JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Breast cancer Prediction using Deep Learning Technique

Pooja malviya
Research Scholar
Computer Science and Engineering
pmalviya3996@gmail.com

Vijay Birchha
Assistant Professor
Computer Science and Engineering
vijaybirchha@gmail.com

Abstract: Breast cancer is the second most prevalent kind of cancer globally, behind lung cancer. Breast cancer is the most common disease affecting women worldwide. There is always the potential for progress and development in the field of medical imaging. Cancer mortality is expected to reduce if the disease is detected and treated early and adequately. Doctors may be able to improve their diagnostic accuracy with the assistance of machine education. Deep learning, also known as neural networking, is one technique for distinguishing between benign and cancerous breasts. As a result, CNN may be used. The Mammograms-MINIDDSM data-collection used in this study consisted of 5358 mammograms, with about 2180 benign and 2998 malignant breast images being obtained. Deep learning in the diagnosis of mammogram cancer has been demonstrated in mammograms in promising experimental findings that encourage the use of in-depth learning techniques based on current characteristics and classification of a range of applications, particularly in the detection of breast cancer demonstrated in mammograms. Though work has to be done, improvements in CNN design and the use of pretrained networks allow for more development, which should result in improved accuracy over the next several years. To extract and categorize features efficiently, proper segmentation is required.

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

I. INTRODUCTION

Oncology is a term that refers to an illness that is characterized by abnormal cell development and fast spread. Most illness cells gather to create a hump or lump known as a tumor, which is identified by the location of its first appearance on the body. Because early-stage breast cancer is seldom symptomatic, screening is essential for getting an accurate diagnosis as soon as feasible. The majority of lumps were found to be non-destructive and of moderate intensity. While most breast cancers develop to the invasive stage, and while breast cancer is often referred to as a single sickness, the disease's course may be classified into up to 21 histological subtypes. According to the American Cancer Society, about 252,710 new cases of invasive cancer were evaluated in women in 2017, whereas 2470 new cases were reviewed in men[1]. Breast cancer incidence and mortality rates rise with age.

Between 2010 and 2014, the average age of discovery of breast cancer was 62. According to the World Health Organization, Pakistan has the highest rate of breast cancer in Asia.

Each year, about 90,000 cases are recorded, with a death rate of 40,000[2]. At this time, it is quite probable that breast cancer will be completely healed of all diseases. Because cancer does not cause pain in its early stages, it is not recognized until the symptoms become severe enough to be hazardous. According to data, Pakistani cancer patients are on average 40 years old. Patient endurance is defined as the percentage of patients who survive a certain period after discovering and accepting a normal life shortly. The rate at which stamina is depleted varies according to the stage of tumor diagnosis.

Recent data[1,] indicate that women with breast cancer have a high capacity for endurance: After five years, the rate had risen to 91 percent of the population. Ten years later, the figure had risen to 86 percent. After 15 years, the rate has decreased to about 80% of the population.

Breast cancer is more likely to attack women and men who have had the illness in the past via relatives, parents, children, or extended families. Mammography is a low-dose x-ray method that enables physicians to see the breasts and their interior architecture. Many sign management techniques may be used to identify breast cancer, including ultrasound imaging, microwave imaging, wavelet change (a timeconsuming portrait of the wavelet), curvelet change (derived from a wavelet change), restricted scale alterations, and images at multiple scales[3].

Various methods, such as fluffy thinking and neuroflaccid frames[3,], weed out potentially dangerous and child-centered courses from the curriculum. As part of a deep learning approach to information architecture, text, images, and voice activities are organized into a computer model. Models are built at different levels of abstraction using a range of datasets and CNN architectures. Advanced learning is used in clinical imaging to detect cancer cells through picture analysis. Without previous preparation, establishing a deep convergence network is challenging due to the massive amount of data collected and processed. For instance, one approach is to calibrate an organization that is well-equipped and prepared.

Deep learning is currently being investigated for various therapeutic applications, including bioinformatics, early detection of Alzheimer's disease, and subatomic imaging. Another area is sub-atomic imaging, which combines patient- and disease-specific sub-atomic data with traditional reading physical images [4, which is a mix of the two]. Other approaches, such as assessing several courses in a single sitting, have been suggested. These methods would need less prior knowledge and fewer wellknown preparation examinations[5].

II. RELATED WORK

Since the 1990s, when improved mammography revealed calcifications, CNN has been utilized in clinical imaging. 'Adaptability' is a critical component of CNN in the pre-prepared CNN. According to the present research, motion learning may be divided into two kinds in clinical imaging. To begin, preparatory organizations are utilized to extract highlights from the layer of a specific organization and then to build another categorization example. Second, the rest of the pre-prepared structure is utilized, except for fully connected levels replaced by another strategic layer.

Numerous techniques such as SVM, wavelet transformations, cosine transformations, and CNN illumination are proposed for removing highlights from this dataset[6]. Multiple experiments were run on this dataset, either by comparing numerous pictures or using various extraction classifiers.

Along with the SIFT, the SVM classification was employed, with components classified into two categories (favorable and threatening) and three categories (child-hearted, benevolent, and damaging)[6]. Mammogram patches were used to enhance the dataset, which was then contrasted. One approach was the 2D-DWT, which classified advanced mammography into four subgroups, using DCT (Discrete curve altering) as the next method for constructing CNNs utilizing SVM and softmax layers. This research used the IRMA knowledge base and found that the mean accuracy for DCT was 81.83 percent and for CT was 83.74 percent. [6].

The time/space signal is transmitted via several highpass and low-pass channels. The resulting wavelet is measured and shifted through one or more signal sections with low or high recurrence. The change in curve recognizes thin lines and confirms the multidimensional elements on wedges[3].

Fluffy reasoning exemplifies many aspects of human brain processes, such as speculation and logical thinking; techniques also assist when a proper numerical representation is unavailable[3]. On the other hand, creating a fluffy framework model is very complex since it needs modification and reconstruction[3]. A mix of neurofluffy reasoning and crisp frames that drive intellectual investigation with flawed sets and asserted principles is a kind of neurofluffy reasoning[3]. With a 98.59 percent accuracy rate[3], the multi-scale curve provides unmatched mass identification findings.

Other methods like C-mean bunching have been employed in conjunction with heritage calculation and have had better results than the division's impact area extraction and placement competency. [7] The water shift was utilized to organize the surface area, and 3D bosom ultrasound images were used to determine and characterize grassy and non-fatty tissues[8]. Bosomthermography may also be utilized in advance to detect breast cancer via K-bunching, where the shadow inspection is recommended for the hot region isolated (disease area)[9]. Tumor regions based on the size of the ribosomal ultrasonography were removed in a later study[10]. First, to separate the tumor from the surrounding tissue, many provisions have been collected from the tumor split, flexible thresholds have been obtained, and repeated classifications have been used to evaluate the facts[10].

An alternative, thick area ID method was conducted on a collection of key recurrence models[11] to investigate stretch and screen differentiated breast cancer. Thirty-two specific picture characteristics have been evaluated. Another technique used to enhance and denoise mammograms are the dyadic wavelet change[12]. This approach showed that tiny characteristics such as microcalcification and modest differential highlights such as masses were produced feasible. Weighted misfortune has been used to characterize and limit the DDSM and MINI DDSM data sets to prevent the locator from sliding towards the ideal district[13].

An audit part suggested methods such as the Calculation of Competent Surface Identification (ESTD) and Surface Examination to emphasize the extraction of mammographic imagery[14]. For example, self-versatile asset allocation networks for bosom malignant development, lead part inspection, and categorization was proposed[14].

III. DATASET

To achieve high accuracy, CNN requires a large amount of training data. As huge data sets were scarce, training and testing were placed using the biggest publicly available internet dataset. The data set used in this research were mammograms of MINI-DDSM[15]. To this end, a total of 5358 images were used. The images are 1372 by 2340 pixels. There were about 2474 pictures of the malignant class and 1940 pictures of the benign class. The course was conducted by randomly dividing the data set into 80% for training and 20% for CNN testing. Images were converted to grayscale before.

Table I: Specifications for the MINI-DDSM dataset, as well as its division into train and test sets in the proposed system

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

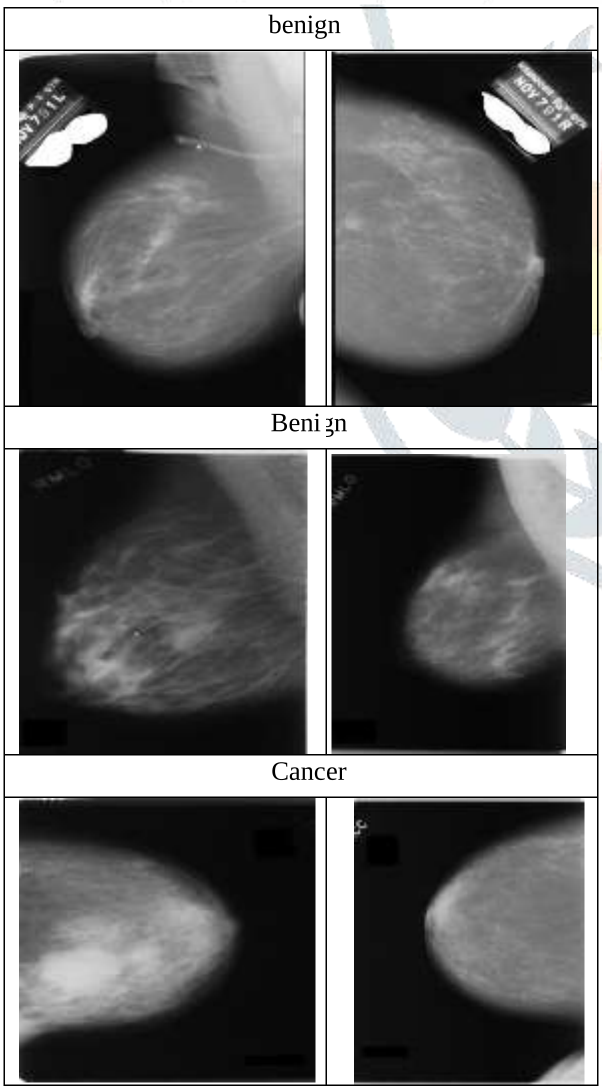


Figure 1 shows a selection of pictures from the MINI-DDSM mammography dataset.

IV. METHODOLOGY

Training was performed on 80 percent of the 5,358 images of the MINI-DDSM mammography. Figure 2 illustrates the method for the proposed system.

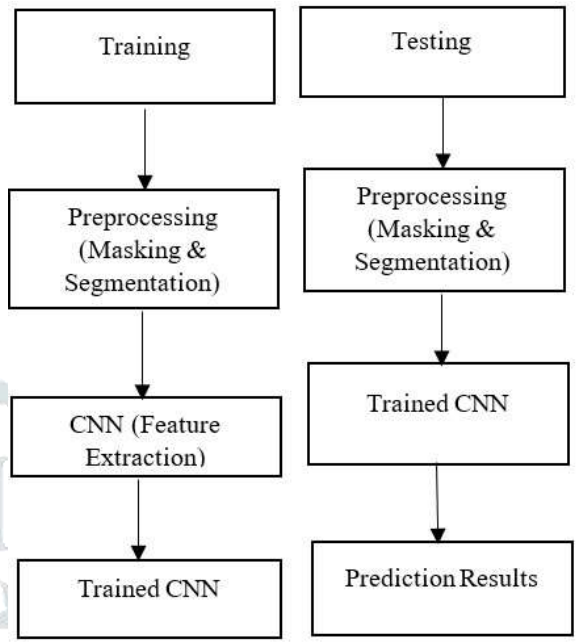


Figure 2 depicts a conceptual level block diagram of the proposed system's training and testing procedures at the highest degree of abstraction.

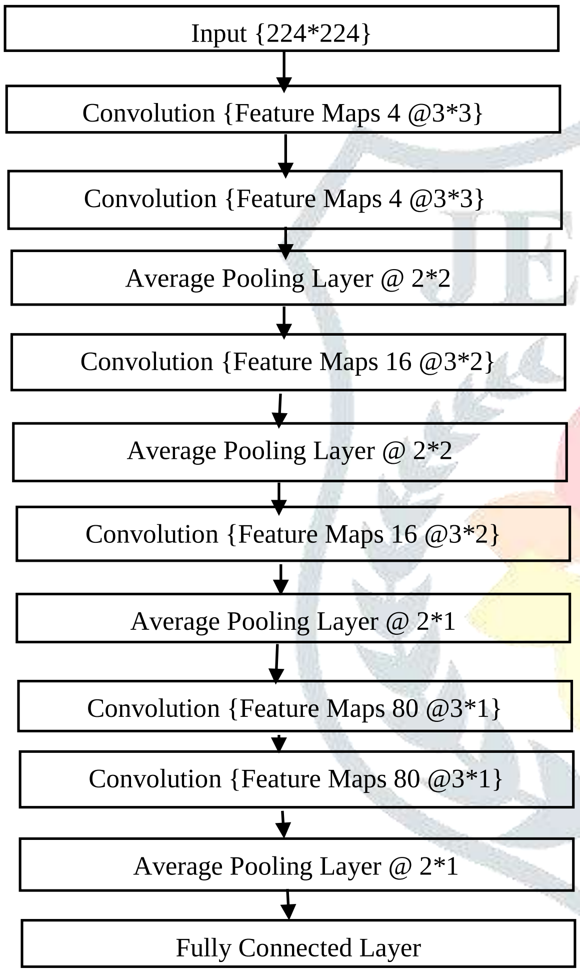
To train for SGDM, stochastic downward momentum is utilized. The best results were obtained by finetuning parameters such as the fundamental learning rate, mini-batch size, and maximum period. Table II is a list of some of the parameters that were utilized.

Our application in this article has been taught from the ground up using a CNN trained from the ground up. It is possible to identify certain patterns or characteristics in a picture using the network layers of a CNN algorithm. The first layers of a CNN recognize characteristics that are big and readily recognized. The next layers are responsible for identifying smaller, more abstract characteristics in the previous levels. The final layer may be extremely thoroughly classified by including all of the characteristics found in the layers that came before it.

According to Figure 3, DCNN comprises seven weighted layers; the first four convolutional and the second three are completely linked, as seen in Figure 3. Images in grayscale are submitted to CNN for broadcast. Each neuron generates a weighted point product whose volume is proportional to the input volume of the immediate area. We utilized 4, 16, and 80 filters (2, 3, 5) and padding (3, 2, 1, 1) at the boundaries of the input layer. The filter size must be specified for filters with a height of 3 and a width of 3, as seen in [3 3] below. It is necessary to shift each filter across the input's width and height.

Two bundling layers are employed, each of which results in downsampling to reduce computation and increase robustness. Layers with filter sizes ranging from 2 to 2 pixels are combined, resulting in a maximum of 4 inputs per area from layers with filter sizes ranging from 2 to 2 pixels.

SoftmaxLayer is a CNN classifier layer that is often used in the final layer. For example, a greater learning rate will result in bigger changes in weight at each step, and the network will learn more quickly as a consequence, and the opposite will be true. The learning rate has an impact on the weight changes at each stage. We utilized a 0.01 study rate for our research.



On the right, you can see the CNN architecture of the proposed system in Figure 3.

Utilization of exclusive CNN development and testing data The original raw data was 1372 by 2340 pixels in size. We started by categorizing the dataset into two categories: those that were favored and those that were threatening, which included 1940 and 2474 preparation pictures as well as additional test images.

Channels of varying widths were used in this project (2, 3, 5). Every preparation and test data set has been split by the physical and organic ratio of 80:20, and different findings have been obtained due to this division. Material that is provided sensibly produces greater results than well-prepared information. The technique that has been proposed is successful and consolidates a large amount of data for the diagnosis of ribosomal disorders. Exceptional outcomes were achieved, as shown in Fig. 6. This is a continuing investigation, and more progress is being made via improved CNN engineering and the use of preprepared organizations, both of which will undoubtedly result in improved accuracy.

Preprocessed CNN training and testing data should be used: The data set was created in advance, and the resolution of the images was reduced from 1372 by 2340 to 512 by 512. Binarization and hiding are two moral methods used to extract the region of interest (ROIs). Morphological techniques are utilized to display and extract image component regions in both video and still images.

The data was then split into seven subclasses, six of which included beautiful pictures of different kinds of cancer. The other two subclasses were just descriptive. Malignancies have been referred to as bending, lopsided, calculating, spicy weight, skyline masses, and a variety of other non-beneficial or even hazardous imagery in the past.

In both training and evaluation, the preprocessed data set was used with and without randomization, and the same three filter sizes were used in both (2, 3, 5). As shown in Fig. 7, the results of all filter sizes were satisfactory.

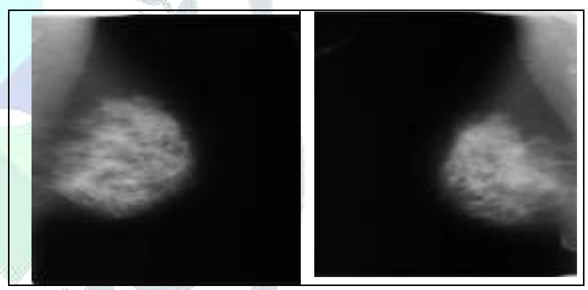
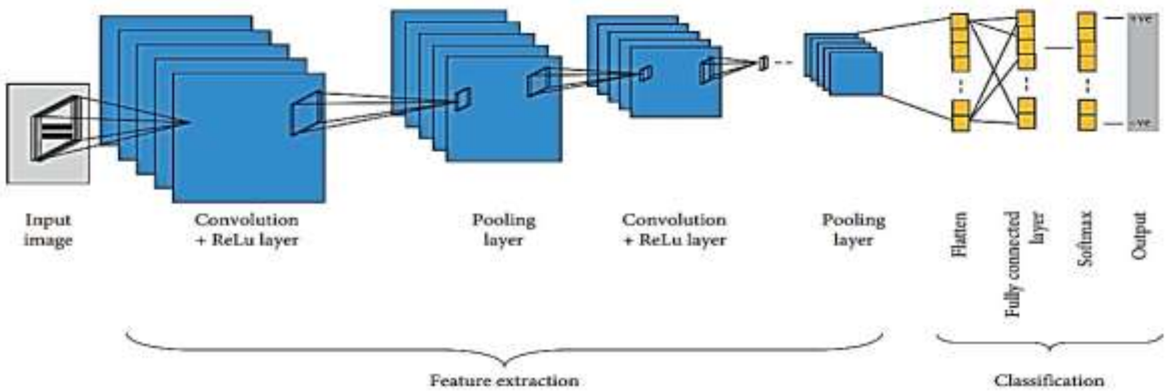


Fig. 4. Morphological operations used on data

Raw pictures were used as input, and morphological closure was used to complete the segmentation process. The morphological closure resulted in worsening of dilatation and noise reduction as a consequence of the closure. A tiny aperture enables the removal of small items, while a little trough is eliminated when the hole is closed. CC-related components were identified in connected areas of binary pictures, indicating that the images were linked. The region with the greatest number of connections was chosen to conceal all of the important areas retrieved. Following the completion of the conversion to zero, the masking procedure illustrated in Fig. 4 was followed. The segmentation stages of preprocessing are shown in Figure 5.



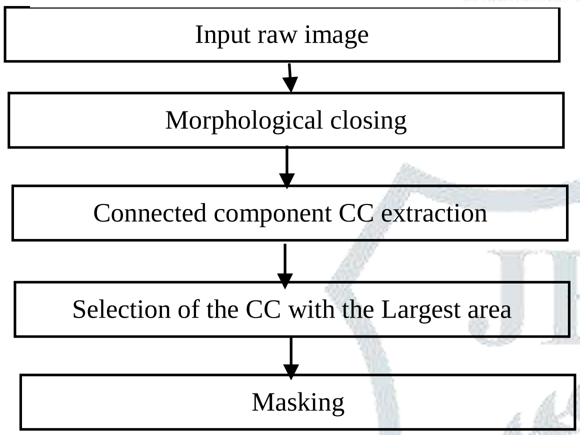


Fig. 5: Pre-processing segmentation stages.

Platform Specifications C. Python was utilized in the implementation. Google Colab for CPU and GPU development.

Figure 6: Typical CNN architecture for automatic detection of IDC breast cancer.

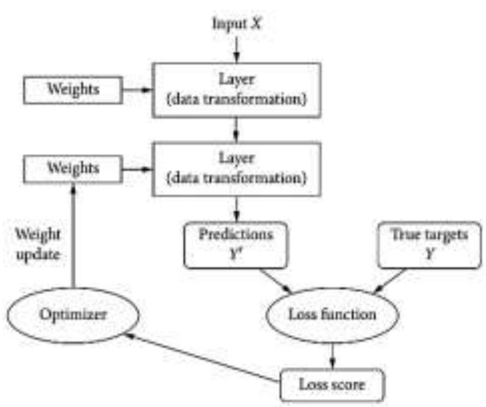


Figure 7: Detailed process of a neural network (NN)

V. RESULTS

The suggested CNN-based breast cancer screening technique has shown satisfactory results in preliminary testing. The dataset was split into seven categories, which were then further subdivided into six groups.

There were two different training and testing techniques used. Initially, the data was divided into two categories: benign and malignant tumors. The third approach was to split malignant groups into six breasts, which were classified as follows: asymmetry, calcification, spicy mass, restricted masses, architectural deformation, variety, and calcification.

Various photographs revealed confusing images, which may be both benign and cancerous. CNN trained and evaluated a data set consisting of 2474 malignant class pictures and 1940 benign class images, used to train and test the algorithm. Raw and preprocessed data were used in the training and testing phases of the study. Pre-processing is used to enhance the neural network's performance and speeds up the learning process in general.

As shown in Fig. 4, the exactness of the crude pictures produced by different channel measurements based on CNN is shown in Fig. 8. At the same time, in Fig. 9, the initial images have been pre-managed by morphological activities to minimize disruption in this area, as demonstrated in Fig. 4. Preprocessed data has overtaken unprocessed pictures in terms of popularity. There will be no further learning or improvement inaccuracy if the model limitations have been learned and fixed. Following Fig. 7[15], the MINI-DDSM dataset had a total accuracy of 65 percent, as shown.

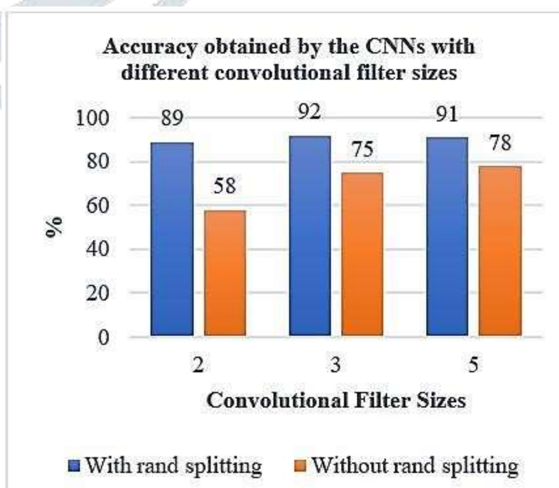


Figure 8 shows the accuracy achieved by CNNs with various convolutional filter sizes on raw pictures from the MINI-DDSM dataset.

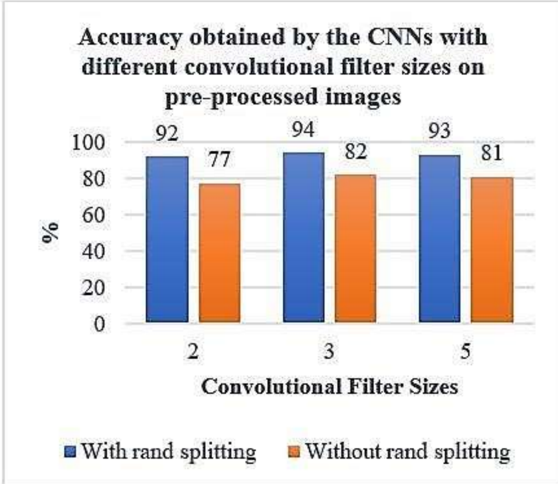


Figure 9 shows the accuracy of CNNs with various convolutional filter sizes on preprocessed pictures from the MINI-DDSM dataset.

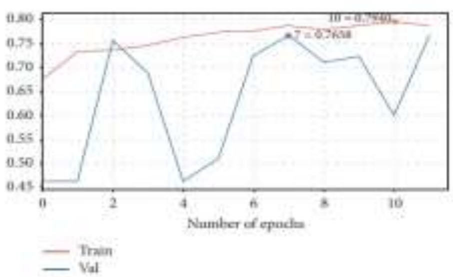


Figure 10: The loss learning curve for Proposed Model.

Table 1: The summary table of Proposed CNN Model.

Layer	Type	Output shape	Param.
conv2d_2	Conv2D	[illegible]	[illegible]
conv2d_3	Conv2D	[illegible]	[illegible]
max_pooling2d_1	MaxPooling2	[illegible]	[illegible]
batch_normalization	BatchNo	[illegible]	[illegible]
dropout_2	Dropout	[illegible]	[illegible]
conv2d_4	Conv2D	[illegible]	[illegible]
conv2d_5	Conv2D	[illegible]	[illegible]
max_pooling2d_2	MaxPooling2	[illegible]	[illegible]
batch_normalization_1	BatchNo	[illegible]	[illegible]
dropout_3	Dropout	[illegible]	[illegible]
conv2d_6	Conv2D	[illegible]	[illegible]
conv2d_7	Conv2D	[illegible]	[illegible]
max_pooling2d_3	MaxPooling2	[illegible]	[illegible]
batch_normalization_2	Batch	[illegible]	[illegible]
dropout_4	Dropout	[illegible]	[illegible]
flatten_1	Flatten	[illegible]	[illegible]
dense_2	Dense	[illegible]	[illegible]
dropout_5	Dropout	[illegible]	[illegible]
dense_3	Dense	[illegible]	[illegible]

Total params: [illegible]
Trainable params: [illegible]
Non-trainable params: [illegible]

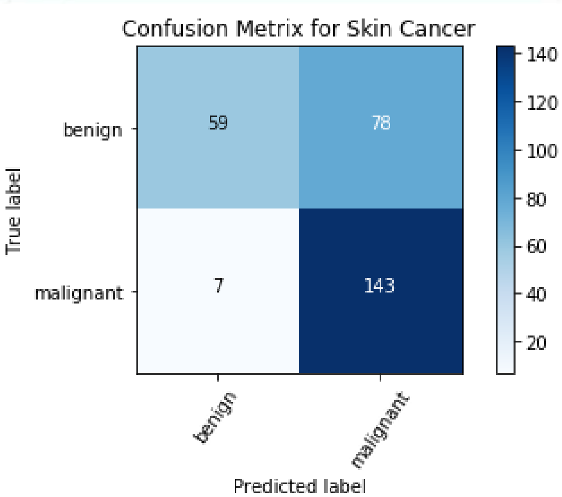


Figure 11: The confusion matrix of Proposed CNN Model

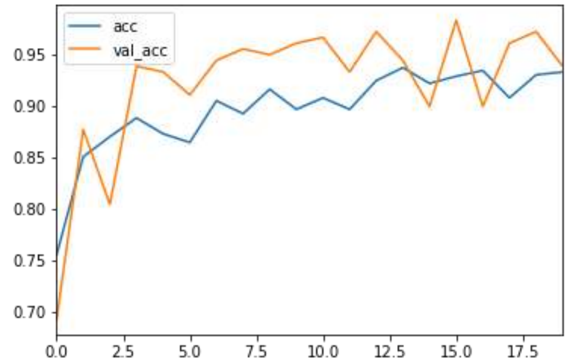


Figure 12: Shows validate accuracy and accuracy of proposed CNN model

Table 2: The metric results of Proposed CNN Model.

Accuracy	Precision	Recall	F1 Score	ROC-AUC
97..6%	68%	94%	79%	0.712

In table 2 represent the proposed result; our model gets 97.6 % accuracies, 68% of precision, 94% recall, 79% F1 score, and ROC-AUC is 0.712.

VI. CONCLUSION

Convolutional neural networks were used to differentiate benign from malignant mammograms in this research. This deep learning method may be used to identify breast cancer by gathering characteristics from segmented malignant groups in the MINI-DDSM mammography dataset and moving them to a benign class. Various filter sizes and preprocessing methods were employed to eliminate noise components that might impair the overall network's accuracy when dealing with the raw data. Additionally, it has been said that appropriate segmentation is necessary for successfully extracting and categorizing characteristics from a dataset. The use of masking and segmentation methods based on morphological principles significantly enhanced classification results.

Reference

[1] "Breast cancer statistics." [Online]. Available: http://www.wcrf.org/int/ cancer-facts-figures/dataspecific-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/ne ws/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] "Breast cancer facts & figures 2017-2018." [Online]. Available: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancerfacts-and-figures/ breast-cancer-facts-and-figures-2017-2018.pdf

[18] K. Vennila, K. Sivakami, and R. Padmapriya, "Detection of mass in digital mammograms," International Journal of Computer Applications, vol. 104, no. 5, 2014.

[19] K. Djaroudib, A. T. Ahmed, and A. Zidani, "Textural approach for mass malignantity segmentation in mammographic images," arXiv preprint arXiv:1412.1506, 2014.

[20] S. Deepa and V. S. Bharathi, "Efficient roi segmentation of digital mammogram images using otsu's n thresholding method," Indian Journal of Automation and Artificial Intelligence, vol. 1, no. 2, pp. 51–56, 2013.

[21] "Filters in the context of convolutional neural networks." [Online]. Available: https://www.quora.com/ What-is-afilter-in-the-contextof-Convolutional-Neural-Networks

[22] T. P. K. Derek C. Rose, Itamar Arel and V. C. Paquit, "Applying deeplayered clustering to mammography image analytics," 2010.

[23] B. R. Spandana Paramkusham*, Kunda.M.M. Rao**, "Early-stage detection of breast cancer using novel image processing techniques, matlab and labview implementation," 2013.

[24] T. I. A. L. C. Muzni Sahar, Hanung Adi Nugroho, "Automated detection of breast cancer lesions using adaptive thresholding and morphological operation," 2016.