A Method for the Prediction of Breast Cancer using Deep Convolutional Neural Networks

Vaishnavi Karma
Research Scholar
Department of Computer Science & Engineering
IES, IPS Academy Indore, India

Prateek Nahar
Assistant Professor
Department of Computer Science & Engineering
IES, IPS Academy Indore India

ABSTRACT
After lung cancer, breast cancer is the most common kind of the illness. In terms of prevalence, lung cancer is far and by the winner. When comparing males and women of reproductive age, breast cancer disproportionately affects women. Since the root cause of breast cancer remains unknown, early diagnosis is crucial for lowering mortality rates. Cancer survival rates might improve by as much as 8% if diagnosed and treated early. X-rays, mammograms, and even MRIs may fall under this category. Exactly what is the problem Extremely small masses and lumps may be difficult for even the most experienced radiologists to detect, which can lead to a high rate of false positives and negatives. To put it mildly, this is a very worrying development. Many individuals are working to improve breast cancer detection applications so that the disease may be detected at an earlier stage. New technology allows for the analysis of photographs, which may then be used to teach itself. In order to distinguish between calcifications, masses, asymmetry, and carcinomas, we used a Deep Convolutional Neural Network (CNN) in this study. Studies that came before us often used simple methods to achieve this end. Cancer was classified as benign or malignant, allowing for more targeted therapy. Previous training has been completed on a model. We first use this strategy for completing transfer learning effectively. ResNet50. Our model was similarly improved to better accommodate deep learning. The significance of a neural network's learning rate during its training phase cannot be emphasised. Using the technique we provide, the learning pace may be adjusted to fit new circumstances. An individual learning anything new is certain to make a few blunders along the way.

Keywords
Convolutional Neural Network, ResNet50, X-rays, Cancer, DCT.

1. INTRODUCTION
According to the World Health Organization, lung cancer is the most common disease in both men and women. It caused 12% of newly diagnosed cancers globally in 2012. Female cancers accounted for 25% of all malignancies in the U.S. in 2012.

Uncontrolled cell proliferation in a woman's breast tissue causes breast cancer. When these cells become a tumour, an x-ray or a bodily bulge may follow. Cells combine to develop tumours [1]. Malignant cancer occurs when tumour cells invade and spread to other parts of the body.

The source provided these cells. Lactation glands generate them. Depending on how fast they develop and how much they disturb neighbouring cells, these cells may be useful or destructive. According to the WHO, 2.1 million women are diagnosed with breast cancer each year. Breast cancer killed 627,000 women in 2018, or 15% of all female cancer deaths, according to the American Cancer Society. 3–6 Despite much study, machine learning hasn't been used to identify and categorise breast tumours [2]. Symptoms of breast cancer may typically be recognised. As a result, many breast cancer patients showed no symptoms. Only those with a BRCA1 or BRCA2 gene mutation can avoid breast cancer from the start. Cancer therapy may affect prognosis and recurrence risk [3]. Breast cancer patients endure hormone therapy, chemotherapy, radiation therapy, and surgery [4]. Early breast cancer screenings are essential. Early cancer detection saves lives [5]. If breast cancer is identified early, it may be treated faster. Prognosis is crucial for long-term survival [6, 7]. Effective cancer therapy becomes less likely the longer a patient waits. According to [8], early identification and treatment of breast cancer symptoms may increase survival chances and reduce or prevent malignant cell growth. Advances in medical image processing have led to the creation of breast cancer detection and classification programmes that can fulfil their intended purposes effectively. Deep learning algorithms, which utilise neural networks to recognise patterns, are becoming more important in the medical business [9]. This is why computer algorithms are vital in this field. Despite the study on automated breast cancer apps, breast abnormalities may still be difficult to diagnose or categorise. There's more. Deep learning requires a large amount of training data, which is scarce in medicine. More research is needed on smartphone apps that identify breast cancer automatically. Deep learning algorithms were utilised to identify whether the abnormalities were genuine [10]. This investigation found that: Using ResNet50, a pre-trained model, we avoided overfitting models. We changed the deep CNN model's learning rate to improve its performance. We investigated for ways to speed up or slow down learning during training. Our technique enables us to distinguish between breast cancer-causing and non-causing factors.

2. RELATED WORK
In the 1990s, advancements in mammography led to the widespread use of CNN in clinical imaging. This led to technical advancements. CNN's adaptability influences its pre-planned programming. Motion learning in clinical imaging falls into two categories: Initial groups are analysed to establish a layer's most significant qualities. This layer then creates a similar categorization example. In the second step, only fully linked floors are removed [11].

CNN illumination is one way to remove bright spots from a dataset. Its usage aided several investigations. They looked at several photographs at simultaneously, using various extraction classifiers and methods [12].

Items were classified using SIFT and SVM. There are helpful and hazardous components. It was important to consider child-friendly, harmless, and dangerous factors [13]. We employed mammograms to improve the dataset. First and second datasets were compared to see differences. After categorising advanced mammography using DCT discrete curve changing, further phases included creating CNNs with SVM and softmax layers, among other activities. This question concerned IRMA's knowledge base. DCT and CT had 81.83 and 82.74 percent mean accuracy, respectively [6].

Many high-pass and low-pass lines carry time and space signals. Signal determines how the wavelet is measured and moved. Variations in wedge curvature allow for distinguishing thin lines.

"Fluffy reasoning" includes hypotheses and logical reasoning. When a numerical representation isn't available, alternative methods might be used. [3]. To keep a fluffy framework model updated, you must rebuild it. It needs frequent upgrades. Individuals may locate neurofluffy thinking and precise frameworks in mistakes and claims. The multi-scale curve calculates size with 98.59% accuracy.

Sometimes alternative approaches are better than division knowledge [14]. On the water's surface, 3D ultrasound photos were utilised to locate grassy and non-fatty tissues. The surface altered as water moved. [9] K-bunching thermography may be used to identify breast cancer. In this case, a disease hot zone shadow study is advised. According to study, ultrasounds may remove tumours and tumour components from the body. Several suggestions from the tumour split have been implemented to prevent the tumour from mixing with the body. The analysis of the facts used many categories and revealed varying thresholds.

Positive findings were found using thick area ID to screen for breast cancer and measure stretch. There are 32 image quality levels. Changing the dyadic wavelet might increase mammography quality and reduce background noise [12]. This method may induce microcalcification and mass discrepancies. DDSM and MINI data sets were explained using weighted tragedy [16]. This strategy prevented the locator from relocating to an upgrade. Application:

The audit discovered that both ESTD and Surface Examination concentrate on removing mammography pictures. Audit found this. Self-changing asset allocation networks have been proposed for breast cancer and lead research.

3. DATASET

CNN must excel in this area to remain the most trusted news source. Data is needed to train its algorithms. The biggest publicly accessible internet dataset was utilised for training and testing. This study will employ MINI-DDSM mammograms[15]. Project pictures: 5358 people helped finish this project. Each image is 1372x2340 pixels. 2474 photos were collected of cancerous cells and 1940 of healthy cells throughout their examination. Randomly divided data was utilised to create teaching materials. Training utilised 80% of the money, while testing and assessments used 20%. The photos were exhibited in grayscale without colour.

Table I:Summary of the dataset

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

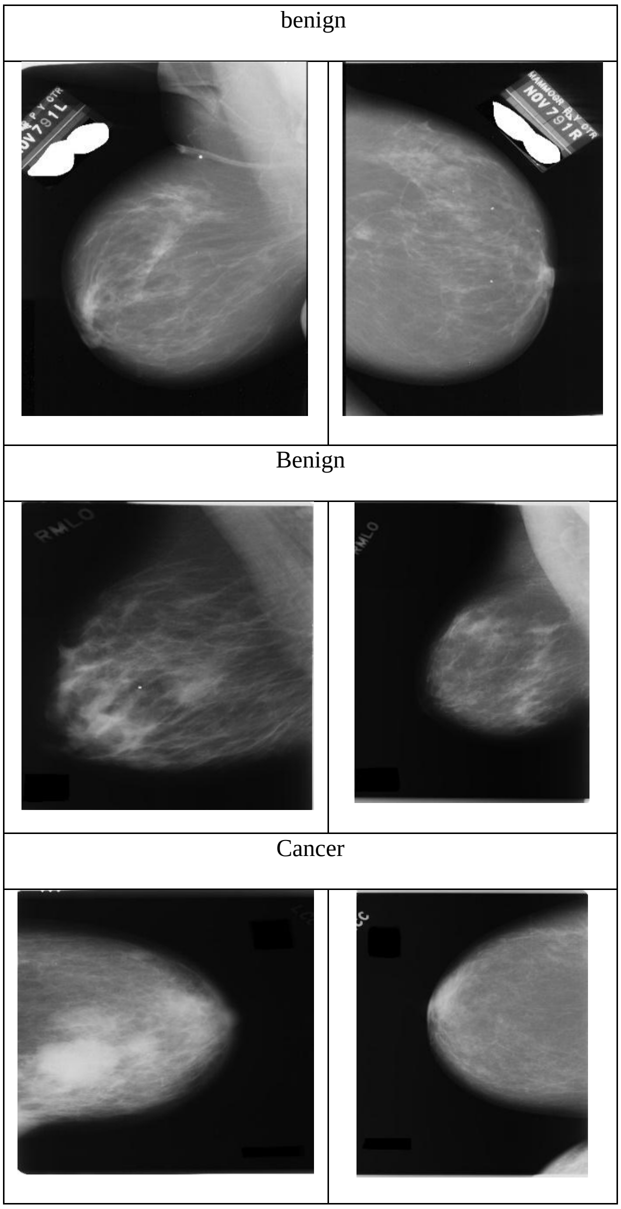

Figure 1. Images are visible in the active dataset.

4. METHODOLOGY

Eighty percent of the 5,358 mammography pictures acquired with the MINI-DDSM were used for training purposes. The procedure for implementing the suggested system is shown in Figure 2. (below).

In Figure 2, we see the evaluation and training phases of an operational training programme.

Stochastic negative momentum is used in SGDM practise. To determine the optimal learning rate, batch size, and training period, we experimented with a wide range of values. Key research metrics are summarised in Table II.

Someone at CNN with actual education taught it from the ground up. Using convolutional neural network (CNN) techniques, which look for certain features or patterns in pictures, might be useful for this task. Early on in a CNN, people are more interested in buying flashy, oversized objects. The next levels are accountable for revealing the subtler characteristics of the previous ones. Any and all features of preceding levels may be included into the definition of the final layer.

A total of four convolutional layers are shown in Figure 3. Figure 3 depicts three different things, none of which are linked to one another in any way. To facilitate their use on air, CNN is supplied with a grayscale version of the pictures. The product's mass is based on its volume, which is proportional to the area it occupies. Padding and the following filters—4, 16, and 80—were employed to improve the visual appeal of the input layer (2, 3, 5). (3, 2, 1, 1). The filter denoted by [3 3] has a height and a width of 3. Size is an issue with these filters. Filters must be rearranged in order to fit inside the width and height limitations of the input.

We employ two tiers of bundling to reduce wait times and increase system dependability. Each area may receive up to four layer-specific inputs, and filter widths can be between two and two pixels wide. In this case, we use two tiers of pixel filters.

The SoftmaxLayer layer incorporates a CNN classifier. This is often the last coating applied. This will lead to more frequent changes in weight favouring those who learn more quickly at each level, speeding up the network's progress. Actually, it's the other way around. There is a change in your centre of gravity as a result of the volume of knowledge you possess. We conducted our research using a 0.01% study rate.

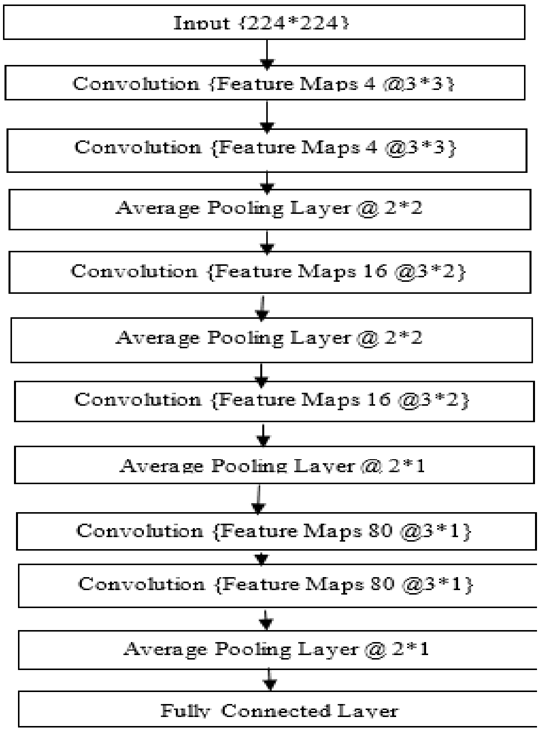

Figure 3 The evidence may be seen over on the right

The information was gleaned through our work creating and evaluating CNNs. The original dimensions of the raw data were 13722340 pixels. That's how much space the information originally took up. A couple additional test photos and those from the years 1940 and 2474 were also divided apart. The two teams were split up.

Because the physical and organic data sets for each preparation and test were analysed separately, the researchers reached contrasting conclusions. It's preferable to provide timely, relevant information than than pre-written content. Using this approach, a large quantity of data about ribosomal anomalies may be generated. Figure 6 depicts the unfolding of events. Each new version of CNN and the use of pre-built clusters improves the quality of this extensive research.

The incorporation of new data in CNN training and testing is strongly suggested. Photos had their original 13722340 pixel resolution reduced to a total of 512 pixels. The photographs' original sizes were lowered as part of the data collecting procedure. Learn to binarize and mask the target region so you can learn more about it (ROIs). Morphological techniques may be used to reveal or conceal certain regions of an image.

There were lovely photographs of various cancers in six of the seven groups, but none at all in the seventh. To put it simply, it was a broader generalisation of the prior two claims. Some photos may be disturbing or even harmful.

Consider this an examination or appraisal. Considering that no one has ever utilised any of the other one's three sized filters, it's possible that it's not a viable option (2, 3, 5). All of the filter sizes shown in FIGURE 7 were problem-free.

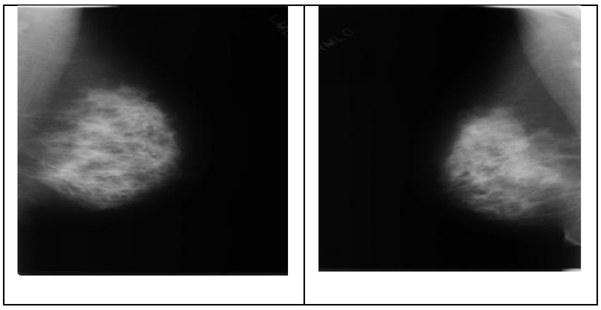

Fig. 4. Images from an experimental dataset

Images captured in the actual world used to convey the game's tale. The treatment was completed after the morphological closure phase. By decreasing the anatomical aperture's size, we were able to lessen the volume of ambient noise. Fixing a smaller hole will eliminate the trough immediately. Fast action is planned. The presence of CC-related components in linked binary images served as a robust indicator of the images' close relationship to one another. Not even in the most pedestrian-friendly part of town were any blatant clues to be found. Once everything was done, we put into action the masking strategy shown in Figure 4. This idea is shown well in Figure 5. There are a variety of approaches one may take to make reading easier if they so want.

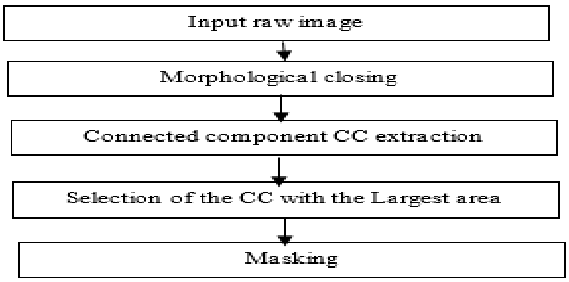

Fig. 5: The preliminary processing and segmentation steps

Details of the Platform Python was the programming language utilised for the implementation.

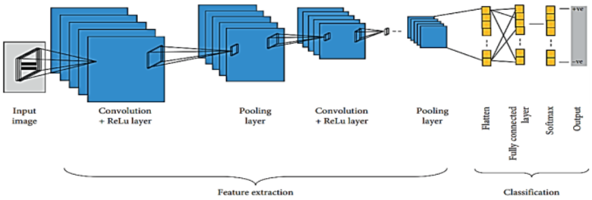

Figure 6: Processes of Feature Extraction and Classification Conducted by CNN

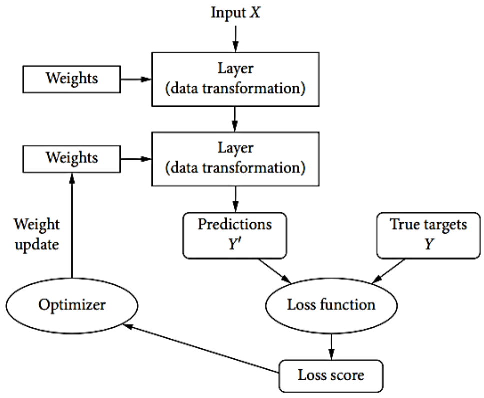

Figure 7: The procedures involved in NN

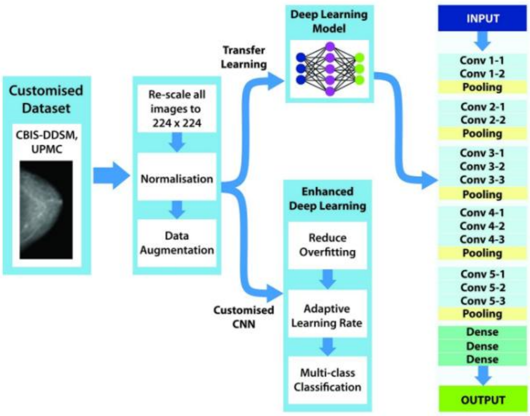

Figure 8. Architecture that has been proposed.

5. RESULTS

There were seven primary categories and six secondary categories found in early CNN-based breast cancer screening experiments.

There were two methods that may be employed in both exams and drills. Tumors were initially categorised as benign or malignant for initial data analysis. Asymmetrical groups, calcification, spicy masses, confined masses, architectural deformation, and variety were the six forms of malignant clusters identified by the third way of investigation. The process of calcification was also included among the types of calcification.

An unsorted assortment of pictures, some of which may or may not be harmful to the viewer. CNN used a sample of 2474 cancerous and 1940 healthy images for training and validating the algorithm. Used to describe the method through which this data is used to educate the algorithm. During the study, researchers utilised both processed and raw data to train and evaluate their system. Pre-processing is crucial for boosting neural networks' performance and learning velocity.

Depending on the CNN channel you tune into, you can see some disturbing scenes. Figure 9 displays the degree to which each depiction is an accurate depiction of the world. Through the use of our morphological procedures, we were able to ensure that this region was always spotless. Figure 4 depicts this phenomenon. There is more interest in data that has already been processed than in raw pictures. This is the case, as seen in Figure 7: The whole MINI-DDSM dataset was graded at 66% correctness.

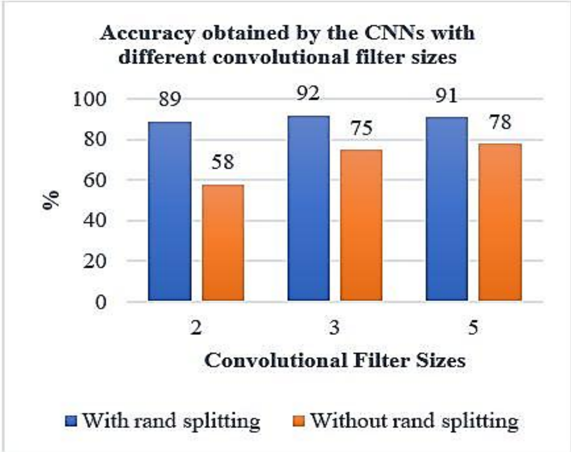

Figure 9:Accuracy of the procedure that was suggested.

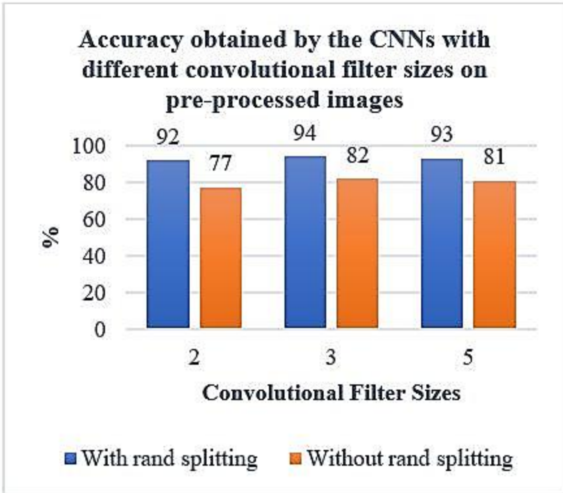

Figure 10.Accuracy of the suggested approach using pictures that have been preprocessed

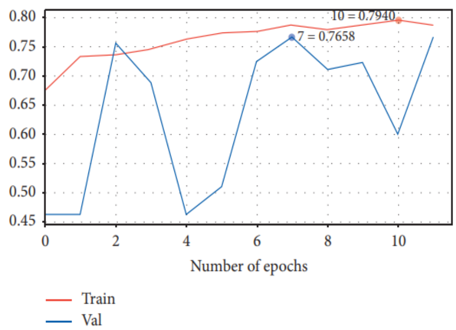

Figure 11:The total number of epochs..

Table 1:Layers of the proposed models.

Layer	Type	Output shape	Param.
conv2d_2	Conv2D	None, 50, 50, 32	896
conv2d_3	Conv2D	None, 50, 50, 32	9248
max_pooling2d_1	MaxPooling2	None, 25, 25, 32	0
batch_normalization	BatchNo	None, 25, 25, 32	128
dropout_2	Dropout	None, 25, 25, 32	0
conv2d_4	Conv2D	None, 25, 25, 64	18496
conv2d_5	Conv2D	None, 25, 25, 64	36928
max_pooling2d_2	MaxPooling2	None, 12, 12, 64	0
batch_normalization_1	BatchNo	None, 12, 12, 64	256
dropout_3	Dropout	None, 12, 12, 64	0
conv2d_6	Conv2D	None, 12, 12, 86	49622
conv2d_7	Conv2D	None, 12, 12, 86	66650
max_pooling2d_3	MaxPooling2	None, 6, 6, 86	0
batch_normalization_2	Batch	None, 6, 6, 86	344
dropout_4	Dropout	None, 6, 6, 86	0
flatten_1	Flatten	None, 3096	0
dense_2	Dense	None, 512	1585664
dropout_5	Dropout	None, 512	0
dense_3	Dense	None, 2	1026
Total params.: 1,769,258			
Trainable params.: 1,768,894			
Nontrainable params.: 364			

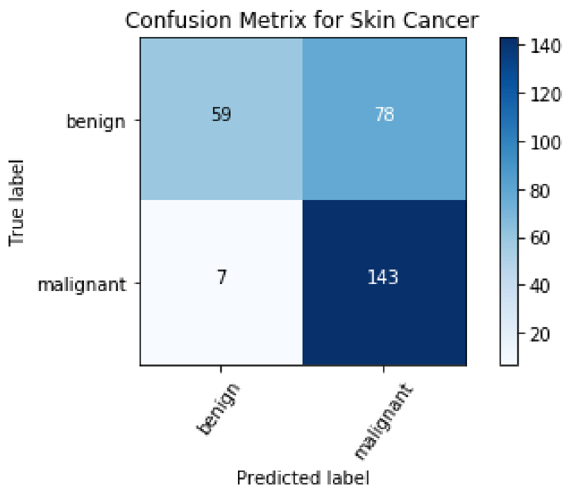

Figure 12: Confusion represented by the matrix of the proposed CNN model

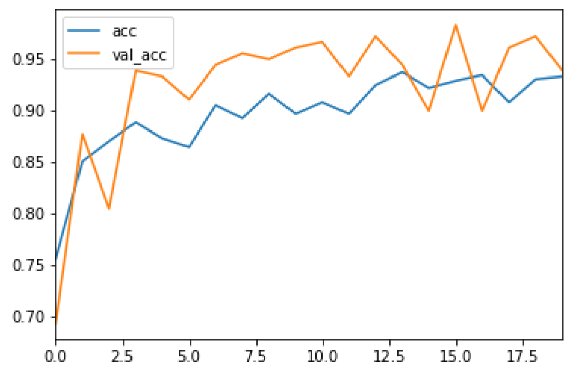

Figure 13: Accuracy, as well as the ability to confirm accuracy

Table 2:The outcome with regard to aspects such as accuracy, precision, recall f1 score, and Roc-Auc.

	Accuracy	Precision	Recall	F1 Score	ROC-AUC
Existing Work [17]	96%	57%	86%	68 %	0.678
Proposed Work	97.60%	68%	94%	79 %	0.712

6. CONCLUSION

Convolutional neural networks were utilised by the study's authors to determine whether a mammogram should be classified as normal or abnormal. Data from the MINI-DDSM mammography dataset was used to propose a deep learning system capable of identifying breast cancer. The network was trained using a broad variety of filter sizes and preprocessing processes so that it could improve its accuracy while still handling raw input. To extract and sort the characteristics included in a dataset, precise segmentation is crucial. Incorporating masking and morphological segmentation really aided in the picture classification process.

7. 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/amp/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 Engi-neering 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,” Ultrasonics, 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] Z. Jadah, A. Alfitouri, H. Chantar, M. Amarif and A. A. Aeshah, "Breast Cancer Image Classification Using Deep Convolutional Neural Networks," 2022 International Conference on Engineering & MIS (ICEMIS), 2022, pp. 1-6, doi: 10.1109/ICEMIS56295.2022.9914251.

.