IJEW

International Journal of Engineering Works
https://www.ijew.io/
ISSN-p: 2521-2419
ISSN-e: 2409-2770
Vol. 12, Issue 08, PP. 170-174, August 2025
https:// doi.org/10.5281/zenodo.16908308
Check for updates

An Intelligent Model for Detection of Breast Cancer based on Convolutional Neural Network

Hussain Ahmad[1], Muhammad Zeb[2], Khalil-ur-Rahman[3]
[1]Department of Electrical Engineering, University of Engineering and Technology Peshawar
[2,3]Department of Electrical and Computer Engineering, COMSATS University Islamabad, Wah campus
hussainahmad770@gmail.com[1], mzebkhan97@gmail.com[2], khaliltaunsvi@gmail.com[3]
Received: 20 February, Revised: June 04, Accepted: 13 August

***Abstract*—**Cancer is among the deadliest diseases afflicting humanity. At present, there exists no successful therapy. Breast cancer is among the most common kinds of cancer. In 2020, the National Breast Cancer Foundation projected that approximately 276,000 fresh patients of invasive breast cancer and 48,000 fresh patients of non-invasive breast cancer were diagnosed in the USA. The patients have a 99% survival rate, as 64% of these cases are detected in initial stage of the disease. Artificial intelligence (AI) has been utilized to detect deadly diseases, which has enhanced the patient's likelihood of survival by enabling early diagnosis and treatment. This research presented convolutional neural network (CNN) for the diagnosis of breast cancer disease automatically. The analysis has been carried out on a real-time invasive ductal carcinoma (IDC) dataset available at Kaggle. The dataset is preprocessed before being fed to CNN. The images is normalized to achieve a better accuracy. The developed model has an accuracy of 90% that is improved by 3% from the previous research paper. Different performances metrics are graphically represented in result section to analyze the model efficiency.

***Keywords*—** Breast Cancer, CNN, Deep learning, Supervised learning.

I. INTRODUCTION

Despite being more prevalent in women, breast cancer is capable of developing in males as well. According to the research findings [1] [2], breast cancer is recognized as the most lethal form of cancer affecting women worldwide. In 2020, the United States witnessed the identification of an estimated 276,000 newly diagnosed cases of invasive breast cancer and 48,000 newly diagnosed cases of non-invasive breast cancer, as reported by the National Breast Cancer Foundation [3]. Detection occurs at an early stage in 64% of these cases, resulting in a 99% survival rate for the patients. There are four fundamental subtypes of breast cancer: benign, in-situ carcinoma, invasive carcinoma, and normal [4]. A benign tumour has minimal impact on the structure of the breast. It is non-toxic and does not meet the criteria for a hazardous malignancy. In cases of in-situ carcinoma, the malignancy is localized exclusively within the lobular system of the mammary duct and does not spread to distant organs. Early detection can mean treatment of this variety, which is non-lethal. Invasive carcinoma, due to its potential for metastasis to other organs, is regarded as the most lethal variant of breast cancer.

Imaging techniques such as X-ray mammography, ultrasound, MRI, thermography, cytopathology, and histopathology may identify breast cancer. When looking for breast cancer, a pathology diagnosis is usually the way to go. To facilitate microscopic examination, the removed tissue is first photographed and then coloured in the lab. Hematoxylin and eosin (H&E) are common hemostaining agents for normal tissues. Breast cancer diagnosis may be accomplished by genetic testing as well as histological image analysis. Early detection and treatment of breast cancer benefit especially from microscopic images of breast tissue called histopathological imaging. Radio-genomics, as described by [5]'s authors, is a new subfield of genomics that examines the multi-scale relationships between imaging studies of patients and their genes' expression patterns. These all techniques are time consuming and less effective.

This paper develop an intellengent model to detect breast cancer automatically using CNN. This model will help in diagnosing the breast cancer in early stages. When breast cancer is found early, it is easy to treat and has a good chance of working. Accordingly, women should begin having screenings at the age of 30 and then annually thereafter. Some of the specific people who gained from this work are:

- Radiologists: This research helps doctors give correct information on time and in the best way possible.
- Woman: Since the suggested model can identify cancer confidently or accurately, it can be potentially manageable to treat.
- Families: Women are the backbone of our society, and once cancer has been diagnosed through screening, they no longer have to worry about whether or not they will respond to therapy.

II. LITERATURE

This section provide a summary of previous work done by researchers to detection cancer automatically using machine learning techniques. The author in [4] provided an account of the diverse approaches employed in the identification of breast cancer through histological image analysis (HIA) utilizing different artificial neural network (ANN) architectures. The research was categorized by the author in accordance with the pertinent dataset. The material was provided in ascending order of chronology. This study found that ANNs were first utilized for HIA in 2012. PNNs constituted the most widely implemented algorithms. But, the majority of feature extraction research employed textural and morphological characteristics. Early detection and diagnosis of breast cancer clearly benefited from deep convolutional neural networks (DCNN), which also produced more successful therapy. Diverse methodologies were employed in order to predict Non-Communicable Diseases (NCDs).

Naturally inspired computing (NIC) algorithms are developed and utilized to detect numerous human diseases. In their work on insect-based cancer and diabetes diagnostic algorithms, the authors of [6] introduced five NIC algorithms. The research discovered that it performed exceptionally well in identifying various forms of cancer (lung, breast, ovarian and prostate). A combination of guided artificial bee colony and neural networks were utilized for diagnosis of breast cancer with greater specificity. The authors also devised a rather successful approach for leukaemia and diabetes diagnosis. Combining NICs with other classifiers produces more accurate and interesting findings, the researchers said. They underlined the need of more work to identify the different stages of cancer.

The authors of [7] illustrated the efficiency of ANN in the classification of an initial-stage cancer diagnosis. The majority of NNs have shown the ability to recognize cancerous cells, according to their research. However, the imaging method needs a large computing capacity for image preprocessing.

The authors in [8] investigated several ML algorithms, pattern recognition, and computational methods for breast cancer detection. Twenty-seven on machine learning, four on ensemble approaches, and eight on deep learning—among other papers on breast cancer research—were evaluated. Most of the study made use of imaging; just a small fraction used genetics. The key ML algorithms that were used to predict breast cancer are random forest, DT, SVM, and genetics. Nevertheless, imaging techniques included numerous algorithms like CNNs and Naive Bayes.

This research [9] presented a DCNN based automated melanoma classifier for accurate classification between benign and malignant melanoma. The DCNN classifier got accuracy rates of 90.42%, 88.23% and 81.41% on the International Skin Imaging Collaboration (ISIC) 2020, 2017 and 2016 datasets, respectively. It was inspired by CAD that this [10] research introduced a new classification model for melanoma. The presented DCNN model utilizing EfficientNet-B6 to diagnose melanoma is evaluated using the most recent public challenge dataset, ISIC 2020, to categorize skin lesions as malignant or benign. This model's accuracy was 97.84 %.

In [11], two hybrid optimum classification approaches for breast cancer classification were devised. Two types of breast cancer tumour samples benign and malignant are taken into consideration. The work is based on using whale optimization and dragon fly algorithms in conjunction with SVM to identify the algorithm's ideal parameters in order to obtain improved breast cancer classification accuracy.

In [12], the author introduced a fresh deep learning (DL) approach to efficiently address imbalanced data in breast cancer datasets. The majority of ML classification algorithms are based on the underlying assumption that that the data's class distributions are homogeneous. Nevertheless, it is crucial to acknowledge that in numerous practical scenarios, the distribution of data often exhibits imbalance. The presence of this imbalance has the potential to result in a range of performance limitations for ML algorithms. The primary focus of the research was the utilization of publicly available structured data for the development of a trained model. In order to better identify and classify cases of malignant breast cancer early on, the study's fundamental premise centered on the viability of using DL algorithms employing structured datasets related to breast cancer. The researchers concluded that applying DL on structured data is a practical approach to handle class imbalances in clinical research.

The research done in [13] posits that the overall survival rate of individuals affected by breast cancer can be considerably impacted by the prompt and accurate identification of the condition.. The authors aimed to develop a comprehensive model driven by machine learning that is capable of autonomous operation. To achieve the research objective, the researchers employed image processing technology to analyze stained histological images of breast cancer tissue. The images were subjected to pre-processing in the first phase, wherein the stain normalization technique was utilized. In addition to performing pre-processing, the utilization of data augmentation techniques is employed as a means to tackle the challenge posed by a limited dataset size. Afterwards, the DCNN model that has been suggested is employed to extract high-level characteristics from the pre-processed images. These extracted features are subsequently utilized as input for a conventional multi-layer perceptron classifier. Within the domain of DCNNs, a number of prominent architectures have arisen that demonstrate efficacy in the extraction of features. The models encompassed in this set are Inception v3, Inception ResNetv2, Xception, as well as two variations of the VGGNet model. The experiment utilized a dataset comprising 400 Breast Cancer Histology (BACH) images obtained from the ICIAR 2018 Grand challenge. The architecture under consideration ultimately achieves a level of accuracy of 92.5%, surpassing that of alternative models [14].

The studies carried out in [15] came to the conclusion that lowering the death rate connected with breast cancer depends mostly on early identification of the condition and correct diagnosis. A new method for extracting edges and a tweaked convolutional recurrent neural network (CRNN) model are introduced in this paper. The goal of these innovations is to make medical imaging a more reliable tool for assessing breast cancer. The suggested Elfa-CRNN model analyses line segments inside the mass as part of its line feature analysis. A total of 250 breast image files were used in the study, all obtained from the Mendeley database. Canny and Sobel, the Elfa-proposed algorithm, achieved an impressive 98% accuracy rate, demonstrating a higher degree of precision. In addition, the Elfa-CRNN model produced results that were in line with expectations, outperforming CRNN, AlexNet, and VGG models with an accuracy rate of 99.75% [16].

The author of [17] investigate the increasing prevalence of breast cancer and emphasize the predominant utilization of conventional X-ray Imaging as the fundamental screening method for breast cancer. The preprocessing of 9000 mammograms involved the implementation of data augmentation methods, contrast-limited adaptive histogram equalization, and a median filter. The classification task was executed by employing a pre-trained dataset and leveraging a CNN. The research results illustrates that the model achieved effectively higher accuracy when using preprocessed images, in contrast to the model that did not utilize preprocessed images.

III. METHODOLOGY

The World is developing day by day and making advancements in every field. The medical field is also developing and introducing advanced disease detection and cure instruments. But these disease detection processes are quite slow. On the other hand, machine learning algorithms are fast and accurate in data classification. This research present a machine learning model to detect breast cancer (Invasive Ductal Carcinoma (IDC)). The workflow of the developed research is discussed below and illustrated in Figure 1.

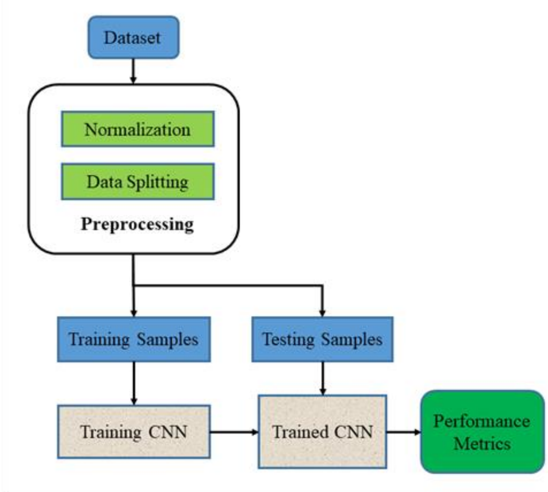

Figure 1. Flow chart of the presented methodology

In order to maintain uniformity across all pixels and mitigate bias, normalization must be applied to the entire image. After normalization, the dataset is label. As the problem is binary classification problem. The binary classes will be designated as 'IDC negative' and ‘IDC positive,' with '0' denoting 'IDC –ive' means healthy and '1' denoting ' IDC +ive' means cancer detected. After labeling, the dataset is divided into three sections: the training dataset, the validation dataset, and the training dataset. The training dataset is given to CNN.

CNNs are often used in computer vision (CV) applications. CNN is a widely used and famous deep learning technique. One major benefit of CNN is its ability to autonomously identify important traits over and time again. Computer vision, face recognition, audio processing, and other areas of CV are just a few of the many sectors that are making use of CNNs. CNN is made up of several layers, convolution layer, pooling layer, and fully connected layers are the basic layers. The convolution operation is define by equation 1.

$$F_{mn} = \sum_{m} \sum_{n} \mathrm{I}_{i-m,\, j-n} * K_{m,n}$$

Where F_{mn} is the feature map extracted from the input image I by applying kernel $K_{m,n}$. After convolution layer, pooling layer is applied. In this model, MaxPool operation is used. ReLU activation function is used to add non linearity to the model for better training. The mathematical equation of ReLU is

$$R(x) = \max(0, x)$$

Typically, the final portion (or layers) of a CNN structure (utilised for classification) comprises fully-connected layers. In this configuration, every neuron within a specific layer is interconnected with each of the neurons from the preceding layer. The CNN architecture employs the last fully-connected layer as the output layer (classifier).

IV. RESULTS AND DISCUSSION

To evaluate the efficacy of a ML model, including neural networks, learning curves are frequently employed. The CNN performance can be analyze by other different parameters i.e. precision, Recall, F1-score, confusion matrix, ROC curve, and speed and efficiency etc. Here are some of the parameters are discussed.

A. *Learning Curves*

The effectiveness of a model can be evaluated exclusively through its accuracy. It is mathematically expressed as the ratio of correct predictions to the total number of predictions made within a particular period of time.

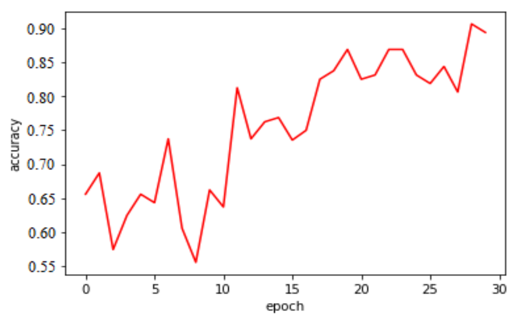

Figure 2. Training accuracy of the model

Training accuracy is generally the accuracy you achieve when we use the model on the training data. This told about how good or bad the model is train. The true prediction ratio of the trained model on the validation dataset is considered as validation accuracy. Figure 2 and Figure 3 illustrate the training and validation accuracy of the model.

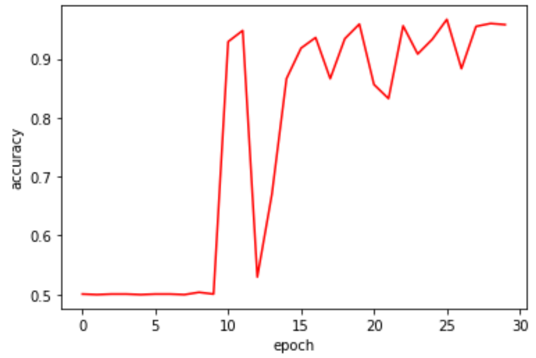

Figure 3. Validation accuracy of the model

B. Loss Curves

Loss is the difference between the predicted value and the labeled value in supervised learning, where the labeled value represents the actual value. Conversely, a greater loss value is associated with a greater difference between the predicted and actual value. Figure 4 displays the training and validation loss curves of the CNN model, which were developed prior to its implementation. As the loss of the model decreases during the training epoch, the model's learning becomes more efficient over time. The model accuracy is indicated by the disparity between the training loss curve and the validation loss curve. The difference in our model is quite less which means the model is well train and the prediction of the model for the new dataset is more accurate.

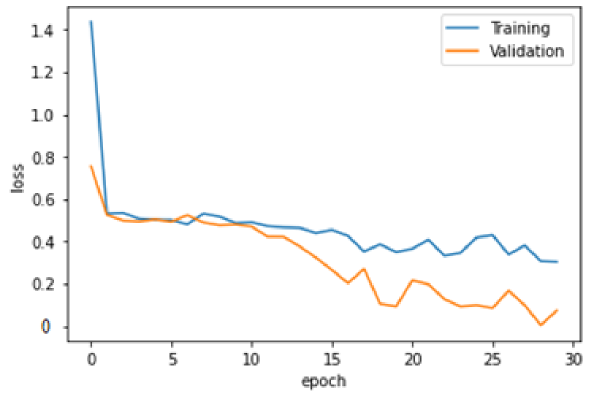

Figure 4. Training and validation loss of the model

C. Confusion Matrix

The confusion matrix encapsulates the efficacy of the model when evaluated against new data, which is conventionally the test dataset. The x-axis of the confusion matrix denotes the predicted classifications, whereas the y-axis illustrates the actual labels assigned to the respective classes. Figure 5 presents the confusion matrix associated with the model.

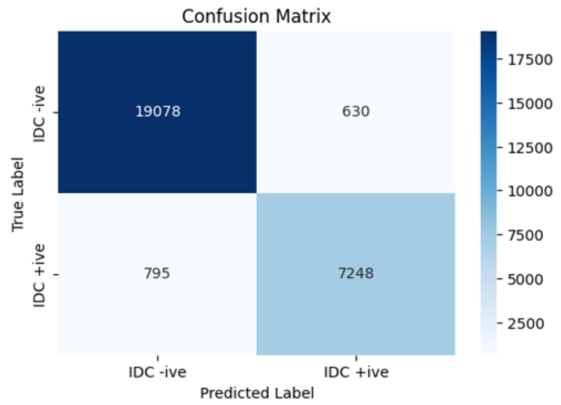

Figure 5. Confusion matrix of the model for the testing samples

D. Comparison

The presented model is working excellently achieving high accuracy than the work done in literature. The work done in [18] achieve a training accuracy of 87% for the same dataset. Table 1 illustrates the comparison of different metrics of both the models.

TABLE I. COMPARISION OF THE PRESENTED MODEL WITH PREVIOUS MODEL

Model	Accuracy	Precision	Recall	F1-score
[18]	87%	0.84	0.84	0.87
Presented CNN Model	90%	0.96	0.95	0.9549

CONCUSLION

In most cases, breast cancer initiates in the mammary ducts and glands. They develop in these areas as epithelial tumors prior to the formation of a cancerous mass. While these masses often appear benign, they have the potential to transform into malignant tumors if they are premalignant. The present research proposes using CNN to automatically detect this malignancy by analyzing the IDC tissue areas. The work is done on real-time publically available dataset at Kaggle. The combination of the layers of CNN model depend on the characteristics of the problem. We tested the different combination and different activation functions to get the bestfit model. The training accuracy and validation accuracy for model is 90% and 95% respectively. The training accuracy is improved by 3% from the previous work done on the dataset. In result chapter, other metrics and results are evaluated to analyze the model's efficiency. Confusion matrix shows the prediction of the model for the testing dataset.

In future, the model accuracy can be increased by the enhancing the dataset, balance the dataset, manually extract features from the images, and try some different combination of CNN layers.

REFERENCES

[1] Y. Qasim, H. Al-Sameai, O. Ali, and A. Hassan, "Convolutional neural networks for automatic detection of colon adenocarcinoma based on histopathological images," in *International Conference of Reliable Information and Communication Technology*, 2020: Springer, pp. 19-28.

[2] A. S. Sakr, "Automatic Detection of Various Types of Lung Cancer Based on Histopathological Images Using a Lightweight End-to-End CNN Approach," in *2022 20th International Conference on Language Engineering (ESOLEC)*, 2022, vol. 20: IEEE, pp. 141-146.

[3] N. A. Abujabal and A. B. Nassif, "Meta-heuristic algorithms-based feature selection for breast cancer diagnosis: A systematic review," in *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2022: IEEE, pp. 1-6.

[4] X. Zhou *et al.*, "A comprehensive review for breast histopathology image analysis using classical and deep neural networks," *IEEE Access,* vol. 8, pp. 90931-90956, 2020.

[5] D. H. Sutanto and M. Abd Ghani, "A Benchmark Feature Selection Framework for Non Communicable Disease Prediction Model," *Advanced Science Letters,* vol. 21, no. 10, pp. 3409-3416, 2015

[6] R. Gautam, P. Kaur, and M. Sharma, "A comprehensive review on nature inspired computing algorithms for the diagnosis of chronic disorders in human beings," *Progress in Artificial Intelligence,* vol. 8, no. 4, pp. 401-424, 2019

[7] M. Mahmood, B. Al-Khateeb, and W. M. Alwash, "A review on neural networks approach on classifying cancers," *IAES International Journal of Artificial Intelligence,* vol. 9, no. 2, p. 317, 2020.

[8] N. Fatima, L. Liu, S. Hong, and H. Ahmed, "Prediction of breast cancer, comparative review of machine learning techniques, and their analysis," *IEEE Access,* vol. 8, pp. 150360-150376, 2020.

[9] R. Kaur, H. GholamHosseini, R. Sinha, and M. Lindén, "Melanoma classification using a novel deep convolutional neural network with dermoscopic images," *Sensors,* vol. 22, no. 3, p. 1134, 2022.

[10] I. Elansary, A. Ismail, and W. Awad, "Efficient classification model for melanoma based on convolutional neural networks," in *Medical Informatics and Bioimaging Using Artificial Intelligence: Challenges, Issues, Innovations and Recent Developments*: Springer, 2021, pp. 15-27.

[11] P. Geetha and S. Umamaheswari, "Choose most efficient features of breast cancer using an SVM classifier for breast cancer diagnosis," in *2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, 2022: IEEE, pp. 1-5.

[12] A. Farhadi *et al.*, "Breast cancer classification using deep transfer learning on structured healthcare data," in *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 2019: IEEE, pp. 277-286.

[13] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Breast cancer diagnosis with transfer learning and global pooling," in *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, 2019: IEEE, pp. 519-524.

[14] R. Watanabe, N. Soga, and H. Noji, "Novel nano-device to measure voltage-driven membrane transporter activity," *IEEE Transactions on Nanotechnology,* vol. 15, no. 1, pp. 70-73, 2015.

[15] C.-M. Kim, R. C. Park, and E. J. Hong, "Breast mass classification using eLFA algorithm based on CRNN deep learning model," *IEEE Access,* vol. 8, pp. 197312-197323, 2020.

[16] S. M. Badawy, A. E.-N. A. Mohamed, A. A. Hefnawy, H. E. Zidan, M. T. GadAllah, and G. M. El-Banby, "Classification of Breast Ultrasound Images Based on Convolutional Neural Networks-A Comparative Study," in *2021 International Telecommunications Conference (ITC-Egypt)*, 2021: IEEE, pp. 1-8.

[17] H.-C. Lu, E.-W. Loh, and S.-C. Huang, "The classification of mammogram using convolutional neural network with specific imagepreprocessing for breast cancer detection," in *2019 2nd international conference on artificial intelligence and big data (ICAIBD)*, 2019: IEEE, pp. 9-12.

[18] S. A. Alanazi *et al.*, "Boosting breast cancer detection using convolutional neural network," *Journal of Healthcare Engineering,* vol. 2021, no. 1, p. 5528622, 2021.

Engr. Hussain Ahmad, born on 15th February 1995, is a postgraduate student in the Department of Electrical Engineering, University of Engineering and Technology, Peshawar, Pakistan. He has completed his BS in Electrical Engineering from COMSATS university Islamabad, Wah campus.

How to cite this article:

Hussain Ahmad, Muhammad Zeb, Khalil-ur-Rahman“An Intelligent Model for Detection of Breast Cancer based on Convolutional Neural Network” International Journal of Engineering Works, Vol. 12, Issue 08, PP. 170-174, August 2025. https:// doi.org/10.5281/zenodo.16908308.