Exploring the Best Parameters of Deep Learning for Breast Cancer Classification System

Andry Chowanda*
Computer Science Department, School of Computer Science, Bina Nusantara University
Jakarta 11480, Indonesia
Email: achowanda@binus.edu

***Abstract*—Breast cancer is one of the deadliest cancers in the world. It is essential to detect the signs of cancer as early as possible, to make the survival rate higher. However, detecting the signs of breast cancer using the machine or deep learning algorithms from the diagnostic imaging results is not trivial. Slight changes in the illumination of the scanned area can significantly affect the automatic breast cancer classification process. Hence, the research aims to propose an automatic classifier for breast cancer from digital medical imaging (e.g., Positron Emission Tomography or PET, X-Ray of Mammogram, and Magnetic Resonance Imaging (MRI) images). The research proposes modified deep learning architecture with five different settings to model automatic breast cancer classifiers. In addition, five machine learning algorithms are also explored to model the classifiers. The dataset used in the research is the Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM). A total of 2,676 mammogram images are used in the research and are split into 80%:20% (2,141:535) for training and testing datasets. The results demonstrate that the model trained with eight layers of Convolutional Neural Networks (CNN) (SET-8) achieves the best accuracy score of 94.89% and 93.75% in the training and validation dataset, respectively.**

***Index Terms*—Best Parameter, Deep Learning, Breast Cancer Classification System**

I. Introduction

BREAST cancer is the top three disease that leads to death in women in the world [1]. Most breast cancer patients are women, but men can also suffer from it with a small percentage. Several symptoms can indicate breast cancer. They are not limited to a lump, bloody discharge, or changes in the texture or shape of the breast or nipple. Moreover, the treatment and survival rate highly depend on the cancer stage. The higher the cancer stage is, the greater the risk of death will be. Hence, it is paramount to identify the sign of cancer as soon as possible. Several tests can be performed to indicate if there is a sign of breast cancer, namely: Magnetic Resonance Imaging (MRI) scan, Computerized Tomography (CT) scan, Positron Emission Tomography (PET) scan, mammogram (X-Ray), ultrasound, or biopsy [1].

Received: Feb. 07, 2022; received in revised form: Apr. 05, 2022; accepted: Apr. 05, 2022; available online: Sep. 14, 2022.
*Corresponding Author

Generally, oncology specialists examine the diagnostic imaging results and determine if there is a risk of cancer in the patients. The results can be no sign of cancer, benign, or malignant. The imaging results also can be interpreted as calcification or a mass. Calcification is generally caused by a small calcium deposit within the breast tissue. In comparison, a mass is the area of dense breast tissue. Both classifications can be a sign of cancer or not. The diagnostic process can be automated by using computer vision and machine (or deep) learning techniques. However, detecting the signs of breast cancer using the machine or deep learning algorithms from the diagnostic imaging results is not trivial. Slight changes in the illumination of the scanned area can greatly affect the automatic breast cancer classification process. Moreover, abundant diagnostic imaging results in breast cancer classification are also required to provide the best results for the automatic breast cancer classification model.

Therefore, the research proposes exploring machine learning and deep learning algorithms to model breast cancer classification from mammogram results. Five classic machine learning algorithms (K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Random Forest (RF), Adaptive Boosting (AdaBoost), and XGBoost (XGB)) and two deep learning architectures and algorithms (Convolutional Neural Network (CNN) and Deep Neural Network (DNN)) are explored in the research. Moreover, there are ten settings proposed. Five settings implement classic machine learning algorithms, and the rest uses deep learning architectures and algorithms.

II. Literature Review

Most works to classify or detect breast cancer implement classical machine learning or deep learning architectures. The classical machine learning techniques are still being implemented in problems or cases where the data are not big enough for the deep learning architectures to learn the features automatically. The previous research has implemented Naive Bayes and K-Nearest Neighbour to model breast cancer classification. The best result is achieved by the model trained with the KNN algorithms (97.51%) [2]. Then, another previous research has proposed a model trained using quadratic SVM in the Wisconsin breast cancer dataset and achieved an accuracy score of 98.1% [3]. Similar results are also achieved by the previous study [4], that the SVM algorithms perform the best among the other traditional machine learning algorithms. Moreover, Tree Augmented Naive Bayes (TAN), Boosted Augmented Naive Bayes (BAN), and Bayes Belief Network (BBN) algorithms are compared to model the breast cancer classifier applied to the Wisconsin breast cancer dataset. The results show that the model trained with TAN is superior in the accuracy score to the other models [5].

Then, a model trained with Principal Components Analysis and Artificial Neural Network to the Wisconsin breast cancer dataset is also proposed. It achieves 95% of accuracy [6]. Meanwhile, another previous study has compared several algorithms and architectures, namely Gated Recreant Unit + SVM, Linear Regression, KNN, Multi-Layer Perceptron, Single Layer Perceptron, and single SVM, to model breast cancer classifier of the Wisconsin breast cancer dataset. The results show that the Multi-Layer Perceptron achieves the best accuracy (99.04%) compared to the other algorithms and architectures [7]. Then, a novel hybrid learning architecture to extract and classify features is proposed based on breast cancer histology images [8]. There are several feature extraction techniques, such as Gray Level Co-occurrence Matrix (GLCM), Local Binary Pattern (LBP), and Local Ternary Pattern (LTP), to extract the features from the input images. The features are trained with several classical machine learning algorithms, such as RF, SVM, and Naive Bayes. The results demonstrate that with the combination architecture, the model trained achieved more accurate results.

Deep learning architectures and algorithms have been implemented to build or train models in several areas [9–15]. Most of the deep learning architectures and algorithms applied to model automatic breast cancer classification are the VGG [12], Residual Network (ResNet) [11], and GoogleNet or InceptionNet [13]. As an example, the breast cancer classification models are trained using several pre-trained CNN models, such as VGG, ResNet, and GoogleNet. The results show that the proposed model is superior to the existing models [16]. Then, previous studies also suggest breast cancer classification models using deep learning architectures, such as ResNet and InceptionNet (GoogleNet) [14, 17–19]. The results indicate that the models trained with the deep learning architectures achieve an accuracy score above 90% (93%) and an AUC improvement to 0.98.

Moreover, a hybrid deep learning model can increase the accuracy of the models and result in an AUC score of 0.70 [15]. Similarly, the proposed hybrid architectures of classical machine learning (e.g., SVM) and deep learning (e.g., VGG-16 and VGG-19) can learn multi-parametric MRI (mpMRI) in breast MRI imaging. Hybrid architectures have the best AUC score of 0.88 [20]. The previous study has also implemented a DNN with three layers (Convolutional and Pooling Layer, Fully Connected Layer, and Classification Layer) to model breast cancer detection using Infrared Thermal Imaging. It results in the best sensitivity score of 78% [21]. Last, the proposed deep learning framework can fuse and select the best features using CNN. It augments the data to provide better and more variation in the dataset. The data are trained in a pre-trained DarkNet- 53 architecture before being fused and trained. The best result is achieved by the proposed architecture with a 99.1% accuracy score [22].

III. Research Method

A. The Architectures and Experimental Settings

Five deep learning architectures are proposed in the research. Four architectures are based on CNN, and one is based on the DNN. Moreover, five classical machine learning algorithms are explored. Those are KNN, SVM, RF, AdaBoost, and XGB. Table I illustrates the settings proposed and explored in the research. The proposed algorithms and architectures are explored and fine-tuned. It results in the ten best settings of algorithms and architectures. The deep learning architectures have six to eight layers. Figure 1 demonstrates the proposed DNN architecture (SET-6). The proposed DNN has six layers with Rectified Linear Unit (RELU) activation functions of 16 units in the first layer, 32 in the second layer, 64 in the third to fifth layers, and 128 in the last layers. In the last layer (sixth layer), a Sigmoid activation function is applied to the layer. Figure 2 shows the proposed CNN architectures (SET-7 to SET-10). The proposed CNN architectures have seven (SET-7 and SET-8) to eight (SET-9 and SET-10) layers. The architectures have two configurations

TABLE I
EXPERIMENTAL SETTINGS.

No.	Name	Architecture	Layers	Dense
1	SET-1	KNN	N/A	N/A
2	SET-2	SVM	N/A	N/A
3	SET-3	RF	N/A	N/A
4	SET-4	AdaBoost	N/A	N/A
5	SET-5	XGB	N/A	N/A
6	SET-6	DNN	6	N/A
7	SET-7	CNN-7	7	256-128
8	SET-8	CNN-7	7	512-256
9	SET-9	CNN-8	8	256-128
10	SET-10	CNN-8	8	512-256

of dense layers. The first configuration has 256 units in the first layer and 128 units in the second layer. Meanwhile, the second configuration has 512 and 256 units in the first and second dense layers, respectively. Then, the parameter settings are determined using a trial-and-error strategy by adding and removing layers and changing the combination of the dense unit layers.

The dataset used in the research is the Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) [23]. The dataset has been widely used in several studies that involve breast cancer classification. A total of 2,676 mammogram images are used. Then, the dataset is split into 80%:20% (2,141:535) for training and testing datasets. The dataset is annotated with mass and calcification. A mass is the area of dense breast tissue, while calcification is generally caused by a small calcium deposit within the breast tissue. The mammography images are cropped and pre-possessed to increase the performance of the models. Figure 3 illustrates the examples of cropped mammography images with the label.

Moreover, the hyper-parameters used in all settings are identical. The batch size is set to 32, the learning rate of 0.01, and the maximum epochs to 200. The optimizer implemented in training is the Adam optimizer, and the loss set to the models is binary cross-entropy. The training implements early stopping and reduces the learning rate when the training model reaches a plateau. The patience hyper-parameter is set to 50. All the models are fine-tuned with the Adam optimizer, and the learning rate is set to 0.0000001.

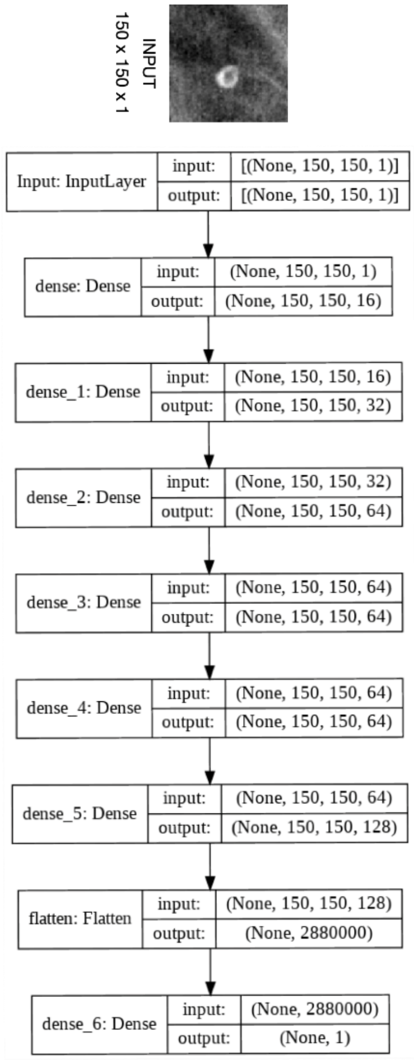

Fig. 1. Proposed Deep Neural Network (DNN) architectures.

IV. RESULTS AND DISCUSSION

Ten training settings are explored and evaluated in the research. Then, five deep learning architectures are proposed in the research. Four architectures are based on the CNN, and one is on the DNN. Moreover, five classical machine learning algorithms are also explored. The algorithms are KNN, SVM, RF, AdaBoost, and XGB. More than 30 hours of training in a Tesla K80 GPU result in the ten best models trained with ten settings of algorithms and architectures. Table II demonstrates the accuracy during the training and validation process. Eight layers of CNN achieve the best performance with 256-128 dense layers (SET-9). Then, the model trained with the architecture with the SET-9 setting achieves the best accuracy score of 94.89% and 93.75% for the training and validation dataset, respectively. The models trained with classical machine learning algorithms (i.e., SET-1 to SET-5) have relatively adequate results. The models get higher than 80% of training accuracy in all classical machine learning algorithms. However, the models trained with traditional machine learning algorithms suffer from over-fitting problems.

Meanwhile, the KNN algorithm obtains an accuracy

Cite this article as: A. Chowanda, “Exploring the Best Parameters of Deep Learning for Breast Cancer Classification System”, CommIT Journal 16(2), 143–148, 2022.

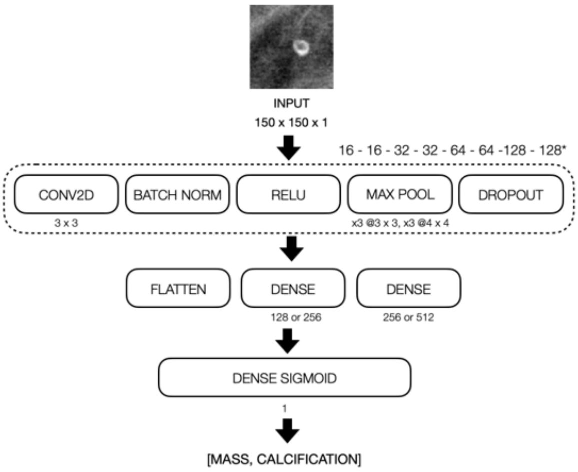

Fig. 2. Proposed Convolutional Neural Networks (CNN) architectures.

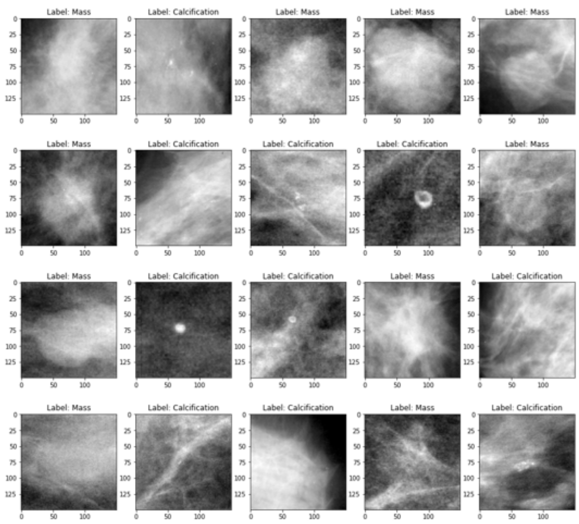

Fig. 3. Cropped dataset examples.

of 81.73% and 69.35% in the training and validation phase, respectively. The SVM algorithm achieves an accuracy of 84.42% and 74.4% in the training and validation phase, respectively. Then, the RF algorithm provides the accuracy of 100.0% and 75.3% in the training and validation phase, respectively. It makes the RF (SET-3) algorithm gets the best results among all the traditional machine learning algorithms. Moreover, the AdaBoost algorithm achieves an accuracy of 81.2% and 67.56% in the training and validation phase, respectively. Then, the XGB algorithm obtains an accuracy of 94.62% in the training and 74.7% in the validation phase.

The models trained with deep learning algorithms and architectures provide more stable results. All the

TABLE II
OVERALL RESULTS.

No.	Name	Training Accuracy	Validation Accuracy
1	SET-1	81.73%	69.35%
2	SET-2	84.42%	74.40%
3	SET-3	100.00%	75.30%
4	SET-4	81.20%	67.56%
5	SET-5	94.62%	74.70%
6	SET-6	75.57%	74.17%
7	SET-7	93.19%	92.70%
8	SET-8	92.81%	93.17%
9	SET-9	94.89%	93.75%
10	SET-10	94.88%	93.75%

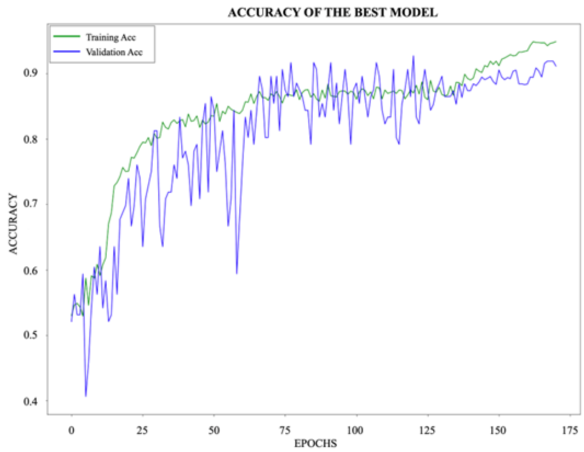

Fig. 4. Accuracy of the best model.

models trained with SET-6 to SET-10 are fined-tuned to enhance the training results. As a result, the models achieve higher than 90% of training accuracy in almost all deep machine learning architectures and algorithms (except for the DNN). The DNN architectures provide an accuracy of 75.57% and 74.17% in the training and validation phase, respectively. The seven layers with 256-128 dense units in CNN have an accuracy of 93.19% in the training and 92.70% in the validation phase. The seven layers with 512- 256 dense units in CNN obtain an accuracy of 92.81% in the training and 93.17% in the validation phase.

Moreover, the eight layers with 256-128 dense units in CNN achieve an accuracy of 94.89% and 93.75% in the training and validation phase, respectively. Next, the eight layers with 512-256 dense units in CNN obtain an accuracy of 94.88% and 93.75% in the training and validation phase, respectively. The best results are achieved by SET-9 (The eight layers with 256-128 dense units in CNN). There is no significant improvement between 256-128 and 512-256 dense units in both seven and eight layers in CNN.

Figure 4 illustrates the history of training and validation accuracy of the best model (i.e., SET-9). The

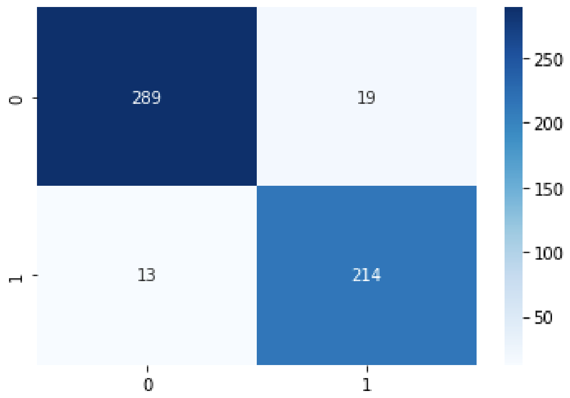

Fig. 5. Confusion matrix in best model.

training and fine-tuning process of SET-9 finishes in a total of 171 epochs. The training is stabilized in approximately 150 epochs. The validation accuracy is quite unstable at the beginning of the training. Nevertheless, the model results in the best accuracy of 94.89% and 93.75% during the training and validation phase, respectively. It shows that the model is most likely not overfitted. Moreover, Fig. 5 shows the confusion matrix of the best model trained with the SET-9 setting. In addition, the model trained with the setting of SET-10 provides quite similar results to the model trained with the SET-9 setting. However, SET-9 has a smaller number of parameters to be trained. Hence, the resources needed for SET-9 to train the model are smaller than SET-10.

In overall, the best model achieves the accuracy score of 94.89% and 93.75% for the training and validation dataset, respectively. The model is trained by using eight layers of CNN architecture. The results achieved in the research surpass the performances from previous research with the same dataset with the research (CBIS-DDSM). The previous research has achieved the best accuracy of 92.53% in classifications using VGGNet [24].

V. Conclusion

Breast cancer is considered one of the deadliest cancers in the world. If the signs are detected as early as possible, the mortality rates of breast cancer can be reduced. Hence, it is paramount to have an automatic system that can help doctors or experts to detect breast cancer signs as soon as possible. The research proposes, explores, and evaluates ten training settings from five classical machine learning algorithms and five deep learning architectures and algorithms. Ten models are trained with ten settings using the CBIS-DDSM dataset. The results show that the models trained with traditional machine learning algorithms suffer from an over-fitting problem, while those trained with deep learning architectures and algorithms perform better. The training setting (SET-9) achieves the best accuracy compared to the other settings. The model results in the best accuracy of 94.89% and 93.75% during the training and validation phase, respectively. There are no significant differences in the accuracy of the models trained with SET-9 and SET-10. However, SET-10 has more parameters to be trained compared to SET-9. Hence, SET-10 needs more resources to train the model compared to SET-9.

The limitation of the research is that the deep learning architectures and algorithms explored is quite limited. There are several deep learning architectures and algorithms that can be explored to improve the model performances. Hence, for future research, more deep learning architectures and algorithms can be studied. The following research can also implement an attention model to capture the Region of Interest within digital imaging. Other features of digital imaging can also be explored and combined for future research. Moreover, combining a fusion of the best features extracted from the dataset can also be explored using deep learning architectures.

References

[1] S. Lei, R. Zheng, S. Zhang, R. Chen, S. Wang, K. Sun, H. Zeng, W. Wei, and J. He, "Breast cancer incidence and mortality in women in China: Temporal trends and projections to 2030," *Cancer Biology & Medicine*, vol. 18, no. 3, pp. 900–909, 2021.

[2] M. Amrane, S. Oukid, I. Gagaoua, and T. Ensari, "Breast cancer classification using machine learning," in *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*. Istanbul, Turkey: IEEE, April 18–19, 2018, pp. 1–4.

[3] O. I. Obaid, M. A. Mohammed, M. K. A. Ghani, S. A. Mostafa, and F. T. AL-Dhief, "Evaluating the performance of machine learning techniques in the classification of wisconsin breast cancer," *International Journal of Engineering & Technology*, vol. 7, no. 4.36, pp. 160–166, 2018.

[4] E. A. Bayrak, P. Kırcı, and T. Ensari, "Comparison of machine learning methods for breast cancer diagnosis," in *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*. Istanbul, Turkey: IEEE, April 24–26, 2019, pp. 1–3.

[5] A. B. Banu and P. Thirumalaikolundusubramanian, "Comparison of Bayes classifiers for breast

cancer classification," *Asian Pacific Journal of Cancer Prevention: APJCP*, vol. 19, no. 10, p. 2917, 2018.

[6] B. Sahu, S. N. Mohanty, and S. K. Rout, "A hybrid approach for breast cancer classification and diagnosis," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 6, no. 20, pp. e2–e2, 2019.

[7] A. F. M. Agarap, "On breast cancer detection: An application of machine learning algorithms on the wisconsin diagnostic dataset," in *Proceedings of the 2nd International Conference on Machine Learning and Soft Computing*, 2018, pp. 5–9.

[8] M. U. Rehman, S. Akhtar, M. Zakwan, and M. H. Mahmood, "Novel architecture with selected feature vector for effective classification of mitotic and non-mitotic cells in breast cancer histology images," *Biomedical Signal Processing and Control*, vol. 71, no. Part B, 2022.

[9] A. Chowanda, "Separable convolutional neural networks for facial expressions recognition," *Journal of Big Data*, vol. 8, no. 1, pp. 1–17, 2021.

[10] A. Chowanda and A. D. Chowanda, "Recurrent neural network to deep learn conversation in Indonesian," *Procedia Computer Science*, vol. 116, pp. 579–586, 2017.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv Preprint arXiv:1409.1556*, 2014.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.

[14] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," *arXiv Preprint arXiv:1606.05718*, 2016.

[15] A. Yala, C. Lehman, T. Schuster, T. Portnoi, and R. Barzilay, "A deep learning mammography-based model for improved breast cancer risk prediction," *Radiology*, vol. 292, no. 1, pp. 60–66, 2019.

[16] S. Khan, N. Islam, Z. Jan, I. U. Din, and J. J. P. C. Rodrigues, "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning," *Pattern Recognition Letters*, vol. 125, pp. 1–6, 2019.

[17] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast cancer multi-classification from histopathological images with structured deep learning model," *Scientific Reports*, vol. 7, no. 1, pp. 1–10, 2017.

[18] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography," *Scientific Reports*, vol. 9, no. 1, pp. 1–12, 2019.

[19] J. Xie, R. Liu, J. Luttrell IV, and C. Zhang, "Deep learning based analysis of histopathological images of breast cancer," *Frontiers in Genetics*, vol. 10, pp. 1–19, 2019.

[20] Q. Hu, H. M. Whitney, and M. L. Giger, "A deep learning methodology for improved breast cancer diagnosis using multiparametric MRI," *Scientific Reports*, vol. 10, no. 1, pp. 1–11, 2020.

[21] L. Tsochatzidis, L. Costaridou, and I. Pratikakis, "Deep learning for breast cancer diagnosis from mammograms–A comparative study," *Journal of Imaging*, vol. 5, no. 3, pp. 1–11, 2019.

[22] K. Jabeen, M. A. Khan, M. Alhaisoni, U. Tariq, Y. D. Zhang, A. Hamza, A. Mickus, and R. Damaševičius, "Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion," *Sensors*, vol. 22, no. 3, pp. 1–23, 2022.

[23] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific Data*, vol. 4, no. 1, pp. 1–9, 2017.

[24] P. Xi, C. Shu, and R. Goubran, "Abnormality detection in mammography using deep convolutional neural networks," in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. Rome, Italy: IEEE, June 11–13, 2018, pp. 1–6.