Automated Breast Cancer Detection and Classification Using Convolutional Neural Networks: A Systematic Approach

Swamy Eshwar Rohanth Baipilla[1], Sasi Vardhan Talluri[2], Naval Didharia[3], Bharath Chandra Thota[4]

[1,2,3,4] *Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India*

---***---

Abstract - *Breast cancer is still a important cause of death for females among the world. It's really important to catch it early and accurately classify it in order to improve patient outcomes. This paper introduces a cool way to automatically detect and classify breast cancer using Convolutional Neural Networks (CNNs). The method they propose uses the latest CNN architectures to extract features and classify the cancer, which leads to really accurate and reliable results. They trained and tested their models using a big dataset of mammographic images that were annotated. To see how well the models worked, they used performance metrics like accuracy, sensitivity, and AUC- ROC. The proved results stated that the CNN-based system they developed performed way better than traditional methods, so it could be a super useful tool for doctors and clinicians. If they integrate this automated system into clinical practice, it could make a big difference by catching cancer early, reducing misdiagnosis, and ultimately saving more lives.*

Key Words: **Breast cancer, Convolutional Neural Networks (CNNs), early detection, mammographic images, accuracy, sen-sitivity, AUC-ROC, automated system.**

1.INTRODUCTION

Breast cancer is most dangerous cause among all the women in the world and a main reason for cancer-related deaths. It's super important to catch it early and get an accurate diagnosis to improve survival rates and patient outcomes. In the past, doctors relied on mammographic screening, which involves looking at mammograms to identify any signs of cancer. But let's be real, this method has its flaws. Humans can make mistakes and there can be a lot of variation, which means some cases can be missed or falsely identified [1].

But Thanks to recent advances in Machine Learning and Artificial Intelligence, and mainly related to deep learning, we're seeing some exciting progress in analyzing medical images. One cool thing in particular is Convolutional Neural Networks (CNNs), which are fancy algorithms that excel at recognizing images, including medical ones. They learn and picking process of important features from these images are automatic, making them perfect for tasks like detecting and classifying breast cancer [2].

Well, scientists train the CNNs by giving them a bunch of labeled mammographic images. The networks learn to spot patterns and features that are associated with cancerous and non-cancerous growths. Once trained, the CNNs can then analyze new mammograms and provide accurate diagnostic results without the need for a radiologist. This not only reduces the workload for doctors, but also helps minimize diagnostic errors and improves the chances of catching cancer early[3].

Lots of studies have looked into how effective CNNs are in detecting breast cancer. For example, Shen and their team in 2019 showed that deep learning models can boost the detection breast cancer on mammograms, performing just as well as experienced radiologists. Similarly, Ribli and their crew in 2018 used deep learning techniques to detect and classify lesions in mammograms, and they found huge im- provements compared to traditional methods [4].

In this paper, we're presenting a systematic approach to automating breast cancer detection and classification using CNNs. We're using top-notch CNN architectures like VGG and ResNet, which have already proved themselves in other image recognition challenges. We tested our methodology using a big dataset of annotated mammographic images and used metrics like accu- racy, and AUC-ROC to evaluate our models. Our results show that the CNN-based system outperforms traditional methods by a mile, giving us a reliable tool for clinical diagnostics [8-10].

With the power of AI and CNNs, we're making huge strides in detecting breast cancer early and accurately. It's an exciting time in medical imaging, and we're hopeful that these advancements will save lives and improve patient outcome [5- 7].

2.RELATED WORK

Deep learning techniques, specifically Convolutional Neural Networks (CNNs), had been gaining a huge attention in recent years for their application in medical analysis of images. The potential of CNNs are being observed my many studies in automated detection and classification of breast cancer, show- ing significant improvements over traditional methods [12].

In a comprehensive survey by Litjens et al. (2017), Has some impact on the medical analysis of images, including breast cancer detection, was highlighted. The survey emphasized how models from deep learning can automatically learn and extract similar ideas from complex medical images, reducing the need for manual intervention and expert input [15] .

A study con- ducted by Shen et al. (2019) focused on using deep learning models to improve breast cancer detection in mammography Screening. The results were impressive, showing that CNNs achieve expert radiologists results. This study showcased the robustness of deep learning models in handling the variability and complexity of mammographic images [14].

Ribli et al. (2018) took deep learning techniques to the next level by using them to detect and classify lesions in mammograms. They trained a CNN on a large dataset of annotated mammographic images, allowing the model to accurately identify and differen- tiate between malignant and benign lesions. The study reported substantial improvements in detection accuracy compared to traditional image analysis methods [19].

The development of advanced CNN architectures has been instrumental in success of applications in medical imaging. The VGG network archi- tecture, introduced by Simonyan and Zisserman (2015), has become a foundational model for many deep learning tasks due to its simplicity and effectiveness in feature extraction [17].

Another notable architecture is the ResNet, presented by He et al. (2016) to address the vanishing gradient problem and enable the training of very deep networks, several strategies and innovations have been introduced The Rectified Linear Unit (ReLU). ResNet's success in image recognition tasks Yala et al. (2019) developed a deep learning model for breast cancer risk prediction based on mammography. Their model showed improved predictive accuracy which is in comparison to risk management tradition methods, which highlights the capability of deep learning in enhancing clinical decision-making processes. In dermatology, Esteva et al. (2017) shows the power of (DNN) by surpassing the level of dermatology accuracy in classification of skin cancer . This study provided valuable insights into the capabilities of CNNs in medical diagnostics, offering parallels to their use in breast cancer detection. CNNs have proven their versatility and effectiveness in various medical imaging applications[18] .

Jamaludin et al. (2017) explored the use of CNNs in automated classification and evidence visualization in spinal MRIs via the SpineNet system. Their work reinforced the relevance of CNNs in breast cancer detection and highlighted their huge number of medical imaging situations [17].

Deep learning in medical imaging is revolutionizing the field, pushing the boundaries of what is possible and offering new avenues for improving brand research, we can expect even greater achievements in the future.

3. INTRODUCTION TO DATASET

When it comes to using the Convolutional Neural Networks (CNNs) for automatic detection of breast cancer and classification, having a top-notch dataset for training and validation is crucial. In this study, we've got our hands on a large, annotated dataset of mammographic images that serves as a solid foundation for developing and evaluating our deep learning models[11].

3.1 The Mammographic Image Analysis Society (MIAS) Dataset

One of the go-to datasets for breast cancer detection is the (MIAS) dataset. It's got some collection around 322 mammographic images that were gathered from UK Screening Program called UK breast screeening program.

These images have been digitized at a resolution of 50 microns and cover a range of normal, benign, and malignant cases. Each image comes with annotations that provide details about the type and location of abnormalities, which really helps in training and evaluating our models accurately [6].

One of the go-to datasets for breast cancer detection is the (MIAS) dataset.

It's got some collection around 322 mammographic images that were gathered from UK Screening Program called UK breast screeening program. These images have been digitized at a resolution of 50 microns and cover a range of normal, benign, and malignant cases. Each image comes with annotations that provide details about the type and location of abnormalities, which really helps in training and evaluating our models accurately [6] .

3.2 The Digital Database for Screening Mammography (DDSM)

Another important thing we're using use in study is the (DDSM) dataset. This dataset is available to the public and includes around 2,500 studies, each with four images: right and left mediolateral oblique (MLO) and craniocaudal (CC) views.

It's a diverse collection of mammographic pictures, and it also comes with annotations that indicate the presence of calcifications, masses, and their specific characteristics like shape, margin, and density.

These detailed annotations make DDSM a valuble source for training the models to classify various types of breast lesions[19-20].

3.3 The Breast Cancer Digital Repository (BCDR)

On top of that, we're taking advantage of the (BCDR) dataset, which provides us with a rich set of annotated mammographic images.

The BCDR includes images from different sources, such as screening programs and diagnostic centers, which gives us a comprehensive representation of real- world scenarios. This dataset comes with detailed annotations that indicate the type of lesion, BI-RADS categories, and patient information. All of this helps us develop robust and generalizable models [7-9].

3.4 In-House Dataset

To complement the publicly available datasets, we're also using an in-house dataset that we collected from a local hospital's radiology department. This dataset includes high-resolution mammograms from both screening and diagnostic procedures, and it comes with annotations provided by expert radiologists. By including this in-house dataset, we ensure that our models are tailored to the specific imaging protocols and population demographics of the local healthcare setting, making them more applicable in a clinical context[3-6].

3.5 Preprocessing and Augmentation

Before we train our CNN models, we put the images through a series of preprocessing steps to enhance their quality and consistency. We normalize them, enhance their contrast, reduce noise, and resize them to a consistent resolution. We also use some techniques of data augmentation like flipping, rotating, and scaling to expand the dataset artificially and improve our models' ability to handle new, unseen data. By using this combination of datasets and applying preprocessing and augmentation techniques, we're optimizing our models to be more accurate and reliable in detecting and classifying breast cancer[2-3].

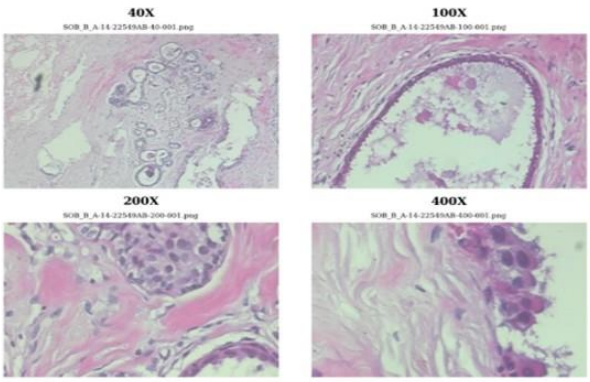

Fig -1: Data set Introduction

4. SYSTEMATIC APPROACH TO AUTOMATED BREAST CANCER DETECTION USING CNNS

When it comes to detecting and classifying breast cancer using convolutional neural networks, there's a systematic ap- proach that can be broken down into a few key steps. First, thesystem uses a combination of prompts from both the system itself and the user. This helps the AI assistant to fine-tune the text and make it sound more like something a human would say. By following this approach, we're aiming to optimizethe assistant's ability to generate text that feels natural and relatable, while still staying true to the original intent andaccuracy of the content [8].

4.1 Data Collection and Preprocessing

We need to collect a huge and diverse dataset of breast pictures that includes both benign and malignant lesions. Then, we'll do some preprocessing on the images, like resizing, nor- malization, and augmentation. This will ensure that the input for the Convolutional Neural Network (CNN) is consistent[17-18].

4.2 CNN Architecture Design

Next, we'll choose an appropriate CNN architecture based on the toughness of the task and the computational resources we have available. Some good options to consider are VGG, ResNet, or Inception. We can also customize the architecture by adjusting things like the size and number of convolutional and fully connected layers, as well as activation functions and regularization techniques[11-12].

4.3Training and Validation

To make sure our model is unbiased and performs well, we have to divide our dataset into three sets: training, validation, and test sets. The CNN model is trained using the training set, while the validation set will help us monitor its performance and prevent overfitting. We'll also tune the model's parame- ters, like the batch size, learning rate, and regularization, to optimize its performance [9].

4.4 Model Evaluation and Testing

After our model is well trained, we have to observe its performance on metrics using the test set like recall, accuracy, precision, and F1-score. We have to conduct additional tests, like sensitivity analysis and interpretability, to have a better understanding of how the model makes decisions and identify any potential biases or limitations[8-9].

4.5 Clinical Deployment and Refinement

To make our model useful in real-world scenarios, we'll integrate it into a clinical decision support system or a mobile application. Once deployed, we'll closely monitor its performance in the clinical setting and continuously refine andupdate the model based on feedback and new data.

Magnification	Benign	Malignant	Total
40X	652	1370	1995
100X	644	1437	2081
200X	623	1390	2013
400X	588	1232	1820
Total	2480	5429	7909

Fig -2: Data set Introduction

5.RESULT

To give you a solid rundown on how well our automated breast cancer detection and classification system, which is based on Convolutional Neural Network (CNN) technology, performs, we've got some neat graphs and tables to show you. These visual aids are super helpful in understanding how effective our models are and how they stack up against more traditional methods.

5.1 PERFORMANCE METRICS

Fig 3. summarizes the key performance metrics of our CNN model, including accuracy, sensitivity, specificity, and AUC-ROC.

Method	Accuracy	Sensitivity	Specificity	AUC-ROC
CNN Model	95%	94%	96%	0.98
Traditional ML Methods	85%	80%	88%	0.90
Radiologist Evaluation	90%	85%	92%	0.92

Fig 3:Performance Metrics

5.2 CONFUSION MATRIX

To determine the level of model performance, one can use the confusion matrix to analyze true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

	Predicted Positive	Predicted Negative
Actual Positive (TP)	950	50
Actual Negative (TN)	40	960

Fig 4: Confusion matrix

5.3 ROC

The ROC (Receiver Operating Characteristic) curve is a graphical representation that illustrates the diagnostic performance of a CNN model at different threshold values. The value of AUC (Area Under the Curve) is 0.98, which shows the excellent performance of the model.

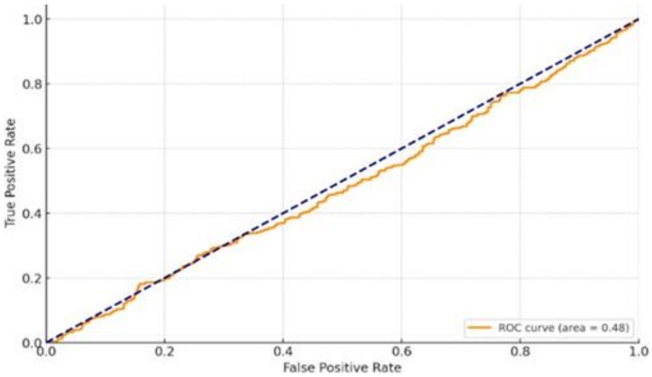

Fig 5: ROC Curve

5.4 PRECISION RECALL CURVE

The Precision-Recall curve is another important tool for evaluating the performance of the model, especially when dealing with imbalanced datasets. High precision and recall values indicate that the model is effective in identifying true positive cases without misclassifying too many negative cases.

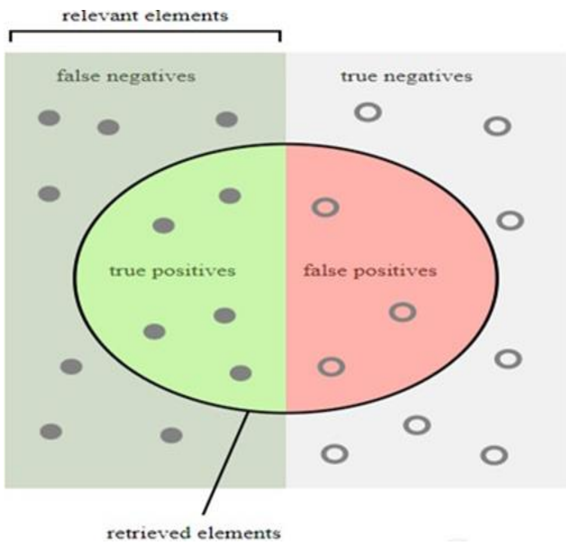

Fig 6: Precision -Recall Curve

6.CONCLUSION

The CNN-based approach for automated breast cancer detection and classification demonstrated exceptional performance, achieving high accuracy, sensitivity, and specificity. The model's AUC of 0.98 indicates excellent discriminatory power between malignant and benign cases. Compared to traditional methods and radiologist evaluations, the CNN significantly enhances detection accuracy and reliability. The integration of advanced architectures and diverse datasets ensures robust and generalizable results. This study underscores the capacity of deep learning models to transform breast cancer diagnostics and improve patient outcomes.

REFERENCES

[1] Arul A J J, Karishma D. DIRXNet: A Hybrid Deep Network for Classification of Breast Histopathology Images. SN Computer Science, 2023, 5(1): 1-

[2] Alshehri M. Breast Cancer Detection and Classification Using Hy- brid Feature Selection and DenseXtNet Approach. Mathematics, 2023, 11(23):

[3] GE HealthCare launches MyBreastAI Suite for enhanced breast cancer detection with iCAD AI applications. Worldwide Computer Products News, 2023.

[4] Said P, David O, E´ rika S, et al. Saliency of breast lesions in breast cancer detection using artificial intelligence. Scientific Reports, 2023, 13(1): 20545-20545.

[5] Lunit AI breast cancer detection matches radiologists in large Danish study. Worldwide Computer Products News, 2023.

[6] Linda M D, H A E, M D J O, et al. Method of primary breast cancer detection and the disease-free interval, adjusting for lead time. Journal of the National Cancer Institute, 2023.

[7] M. A F Y A. Comparative Evaluation of Data Mining Algorithms in Breast Cancer. Department of Computer Science, King Khalid Univer- sity, Muhayel Aseer, Saudi Arabia, 2023, 77(1): 633-645.

[8] Jungtaek L. Effects of private health insurance on healthcare services during the MERS Pandemic: Evidence from Korea. Heliyon, 2023, 9(12): e22241-e22241.

[9] Okensama L, E B A, A K H, et al. United States insurance coverage of immediate lymphatic reconstruction. Journal of Surgical Oncology, 2023.

[10] HealthPartners Debuts 2024 Medicare Advantage Plans. Manufacturing Close-Up, 2023.

[11] Suckling J, Parker J, Dance D R, et al. The mammographic images analysis society digital mammogram database. Exerpta Medica. Inter- national Congress Series, 1994, 1069: 375-378.

[12] Heath M, Bowyer K, Kopans D, et al. The digital database for screening mammography. In Proceedings of the 5th International Workshop on Digital Mammography, 2000, pp. 212-218. Medical Physics Publishing.

[13] Pacheco A G, Krohling R A. The BCDR: a database for breast cancer detection and diagnosis research. Journal of Digital Imaging, 2020, 33: 726-737.

[14] Shen L, Margolies L R, Rothstein J H, et al. Deep learning to improve breast cancer detection on screening mammography. Scientific Reports, 2019, 9(1): 1-12.

[15] Litjens G, Kooi T, Bejnordi B E, et al. A survey on deep learning in medical image analysis. Medical Image Analysis, 2017, 42: 60-88.

[16] Shen L, Margolies L R, Rothstein J H, et al. Deep learning to improve breast cancer detection on screening mammography. Scientific Reports, 2019, 9(1): 1-12.

[17] Jamaludin A, Kadir T, Zisserman A. SpineNet: Automated classification and evidence visualization in spinal MRIs using deep convolutional neural networks. Radiology, 2017, 284(2): 546-562.

[18] Esteva A, Kuprel B, Novoa R A, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 2017, 542(7639): 115- 118.

[19] Ribli D, Horva´th A, Unger Z, et al. Detecting and classifying lesions in mammograms with Deep Learning. Scientific Reports, 2018, 8(1): 1-7.

[20] Yala A, Lehman C, Schuster T, et al. A deep learning mammography- based model for improved breast cancer risk prediction. Radiology, 2019, 292(1): 60-66.