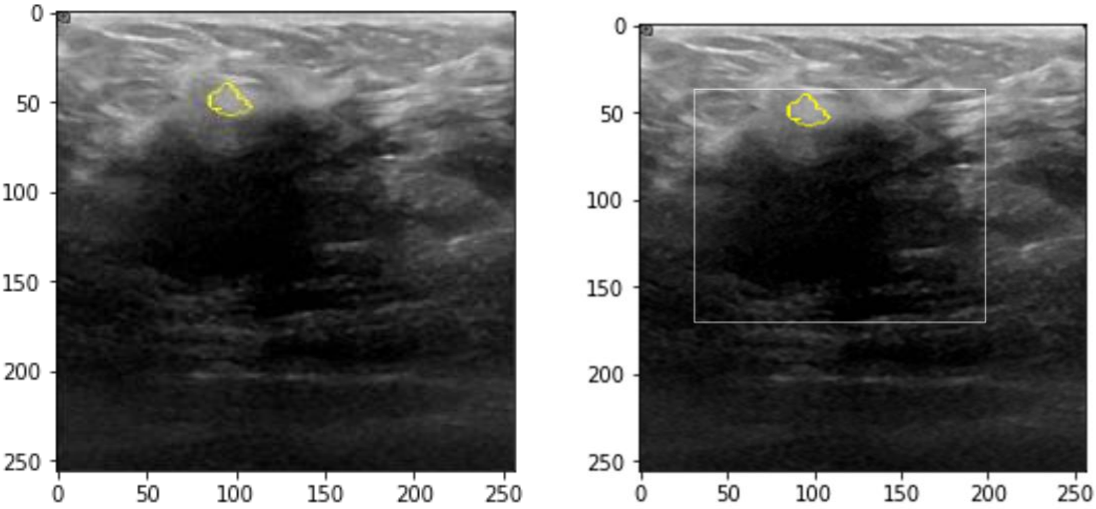


Figure 6. Summary of training and validation process: the training result is illustrated with blue dot, while the validation result is illustrated with blue line.

The trained model was then used as an initial model in the explaination step. In this step we performed LIME. LIME worked by finding the features that were similar to the target. The result of this step was the boundary area representing the importance features defining the cancer. Result of LIME can be seen in Figure 7(a). The visualization of LIME is indicated by the yellow area in Figure 7(a). According to Figure 7(a), the yellow area represents the important features indicating the level of breast cancer. The result of LIME was then compared to the doctor visualization. The comparison is represented in the Figure 7(b). According to Figure 8, it can be inferred that the LIME result is covered by doctor visualization. It was indicated that LIME was suitable enough in visualizing the important features of breast cancer severity. However, the area that was covered by LIME was quite small. It can be solved by adding more parameters when the data was trained and analyzed using LIME. This suggestion was difficult to be conducted in our study regarding to the limitation of our training device.



(a) (b)

Figure 7. Results of: (a) Visualization result of interpretable AI using LIME and (b) comparison result between doctor visualization (represented by white line) and LIME visualization (represented by yellow line).

4. CONCLUSION

In this research work, we evaluated the effectiveness of proposed convolutional neural network combined with LIME as an interpretable method for identifying the malignancy of breast cancer. Our proposed method consisted of four steps which were pre-processing step, training and testing of CNN, interpretable step, and evaluation. According to the result, our proposed method obtained training accuracy of 0.9795, training loss of 0.0675. The validation process obtained validation accuracy of 0.8000 and validation loss of 0.5096. While, in the testing process, our proposed method achieved accuracy of 0.7923. In the interpretable process using LIME, the LIME result is covered by doctor visualization. It was indicated that LIME was suitable enough in visualizing the important features of breast cancer severity. Regarding to the results, our proposed method has a potensial to be implemented as an early detection method for classifying malignancy of breast cancer in order to help the doctor in the screening process.

REFERENCES

[1] N. Goyal and M. Chandra Trivedi, "Breast cancer classification and identification using machine learning approaches," Mater. Today Proc., 2020, doi: https://doi.org/10.1016/j.matpr.2020.10.666.

[2] I. Hirra et al., "Breast Cancer Classification From Histopathological Images Using Patch-Based Deep Learning Modeling," IEEE Access, vol. 9, pp. 24273–24287, 2021, doi: 10.1109/ACCESS.2021.3056516.

[3] P. Kaur, G. Singh, and P. Kaur, "Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification," Informatics Med. Unlocked, vol. 16, p. 100239, 2019, doi: https://doi.org/10.1016/j.imu.2019.100239.

[4] C. Dromain, B. Boyer, R. Ferré, S. Canale, S. Delaloge, and C. Balleyguier, "Computed-aided diagnosis (CAD) in the detection of breast cancer," Eur. J. Radiol., vol. 82, no. 3, pp. 417–423, 2013, doi: https://doi.org/10.1016/j.ejrad.2012.03.005.

[5] T. Ha, Y. Jung, J. Y. Kim, S. Y. Park, D. K. Kang, and T. H. Kim, "Comparison of the

diagnostic performance of abbreviated MRI and full diagnostic MRI using a computer-aided diagnosis (CAD) system in patients with a personal history of breast cancer: the effect of CAD-generated kinetic features on reader performance," Clin. Radiol., vol. 74, no. 10, pp. 817.e15-817.e21, 2019, doi: https://doi.org/10.1016/j.crad.2019.06.025.

[6] C. Kaushal, S. Bhat, D. Koundal, and A. Singla, "Recent Trends in Computer Assisted Diagnosis (CAD) System for Breast Cancer Diagnosis Using Histopathological Images," IRBM, vol. 40, no. 4, pp. 211–227, 2019, doi: https://doi.org/10.1016/j.irbm.2019.06.001.

[7] B. Zeimarani, M. G. F. Costa, N. Z. Nurani, S. R. Bianco, W. C. D. A. Pereira, and C. F. F. C. Filho, "Breast Lesion Classification in Ultrasound Images Using Deep Convolutional Neural Network," IEEE Access, vol. 8, pp. 133349–133359, 2020, doi: 10.1109/ACCESS.2020.3010863.

[8] J. Zheng, D. Lin, Z. Gao, S. Wang, M. He, and J. Fan, "Deep Learning Assisted Efficient AdaBoost Algorithm for Breast Cancer Detection and Early Diagnosis," IEEE Access, vol. 8, pp. 96946–96954, 2020, doi: 10.1109/ACCESS.2020.2993536.

[9] G. Latif, M. O. Butt, F. Y. Al Anezi, and J. Alghazo, "Ultrasound Image Despeckling and detection of Breast Cancer using Deep CNN," in 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), 2020, pp. 1–5, doi: 10.1109/RIVF48685.2020.9140767.

[10] B. Gong et al., "BI-Modal Ultrasound Breast Cancer Diagnosis Via Multi-View Deep Neural Network SVM," in 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020, pp. 1106–1110, doi: 10.1109/ISBI45749.2020.9098438.

[11] A. Hijab, M. A. Rushdi, M. M. Gomaa, and A. Eldeib, "Breast Cancer Classification in Ultrasound Images using Transfer Learning," in 2019 Fifth International Conference on Advances in Biomedical Engineering (ICABME), 2019, pp. 1–4, doi: 10.1109/ICABME47164.2019.8940291.

[12] M. Wei, Y. Du, X. Wu, and J. Zhu, "Automatic Classification of Benign and Malignant Breast Tumors in Ultrasound Image with Texture and Morphological Features," in 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID), 2019, pp. 126–130, doi: 10.1109/ICASID.2019.8925194.

[13] F. A. González-Luna, J. Hernández-López, and W. Gomez-Flores, "A Performance Evaluation of Machine Learning Techniques for Breast Ultrasound Classification," in 2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), 2019, pp. 1–5, doi: 10.1109/ICEEE.2019.8884547.

[14] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K.-R. Müller, "Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications," Proc. IEEE, vol. 109, no. 3, pp. 247–278, 2021, doi: 10.1109/JPROC.2021.3060483.

[15] Y.-H. Wu et al., "JCS: An Explainable COVID-19 Diagnosis System by Joint Classification and Segmentation," IEEE Trans. Image Process., vol. 30, pp. 3113–3126, 2021, doi: 10.1109/TIP.2021.3058783.

[16] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[17] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," Data Br., vol. 28, p. 104863, 2020, doi: https://doi.org/10.1016/j.dib.2019.104863.

[18] H. A. Nugroho, Zulfanahri, E. Frannita, I. Ardiyanto, and L. Choridah, "Computer Aided Diagnosis for Thyroid Cancer System based on Internal and External Characteristics," J. King Saud Univ. - Comput. Inf. Sci., Jan. 2019, doi: 10.1016/j.jksuci.2019.01.007.

[19] E. R. Davies, "Chapter 3 - Image filtering and morphology," E. R. B. T.-C. V. (Fifth E. Davies, Ed. Academic Press, 2018, pp. 39–92.