International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Machine Learning Approaches for Automatic Lesion Detection in Mammography Images

[1]Indrajeet Kumar, [2]Dr. Rashmi Gudur

Submitted:22/03/2023 **Revised:**27/05/2023 **Accepted:**12/06/2023

Abstract: Mammography is the major diagnostic tool for detecting breast cancer early, alerting the patient to abnormalities long before she would notice them physically. Using digital mammography pictures, the Computer Aided Diagnosis (CAD) technology detects breast abnormalities. To forecast the required items, deep learning techniques learn the image's characteristics using a small set of expert-annotated data. In recent years, the accuracy of convolutional neural networks (CNN) has soared in a variety of image processing tasks, including image detection, identification, and classification. This work offers an automated approach for detecting and classifying breast cancer lesions in mammograms, using the state-of-the-art object detection deep learning technique Faster R-CNN. In order to train the Faster R-CNN network, the proposed CAD system employs a total of 330 mammography pictures, 121 of which have been annotated. Using the testing dataset, the suggested method achieved a mAP (mean Average Precision) of 0.857.

Keywords: *Breast Cancer, Mammogram, Faster R-CNN, Breast lesion detection, Breast lesion classification, Deep Learning Network for Faster R-CNN*

1. Introduction

Today Worldwide, a growing number of women are being diagnosed with breast cancer (BC). Numerous women lose their lives every year to this terrible illness. More than 5,000 women every year lose their lives to BC, according to data from the World Health Organization (WHO). It has been noted that most instances of BC are diagnosed late. As a result, patients' odds of survival plummet to between 10 and 40 percent. If discovered early and basic therapy is given promptly, the survival rate may be enhanced from 10-40% to 80% [1][2]. The initial stage of breast cancer may be discovered with early diagnosis and screening. Breast cancer early detection is aided by public and medical understanding of the disease's symptoms. Clinical breast examination by a doctor and other medical experts is the first step in the screening process, whereas mammography is the second. A mammogram is a picture of the breast produced by a low-dose X-ray examination of a woman's breast [3]. Mammography has been shown to be an efficient and successful means of screening for breast cancer. Because it detects even subtle changes in the breast. It's not only the physicians and nurses that could miss these shifts; occasionally the patient is unaware of them, too.

Mammography is a cutting-edge technique used for the early detection of breast cancer. Breast mammography penumbras are the primary target of this method because of their high tissue density. Therefore, it is a useful method for detecting breast masses and microcalcifications (MCC) in women. Here are some of digital mammography's benefits:

To get the best possible outcomes in the diagnosis of BC, clinical images are shown digitally. Using an image processing approach, the mammogram's display quality may be improved by adjusting the picture's edges, contrast, and density.

ii) CAD, or computer-aided diagnosis: Several CAD and CAD methodologies are now being developed. In the future, these methods will aid the radiologist's diagnostic procedure by producing more cost-effective and reliable findings. For remote consultation with other doctors, the present technology transforms digital mammograms across phone lines or a network, making data accessible in distance modes and for future references. iii. Image Archiving and Management. The radiologists will have a far better chance of finding the masses if they use all three methods at once. A significant drawback of mammography screening is that radiologists miss many instances of early-stage breast cancer. Research shows that 15 percent of breast cancers that were detectable in earlier screening were missed by radiologists during mammography [9]. Many early-stage instances of breast cancer are overlooked when just the most subtle

1Asst. Professor, Department of Comp. Sc. & Info. Tech. Graphic Era Hill University, Dehradun Uttarakhand 248002,
ikumar@gehu.ac.in
2Krishna Institute of Medical Sciences, Krishna Vishwa Vidyapeeth "Deemed To Be University" Karad Malkapur, Karad (Dist. Satara), Maharashtra, India. PIN – 415539
anandgudur@gmail.com

symptoms are used as a yardstick, and when mistakes in perception are also included in. In the eye tracking studies, these mistakes have been broken down into three classes. ii) Detection Error; Radiologists can identify lesions for a very little time, say, a single second. In rare circumstances, radiologists are unable to determine whether a growth is malignant or noncancerous despite extensive testing. Search and detection mistakes occur when radiologists miss the existence of a tumor, whereas interpretation errors occur when radiologists correctly identify a tumor but decide no treatment is necessary. Depending on the specific characteristics, an abnormal tissue in a woman's breast may be either benign or malignant [11]. Breast tissue evaluated by a biopsy may definitively be classified as benign or malignant. Suspect breast cancer tissue is removed surgically and examined by a pathologist. Although digital mammography has been widely praised for its ability to identify breast cancer, there is still a need for highly trained radiologists who can distinguish between healthy tissue and malignant growths [12].[13]. Medical diagnostic tests are evaluated by their sensitivity, specificity, and accuracy in distinguishing between malignant and benign samples. A false-positive (FP) instance is one in which a radiologist incorrectly diagnoses a benign condition as cancer because of subpar imaging [14]. Breast cancer lesions and microcalcifications (MCCs) are key diagnostic indicators. Thirty percent to fifty percent of people with cancer have these symptoms. Detecting large lesions might be challenging because to [15].

Lesions in a woman's breast are notoriously difficult to diagnose because of their varied appearance, size, and location inside the thick glandular tissue.

Lesions, according to their heuristic properties, often seem indistinguishable from normal tissues and sometimes are mistaken for them. It's difficult to see MCC on a mammogram because it looks as a cluster of granular white spots rather than a single, discrete calcium deposit. They range in diameter from around 0.1 millimeters to about 1 millimeter, with a mean of about 0.3 millimeters. Those things are classified as irregular, granular, or linear.

The foundations of computer-assisted diagnostics may be traced back to image processing and machine learning techniques. Digital mammography requires a pre-processing approach [18] to reduce artifacts, get rid of noise, and enlarge breast pictures so that mass lesions and MCC may be better seen. In the computer-aided design (CAD) process, machine learning (ML) techniques are used to classify patterns. Attributes obtained from the ROI are employed in the diagnostic model's training and testing phases of categorization. The sample at hand is classified as benign or malignant using a classifier trained on samples with known anomalies (MCC and mass). Over the last decade, several different categorization strategies have been used for distinguishing between benign and malignant tumors in breast imaging.

2. Literature Review

Kosmia Loizidou et.al(2022) The information used in this analysis comes from a confidential dataset. We collected 196 pictures from 49 patients (2 views per breast at 2 time points per patient) with detailed annotations of mass sites and instances of malignancy verified by biopsy. In all, 96 features were recovered for categorization using 5 feature selection methods. This study shows that subtracting digital mammograms taken at different times may reliably distinguish between benign and malignant breast tumors using machine learning.

U Supriya et.al (2022) The incidence of breast cancer continues to rise worldwide at an alarming rate. Women significantly outnumber males in the prevalence of breast cancer. If caught and treated early, breast cancer is manageable. One method for finding breast cancer early is mammography. Six optimal deep learning algorithms (are tested for their ability to detect breast cancer in mammography images, and InceptionV3 is shown to be the most effective.

Aditi Kajala et.al(2020) This study provides a high-level overview of how machine learning techniques may be utilized to improve the accuracy and timeliness with which breast cancer diagnoses are made. The examined studies primarily aim to aid in proper diagnosis and categorization.

G. Umarani Srikanth et.al (2019) A compression, feature, clustering, regression, and classification technique based on Extreme Learning Machine's optimization-based learning paradigm. To fix these problems and improve generalization efficiency, the authors suggest using fake hidden nodes and the Unique minimal method. In this study, we provide a CAD system that uses deep, morphological, textural, and density-based characteristics learned by a Convolutional Neural Network (CNN). With the use of mammography scans, CAD systems can determine if a patient has malignant or benign breast tissue and, if so, at what stage of the disease it is. Using this approach, the authors of this research hope to determine whether or not a certain stage of breast cancer tissue poses a health risk.

Identification of Malignant Lesions in Mammograms Image by using Supervised Deep Learning ICNN Model

Over the last two decades, CAD systems have been developed to assist radiologists in the analysis of mammography screening. Recent CAD systems used for screening mammography analysis have shown subpar results and need to be refined to provide better screening results. Since 2012, a convolutional neural network (CNN) system has been utilized to analyze screening mammograms, with impressive results in the field of medical intellectual property. There is much hope that CNN will revolutionize the field of medical image analysis. Based on one of the most effective lesion diagnosis methods, we offer an enhanced version of a convolutional neural network (ICNN) model. This model may be seen as an aid to the radiologist since it can detect both normal and pathological lesions in a mammographic picture automatically. On the open MIAS database, the ICNN model achieves top classification results. With 10 fold cross validation, it successfully distinguishes between malignant and benign lesions on mammograms with an impressive accuracy rate of 92.05%.

Most women will get breast cancer (BC) at some point in their lives. Breast cancer is the second leading cause of death among women worldwide. Women's breast cancer mortality may be reduced by 38-48% with regular mammography screening. Twenty-five of the European Union's 28 members use digital screening technologies to find breast cancer early, when it is most treatable.Digital mammograms are obtained from both breasts at opposite angles throughout the screening process. Radiologists evaluate the mammograms to see whether the tumors are cancerous or not. Radiologists may use a computer-aided detection (CAD) system to help them find suspicious lumps in mammograms, which might indicate the presence of cancer. More in-depth study is needed to improve the accuracy of the model for identifying breast cancer bio markers in mammography pictures, according to studies on the efficacy of CAD systems. Digital mammography, advancements in machine learning and intellectual property procedures, and other recent innovations have opened up a significant new window of opportunity for early and precise diagnosis of breast cancer.

Our study's goal is to build a better CNN system by integrating new augmentation techniques, batch normalization, and dropout settings to boost the precision of a CNN-based model. In conclusion, DL networks are increasingly being used to address medical analytical difficulties on massive mammography datasets, such as the categorization of distinct kinds of mammograms. As a result, we propose a unique method for classifying the various kinds of masses shown in mammographic pictures using an effective ICNN architecture that addresses the aforementioned three CNN-related difficulties. The suggested model and its associated methodologies are described in this section. In this study, we use both malignant and benign mammograms to prove that the suggested model works.

3. Materials and Methods

Dataset Specifics From the MIAS database, 2,000 normal and cancerous breast mammograms have been categorized.

In Table 1 the quantity of images of each category is being described

Types of images	Total number in original	Total number after augmentation
Normal	208	1,000
Cancer	114	1,000

Table 1: Details of the images considered for study and taken from MIAS database

Mammograms are all 1024 pixels on the widest side. Due of the enormous file sizes, these photographs have been preprocessed and resized to 256 256 pixels before being used as input in our suggested method.

Over fitting may occur when just a little quantity of data is utilized to train a deep neural network. Training mammograms undergo image processing to prevent the overfitting problem. Therefore, we have modified the original mammograms by altering their size, orientation, and rotation. As you can see in Table 1, we have augmented our data such that the two groups of cases are equally represented.

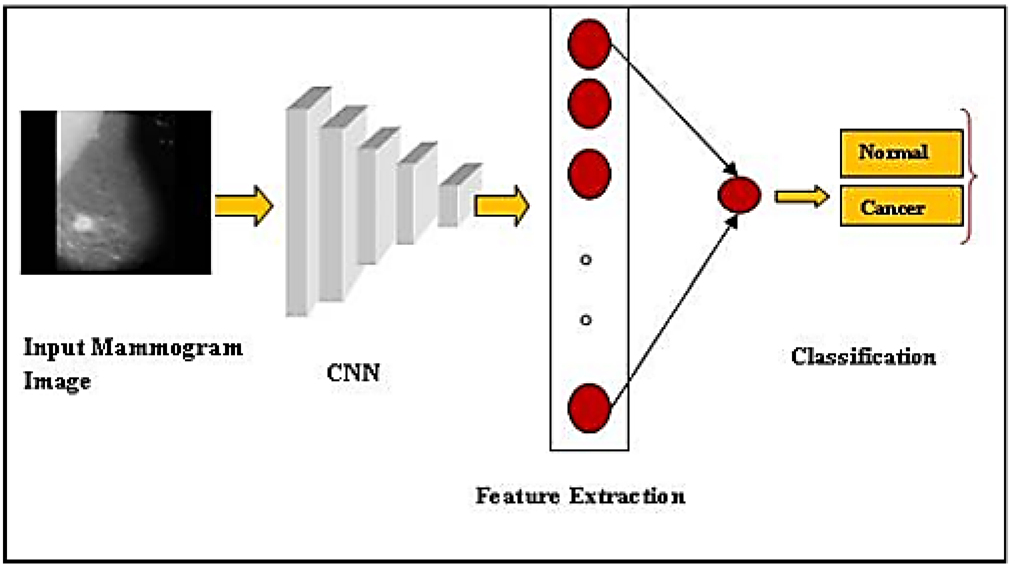

Fig. 2. Framework of the proposed model

Proposed Network Architecture

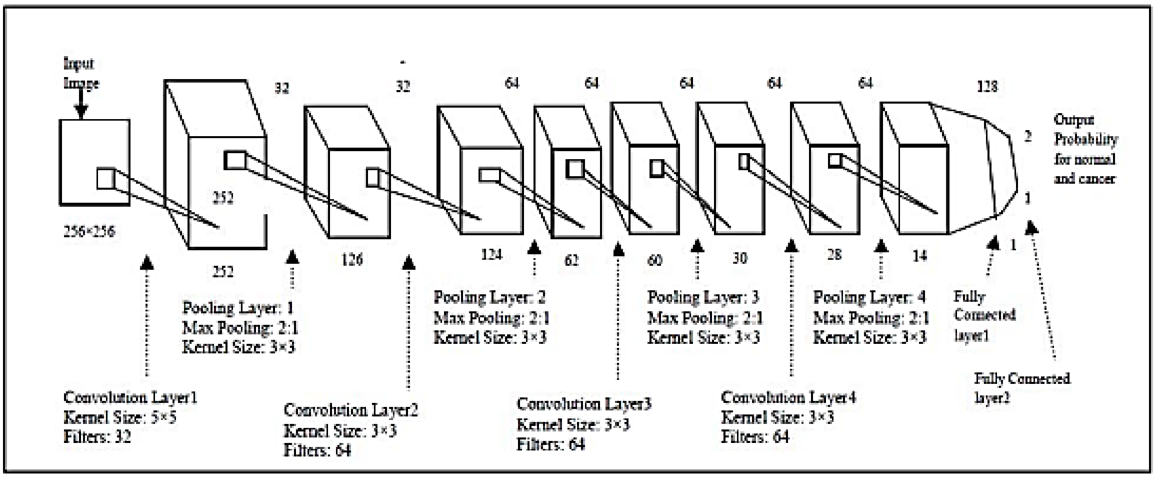

Fig 3. ICNN system architecture

Figure 2. depicts the suggested model architecture. It is built from 6 layers total: 2 completely linked layers, 2 pooling layers, and 4 convolution layers. The first layer of the ICNN is fed input. The first convolutional phase employs 32 filters with a 5 5 1 kernel to generate a feature map of 256 256 32-pixel cells. The feature map's matrix size is decreased from 256 256 256 to 128 128 32 thanks to a resampling operation in the first max pooling layer that yields the maximum value for each pair of pixels in a 3 3 kernel. Two completely linked layers follow the four convolution layers and four pooling layers; these are a multilayer perception layer and a softmax layer. In the last layer, the cancer probabilities are calculated using a softmax function. Each convolution layer's RLU activation function helps ICNN deal with vanishing gradient problems. To avoid focusing on the overfitting problem, we have employed the dropout strategy during training (a dropout of 0.5 is typical for mammography). Over fitting may be avoided and network performance can be enhanced by using dropout methods.

4. Discussion of Experimental Outcomes

The proposed ICCN system is intended for use in the detection of breast cancer. This technology will provide radiologists with more resources and information. The enhancement of mammography, the detection accuracy, and the comparison with other models are the three main focuses of this model's outcomes.

In this model, 70% of the dataset was utilized for training and 30% was used for testing in accordance with conventional procedures [168][169]. The ICNN model for mammography classification was trained with both raw data and additional information. Table 2 displays the degree to which the original and enhanced photos maintain their correctness and the amount by which they degrade. The training accuracy and validation accuracy of ICNN have been attained as 99.48% and 92.05% correspondingly with the usage of augmented pictures and the 10-fold cross validation approach. Additionally, it produces a training loss of 0.0186% and a validation loss of 0.3584%. Therefore, ICNN's augmented accuracy rate is preferable. Table 3 displays the training time required by the model at 5437.6 seconds.

Table 2. Different types of accuracy and loss measures in percentage for original and augmented images

	Original Image	Augmented Image
Training Accuracy	97.88%	99.48%
Validation Accuracy	50.03%	92.05%
Training Loss	1.01%	0.0186%
Validation Loss	8.02%	0.3584

Table 3. Training options for pre-trained ICNN models

Training Option	Configuration	Input image size	Training Time	Data set
Optimizer	**Adadelta** is a popular Gradient Descent	256×256	5437.6 seconds	MIAS
Mini Batch Size	40			
Momentum Value	0.9			
Maximum Epochs	100			
Initial Learning Rate	0.001			
Execution Environment	GPU			

Achievement and accuracy rate:

Figure 4 depicts the success and failure rates for both the training and testing sets. After 60 epochs, the accuracy rates in training and testing are growing linearly or staying constant. For the training set, the loss rate gradually decreases for the first 60 epochs before leveling out. However, throughout testing, it varies and eventually drops to 100 epochs, after which the variation becomes quite minimal.

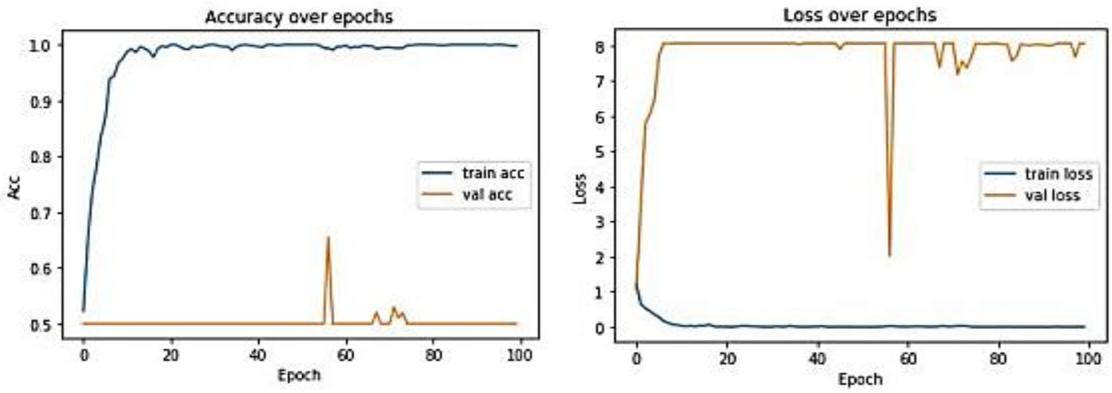

Fig 4. Accuracy and loss of the original images used in ICNN

5. Conclusion

Breast cancer lesions were classified using a supervised deep learning computer-aided diagnosis (CAD) system as presented in the paper. Results from experiments indicate that DL may significantly improve breast mammograms. The accuracy of each CNN model is compared to the others. Experiments with the BC cancer data set using the DL approach ICNN have shown an accuracy of 92.05%. CNN's quicker learning techniques and its ability to automatically extract relevant variables allow for superior performance in the prediction and prognosis of malignant lesions in mammographic images. The depth architecture of DL may be used in the detection of breast cancer in the future.

References

[1] K. Loizidou, G. Skouroumouni, G. Savvidou, A. Constantinidou, C. Nikolaou and C. Pitris, "Benign

and Malignant Breast Mass Detection and Classification in Digital Mammography: The Effect of Subtracting Temporally Consecutive Mammograms," *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, Ioannina, Greece, 2022, pp. 1-4, doi: 10.1109/BHI56158.2022.9926810.

[2] U. Supriya, R. Madhumathi and R. Sulthana, "An Analysis of Deep Learning Models for Breast Cancer Mammography Image Classification," *2022 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)*, Chennai, India, 2022, pp. 1-5, doi: 10.1109/ICDSAAI55433.2022.10028931.

[3] A. Kajala and V. K. Jain, "Diagnosis of Breast Cancer using Machine Learning Algorithms-A Review," *2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3)*, Lakshmangarh, India, 2020, pp. 1-5, doi: 10.1109/ICONC345789.2020.9117320.

[4] G. U. Srikanth and H. A., "Survey on Breast Cancer Based on Extreme Learning Machine Features," *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, Chennai, India, 2019, pp. 1-5, doi: 10.1109/ICIICT1.2019.8741497.

[5] Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C., & Parkin, D. M. (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. International journal of cancer, 127(12), 2893-2917.

[6] Swaminathan, R., Lucas, E., & Sankaranarayanan, R. (2011). Cancer survival in Africa, Asia, the Caribbean and Central America: database and attributes. IARC Sci Publ, 162, 23-31.

[7] Després, J. P., Alméras, N., & Gauvin, L. (2014). Worksite health and wellness programs: Canadian achievements & prospects. Progress in Cardiovascular Diseases, 56(5), 484-492.

[8] Richman, A. R., Torres, E., Wu, Q., & Kampschroeder, A. P. (2020). Evaluating a Community-Based Breast Cancer Prevention Program for Rural Underserved Latina and Black Women. Journal of Community Health, 1-6.

[9] Siegel, R. L., Miller, K. D., & Jemal, A. (2019). Cancer statistics, 2019. CA: a cancer journal for clinicians, 69(1), 7-34.

[10] Howlader, N., Noone, A. M., Krapcho, M., Garshell, J., Miller, D., Altekruse, S. F., ... & Mariotto, A. (2019). SEER cancer statistics review,1975-2016. Bethesda,MD: National Cancer Institute, 2019.

[11] DeSantis, C. E., Ma, J., Gaudet, M. M., Newman, L. A., Miller, K. D., Goding Sauer, A., ... & Siegel, R. L. (2019). Breast cancer statistics, 2019. CA: a cancer journal for clinicians, 69(6), 438-451.

[12] Goergen, S. K., Evans, J., Cohen, G. P., & MacMillan, J. H. (1997). Characteristics of breast carcinomas missed by screening radiologists. Radiology, 204(1), 131- 135.

[13] Timp, S., & Karssemeijer, N. (2004). A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography. Medical physics, 31(5), 958-971. -86

[14] Hadjiiski, L., Sahiner, B., Chan, H. P., Petrick, N., & Helvie, M. (1999). Classification of malignant and benign masses based on hybrid ART2LDA approach. IEEE transactions on medical imaging, 18(12), 1178-1187.

[15] Khuzi, A. M., Besar, R., Zaki, W. W., & Ahmad, N. N. (2009). Identification of masses in digital mammogram using gray level co-occurrence matrices. Biomedical imaging and intervention journal, 5(3).

[16] Vaqur, M. ., Kumar, R. ., Singh, R. ., Umang, U., Gehlot, A. ., Vaseem Akram, S. ., & Joshi, K. . (2023). Role of Digitalization in Election Voting Through Industry 4.0 Enabling Technologies. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2), 123–130. https://doi.org/10.17762/ijritcc.v11i2.6136

[17] Giger, M. L., & Suzuki, K. (2008). Computer-aided diagnosis. In Biomedical information technology (pp. 359-XXII). Academic Press.

[18] Fogel, D. B., Wasson III, E. C., Boughton, E. M., & Porto, V. W. (1998). Evolving artificial neural networks for screening features from mammograms. Artificial Intelligence in Medicine, 14(3), 317-326.

[19] Sickles, E. A. (1986). Mammographic features of 300 consecutive nonpalpable breast cancers. American Journal of Roentgenology, 146(4), 661-663.

[20] Hussain Bukhari, S. N. . (2021). Data Mining in Product Cycle Prediction of Company Mergers . International Journal of New Practices in Management and Engineering, 10(03), 01–05. https://doi.org/10.17762/ijnpme.v10i03.127

[21] Mudigonda, N. R., Rangayyan, R. M., & Desautels, J. L. (2001). Detection of breast masses in mammograms by density slicing and texture flow-field analysis. IEEE Transactions on Medical Imaging, 20(12), 1215-1227