

# A Review of Breast Cancer using Machine Learning

M.Tech. Scholar Adarsh Gupta, Prof. Sachin Mahajan

Department of Computer Science & Engineering  
JIT Borawan, Khargone, MP, India  
guptaadarsh527@gmail.com, Sachinmahajan248@gmail.com

**Abstract-** Breast cancer is, after lung cancer, the most prevalent form of the disease in the globe. Women are the demographic most likely to be affected by this condition. Breast cancer is the most common kind of cancer to result in a woman's death if she is of childbearing age. Because there is always more to learn and there is room for improvement in every line of work, medical imaging is not an exception to this rule. It is expected that the death rate associated with cancer would decrease if it is discovered early and effectively treated. The diagnosis accuracy of persons working in the health care profession may be improved via the use of machine learning techniques. The technique known as deep learning has the potential to differentiate between breasts that are healthy and those that have cancer (also known as neural networking). This method might be used to differentiate between healthy breast tissue and breast tissue affected by illness. Long-term research on the topic aimed, among other things, to examine breast cancer and screening practices among Indian women. This was one of the primary goals of the inquiry. A literature study was carried out with the assistance of several databases along with additional sources. Participants in the study were instructed to use phrases linked to breast cancer such as "breast carcinoma" and "breast cancer awareness," in addition to terms such as "knowledge" and "attitude," as well as the gender neutral term "women." In addition, India had a role in the study that was done. This search does not look for articles that have been published in the English language in the last 12 years.

**Keywords-** Breast Cancer, CNN, Mammograms-MINI-DDSM, Machine Learning.

## I. INTRODUCTION

The incidence of cancer is rising as the population ages and more individuals are diagnosed with the disease. Smoking, which is associated with an increased risk of cancer in developing nations, is widely practiced. It is the most prevalent cancer in women and the most frequent cancer overall. A staggering 23% of all cancer diagnoses and 14% of all cancer-related fatalities in women are due to breast cancer. Every American woman has a one in eight risk of being diagnosed with breast cancer. To put it simply, the number of women developing breast cancer is rising in many nations with lower per capita incomes and higher life expectancies. [2]

In oncology, aberrant cell development and the fast spread of the illness are the hallmarks of a certain kind of disease. One of the most frequent symptoms of sickness is the accumulation of cells in the body, resulting in a bump or lump. Most often, this occurs via the development of tumours on the body, which may be recognized by their initial location.

The signs of breast cancer are rare in the early stages, thus routine screening is essential to ensure an early and correct diagnosis. There were a large number of moderately intense yet non-destructive lumps discovered. That whatever has gone before has come before. Breast cancer may be divided into as many as 21 distinct subtypes. There

are as many as 21 distinct forms of breast cancer, even though the majority of them do not advance beyond the non-invasive stage. On the other hand, there were only 2470 new instances of invasive cancer discovered among males in 2017[1]. Breast cancer mortality rates increase as you become older. The average age of those diagnosed with breast cancer between 2010 and 2014 was 62. WHO claims Pakistan has Asia's highest incidence of breast cancer.

There are around 90,000 new cases and 40,000 deaths per year. There is a strong possibility at this time that the individual will completely recover from breast cancer. When cancer is in its early stages, it doesn't cause any discomfort, therefore it isn't discovered until the symptoms worsen. According to the information currently available, cancer patients in Pakistan are, on average, 40 years old.

Patient endurance is the proportion of patients who survive for a specific period of time after discovering they will be able to return to their regular life shortly. The tumor's stage impacts how rapidly stamina is depleted throughout treatment. [1,] Women with breast cancer, according to recent studies, have a lot of stamina. Up until recently, it was estimated that just 89% of Americans used social media in some way. There has been an increase in the percentage since then. After 15 years, the proportion of the population has reduced to roughly 80%.

According to the American Cancer Society, having a close relative with breast cancer increases one's risk of developing the illness one's self. Both sexes are affected by this. Breasts may be examined using a low-dose x-ray method called mammography. It is possible to detect breast cancer in a variety of methods. Examples of these approaches include wavelet, curvelet, and restricted scale transformations, as well as images captured at various scales. These are steps three and four.

Neuro-flooding frames and fuzzy reasoning are used by teachers in an effort to prevent students from taking courses they would find harmful. [3] An information architecture computer model is built using deep learning methods. This paradigm incorporates text, graphics, and audio into a single entity. Different datasets and CNN architectures are utilized to build models at various degrees of abstraction. Cancer cells may be detected in images using sophisticated learning techniques in clinical imaging. In order to build an extensive convergence network, it's necessary to acquire and analyze vast amounts of data. Calibrate a well-equipped and prepared group utilizing the procedure above.

Deep learning is currently being used in bioinformatics, Alzheimer's disease diagnosis, and subatomic imaging. Another area of study that combines the two is molecular imaging [4]. In this sector, physical image readers and subatomic data collectors collaborate. Taking a lot of lessons at once isn't the only method to learn. A lot less information and pre-testing would have to be done because of this.

Many individuals in underdeveloped nations don't go to the doctor since they don't know much about health care. When it comes to spotting breast cancer early and having it looked properly, there is a lack of knowledge. Breast cancer screening should no longer be done at random on women. As a result, programmes that are community-based and coordinated should be implemented immediately. 10 To detect cancer early, there are a variety of methods available. In addition to breast self-examination (BSE), clinical breast examination (CBE), and mammography, there are a number of other procedures.

Even though there have been several public awareness programmes, breast cancer prevention and treatment are not addressed in the country's health care system. Women may not be aware of breast cancer, but one research found that their views about the condition are favorable. Breast cancer screenings are new to the people who work for them. Women in India need to know more about breast cancer as soon as possible, since it may be deadly. Research on Indian women's knowledge, attitudes, and practices related to breast cancer screening was the overarching purpose of this research.

## II. RELATED WORK

Calcifications in the breasts were discovered thanks to improved mammography in the 1990s. Using CNN in clinical imaging has become common since then. Changing with the times is a major component of CNN's success. According to this research, there are two forms of motion learning in clinical imaging. Another example of how preliminary organizations are categorized is shown using highlights from a single organization's layer. Another strategic layer will be created when the pre-made structure is exhausted, save for the completely connected levels.

Highlights may be removed from this dataset using SVM and wavelet techniques as well as cosine transforms and CNN illumination [22-30]. Many different kinds of tests were run on this dataset, including large-scale picture comparisons and the use of various extraction classifiers. It was used in conjunction with SIFT. Components were classified into two categories (good and negative), as well as three (benevolent, childlike, and destructive) based on their qualities. We employed mammography patches to expand the data collection.

A comparison with another dataset would show us how well it worked. To solve this problem, the 2D-DWT was used to partition advanced mammography into four groups before generating CNNs using SVMs utilizing softmax layers and DCT to produce CNNs (Discrete curve altering) [30-36]. Before employing DCT to create CNNs with SVM and softmax layers, 3D-DWT was another method of dividing advanced mammography into four categories. Data from the IRMA knowledge base was utilized in this investigation, which indicated that 81.83% of DCTs were accurate and 83.74% of CTs were accurate. [6].

Channels of high and low pass are used to transmit it. Measurement and movement of the wavelet are carried out after it has been created. Curvature changes are used to identify more intricate regions on wedges and to detect shorter lines [3]. [4] For example, fuzzy reasoning might demonstrate conjecture and logical reasoning [3-4]. Even if there is no accurate method to display a number, fuzzy reasoning may still be applied. It is harder to construct a fluffy model since it needs to be rebuilt once modifications are made. Starting with an incorrect set of facts and stated principles, and then doing more investigation, is known as "neurofluffy reasoning." [3] In mass identification, the multi-scale curve holds a record of 98.59 percent accuracy that has never been broken.

C-mean bunching, for example, has been proven to be more successful than the division's impact area extraction and placement capabilities when used to heritage estimations. [7] The surface area of the breast was organized using 3D ultrasound pictures to identify and categories grassy and non-fatty tissue. If you utilise K-

bunching, breast thermography may detect breast cancer in its early stages. [9] In this instance, shadow investigations of the illness region should be done. [10] Researchers found that ribosome-sized tumour zones were wiped off by the findings of this investigation. Many options were available when it came to the tumour split, including configurable criteria and various classifications [10]. In fact-checking, the same kinds of items were used again. In addition, the tumour split has led to a variety of regulations, flexible criteria and alternative methods of looking at the facts[10].

Researchers employed an alternate strategy to locate thick spots on a collection of important recurrence models[11]. A total of 32 distinct categories of visual quality were examined in this experiment. We'll go into more depth about this method of improving and denoising mammography in the following sections. Furthermore, it was discovered that micro calcification and minor colour changes, such as masses, were possible. In order to ensure that our locator didn't end up in a favourable area, we employed weighted misfortune.

ESTD and Surface Examination were used in an audit segment to demonstrate several techniques. They served as a visual aid for demonstrating the process of obtaining mammographic pictures. [14] Factors for breast cancer development were shown in an audit part that included lead component inspection and classification [14].

### III. DATASET

Table 1. MINI-DDSM dataset.

		Class	
		Benign	malignant
Images	Training Samples (80%)	1940	2474
	Test Samples (20%)	420	524

A CNN requires a lot of training data in order to get decent results. Because of the lack of huge datasets, training and testing were conducted using the biggest publicly accessible internet dataset. The MINI-DDSM was utilized to collect the majority of the data in this investigation. A total of 5358 images were required to meet this objective. To put that into perspective, each image is 1372 by 2340 pixels wide [15]. There were about 2474 photos of malignancy and around 1940 photographs of healthy subjects. CNN was utilized to educate and to test throughout the course. Images captured in grayscale need to be converted into another colour format before being used [16].

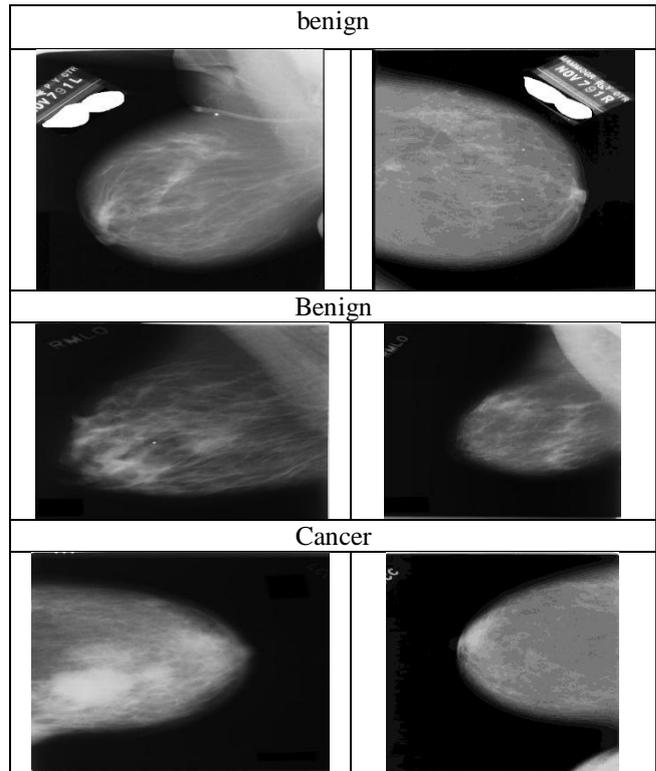


Fig 1. Display MINI-DDSM mammography dataset.

### IV. LIMITATIONS

There are some problems with this study, such as the fact that it only looked at research that had already been done [17]. This could have led to a bias toward studies that had been published in academic journals. We did everything we could to find and include all of the important articles, but some could not be included because of our exclusion rules [18].

Because India has a wide range of socioeconomic backgrounds, cultural backgrounds, geographic areas, and other factors that may make people look different than they do, the review's findings may not be the same across the country [19]. People in rural areas may not be able to speak or read the same languages as people in cities, so some statistics may be wrong [20]. Women may have been afraid to talk about their breast cancer with someone they didn't know, or they may have given out the wrong information because of social or cultural pressures to do so [21].

### V. CONCLUSION

More than half of the study population had adequate information and a positive attitude toward breast cancer screening and early detection measures, but there is a big gap in knowledge and attitudes that needs to be addressed

in practice, according to the results of the systematic review.

To prevent breast cancer, women in India must be made aware of the many risk factors. There is an urgent need for effective national and state-level cancer literacy programmes, as well as collaborations with community-based organizations and the health care system. Since breast cancer incidence and death are significantly reduced in low-income nations like India, the government should concentrate on breast cancer early detection techniques as soon as possible in health centres and health posts.

In order to identify the early stages of breast cancer and begin treatment as soon as possible, early detection and screening methods such as breast self-examination (BSE), cervical cancer screening (CBE), and mammography (mammography) are available.

## REFERENCES

- [1] "Breast cancer statistics." [Online]. Available: <http://www.wcrf.org/int/cancer-facts-figures/data-specific-cancers/breast-cancer-statistics>
- [2] "Over 40,000 die of breast cancer every year in pakistan." [Online]. Available: <http://www.google.com.pk/am p/s/www.dawn.com/news/amp/1319675>
- [3] S. Bagchi and A. Huong, "Signal processing techniques and computer-aided detection systems for diagnosis of breast cancer—a review paper," *Indian Journal of Science and Technology*, vol. 10, no. 3, 2017.
- [4] F. A. Jaffer and R. Weissleder, "Molecular imaging in the clinical arena," *Jama*, vol. 293, no. 7, 2005.
- [5] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of alzheimer's disease with deep learning," *IEEE*, 2014
- [6] J. C. Tobias Charistian Cahoon, Melanie A. Sutton, "Three-class mammogram classification based on descriptive cnn features," 2000.
- [7] S. Sharma, M. Kharbanda, and G. Kaushal, "Brain tumor and breast cancer detection using medical images," *International Journal of Engineering Technology Science and Research*, vol. 2, 2015.
- [8] P. Gu, W.-M. Lee, M. A. Roubidoux, J. Yuan, X. Wang, and P. L. Carson, "Automated 3d ultrasound image segmentation to aid breast cancer image interpretation," *Ultrasonic's*, vol. 65, 2016.
- [9] P. Hankare, K. Shah, D. Nair, and D. Nair, "Breast cancer detection using thermography," *Int. Res. J. Eng. Technol*, vol. 4, 2016.
- [10] W. K. Moon, I.-L. Chen, J. M. Chang, S. U. Shin, C.-M. Lo, and R.-F. Chang, "The adaptive computer-aided diagnosis system based on tumor sizes for the classification of breast tumors detected at screening ultrasound," *Ultrasonics*, vol. 76, 2017.
- [11] F. Strand, K. Humphreys, A. Cheddad, S. Törnberg, E. Azavedo, J. Shepherd, P. Hall, and K. Czene, "Novel mammographic image features differentiate between interval and screen-detected breast cancer: a case-case study," *Breast Cancer Research*, vol. 18, no. 1, 2016.
- [12] A. Mencattini, M. Salmeri, R. Lojacono, M. Frigerio, and F. Caselli, "Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing," *IEEE transactions on instrumentation and measurement*, vol. 57, no. 7, 2008.
- [13] S. Hwang and H.-E. Kim, "Self-transfer learning for fully weakly supervised object localization," *arXiv preprint arXiv: 1602.01625*, 2016.
- [14] Z. Mohammadzadeh, R. Safdari, M. Ghazisaeidi, S. Davoodi, and Z. Azadmanjir, "Advances in optimal detection of cancer by image processing; experience with lung and breast cancers," *Asian Pacific journal of cancer prevention: APJCP*, vol. 16, no. 14, 2015.
- [15] "Mammographic image analysis society digital mammogram database." [Online]. Available: <http://peipa.essex.ac.uk/info/MINI-DDSM.html>
- [16] M. M. Jadoon, Q. Zhang, I. U. Haq, S. Butt, and A. Jadoon, "Threeclass mammogram classification based on descriptive cnn features," *BioMed research international*, 2017.
- [17] "Breast cancer facts & figures 2017-2018." [Online]. Available: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf>
- [18] K. Vennila, K. Sivakami, and R. Padmapriya, "Detection of mass in digital mammograms," *International Journal of Computer Applications*, vol. 104, no. 5, 2014.
- [19] K. Djaroudib, A. T. Ahmed, and A. Zidani, "Textural approach for mass malignancy segmentation in mammographic images," *arXiv preprint arXiv: 1412.1506*, 2014.
- [20] S. Deepa and V. S. Bharathi, "Efficient roi segmentation of digital mammogram images using otsu's n thresholding method," *Indian Journal of Automation and Artificial Intelligence*, vol. 1, no. 2, pp. 51–56, 2013.
- [21] "Filters in the context of convolutional neural networks." [Online]. Available: <https://www.quora.com/What-is-a-filter-in-the-context-of-Convolutional-Neural-Networks>
- [22] Gupta, P., Shukla, M., Arya, N., Singh, U., Mishra, K. (2022). Let the Blind See: An AIoT-Based Device for Real-Time Object Recognition with the Voice Conversion. In: Al-Turjman, F., Nayyar, A. (eds) *Machine Learning for Critical Internet of Medical Things*. Springer, Cham. [https://doi.org/10.1007/978-3-030-80928-7\\_8](https://doi.org/10.1007/978-3-030-80928-7_8).
- [23] Patidar, M., Singh, U., Shukla, S.K. et al. An ultra-area-efficient ALU design in QCA technology using synchronized clock zone scheme. *J Supercomput* (2022). <https://doi.org/10.1007/s11227-022-05012-2>

- [24] Taiwade, N. Gupta, R. Tiwari, S. Kumar and U. Singh, "Hierarchical K-Means Clustering Method for Friend Recommendation System," 2022 International Conference on Inventive Computation Technologies (ICICT), 2022, pp. 89-95, doi: 10.1109/ICICT54344.2022.9850852.
- [25] R. Baghel, P. Pahadiya and U. Singh, "Human Face Mask Identification using Deep Learning with OpenCV Techniques," 2022 7th International Conference on Communication and Electronics Systems (ICES), 2022, pp. 1051-1057, doi: 10.1109/ICES54183.2022.9835884.
- [26] M. Ranjan, A. Shukla, K. Soni, S. Varma, M. Kuliha and U. Singh, "Cancer Prediction Using Random Forest and Deep Learning Techniques," 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), 2022, pp. 227-231, doi: 10.1109/CSNT54456.2022.9787608.
- [27] Singh, Upendra, Gupta, Puja, and Shukla, Mukul. 'Activity Detection and Counting People Using Mask-RCNN with Bidirectional ConvLSTM'. 1 Jan. 2022 : 6505 – 6520.
- [28] Gupta, P., Shukla, M., Arya, N., Singh, U., Mishra, K. (2022). Let the Blind See: An AIoT-Based Device for Real-Time Object Recognition with the Voice Conversion. In: Al-Turjman, F., Nayyar, A. (eds) Machine Learning for Critical Internet of Medical Things. Springer, Cham. [https://doi.org/10.1007/978-3-030-80928-7\\_8](https://doi.org/10.1007/978-3-030-80928-7_8)
- [29] Patidar, M., Singh, U., Shukla, S.K. et al. An ultra-area-efficient ALU design in QCA technology using synchronized clock zone scheme. *J Supercomput* (2022). <https://doi.org/10.1007/s11227-022-05012-2>
- [30] A. Taiwade, N. Gupta, R. Tiwari, S. Kumar and U. Singh, "Hierarchical K-Means Clustering Method for Friend Recommendation System," 2022 International Conference on Inventive Computation Technologies (ICICT), 2022, pp. 89-95, doi: 10.1109/ICICT54344.2022.9850852.
- [31] Jindal, R, Kumar, N, Patidar, S. IoT streamed data handling model using delta encoding. *Int J Commun Syst.* 2022; 35( 13):e5243. doi:10.1002/dac.5243
- [32] S. Patidar, A. Jain and A. Gupta, "Comparative Analysis of Machine Learning Algorithms for Heart Disease Predictions," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), 2022, pp. 1340-1344, doi: 10.1109/ICICCS53718.2022.9788408.
- [33] B. Jain, S. Patidar and D. Sudershan, "Heterogeneous Software Defect Prediction using Generative Models," 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), 2022, pp. 367-372, doi: 10.1109/CSNT54456.2022.9787607.
- [34] Singh, Shani Pratap and Shukla, Jayesh and Sharma, Shaili and Daga, khushhal and Bhalavi, Brahman Singh and Singh, Upendra, Face Mask Detection using Multi-Stage CNN Architecture (July 10, 2021). Proceedings of the International Conference on IoT Based Control Networks & Intelligent Systems - ICICNIS 2021, Available at SSRN: <https://ssrn.com/abstract=3884022> or <http://dx.doi.org/10.2139/ssrn.3884022>
- [35] A. Saxena, A. Vyas, L. Parashar and U. Singh, "A Glaucoma Detection using Convolutional Neural Network," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 815-820, doi: 10.1109/ICESC48915.2020.9155930.
- [36] A. Bamne, N. Shrivastava, L. Parashar and U. Singh, "Transfer learning-based Object Detection by using Convolutional Neural Networks," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 328-332, doi: 10.1109/ICESC48915.2020.9156060.