International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

Classification of Cancer Images using Deep Learning

Ms. Mrudula Bapat[1], Ms. Prachi Jirimali[2], Ms. Shravani Tapole[3], Prof. Ujwala Gaikwad[4]

[1]Final Year Student, Department of Computer Engineering, Terna Engineering College, Nerul, India.
[2]Final Year Student, Department of Computer Engineering, Terna Engineering College, Nerul, India.
[3]Final Year Student, Department of Computer Engineering, Terna Engineering College, Nerul, India.
[4]Professor, Department of Computer Engineering, Terna Engineering College, Nerul, India.

Abstract - *This paper presents a deep learning approach for breast cancer detection and classification of breast histopathology images into invasive ductal carcinoma IDC or non-invasive ductal carcinoma Non-IDC type. Breast cancer is one of the most common and deadly cancers in women worldwide. It remains a challenge to cure this disease and therefore it is important that it is diagnosed in time. It is proven that early detection can increase the survival rate of the person. Deep learning plays a vital role in the automatic detection and early detection of cancer. This paper is going to focus on Invasive Ductal Carcinoma (IDC) which is the most common subtype of all breast cancers. Invasive breast cancer detection can be a time consuming and challenging task. Using the breast histopathology images, an image classifier would be built which would be able to predict whether the person is suffering from Invasive Ductal Carcinoma (IDC) type of breast cancer or not. For the image classification, this paper presents developing of a Convolutional Neural Network (CNN) model. CNN uses a data-driven approach to automatically learn feature representations for images.*

Key Words*:* **image classification, cancer, deep learning, neural network**

1. INTRODUCTION

Cancer refers to a large number of diseases characterized by the event of abnormal cells that divide uncontrollably and have the power to destroy normal body tissue.[1] Cancer may often spread throughout your body. Many different types of cancer are Skin cancer, Breast cancer and Brain tumor. Many different technologies are adopted for detection of various cancer symptoms, like for Brain tumors, magnetic resonance images of brains are used for detection, while in Skin cancer dermatologists level classification using Deep learning. Breast cancer is the most common invasive cancer in women and it is also the second leading cause of cancer death in women. [2] Breast cancer develops from cells lining the milk ducts and slowly grows into a lump or a tumor. Clinicians assumed that it takes a significant amount of time for a tumor to grow 1 cm in size starting from a single cell. A malignant tumor can spread beyond the breast to other parts of the body via the lymphatics or the bloodstream. Breast cancer can be either invasive or non-invasive. Invasive cancer spreads from the milk duct or lobule to other tissues in the breast. Non-invasive ones are unable to invade other breast tissues. Invasive ductal carcinoma (IDC), also known as infiltrating ductal carcinoma, is cancer that began growing in a milk duct and has invaded the surrounding fibrous or fatty tissue or tissue outside of the duct. IDC is the most common form of breast cancer. It represents 80 percent of all breast cancer diagnoses.[2] Invasive breast cancer detection is a challenging and a time-consuming process. It involves a pathologist examining the tissue slide under a microscope. The pathologist needs to visually scan large regions where there's no cancer in order to ultimately find malignant areas. Therefore, it is important to develop a system that helps in the early detection of cancer which can ultimately save more lives. In a TEDx 2014 talk, CEO and Founder of Enlitic Jeremy Howard said that if you detect cancer early, a person's probability of survival is 10 times higher. [3]

Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which takes in an input image, assigns importance (learnable weights and biases) to various aspects/objects in the image and is able to differentiate one from the other.[4] It can be used to predict which class the image belongs to. The architecture of a CNN is analogous to that of the connectivity pattern of neurons in the human brain. The role of the CNN is to reduce the images into a form which is simpler to process, without losing features which are critical for getting a good prediction.

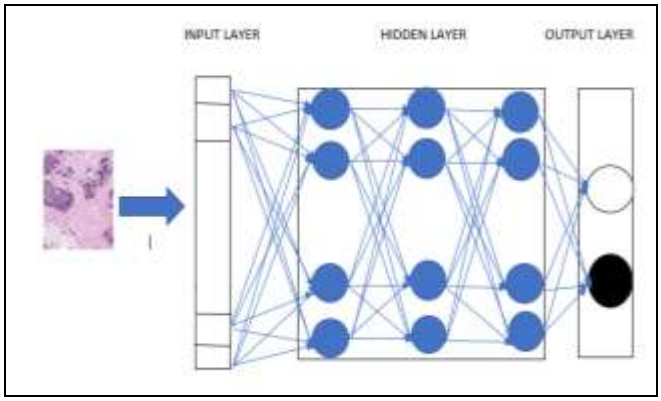

Fig -1: Convolutional Neural Network

2. LITERATURE SURVEY

Andre Esteva, et. Al. [5], the paper studies the development of Deep CNN (Convolutional Neural Network) and to match its image classification performance with the performance of the dermatologists. They used a GoogleNet Inception v3 CNN architecture that was pretrained on approximately 1.28 million images from the 2014 ImageNet Visual Recognition Challenge. Next, they trained it on the dataset using transfer

learning. They used Google's TensorFlow30 deep learning framework to train, validate and test the network. In this task, the CNN got 72.1 ± 0.9% (mean ± s.d.) overall accuracy (the average of individual inference class accuracies) and two dermatologists attained 65.56% and 66.0% accuracy on a subset of the validation set. The CNN achieved 55.4 ± 1.7% overall accuracy whereas the same two dermatologists attained 53.3% and 55.0% accuracy. The CNN was tested against atleast 21 dermatologists at keratinocyte carcinoma and melanoma recognition.

Saurabh Yadav [6] gives a brief history and introduction to using CNN in medical image analysis. The highlight of the article was the increasing research done in this field in the recent couple of years, starting from 2015. The use of CNNs has gained momentum today because of improved training algorithms in Deep Learning, introduction of GPUs etc. It can be used in classification, segmentation, detection, registration and Content based image retrieval (CBIR).

Zhang S, et. Al. [7], the research paper discusses the building of CNN using two datasets-Colorectal polyps and Lung Nodules. The dataset they used were relatively small (both less than 70 subjects) but both these datasets were pathologically proven datasets. The features used were: Local Binary Pattern (LBP) is a very famous and efficient texture operator. It labels the pixels of an image via thresholding the neighborhood of each pixel and considering the result as a binary number, Gradient Feature, histogram of oriented gradients (HOG). Two CNN models were constructed for the two datasets, Multi-Channel-Multi-Slice-2D CNN Model and Voxel-Level-1D CNN Model was constructed. The conclusion drawn was thus – The MCMS-2D CNN model achieved relatively good classification performance in small datasets, and the model is very consistent across different datasets. However, the V-1D CNN model worked better for the small and unbalanced dataset. In general, the results demonstrated that the proposed models have their own advantages on studying small datasets.

Salim Ouchtati, et. Al. [8], the research paper proposed a novel method for the classification of Magnetic Resonance (MR) images of the brain. The three processing steps that compose the proposed system have been described. The proposed system is composed of three steps: 1) Preprocessing step: The preprocessing operations are classical operations in image processing, they are used to "clean" and "prepare" the MR image of the brain for the subsequent processing steps. 2) Features Extraction step: In this paper, the method that used to extract the features is based on the calculating of the central moments of order k, this method was used in several studies of pattern recognition. 3) Brain tumors classification: This step consists to affect the brain image fed into the input of the proposed system to an appropriate class. The output layer neurons number is equal to the number of the classes brain tumor to be recognized, while the number of hidden layers and the number of neurons per hidden layer are determined by groping. The obtained results are very encouraging and very promising; the system arrives to properly affect 88.333% of the images of the databases. This analysis allows to identifying the reasons for the misclassification and therefore to propose the necessary solutions.

V. Vishrutha and M. Ravishankar [9], this research paper discusses how early detection of breast cancer benefits. It gives idea about an automated computer aided system that help radiologists in breast cancer detection and classification in digital mammograms. A tumour here can be classified as benign or malignant. It includes four stages starting from image pre-processing, ROI selection, feature extraction and selection and classification. The model achieves accuracy of 92% and is able to detect early stage breast cancer and classify them as benign and malignant using SVM Classifier.

Poorti Shani and Neetu Mittal [10], this research paper discusses about how this approach works best for early detection which can be lifesaving and make practitioners task easy. Here mammogram imaging technique and its grayscale conversion is done and then histogram analysis of original mammogram and MRI images followed by image segmentation process, edge detection to extract the tumorous portion out of rest of mammogram and MRI images. The entropy value of edge detection method and thresholding value of MRI are 0.0987 and 0.0763, respectively, and the values of edge detection method and thresholding value of MRI are 0.1002 and 0.040. The paper resulted in comparison of two modalities out of which edge detection method gives better results for MRI and mammogram images.

Maleika Heenaye- Mamode Khan [11], this paper discusses about efficient algorithms like ANN Artificial Neural Network, BNN Bayesian Neural Network developed to detect texture features or descriptor features that can detect the presence of abnormalities in the breast. Here the process involves the development of a breast cancer application, the four stages namely image capture, image enhancement and segmentation, feature extraction and selection and image classification have to be performed. Two databases on which working is done are the BCDR database and another online database is the mini-MIAS dataset. Thus, three modalities with three databases were compared like SVM, Bayesian and segmentation Technique on Mammograph, BCDR ultrasound, BCDR mammogram. And resulting SVM gives 100% sensitivity and specificity for BCDR ultra sound images.

Stanitsas, et. Al. [12] discuss active learning applied on CNN. They used Active learning for the selection of data samples to train CNN for Cancerous Tissue Recognition (CTR). They performed the experiment on three datasets namely - (i) Breast cancer, (ii) Prostate cancer, and (iii) Myometrium tissue samples. The results they obtained on these datasets showed that active learning is useful and leads to faster training of CNN. The absolute performance on the CTR datasets they obtained was 93.4%, 94.1% and 89.6% for

breast cancer, myometrial leiomyomas and prostate cancer respectively.

3. PROPOSED METHODOLOGY

Invasive Ductal Carcinoma (IDC) is the most common phenotypic subtype of all Breast cancers (BCa) comprising nearly 80% of them. [5] Breast cancer affecting women is known as high mortality unless diagnosed in time. So, our prime target is to classify whether the image has IDC type of breast cancer or not because it would help in automatic and early detection of breast cancer which can be helpful for complete removal of the cancer.

We will create an Image Classifier which can distinguish whether a patient has IDC subtype of breast cancer. We will use Convolutional neural networks algorithm for classification of the images. TensorFlow is also used as it is an end-to-end open source platform for machine learning. To achieve our goal, we will use one of the famous deep learning algorithms out there which is used for Image Classification i.e. Convolutional Neural Network (or CNN).

CNN has the following steps- convolution, pooling, flattening, full connection. We will apply the above steps to the histopathology images dataset.[13] Convolution is used is to extract the high-level features such as edges, from the input image. The next layer which is the pooling layer is responsible for reducing the spatial size of the Convolved Feature. Flattening, which is that subsequent layer converts all the resultant 2 dimensional arrays into a single long continuous linear vector. The fourth step is to create a fully connected layer, and to this layer we are going to connect the set of nodes we got after the flattening step, these nodes will act as the input layer to these fully-connected layers. The next step is to initialise our output layer, which should contain only one node, as it is binary classification. This single node will give us a binary output of whether a person is suffering from or not.

The next step is to fit our CNN to the image dataset that we have downloaded from the Kaggle website [13]. But before we do that, we separated the images into folders for training and testing. We used total 2000 images from the dataset for the purpose. The next step was to fit the model over the training data. After the training is complete, we made new predictions on the trained model. For this we used the test images. Then we are using predict() method on our classifier object to make the prediction. As the prediction will be in a binary form, we will be receiving either a 1 or 0, whether the histopathology image depicts IDC type of breast cancer or not.

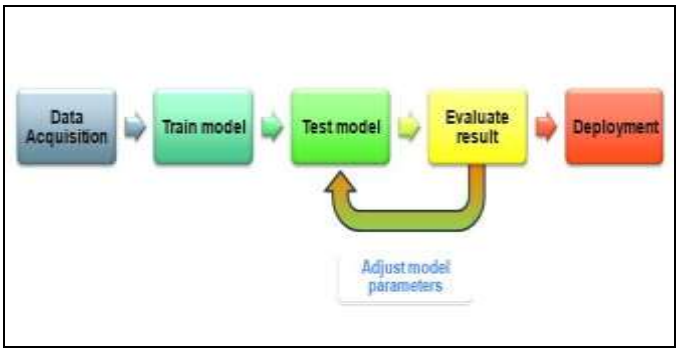

Fig -2: Block Diagram

4. RESULTS

This paper focuses on earlier diagnosis of breast cancer, as a detection brings about the success of about 93.75% accuracy by the use of CNN algorithm for classification. By calculating true positives, false positives, false negatives, true negatives we calculated a set of performance measures for IDC detection. The sensitivity is 95.66% and the specificity is 75.66%.

By the use of Tensorflow, an open-source software library, training and the computational time was reduced as compared to other models. This project thus ensures the greater level of detection of breast cancer at earlier stage, by which mortality rate of cancer affected person can be reduced and earlier diagnosis would increase the lifetime of a patient by giving them the right treatment at the right stage.

5. CONCLUSION

The paper studies various classification methods on different types of cancer and aims to classify breast histopathology images. The dataset used was breast histopathology images obtained from Kaggle website.[12] The system would make use of Convolutional Neural Network- a Deep learning model for the purpose of image classification. Building the model with good accuracy will help in providing automatic and easy diagnosis of breast cancer. A model will give better results in Breast cancer image classification and will help in the medical domain to provide cost-effective diagnosis of breast cancer. Thus,the paper aims to present the system to create an Image Classifier which would be able to distinguish whether a patient has Invasive Ductal Carcinoma (IDC) a common subtype of breast cancer or not.

6. REFERENCES

[1] https://www.hopkinsmedicine.org/breast_center/

[2] https://www.breastcancer.org/symptoms/types/idc/

[3] https://emerj.com/ai-sector-overviews/deep-learning-in-oncology/

[4] https://medium.com/@priyankapatel2205/convolutional-neural-nets-1813eee0510

[5] Andre Esteva, Brett Kuprel, Roberto a. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau and Sebastian Thrun, "Dermatologist-level classification of skin cancer with deep neural networks", Nature vol 542.

[6] Brief Intro to Medical Image Analysis and Deep Learning, https://medium.com/@saurabh.yadav919/brief-intro-of-medical-image-analysis-and-deep-learning.

[7] Zhang S, Han F, Liang Z, Tan J, Cao W, Gao Y, Pomeroy M, Ng K, Hou W, "An Investigation of CNN Models for Differentiating Malignant from Benign Lesions Using Small Pathologically Proven Datasets", Computerized Medical Imaging and Graphics (2019), doi: https://doi.org/10.1016/j.compmedimag.2019.101645.

[8] Salim Ouchtati, Jean Sequeira, Belmeguenai Aissa, Rafik Djemili, Mohamed Lashab "Brain Tumors Classification From MR images Using a Neural Network and the Central Moments", 2018 IEEE.

[9] Vishrutha and M. Ravishankar, "Early Detection and Classification of Breast Cancer "– Vol. 1, Advances in Intelligent Systems and Computing 327, DOI: 10.1007/978-3-319-11933-5_45.

[10] Poorti Sahni and Neetu Mittal," Breast Cancer Detection Using Image Processing Techniques", Springer Nature Singapore Pte Ltd. 2019.

[11] Maleika Heenaye- Mamode Khan, "Automated Breast Cancer Diagnosis using Artificial Neural Network ",2017 IEEE.

[12] Stanitsas, Panagiotis & Cherian, Anoop & Truskinovsky, Alexander & Morellas, Vassilios & Papanikolopoulos, Nikolaos. (2017) "Active convolutional neural networks for cancerous tissue recognition" 1367-1371. 10.1109/ICIP.2017.8296505.

[13] https://www.kaggle.com/paultimothymooney/breast-histopathology-images