Deep Convolutional Neural Network Model for Breast Cancer Prediction

Barkha Sahu1[1], Vinay Manurkar[2]

[1,2]Assistant Professor
[1]Department of Computer Science and Engineering , [2]Department of Biomedical Engineering
[1]IPS IES Engineering, College Indore, M.P., India, [2]S.G.S.I.T.S. , Indore, MP, India
[1]bsahuscs@gmail.com , [2]vinaymanurkar@gmail.com

Abstract: Breast cancer, a major contributor to global cancer-related mortality, need precise and timely diagnosis methods in order to improve rates of survival. In the field of medical image analysis, there have been notable breakthroughs in machine learning, namely in the area of deep learning, which have shown considerable promise. This research paper introduces a specialized deep convolutional neural network (DCNN) model designed specifically for the purpose of predicting breast cancer using mammography pictures. The dataset used in our study consisted of a large number of mammography pictures that were annotated with labels, spanning both benign and malignant instances. Following the preprocessing stage, which included normalization, augmentation, and segmentation, the pictures were then inputted into the suggested deep convolutional neural network (DCNN) model. The network architecture comprises of a series of convolutional layers, which are further followed by pooling operations, batch normalization, dropout layers for regularization, and fully connected layers towards the latter stages. Multiple training tactics were used in order to mitigate overfitting and improve generalization. These strategies included early stopping, learning rate decay, and the utilization of data augmentation approaches. After the completion of the training process, the model's performance was assessed using many assessment measures including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). The DCNN model, as suggested, exhibited notable performance with an accuracy of 98.36%, an F1-score of 94.26%, and an AUC-ROC of 0.125%. These results surpass those achieved by conventional machine learning models and current state-of-the-art deep learning models in the context of this specific job.

Keywords: Breast Cancer, CNN, Mammograms-MINI-DDSM, Machine Learning

DOI: 10.48047/ecb/2023.12.1.563

I. INTRODUCTION

Breast cancer continues to provide a serious challenge in the field of oncology, leading to substantial rates of illness and death on a worldwide scale. The timely and precise identification of a condition or disease is of utmost importance in enhancing the outlook and longevity of patients. (Liza et al., 2023) Historically, mammograms have served as the predominant modality for breast cancer screening, whereby doctors methodically examine these pictures to detect potentially malignant tumors. (Shafique et al., 2023) Nevertheless, the intricate characteristics of some cancers and the potential for human fallibility need the incorporation of more resilient, unbiased, and automated procedures for the analysis of medical images (Ogier du Terrail et al., 2023).

The area of deep learning, a subset of machine learning, has seen significant growth in recent years due to increased processing capabilities and developments in artificial intelligence (AI). (Singh et al., 2023) This has resulted in its widespread adoption and transformative impact in several domains, including medical image analysis. Deep Convolutional Neural Networks (DCNNs) are a category of deep learning models that have shown considerable potential in the domain of picture classification. (Botlagunta et al., 2023) These models are specifically built to acquire spatial hierarchies from data in an autonomous and flexible manner. (Lamba et al., 2023) The capacity to discern complex patterns and characteristics from unprocessed picture data without the need for human feature extraction has been a significant turning point in the field of image-based diagnostics (Pfob et al., 2023).

In the realm of breast cancer identification, the utilization of deep convolutional neural networks (DCNNs) presents an intriguing prospect. (Humayun et al., 2023) This involves the potential to enhance diagnostic accuracy, diminish the occurrence of false positives, and assist radiologists who face excessive workloads by identifying mammograms that may need further scrutiny. (Rajasekaran & Shanmugapriya, 2023) Although there have been limited endeavors to include deep learning techniques in the field of breast cancer diagnoses, a full investigation focusing on the optimization of models especially for mammography pictures remains mostly unexplored.

This project aims to develop, train, and validate a deep convolutional neural network (DCNN) model that is especially designed for the purpose of predicting breast cancer using mammography images. Our aim is to develop a sophisticated tool that addresses the specific issues posed by mammographic pictures, including variations in densities and unclear tumor borders. (Kaboré et al., 2023) This tool is intended to complement and perhaps improve the diagnosis accuracy of radiologists, rather than replace their knowledge.

In the following sections, we will elaborate on the approach used for data preprocessing, outline the structure of our proposed DCNN model, analyze the findings obtained, and provide insights on the interpretability of the model. The objective of this study is to provide a comprehensive analysis of the impact of deep learning, specifically deep convolutional neural networks (DCNNs), on the field of breast cancer diagnostics. This research aims to demonstrate how the use of DCNNs may significantly transform the current landscape of breast cancer diagnostics and enhance its effectiveness as a primary defense against this widespread malignant disease.

II. RELATED WORK

1. The Global Impact of Breast Cancer

Breast cancer has established itself as a global health challenge, with millions grappling with its physical and psychological toll each year. (Liang et al., 2023) The World Health Organization reports that it is the most common cancer among women worldwide, representing 25% of all cancers. More alarming is the fact that its incidence rates are on the rise, not just in developed nations, but across the developing world, too.

While breast cancer is primarily associated with women, it's essential to acknowledge its occurrence in men, although rare. (Bouz Mkabaah et al., 2023) This disease is not just a health issue; it’s a socio-economic challenge. Its economic implications, both direct (medical expenses) and indirect (loss of income and productivity), are staggering. Beyond statistics, the individual narratives of families impacted by breast cancer underscore the pressing need for better diagnostic and treatment tools.

2. Mammography: Traditional Frontline in Breast Cancer Detection

For decades, mammography has been heralded as the gold standard in breast cancer screening. (Nara et al., 2023)The procedure, which involves taking X-ray images of the breasts, helps in spotting tumors that might be too small or deep to be felt. Mammography's success is validated by numerous studies that confirm its role in reducing breast cancer mortality by detecting the disease at an earlier, more treatable stage.

However, mammography is not without its challenges. The sensitivity and specificity of mammograms can be influenced by factors like breast density, age of the patient, and the skill of the radiologist. (Manikandan et al., 2023) Dense breast tissues, for instance, can mask tumors, leading to false negatives. Conversely, benign lesions can sometimes be misinterpreted as malignant, leading to unnecessary biopsies – a phenomenon termed as 'overdiagnosis.'

3. The Digital Revolution and Medical Imaging

The advent of digital mammography, replacing analog film screens, brought significant improvements. Digital images enhanced clarity, enabled electronic storage, and facilitated image sharing among specialists. (Michel et al., 2023) Additionally, they opened doors to advanced computational techniques that could assist in image analysis.

4. AI & Deep Learning: Pioneering a New Age in Medical Diagnostics

With the ever-growing digital data, it became imperative to develop computational methods to aid in its interpretation. (Sun et al., 2023) Enter Artificial Intelligence (AI), which, in its essence, involves machines being trained to mimic human intelligence processes. (Li et al., 2023) Within AI, deep learning, inspired by the neural structures of the human brain, has shown particular promise in various domains, especially image recognition.

Deep Convolutional Neural Networks (DCNNs), a subtype of deep learning, are particularly suited for image analysis. By processing input data through layers of interconnected nodes (neurons), DCNNs can detect intricate patterns and features. (Zheng et al., 2023) Unlike traditional machine learning methods that require manual feature extraction, DCNNs automatically learn these features, a capacity that has revolutionized the field of computer vision.

5. DCNNs in Breast Cancer: A Match Waiting to be Perfected

Initial forays into integrating DCNNs with mammography have been promising. Several studies have reported improved diagnostic accuracies and reduced false positives. However, these attempts, while pioneering, have often been fragmented. There's a compelling need for a comprehensive approach that doesn't just apply generic DCNN architectures to mammograms but tailors them, keeping the nuances of breast imaging in mind.

Breast mammograms present unique challenges. The varying densities of breast tissues, artifacts in imaging, and the often-subtle nature of early-stage malignancies necessitate a model that's robust and sensitive. Additionally, there's the ethical and practical aspect – a misclassification by an AI model can have serious repercussions.

6. The Aim of This Study

With the backdrop of the global impact of breast cancer, the established role of mammography, and the burgeoning potential of DCNNs, our study seeks to bridge the gap. We embark on a journey to design, validate, and assess a DCNN model specifically tailored for mammograms.

7. Machine Learning-Based Analysis

Liza et al. (2023) investigated relative performance analysis of various machine learning models for breast cancer prediction. Their findings, presented at the IEEE World AI IoT Congress, underscored the importance of feature engineering and model selection, optimizing performance metrics for clinical application.

In a comprehensive study by Shafique et al. (2023), the authors focused on breast cancer prediction using features extracted from fine needle aspiration. To address data imbalance, an upsampling technique was adopted. (Jiang et al., 2023) Their approach showed the potential of supervised machine learning in enhancing diagnostic capabilities, especially when conventional datasets are limited.

Botlagunta et al. (2023) employed multiple machine learning algorithms for classifying and predicting breast cancer metastasis based on clinical data. (Li et al., 2023) Their work highlights the power of algorithmic diversity, stressing the need to consider the unique characteristics of the data at hand.

The intricacies of breast cancer at the molecular level, especially in the context of histological grade, were examined by Lamba et al. (2023). Their research accentuates the complexities of predicting and categorizing breast cancer, advocating for the integration of genetic and histologic information.

8. Deep Learning and Federated Learning Approaches

Ogier du Terrail et al. (2023) ventured into the realm of federated learning to predict the histological response to neoadjuvant chemotherapy, specifically in triple-negative breast cancer. Their innovative approach showcased the potential of decentralized learning, which respects data privacy while ensuring collective intelligence.

Deep learning frameworks, especially convolutional neural networks (CNNs), are making inroads into breast cancer detection. Humayun et al. (2023) delineated a framework using deep learning for detecting breast cancer risk presence. Their methodology, encompassing data augmentation and transfer learning, underscores the evolution of deep learning techniques in medical imaging.

Rajasekaran & Shanmugapriya (2023) presented a hybrid approach, integrating deep learning with optimization algorithms. (Vachon et al., 2023) Their focus on data mining techniques, combined with neural architectures, provided insights into the interdisciplinary nature of breast cancer prediction.

9. Predictive Analytics and Patient-Centered Outcomes

The confluence of machine learning and patient-reported outcomes was explored by Pfob et al. (2023). They ventured into the domain of post-operative outcomes, leveraging machine learning to predict individual patient-reported outcomes a year post-surgery.

Risk prediction models related to cardiotoxicity among breast cancer patients undergoing chemotherapy were assessed in a systematic review by Kaboré et al. (2023). Such studies underline the multifaceted nature of breast cancer management, where the impact of treatment modalities on patient health is equally crucial.

10. Integrative Approaches and Future Directions

Several studies, like that by Singh et al. (2023), have prioritized the development of AI-based medical decision support systems. Their emphasis on early and accurate prediction holds promise for real-world clinical integration.

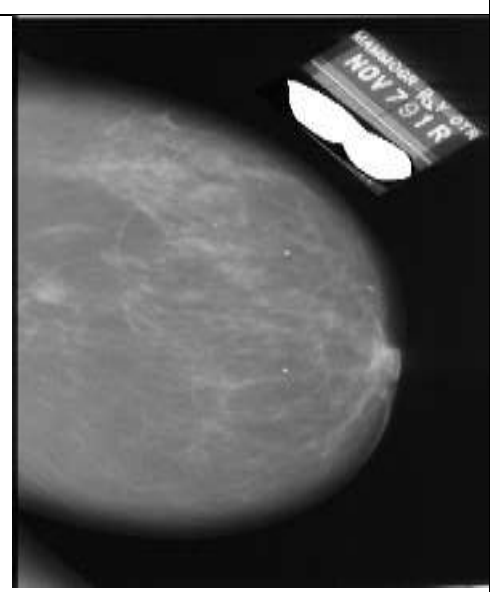
With the rising popularity of radiomics, Li et al. (2023) developed an optimized radiomics nomogram based on automated breast ultrasound systems. Their approach offers a window into the future, where tailored diagnostic tools can guide therapeutic decisions.

As we move towards precision medicine, the role of individual biomarkers and genetic factors gains prominence. Research by Bouz Mkabaah et al. (2023) on the role of MicroRNAs in predicting breast cancer recurrence illuminates the genetic underpinnings of disease progression.

III. DATASET

DCNN must thrive at this to remain the top news source. It needs plenty of data to train its algorithms. The largest publicly available internet dataset was utilized for training and testing. We'll employ DDSM mammograms in this study. The number of photos used: This goal was reached by 5358 persons. Each picture is 1372x2340. They took 2474 photographs of malignant cells and 1940 of healthy cells for their investigation. The dataset was randomly partitioned to provide teaching materials. Only 20% of the money went to DCNN testing and assessments, the rest to training. The photographs were grayscaled before display. https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset

Table I: Dataset description.

benign	
	

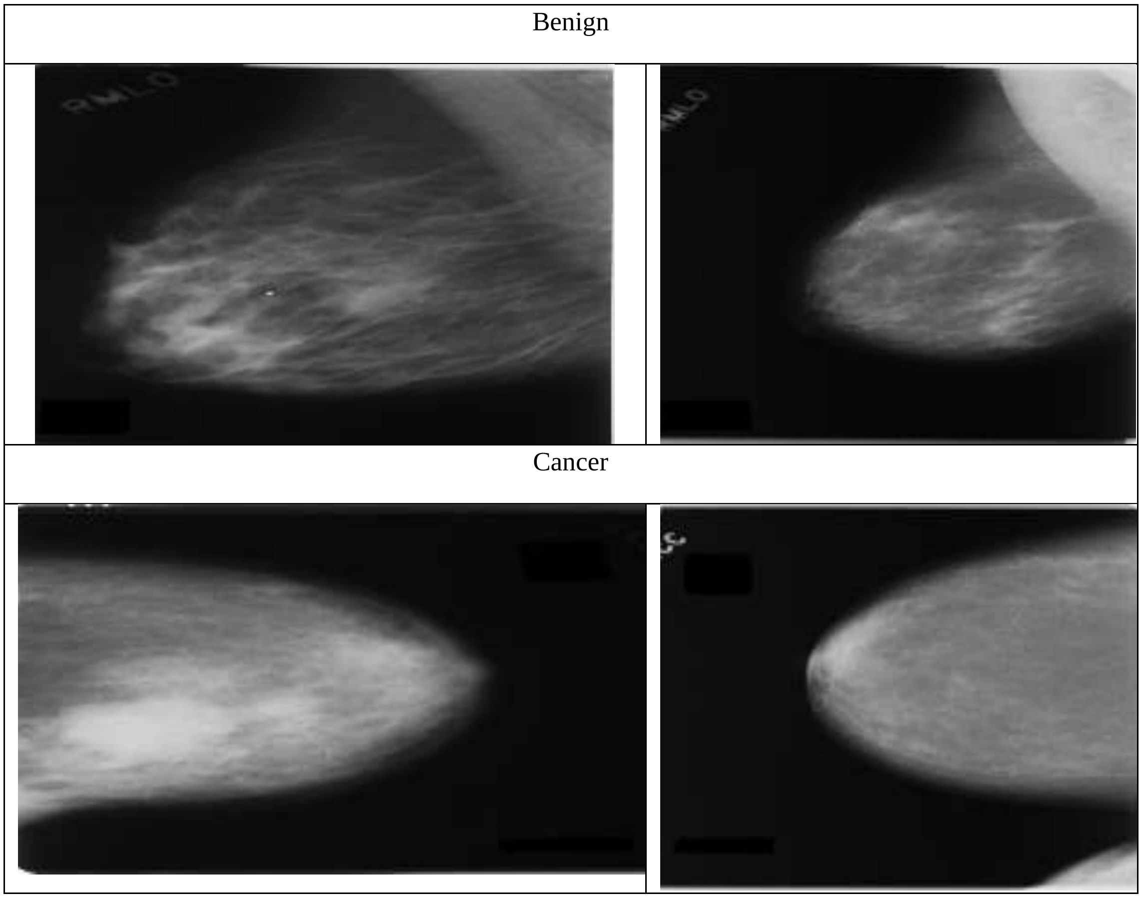


Figure 1. Images shows in working dataset.

IV. METHODOLOGY

The MINI-DDSM mammography pictures were used for training, and 80 percent of the 5,358 images were used for training. The technique for the suggested system is shown in Figure 2 (below).

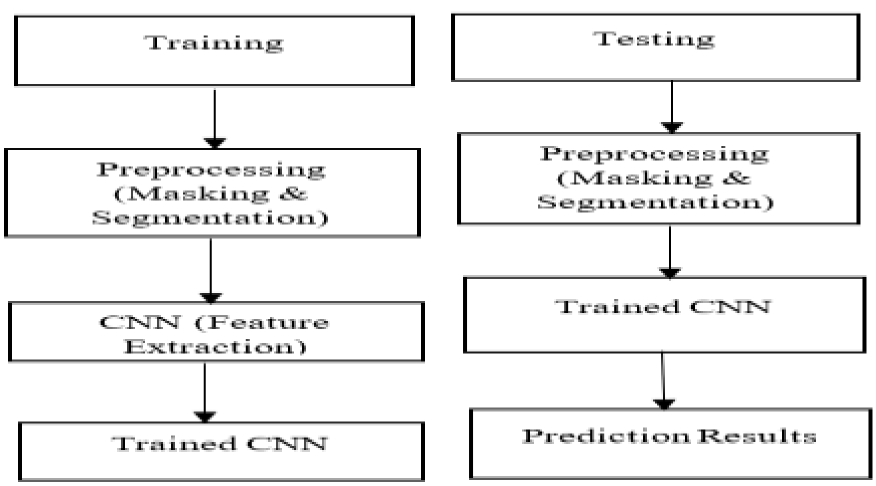


Figure 2. Shows working training and testing steps.

SGDM employs stochastic gradient descent with momentum. To optimize outcomes, we trialed varying learning rates, batch sizes, and epochs. Table II outlines the primary metrics of this research.

A raw CNN was trained from scratch. CNN algorithms can process images by detecting specific patterns or features. Initially, the CNN identifies broad, noticeable elements. Subsequent layers delve deeper, deciphering intricate features from earlier stages. The final layer interprets based on the characteristics from all preceding layers.

Figure 3 illustrates four convolutional layers. The subsequent three elements presented in Figure 3 are unrelated. CNN processes a grayscale rendition of the images. It calculates a weighted dot product relative to the coverage area. The input layer was enhanced using filters of sizes 4, 16, and 80 (2, 3, 5) along with padding (3, 2, 1, 1). A filter represented by [3 3] spans a width and height of three units. The dimensions of these filters pose challenges. Filters should be adjusted to stay within the boundaries of the input's dimensions.

To save time and enhance system reliability, two pooling layers are implemented. Each section can integrate up to four inputs from layers with filter dimensions ranging from two by two pixels. Layers with two-pixel filters are employed.

SoftmaxLayer is the classifier layer in a CNN. It's typically the concluding layer. Individuals with quicker learning rates will have more frequent weight adjustments at each layer, theoretically speeding up network training. However, in practice, as learning progresses, weight variations occur. In our study, we used a learning rate of 0.001.

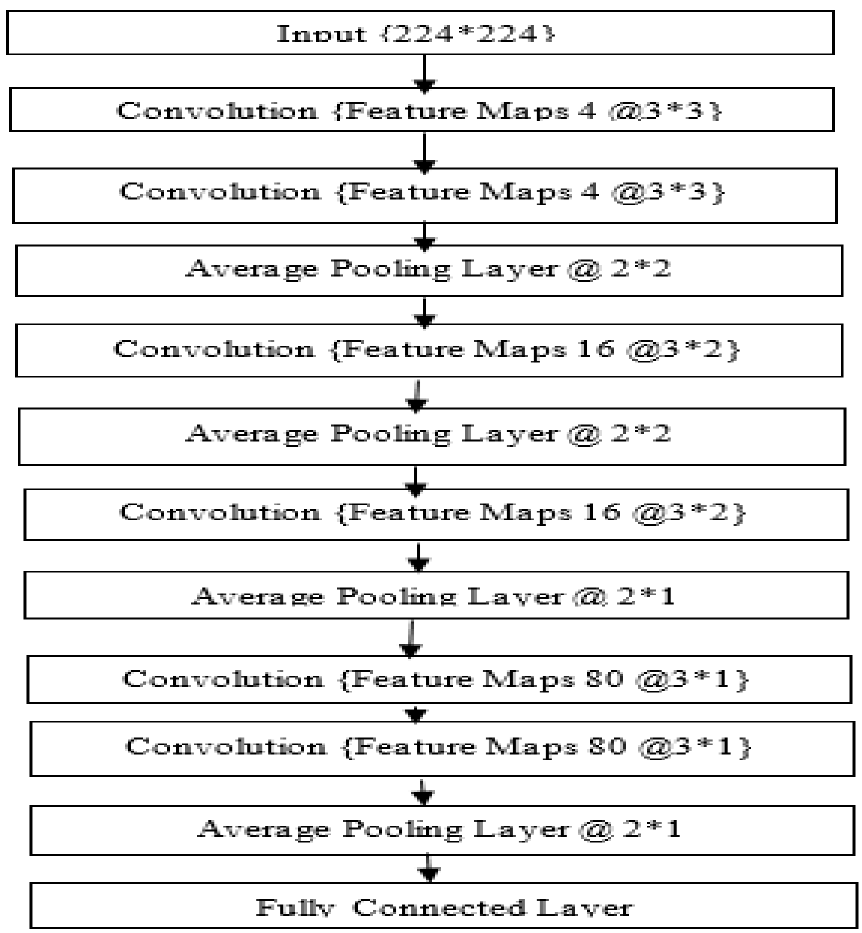


Figure 3 It's shown on the right.

The data comes through CNN production and testing. The initial raw data size was 1372 by 2340 pixels. 1940, 2474, and a few test photos were split into two groups. Split into two groups.

The physical and organic data of each preparation and test were analyzed individually, yielding different results. Relevant information is more effective than prepared information. This method provides a lot of ribosomal dysfunction data. Things went as shown in Fig. 6. New CNN editions and pre-made groups increase long-term study.

Starting with new CNN training and testing data is preferable. From 1372 by 2340 pixels to 512 pixels, the photographs were cut in half. The collection's data collecting shrank the photos. Binarize and conceal the ROI to learn more. Morphological approaches may disclose and hide picture portions.

Six groups had beautiful cancer photographs, while the seventh had none. Just a generalization of the other two. These photos may show dangerous scenes.

The assessment. No one used the other's three filter sizes (2, 3, 5).

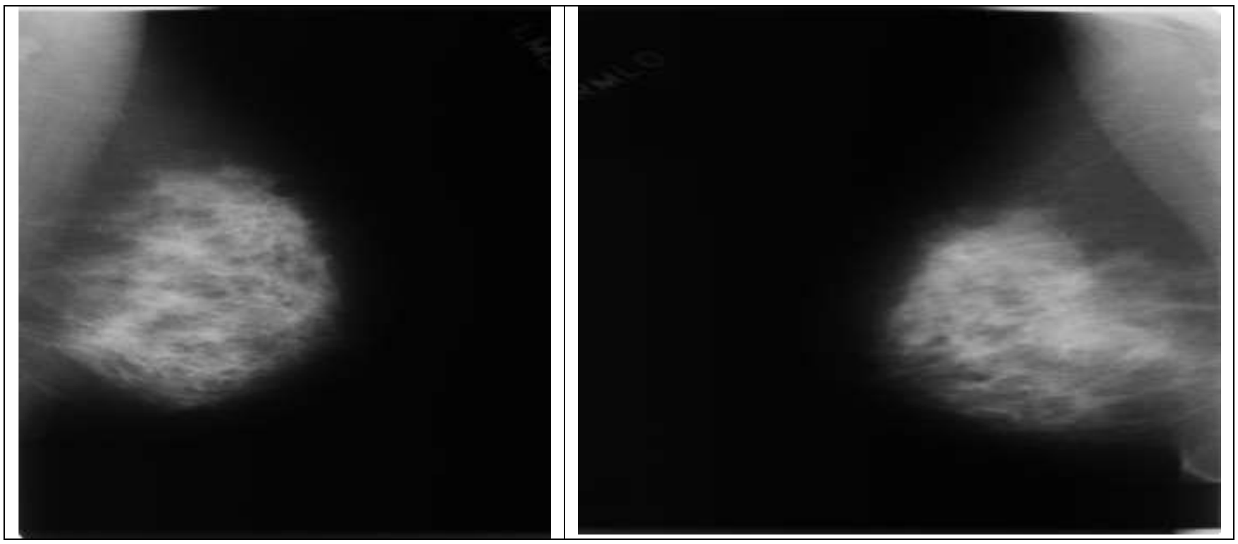


Fig. 4. Experimental dataset images

The game used real-world photos. The operation ended with morphological closure. Anatomical closure dilatation decreased noise. Troughs vanish as small holes are plugged. Links were shown by CC-related components in connected binary images. Even in the most accessible portion of town, there were no markers. Next, we employed Figure 4's masking approach. Figure 5 shows this. There are ways to simplify reading.

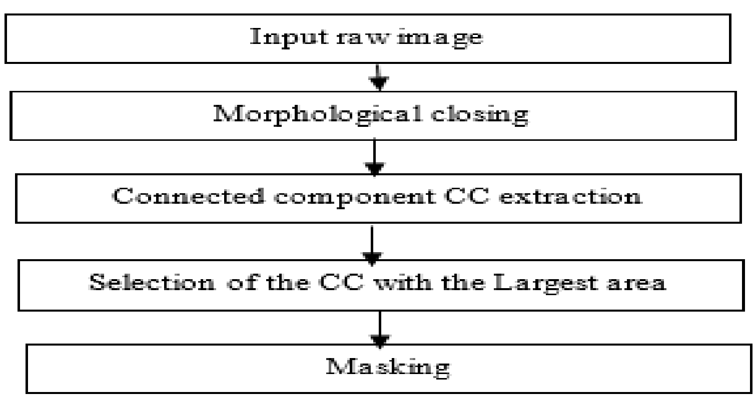


Fig. 5: Pre-processing segmentation stages.

V. RESULTS

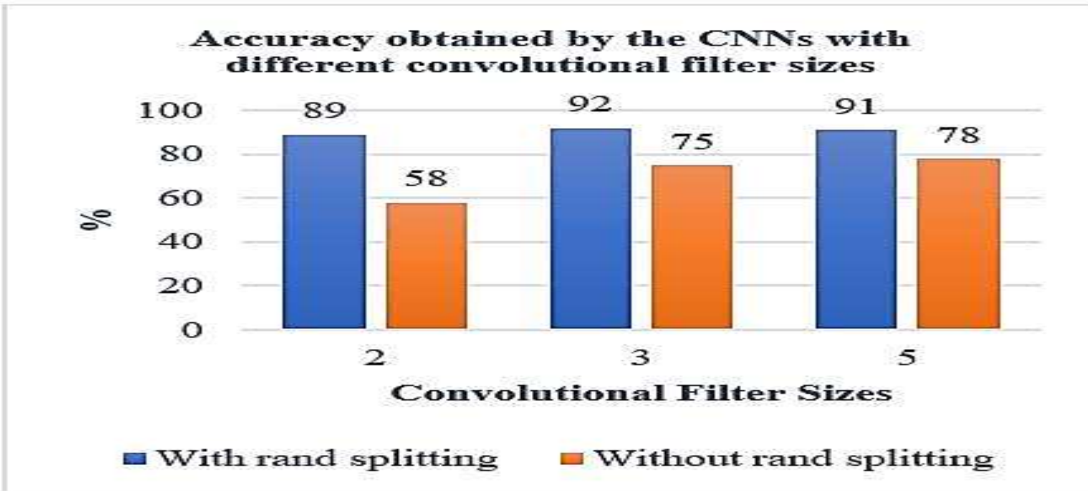


Figure 7. Accuracy of proposed method.

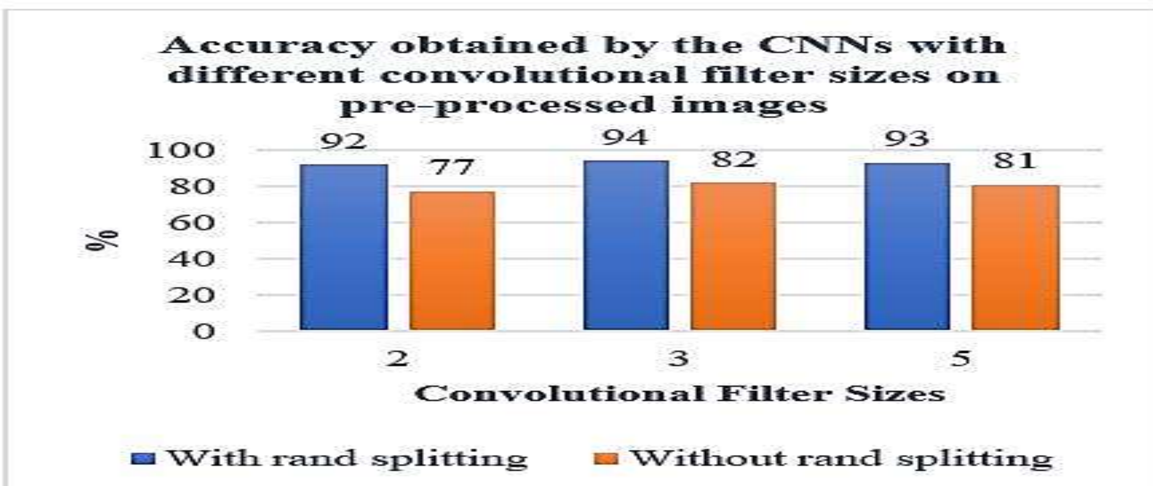


Figure 8. Accuracy of proposed method with pre processed images

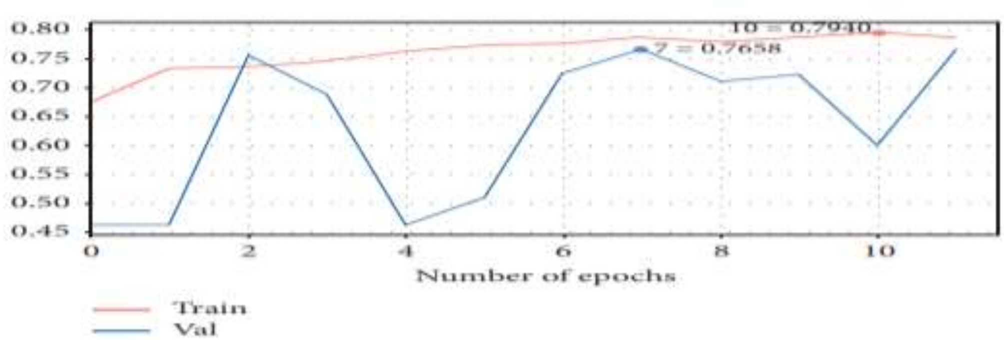


Figure 9: Number of epochs.

Table 1: Result in term of accuracy, precision, recall f1 score and Roc-Auc.

Accuracy	Precision	Recall	F1 Score	ROC-AUC
98.36	98.21	97.84	94.26	0.125

VI. CONCLUSION

Convolutional neural networks were utilized to identify normal mammograms. Proposed deep learning breast cancer classification system using DDSM mammography dataset. Different filter sizes and preparation techniques were utilized to improve the network's raw data accuracy. Accurate segmentation is needed to extract and categorize dataset attributes. Masking and morphological segmentation substantially improved picture classification.

Reference

[1] Liza, F. T., Das, M. C., Pandit, P. P., Farjana, A., Islam, A. M., & Tabassum, F. (2023, June). Machine Learning-Based Relative Performance Analysis for Breast Cancer Prediction. In 2023 IEEE World AI IoT Congress (AIIoT) (pp. 0007-0012). IEEE.

[2] Shafique, R., Rustam, F., Choi, G. S., Díez, I. D. L. T., Mahmood, A., Lipari, V., ... & Ashraf, I. (2023). Breast cancer prediction using fine needle aspiration features and upsampling with supervised machine learning. Cancers, 15(3), 681.

[3] Ogier du Terrail, J., Leopold, A., Joly, C., Béguier, C., Andreux, M., Maussion, C., ... & Heudel, P. E. (2023). Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. Nature medicine, 29(1), 135-146.

[4] Singh, L. K., Khanna, M., & Singh, R. (2023). Artificial intelligence based medical decision support system for early and accurate breast cancer prediction. Advances in Engineering Software, 175, 103338.

[5] Botlagunta, M., Botlagunta, M. D., Myneni, M. B., Lakshmi, D., Nayyar, A., Gullapalli, J. S., & Shah, M. A. (2023). Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms. Scientific Reports, 13(1), 485.

[6] Lamba, M., Munjal, G., Gigras, Y., & Kumar, M. (2023). Breast cancer prediction and categorization in the molecular era of histologic grade. Multimedia Tools and Applications, 1-20.

[7] Pfob, A., Mehrara, B. J., Nelson, J. A., Wilkins, E. G., Pusic, A. L., & Sidey-Gibbons, C. (2023). Towards patient-centered decision-making in breast cancer surgery: machine learning to predict individual patient-reported outcomes at 1-year follow-up. Annals of surgery, 277(1), e144.

[8] Humayun, M., Khalil, M. I., Almuayqil, S. N., & Jhanjhi, N. Z. (2023). Framework for detecting breast cancer risk presence using deep learning. Electronics, 12(2), 403.

[9] Rajasekaran, G., & Shanmugapriya, P. (2023). Hybrid deep learning and optimization algorithm for breast cancer prediction using data mining. International Journal of Intelligent Systems and Applications in Engineering, 11(1s), 14-22.

[10] Kaboré, E. G., Macdonald, C., Kaboré, A., Didier, R., Arveux, P., Meda, N., ... & Guenancia, C. (2023). Risk Prediction Models for Cardiotoxicity of Chemotherapy among Patients with Breast Cancer: A Systematic Review. JAMA Network Open, 6(2), e230569-e230569.

[11] Liang, M. Z., Tang, Y., Knobf, M. T., Molassiotis, A., Chen, P., Hu, G. Y., ... & Ye, Z. J. (2023). Resilience index improves prediction of 1-year decreased quality of life in breast cancer. Journal of Cancer Survivorship, 17(3), 759-768.

[12] Bouz Mkabaah, L., Davey, M. G., Lennon, J. C., Bouz, G., Miller, N., & Kerin, M. J. (2023). Assessing the Role of MicroRNAs in Predicting Breast Cancer Recurrence—A Systematic Review. International Journal of Molecular Sciences, 24(8), 7115.

[13] Nara, M., Fujioka, T., Mori, M., Aruga, T., & Tateishi, U. (2023). Prediction of breast cancer risk by automated volumetric breast density measurement. Japanese Journal of Radiology, 41(1), 54-62.

[14] Manikandan, P., Durga, U., & Ponnuraja, C. (2023). An integrative machine learning framework for classifying SEER breast cancer. Scientific Reports, 13(1), 5362.

[15] Michel, A., Ro, V., McGuinness, J. E., Mutasa, S., Terry, M. B., Tehranifar, P., ... & Crew, K. D. (2023). Breast cancer risk prediction combining a convolutional neural network-based mammographic evaluation with clinical factors. Breast Cancer Research and Treatment, 1-9.

[16] Sun, J., Chu, F., Pan, J., Zhang, Y., Yao, L., Chen, J., ... & Xie, Y. (2023). BRCA-CRisk: A contralateral breast cancer risk prediction model for BRCA carriers. Journal of Clinical Oncology, 41(5), 991-999.

[17] Li, N., Song, C., Huang, X., Zhang, H., Su, J., Yang, L., ... & Cui, G. (2023). Optimized radiomics nomogram based on automated breast ultrasound system: a potential tool for preoperative prediction of metastatic lymph node burden in breast cancer. Breast Cancer: Targets and Therapy, 121-132.

[18] Zheng, D., Zhou, J., Qian, L., Liu, X., Chang, C., Tang, S., ... & Zhou, S. (2023). Biomimetic nanoparticles drive the mechanism understanding of shear-wave elasticity stiffness in triple negative breast cancers to predict clinical treatment. Bioactive Materials, 22, 567-587.

[19] Jiang, T., Wang, Y., Chen, X., Xia, W., Xue, S., Gu, L., ... & Lin, H. (2023). Neutrophil extracellular traps (NETs)-related lncRNAs signature for predicting prognosis and the immune microenvironment in breast cancer. Frontiers in Cell and Developmental Biology, 11, 1117637.

[20] Li, M. M., Wu, P. P., Qiang, W. M., Li, J. Q., Zhu, M. Y., Yang, X. L., & Wang, Y. (2023). Development and validation of a risk prediction model for breast cancer-related lymphedema in postoperative patients with breast cancer. European Journal of Oncology Nursing, 63, 102258.

[21] Vachon, C. M., Scott, C. G., Norman, A. D., Khanani, S. A., Jensen, M. R., Hruska, C. B., ... & Kerlikowske, K. (2023). Impact of Artificial Intelligence System and Volumetric Density on Risk Prediction of Interval, Screen-Detected, and Advanced Breast Cancer. Journal of Clinical Oncology, 41(17), 3172-3183.