Breast Cancer Detection and Classification Techniques

Barış Kaplan
bkaplan18@ku.edu.tr

Canberk Eker
ceker16@ku.edu.tr

1. Introduction

Despite the centuries, cancer remains a lethal threat to people worldwide. Among all cancer types, breast cancer poses one of the most significant threats to people, especially for women. This is supported by the fact that approximately 2 million women with breast cancer cases are detected and 685000 women are dead out of them worldwide in 2020 [1]. Therefore, early diagnosis of breast cancer is very crucial since identifying the disease early may decrease the mortality rates and save the lives of millions. Early detection of breast cancer has become easier with the rapid technological developments and use of state-of-the-art techniques that will be discussed in this survey. While analyzing, we will divide them into mainly three groups: machine learning, deep learning, and hybrid approaches.

2. Techniques of Detecting and Classifying Breast Cancer

2.1 Machine Learning Based Techniques

2.1.1 AdaBoost, SR, and NSCT

One study proposed an approach using the AdaBoost algorithm on top of Super Resolution (SR) and Non-Subsampled Contourlet Transform (NSCT) preprocessing algorithms [5]. The study can be divided into three primary subsections: preprocessing, feature extraction, and classification. In preprocessing, thresholding and erosion methods were used to remove noise and obtain the region of interest, while median filtering addressed salt-pepper noise. NSCT and SR algorithms were then employed to achieve shift-invariance, enhance image quality, and improve the resolution & prediction of high-frequency image segments [5]. The feature extraction step involved classifying mass and calcification structures based on densities, shapes, sizes, and scattering, instead of utilizing a high-dimensional matrix of feature vectors [5]. While doing so, this research used histogram thresholding based on pixel intensities. This approach reduced computational complexity and increased accuracy. The study revealed that lesion shape was the most significant determinant and mostly extracted shape-related features such as spread and central moment using boundary, density, and regional descriptors [5]. For classification, the AdaBoost algorithm was utilized, surpassing alternative methods with a mean accuracy of 91.43%.

2.1.2 PCA, SVM, and LDA

ML techniques like principal component analysis (PCA) are utilized in dimensionality reduction, breast feature extraction, and auto-covariance coefficient optimization [2]. One study applied PCA in them and obtained higher accuracies compared to traditional methods. Then, this study used linear stepwise functions to select features, which identify the depth-to-width ratio and average margin orientation of gray-scale gradients as effective determinants for severity classification in breast tissues [2]. Next, various ML methods including SVM, decision tree, logistic regression, and LDA are experimented with to classify breast tissues as lesion or non-lesion. The study found that based on average orientation values, LDA and logistic regression performed better for linearly separable data, while decision trees and SVM showed better performance for more heterogeneous datasets [2].

2.2 Deep Learning Based Techniques

2.2.1 GoogleNet, ResNet, and VGGNet

There are also several deep learning methods used for the detection and classification of breast cancer. For example, in one study, a deep learning framework using transfer learning is mentioned [3]. The proposed method firstly preprocessed the breast cytology images for testing then applied data augmentation upon it to handle overfitting. Next, based on the transfer learning, this study experimented with three different CNN architectures named ResNet, GoogLeNet, and VGGNet in the feature extraction step [3]. While applying transfer learning, the background knowledge of ImageNet is used repetitively for different CNN models proposed. With GoogLeNet, the researchers achieved a small model with a combined convolution filter and thus minimized time complexity [3]. With VGGNet, this study presented the improved version of the AlexNet architecture, while with ResNet, adopted a deep CNN structure that effectively reduces training time and handles the degradation issue through concatenating varying sizes of convolutions [3]. The features extracted by these architectures are then provided to the connected layer to be used in the classification part. By taking the augmented test data and combined features as inputs and using the average pooling technique, the classification of the breast cells as benign or malignant is performed. As the experiment results, the proposed approach achieved higher precision (% 95) and accuracy (% 97.52) compared to each mentioned CNN structure and alternative method.

2.2.2 RetinaNet, YOLO, AlexNet, FrCN, and DBN

Besides CNN, various deep learning architectures such as RetinaNet, YOLO, AlexNet, FrCN (fully resolutional CNN), and DBN have been utilized for breast cancer detection and classification [4]. In a recent study, YOLO was used for mass detection and FrCN for pixel-level segmentation of masses. RetinaNet was employed for the detection and segmentation of breast masses; while YOLO, AlexNet, and DBN were experimented for breast cancer classification [4]. The experiment results showed that RetinaNet and YOLO achieved higher accuracies compared to the other methods and traditional CNNs, effectively addressing localization and class imbalance issues encountered in previous architectures [4].

2.2.3 ANN architecture

The significance of early detection and developments to speed it are covered in the article titled "Neural Network Architecture for Breast Cancer Detection and Classification"[6]. The paper suggests an artificial neural network (ANN) architecture that focuses on a low-complexity structure for the diagnosis of breast cancer. Finding the most appropriate activation function will help determine if breast cancer is benign or malignant. This study makes use of the Breast Imaging Reporting and Data System (BI-RADS), which offers a collection of mammographic pictures that have been tagged and categorized by medical professionals. On this dataset, experiments are carried out to establish the ideal activation function for the ANN design. The outcomes show how well ANN performs in categorizing breast cancer. It has been demonstrated that using a low-complexity structure makes it possible to decrease classification errors. This research stresses the availability and efficiency of artificial neural networks in the diagnosis of breast cancer.

2.2.4 CNN and MIAS database

In a different study, convolutional neural networks (CNN) are suggested as a deep learning technique for breast cancer detection utilizing histological photos. This study uses the Mammographic Image Analysis Society (MIAS) database to show how deep learning technology can be used to identify breast cancer. High-level domains including computer vision, image processing, medical diagnosis, and natural language processing frequently use deep learning techniques. Deep learning technology is used in this work to diagnose breast cancer with 98% accuracy by using the MIAS database[9]. The article of this study is broken up into three sections: gathering data and preparing it using different algorithms; dividing the dataset into training and testing sets; and creating graphs to visualize the data. After applying the model to the training dataset, it achieves 98% accuracy [9]. This study reveals that deep learning technology may accurately diagnose breast cancer when used with the MIAS database. A dataset of 200 photos with 12 features is used in the investigation. For the purpose of diagnosing breast cancer, the 12 features obtained after preprocessing are used.

2.2.5 DDSM method

In a study on the use of deep learning techniques to identify breast cancer in screening mammograms. This study uses a comprehensive neural network model and an end-to-end training strategy to identify breast cancer. This study makes use of the digital mammography databases INbreast and the Digital Database for Screening Mammography (CBIS-DDSM)[8]. Lesion annotations are used for training in the initial step; however, in later phases, just image-level labels are needed. This method eliminates the need for infrequent lesion annotations and enables the use of larger datasets. The findings show that deep learning techniques identify breast cancer with great accuracy. The deep learning model demonstrates high Area Under the Curve (AUC) values and produces successful results in terms of sensitivity and specificity in tests conducted on the CBIS-DDSM and INbreast datasets. The usefulness of deep learning techniques for breast cancer screening is highlighted in this study, as well as the potential for creating clinical tools that will lessen false positives and false negatives.

2.2.6 Her2Net method

One of the recent studies describes an additional technique for detecting breast cancer. It offers a powerful deep learning framework for recognizing, segmenting, and categorizing human epidermal growth factor receptor-2 (HER2) stained cell membranes and nuclei from breast cancer images. Pathologists have long struggled with the error-prone, expensive, and time-consuming manual measurement of HER2. The article suggests a HER2 deep neural network (Her2Net) based on deep learning to overcome this problem. Convolutional and deconvolutional sections, several convolutional layers, maximum pooling layers, spatial pyramid pooling layers, deconvolution layers, upsampling layers, and Trapezoidal Long Short-Term Memory (TLSTM) are all included in the proposed Her2Net architecture. For classification and error estimates, a fully connected layer and softmax layer are also utilized [10]. The classification findings are then used to calculate HER2 scores. The application of TLSTM and the creation of a deep learning framework for the detection, segmentation, and classification of cell membranes and nuclei, as well as HER2 score, constitute the proposed Her2Net system's primary contributions. Her2Net achieves 96.64% sensitivity, 96.79% recall, 96.71% F-score, 93.33% accuracy, 93.08% negative predictive value, and a false positive rate of 6.84%. These findings show that the suggested Her2Net for HER2 scoring in breast cancer evaluation achieves excellent accuracy and wide application.

2.3 Hybrid Techniques

Hybrid approaches combine multiple techniques to enhance breast cancer diagnosis accuracy. One study utilized histogram thresholding, active contour models, neural networks (NN), and Markov random field (MRF) for boundary determination of breast lesions [2]. Histogram thresholding segments breast tissue based on pixel intensity, while active contour models refine tissue boundaries. MRF techniques improve segmentation by considering spatial relationships between pixels, and neural networks extract features for classification [2]. These integrated techniques aimed to improve breast cancer detection by utilizing Computer-Aided Diagnosis (CAD) which is based on ultrasound images. Through CAD, the study employs boundary refinement, segmentation, and feature extraction [2].

Another study proposed a hybrid approach for breast cancer diagnosis using thermal imaging [7]. The system consists of four main subsections: image preprocessing, image segmentation, feature extraction, and classification. The study used both RGB and grayscale formats to capture finer-grained details in lesions. Median filtering was applied for image preprocessing [7]. The region-growing algorithm was used for image segmentation. Hierarchical and textural features were extracted from the segmented breast images. By following a comparative analysis; CNN, SVM, and Random Forest algorithms were trained and improved using the extracted features [7]. These algorithms were then used to classify breast regions as cancerous or non-cancerous. The study found that CNN achieved higher accuracy (99.67%) compared to Random Forest (90.55%) and SVM (89.84%) models [7]. Therefore, this study recommended CNN for classification under thermographic image modality.

3. Conclusion

In this survey; we mentioned machine learning, deep learning and hybrid techniques to diagnose and categorize breast cancer. Different facets, value, and potential of CAD techniques and technological developments in breast cancer diagnostics are highlighted in each study. Among all, using deep learning algorithms is thought to be a more viable and accurate strategy for early detection and precise categorization of breast cancer.

References

[1] "Breast Cancer - Statistics." *Cancer.Net*, 23 Feb. 2023, Retrieved from: www.cancer.net/cancer-types/breast-cancer/statistics#:~:text=Worldwide%2C%20female%20breast%20cancer%20is,United%20States%20after%20lung%20cancer.

[2] H.D. Cheng, Juan Shan, Wen Ju, Yanhui Guo, Ling Zhang, Automated breast cancer detection and classification using ultrasound images: A survey, Pattern Recognition, Volume 43, Issue 1, 2010, Pages 299 - 317, ISSN 0031 - 3203, https://doi.org/10.1016/j.patcog.2009.05.012.
(Retrieved from: https://www.sciencedirect.com/science/article/pii/S0031320309002027)

[3] SanaUllah Khan, Naveed Islam, Zahoor Jan, Ikram Ud Din, Joel J. P. C Rodrigues,
A novel deep learning based framework for the detection and classification of breast cancer using transfer learning, Pattern Recognition Letters, Volume 125, 2019, Pages 1-6, ISSN 0167-8655, Retrieved from: https://doi.org/10.1016/j.patrec.2019.03.022.

[4] Hamed, G., Marey, M. A. E. R., Amin, S. E. S., & Tolba, M. F. (2020). Deep learning in breast cancer detection and classification. In *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)* (pp. 322-333). Springer International Publishing.
(Retrieved from: https://link.springer.com/chapter/10.1007/978-3-030-44289-7_30)

[5] Fatemeh Pak, Hamidreza Rashidy Kanan, Afsaneh Alikhassi, Breast cancer detection and classification in digital mammography based on Non-Subsampled Contourlet Transform (NSCT) and Super Resolution, Computer Methods and Programs in Biomedicine, Volume 122, Issue 2, 2015, Pages 89-107, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2015.06.009.
(Retrieved from: https://www.sciencedirect.com/science/article/pii/S0169260715001765)

[6] H. Jouni, M. Issa, A. Harb, G. Jacquemod and Y. Leduc, "Neural Network architecture for breast cancer detection and classification," 2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET), Beirut, Lebanon, 2016, pp. 37-41, doi: 10.1109/IMCET.2016.7777423.

[7] Allugunti, R. V. (2021). Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. International Journal of Engineering in Computer Science 2022; 4(1): pp. 49-56, doi: https://doi.org/10.33545/26633582

[8] Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., & Sieh, W. (2019). Deep learning to improve breast cancer detection on screening mammography. *Scientific reports*, *9*(1), 12495.

[9] N. Khuriwal and N. Mishra, "Breast Cancer Detection From Histopathological Images Using Deep Learning," 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India, 2018, pp. 1-4, doi: 10.1109/ICRAIE.2018.8710426.

[10] M. Saha and C. Chakraborty, "Her2Net: A Deep Framework for Semantic Segmentation and Classification of Cell Membranes and Nuclei in Breast Cancer Evaluation," in IEEE Transactions on Image Processing, vol. 27, no. 5, pp. 2189-2200, May 2018, doi: 10.1109/TIP.2018.2795742.