TECH-CHRONICLE
AN INTERNATIONAL E-JOURNAL ON
EMERGING TRENDS IN SCIENCE, TECHNOLOGY AND MANAGEMENT

Breast Cancer Detection Using Convolutional Neural Network

Abhishek Anil Borse#, Prof. Jayant Adhkari*, Prof. Jayant Rohankar*

#Department of Information Technology, RTMNU
(Artificial Intellegence and Machine Learning)
[1]abhishek.borse10@gmail.com
[2]jayant.cse@tgpcet.com
[3]jayant.mtech@tgpcet.com

***Abstract*— Breast Cancer (BC) is one of the most common cancer among women worldwide. Early detection of breast cancer is critical to increase the chances of survival. Deep learning has emerged as a promising technique for breast cancer detection from medical images. In this paper, we present a novel deep learning framework for breast cancer detection using histopathology images. Without directly applying the Deep Learning techniques , we first performed some data preprocessing and image processing on images. The proposed framework consists of a deep convolutional neural network (CNN) DenseNet121 that is trained on a large dataset of histopathology images. The CNN is designed to automatically learn features from images and classify them into malignant or benign categories. We evaluate our proposed framework on a publicly available dataset and achieve a high classification accuracy of 90%. Our results demonstrate the potential of deep learning for breast cancer detection and its potential to assist radiologists in their clinical decision-making process. This work could have significant implications for improving breast cancer screening and diagnosis, ultimately leading to better patient outcomes.**

***Keywords*— Breast Cancer, Histopathology images, CNN, Deep Learning, Densenet121, Benign, Malignant.**

I. Introduction

Breast Cancer (BC) is a type of cancer that develops in the breast tissue. It occurs when abnormal cells in the breast grow and divide uncontrollably, forming a tumor. Breast cancer is one of the most common cancers among women worldwide, and it can also occur in men, although it is rare. Early detection of breast cancer is important, as it can increase the chances of successful treatment and improve the long-term outcomes for patients. Breast cancer can be detected through various methods, including mammography, clinical breast exams, and self-exams. Mammography is a common screening method for breast cancer, in which X-rays are used to produce images of the breast tissue. Clinical breast exams are physical exams performed by a healthcare provider to check for lumps or other abnormalities in the breast tissue. Self-exams involve a woman checking her own breasts for lumps or other changes.In addition to these methods, there are also several new and emerging technologies for breast cancer detection, including molecular imaging, magnetic resonance imaging (MRI), and genetic testing. These technologies can help to identify breast cancer at an earlier stage, and may also be used to monitor the progression of the disease and the effectiveness of treatment. In recent years, deep learning, a subset of artificial intelligence, has shown promising results in detecting breast cancer from medical images such as mammograms and Biopsy slide images. Deep learning models can analyze large amounts of data and learn patterns that may not be discernible to the human eye, thereby improving the accuracy and speed of breast cancer detection. Deep learning-based breast cancer detection has the potential to revolutionize breast cancer screening and diagnosis. It can aid in the early detection of breast cancer, thereby improving treatment outcomes and saving lives. Deep learning is a sub-set of machine learning. Deep is an unsupervised learning that learns from the data. The data may be unstructured or unlabeled. Deep neural network contains more than two hidden layers then it is called a deep network. Basically, the first layer is the input layer and the second is the output layer. The intermediate layer is called a hidden layer that has more layers as compared to a neural network. The node containing the layer is called neurons. The difference between machine learning and deep learning is that deep learning is closer to its goal as compared to machine learning. In this paper, Convolution Neural Network (CNN) is used for the classification of the breast cancer dataset. Convolutional Neural Network is used to classify the images. It takes the images of the breast cancer dataset as an input. CNN takes the images as an input associated with their corresponding weights. The weights are adjusted to minimize the error and enhance the performance. CNN contains many layers such as convolution layer, pooling layer, ReLU layer (activation function layer) and fully connected layer. In the convolution layer, a feature map is used to extract the features of the given image and makes the original image more compact. In addition,pooling layer is used to reduce the dimensions of the image. ReLU layer is used as an activation function in which it checks if the value of the activation function lies in a given range or not. Fully connected layer is the last layer of the model. It combines the results of all layers and applies the classification function to give the probability to each class of the output for image classification. Here in this paper we are using feature extractor DenseNet network to extract feature from the biopsy slide images to detect the breast cancer in early stage by classifying the images as cancerous and non cancerous.

II. Related Work

This section gives the information about the related work of the research that has been already done. Basically two techniques are used to detect breast cancer. First one is machine learning and the second is deep learning. There is a lot of research that is conducted through machine learning. But machine learning techniques have some problems that are removed through deep learning.

I. Prediction of Breast Cancer Using SVM Algorithm[1]

In this, author proposed an approach using Support Vector Machine (SVM) on different datasets like Wisconsin Breast Cancer (WBC) dataset the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, the Surveillance, Epidemiology, and End Results (SEER) dataset. The outcomes of SVM consist of accuracy and precision. It is used for classification, which trains models to categorize cancer patients according to their diagnosis. The system also focuses on providing a detailed description of the tumor for each patient diagnosed. Here the author uses K-fold cross validation technique for data validation, where the accuracy of various machine learning algorithmson the data is tested.survival. The proposed system classifies the tumors into malignant or benign using features obtained from cell images with 99% accuracy.

II. Hybrid Approach to predict Breast Cancer using Machine Learning Techniques[2]

In this, the author proposed a model based on a hybrid approach using machine learning. It implemented this approach using MRMR feature selection with four classifiers to find out the best results. The author used the four classifiers SVM, Naïve Bays, Function tree and End Meta and did the comparison between all. It found that SVM was a good classifier. To find out the better results. This model is being trained and tested on datasets like Wisconsin Diagnostic Breast Cancer dataset, Wisconsin Breast Cancer Data and Breast Tissues data. The proposed system classifies the tumors into malignant or benign using features obtained from cell images with 98% accuracy.

II. An Automatic Detection of Breast Cancer Diagnosis and Prognosis Based on Machine Learning Using Ensemble of Classifiers

In this, the author proposed an approach where we build an architecture using 4 different ML models. This architecture is composed of four different ML-based classifiers named SVM, LR, NB, and DT. They are stacked and then further trained as an ensemble. After training, the ANN model is used for the outcome. The performance is compared with the several ML classifiers individually with and without up- sampling techniques. We also compared the performance of the proposed ensemble model with other ensemble models. An ensemble model is stacked, and predictions are concatenated and then fed to the ANN model for final prediction. Here we have analyzed classification results using a 10-fold cross-validation technique. and this experiment is performed on the Breast cancer Wisconsin (Diagnosis) andBreast cancer Wisconsin (Prognosis) databases.

V. Breast Cancer Classification From Histopathological Images Using Patch-Based Deep Learning Modeling

In this study, the author proposed a Pa-DBN-BC model for the classification of breast cancer on the histopathology images. The Pa-DBN-BC model comprises four main phases which are the preprocessing, patch generation, DBN, and classification phase where the model works on equal size patches of images. In preprocessing we can crop our input data around the region of interest (ROI) without worrying about the limitation of keeping the images of equal sizes. The proposed model works in an unsupervised fashion for the extraction of features from the input histopathology image patches in the form of feature vectors. The extracted features matrix is then transferred to the backpropagation neural network which is a supervised learning phase and it comprises the conjugate gradient. A model is formed by the feature matrix of images in the training phase and the final stage is the classification stage which discriminates between cancerous and non-cancerous

TECH-CHRONICLE
AN INTERNATIONAL E-JOURNAL ON
EMERGING TRENDS IN SCIENCE, TECHNOLOGY AND MANAGEMENT

regions. a publicly available dataset based on histopathology images is used. The dataset includes histopathology images from the four different data cohorts, Hospital of the University of Pennsylvania (HUP), Case Western Reserve University (CWRU), Cancer Institute of New Jersey (CINJ), and The Cancer Genome Atlas (TCGA) and their corresponding binary masks of invasive breast cancer regions annotated by pathologists. In this experiment the model has achieved 86% of the accuracy.

V. Breast Cancer–Detection System Using PCA, Multilayer Perceptron, Transfer Learning, and Support Vector Machine[5]
In this study, the author has proposed an approach of combining 4 different approaches to achieve classification. approaches are Principal Component Analysis (PCA), Multilayer Perceptron (MP), Transfer Learning (TL), and Support Vector Machine (SVM). The experiment performed k-fold cross-validation 50 times on average to enhance the generalizability of the model. and the highest accuracy achieved during 10-fold cross-validation was 86.97%, as applied to the BCCD data set.

VI. Optimized Deep Neural Networks Architecture Model For Breast Cancer Diagnosis
In this study, the author proposed Deep Neural Network (DNN) architecture for breast cancer prediction. Here the author has used the Mammographic Mass dataset obtained from UC Irvine Machine Learning Repository .In this experiment, three types of NN models were used: NN1HL (Neural Networks with One Hidden Layer), DNN4HL (Deep Neural Networks with Four Hidden Layers), and DNN8HL (Deep Neural Networks with Eight Hidden Layers) additionally with dropout layer. ReLu is used as an activation function on all the models and sigmoid is used as output activation function. In this experiment the author has achieved 89% accuracy.

III. Proposed Architecture

The proposed architecture design to classify breast cancer in the early stage using biopsy slide image, Here we are using image samples that are generated from breast tissue biopsy slides stained with hematoxylin and eosin (HE).

This system uses Deep learning CNN architecture to classify the images as Benign and Malignant. We are using the DenseNet121 architecture for feature extraction and classification. This architecture utilizes dense connections between layers, through Dense Blocks, where we connect all layers (with matching feature-map sizes) directly with each other.

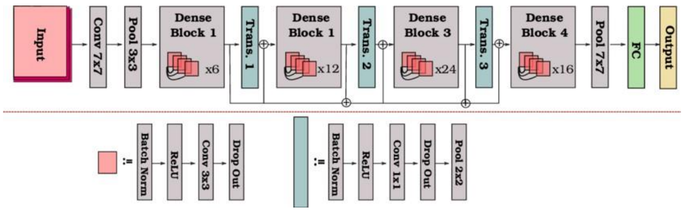

Fig.1 DenseNet121 Architecture

IV. Data Collection

The dataset which we are using in this paper is the open source data available at Kaggle https://www.kaggle.com/ambarish/breakhis.

It has 7909 microscopic images of Breast tumor tissue collected from the 82 patients using different magnifying factors (40X, 100X, 200X, and 400X). It contains 2480 benign and 5429 malignant samples. All the images are of size *700 x 460* pixels, 3-channel RGB images with 8-bit depth in each channel in PNG format. We have divided the dataset as 10% images of the dataset used for testing purposes and 90% images are used in the training.

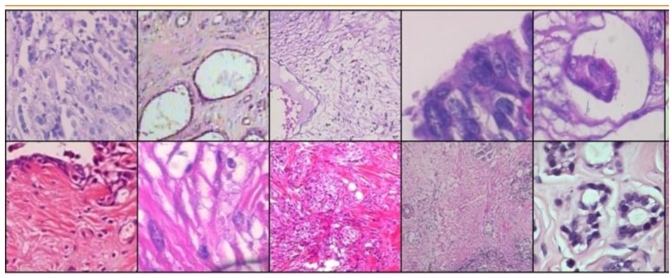

Fig. 2 Dataset Images

The dataset BreaKHis is divided into two main groups: benign tumors and malignant tumors. Histologically benign is a term referring to a lesion that does not match any criteria of malignancy-e.g., marked cellular atypia, mitosis, disruption of basement membranes, metastasize, etc. Normally, benign tumors are relatively "innocents", present slow growing and remain localized. Malignant tumor is a synonym for cancer: lesions can invade and destroy adjacent structures (locally invasive) and spread to distant sites (metastasize) to cause death.

The dataset currently contains four histological distinct types of benign breast tumors: adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenona (TA); and four malignant tumors (breast cancer): carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC) and papillary carcinoma (PC).

For data preprocessing we have applied data transform to make our data useful for training. The data transform that we have applied in this study are RandomResizedCrop and Normalization.

V. Proposed Network

The network which we are using is the DenseNet121 architecture. The training dataset were trained on this architecture to classify the images as Benign and Malignant. The model is trained on the training set for 15 epochs, with an evaluation on the validation set after each epoch. Depending on the accuracy value of the model, the weights are saved after each epoch to keep the best model, which is then evaluated on the test set.

The model is trained with 16 batch size with learning rate of 0.001, using Adam Optimizer and Cross entropy loss for calculation of Loss during training the model.

For Visualization of model performance we have used Tensorboard. This gives the details about the training that How our model is performing after each step.

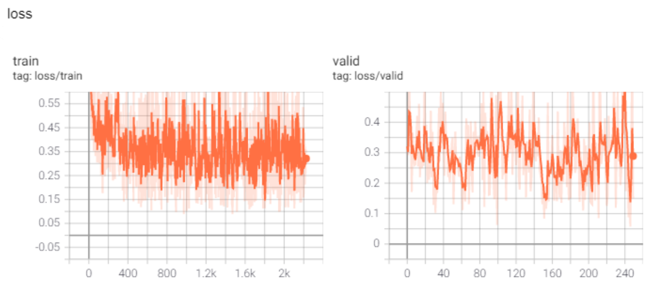

Fig. 3 Train Validation Lose

VI. Result

After testing our model on testing dataset we have achieved the accuracy of 90% with DenseNet121. To analyze the model performance I have measured it on the parameters like accuracy, precision, recall and confusion matrix.

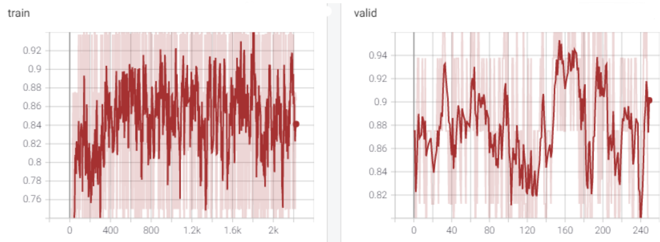

Fig. 4 Train Validation Accuracy

VII. Conclusion and Future Scope

It is concluded that early detection of Breast Cancer helps in curing the disease. Cancer is the collection of the various diseases so; medicinal effects may be different at the same stage of cancer. In this paper, detection of breast cancer is done through Convolution Neural Network with DenseNet121 Architecture. Data sets are collected from the Kaggle Open Source Repository. In the initial stage, 82 medical cases are taken and then the classification of breast cancer is done to determine benign and malign. Though cancer is predicted and classified through various approaches, automatic detection of disease is still difficult to approach due to different categories of cancer. In Future Scope, various new deep learning algorithms are required to be implemented for the detection of different stages and categories of breast cancer simultaneously.

References

[1] Prediction of Breast Cancer Using SVM Algorithm Deepika S and Kapilaa Ramanathan Devi N Department of Information Technology Sri Venkateswara College of Engineering, Sriperumbudur, Tamil Nadu, India – 602117 Assistant Professor Department of Information Technology Sri Venkateswara College of Engineering, Sriperumbudur, Tamil Nadu, India - 602117

[2] Hybrid approach to predict breast cancer using machine learning techniques," International Journal of Computer Science Engineering, vol. 5, no. 3, pp. 125- 136, 2016.

[3] An Automatic Detection of Breast Cancer Diagnosis and Prognosis Based on Machine Learning Using Ensemble of Classifiers Usman Naseem, Junaid Rashid Department of Information Technology, Sydney International School of Technology and Commerce, Sydney, NSW 2000, Australia

[4] Breast Cancer Classification From Histopathological Images Using Patch-Based Deep Learning Modeling Irum Hirra, Mubashir Ahmad Department of Computer Science and IT, The University of Lahore, Sargodha Campus, Sargodha 40100, Pakistan

[5] Breast Cancer–Detection System Using PCA, Multilayer Perceptron, Transfer Learning, and Support Vector Machine Huan-Jung Chiu, Tzuu-Hseng,Ping-Huan Kuo AI Robots Laboratory, Department of Electrical Engineering, National Cheng Kung University, Tainan 70101, Taiwan Department of Intelligent Robotics, National Pingtung University, Pingtung 90004, Taiwan

[6] Optimized Deep Neural Networks Architecture Model For Breast Cancer Diagnosis G. Kanimozhi, P. Shanmugavadivu Department of Computer Science and Applications, The Gandhigram Rural Institute (Deemed to be University), Gandhigram, Tamil Nadu, India.

[7] Using convolutional neural networks for the classification of breast cancer images. Eric Bonnet Centre National de Recherche en G ́enomique Humaine Institut de Biologie Fran ̧cois Jacob Commissariat `a l'Energie Atomique Universit ́e Paris–Saclay, Evry, France eric.bonnet@cea.fr June 2021

[8] Predicting Axillary Lymph Node Metastasis in early breast cancer using deep learning on primary tumor biopsy slides. Feng Xu, Chuang Zhu, Wenqi Tang, Ying Wang, Yu Zhang , arXiv:2112.02222v4 [eess.IV] 8 Jun 2022.

[9] Breast Cancer Prediction using Supervised Machine Learning Algorithms , Mamta Jadhav, Zeel Thakkar, Prof. Pramila M. Chawan, Oct 2019. [4] Using convolutional neural networks for the classification of breast cancer images , Eric Bonnet , 31 Aug 2021.

[10] BCI: Breast Cancer Immunohistochemical Image Generation through Pyramid Pix2pix, ShengJie Liu, Chuang Zhu, Feng Xu, Xinyu Jia, Zhongyue Shi, Mulan Jin 25 Apr 2022.

[11] Breast cancer detection using artificial intelligence techniques: A systematic literature review, Ali Bou Nassif, Manar Abu Talib, Qassim Nasir, Yaman Afadar, Omar Elgendy, Artificial Intelligence in Medicine, Elsevier, Vol 127, May 2022.

[12] Preprocessing of Breast Cancer Images to Create Datasets for Deep-CNN ABHIJITH REDDY BEERAVOLU 1 , SAMI AZAM 1 , MIRJAM JONKMAN 1 , (Member, IEEE), BHARANIDHARAN SHANMUGAM1 , (Member, IEEE), KRISHNAN KANNOORPATTI 1 , AND ADNAN ANWAR 2 , (Member, IEEE) 1College of Engineering, IT, and Environment, Charles Darwin University, Casuarina, NT 0810, Australia 2Centre for Cyber Security Research and Innovation (CSIR), School of IT, Deakin University, Geelong, VIC 3216, Australia Corresponding author: Sami Azam (sami.azam@cdu.edu.au)

[13] Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques Md. Milon Islam1 · Md. Rezwanul Haque1 · Hasib Iqbal1 · Md. Munirul Hasan2 · Mahmudul Hasan3 · Muhammad Nomani Kabir2 Received: 11 August 2020 / Accepted: 18 August 2020 / Published online: 1 September 2020 © Springer Nature Singapore Pte Ltd 2020

[14] Breast Cancer Detection and Classification using Analysis and Gene-Back Proportional Neural Network Algorithm Amandeep Kaur, Prabhjeet Kaur

[15] Augmentation of Classifier Accuracy through Implication of Feature Selection for Breast Cancer Prediction Deepa B G, Senthil S , Gupta Rahil M, Shah Vishakha R

[16] EARLY DETECTION OF BREAST CANCER USING SVM CLASSIFIER TECHNIQUE Y.Ireaneus Anna Rejani+ , Dr.S.Thamarai Selvi* + Noorul Islam College of Engineering, Tamilnadu, India. *Professor&Head, Department of Information and technology, MIT, Chennai, Tamilnadu, India.

[17] Preprocessing of Breast Cancer Images to Create Datasets for Deep-CNN ABHIJITH REDDY BEERAVOLU, SAMI AZAM, MIRJAM JONKMAN, (Member, IEEE), BHARANIDHARAN SHANMUGAM , (Member, IEEE), KRISHNAN KANNOORPATTI, AND ADNAN ANWAR , (Member, IEEE)

[18] An Automatic Detection of Breast Cancer Diagnosis and Prognosis Based on Machine Learning Using Ensemble of Classifiers USMAN NASEEM , JUNAID RASHID, LIAQAT ALI, JUNGEUN KIM, QAZI EMAD UL HAQ, MAZHAR JAVED AWAN, AND MUHAMMAD IMRAN

[19] Breast Cancer Prediction using Naïve Bayes Classifier Megha Rathi, Arun Kumar Singh

[20] Prediction of Breast Cancer Using SVM Algorithm Deepika S* and Kapilaa Ramanathan Devi N

[21] Breast Cancer Detection using Artificial Neural Networks Md Haris Uddin Sharif1

[22] BREAST CANCER PREDICTION USING MACHINE LEARNING Ramik Rawal