BREAST CANCER DETECTION USING CNN FOR MAMMOGRAM IMAGING SYSTEM

Rashmi Kolambe[1], Dhanashree Dashpute[2], Chaitanya Shah[3], Dr.S.R.Ganorkar[4]

[1] *Student, Electronics & Telecommunication, Sinhgad College Of Engineering Pune ,Maharashtra, India*
[2] *Student, Electronics & Telecommunication, Sinhgad College Of Engineering Pune ,Maharashtra, India*
[3] *Student, Electronics & Telecommunication, Sinhgad College Of Engineering Pune ,Maharashtra, India*
[4] *HOD, Information Technology, Sinhgad College Of Engineering Pune, Maharashtra , India*

ABSTRACT

Breast cancer detection using CNN for mammogram imaging system is proposed to classify mammogram images into normal, benign (non-cancerous abnormality), and malignant (cancerous abnormality). Breast Cancer Detection Using Convolutional Neural Networks (BCDCNN) is aimed to speed up the diagnosis process by assisting specialists in the diagnosis and classification the breast cancer. A series of mammogram images are used to carry out pre-processing to convert a human visual image into a computer vision image and adjust suitable parameters for the CNN classifier. After that, all changed images are assigned to the CNN classifier as a training source. The CNN classifier will produce a system to recognize the mammogram image. By comparing the BCDCNN method with Mammogram Classification Using Convolutional Neural Networks (MCCANN), BCDCNN has improved the accuracy of classification on mammogram images. Therefore, the results show that the proposed method has higher accuracy.

Keyword: - *Breast cancer, Mammogram, Convolution Neural Network, Classification, Diagnosis, Image.*

1. INTRODUCTION

Breast Cancer is one of the major causes of female deaths worldwide. Cancer is a disease raised by the changes occurred in cells and spread uncontrollable. Mostly cancer cells form a lump or mass which is called a tumour and is named after the part of the body in which it originates. Breast cancer usually produces no pain at its early stage when it is easily treated, that's why screening is important for early detection.

Diagnostic medical imaging is the most fundamental approach in the act of the current solution. There are a lot of methods to obtain the medical image, such as Magnetic Resonance Imaging (MRI), mammography, X-ray, and Ultrasound. Normally, mammography and MRI are used for breast cancer diagnosis. Besides, the mammogram is a fast procedure that takes only about 20 minutes. The whole process is just an extremely small measure of radiation exposure from a mammogram which is safe than other treatments. In the present-day scenario, to observe breast cancer mammograms are used and they are known to be the most effective scanning technique. In this paper, the detection of cancer cells is done by the CNN technique. A new approach to of image recognition system for breast cancer detection is proposed This image recognition system is using convolutional neural network, which is a type of ANN and it is designed for to recognize visual imagery. This system is used to classify and detects abnormalities in mammograms image. In general, a mammogram image can be classified as cancerous or non-cancerous.

A series of mammogram images are used to carry out preprocessing to convert a human visual image into a computer vision image and adjust suitable parameters for the CNN classifier. After that, all changed images are assigned to the CNN classifier as a training source. The CNN classifier will produce a system to recognize the mammogram image. By comparing the BCDCNN method with Mammogram. Classification Using Convolutional Neural Networks (MCCANN), BCDCNN has improved the accuracy of the classification on mammogram

images. Therefore, the results will show higher accuracy.

Literature Review

[1] Siyabend Turgut et al. (IEEE 2018)"Microarray Breast Cancer Data Classification Using Machine Learning Methods" The paper uses microarray breast cancer data for classification of the patients using machine learning methods. In the first step, eight different machine learning algorithms are applied to the dataset and the results of classification were data. Then in the second step, two different feature selection methods such as Recursive Feature Elimination (RFE) and Randomized Logistic Regression (RLR) were applied on the microarray breast cancer dataset and 50 features were chosen as stop criterion. Again, the same eight machine learning algorithms were applied on the modified dataset. The results of the classifications are compared with each other and with the results of the first step. The methods applied are SVM, KNN, MLP, Decision Trees, Random Forest, Logistic Regression, Ad boost and Gradient Boosting Machines.

[2] Saria Eltalhi1 , Huda Kutrani 2 (April 2019) "Breast Cancer Diagnosis and Prediction Using Machine Learning and Data Mining Techniques"Breast cancer is the second cause of death among women. Data mining and machine learning have been widely used in the diagnosis of breast cancer and on the early detection of breast cancer. The aim role of this research is to review of machine learning and data mining techniques in breast cancer detection and diagnosis. The studies that concentrated on diagnoses and prognoses breast cancer using WEKA tool.

[3]Sanjana Balasubramanian(April 2021) "Breast cancer prediction using machine learning" The project uses Mammography which is the main test used for screening and early diagnosis, and its analysis and processing are the keys to improving breast cancer prognosis. To detect breast cancer in mammogram, image segmentation is performed with the help of Fuzzy C-means (FCM) technique. Further those segmented regions features are extracted, and it is trained completely, finally trained images are classified by the efficient classifier of different classes in mammogram. Texture features are extracted using a feature extraction technique like Multi-level Discrete Wavelet Transform, Principal Component Analysis (PCA), Gray-level Co-occurrence Matrix (GLCM). Morphological operators are used to distinguish masses and microcalcifications from the background tissue and KNN algorithm is used for classification.

[4] Srwa Hasan Abdulla 1, a Ali Makki Sagheer 2, b , Hadi Veisi 3,c (September 2021) "Breast Cancer Classification Using Machine Learning Techniques". Artificial intelligence (AI) has been used for easy diagnosis, rapidly, and accurately breast tumors. Machine learning algorithms such as Support Vector Machine (SVM), K Nearest Neighbour (K-NN), and Random Forest (RF) are used to classify medical images into malignant and benign. Moreover, deep learning has been employed recently for the same purpose, among them, Convolutional Neural Network (CNN) is one of the most popular techniques.

[5]Manav Mangukiya1 , Anuj Vaghani2 , Meet Savani3 (February 2022)"Breast Cancer Detection with Machine Learning".In this paper, Data Visualization and performance comparisons between different machine learning algorithms: Support Vector Machine (SVM), Decision Tree, Naive Bayes (NB), K Nearest Neighbours (k-NN), Adaboost, XGboost and Random Forest conducted on Wisconsin breast cancer Dataset. The main role is to evaluate the accuracy in the classification of data in classes of efficiency and effectiveness of each algorithm in classes of accuracy, precision, sensitivity and specificity. Our aim is to review various Techniques To detect early, efficiently and accurately Using Machine Learning. Experimental results show that XGboost offers the highest accuracy.

2. METHODOLOGY

Proposed block diagram of Breast Cancer detection Using CNN for Mammogram Imaging System is shown in Fig.1.

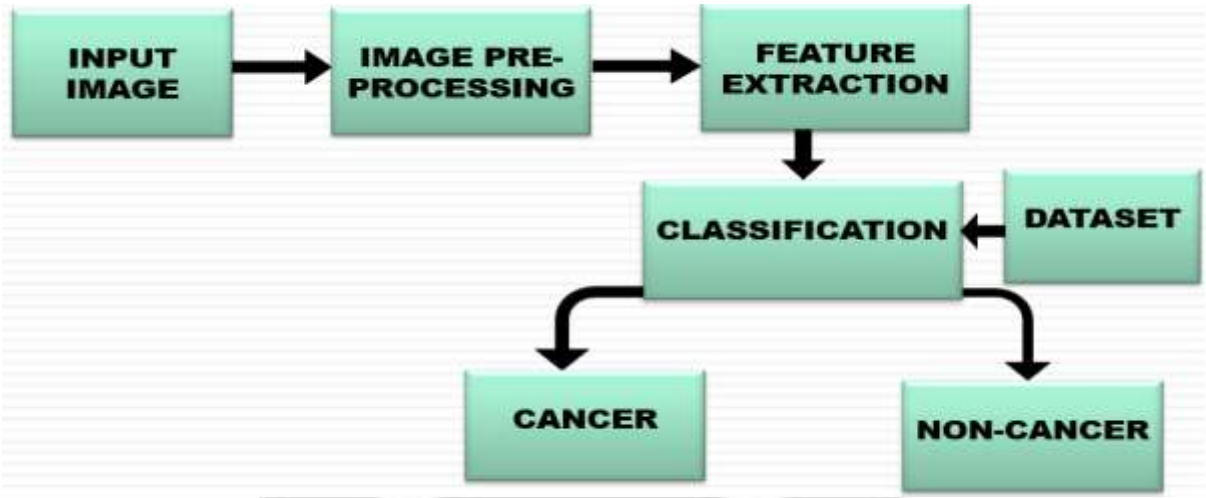

Fig .1:- Block diagram of the Breast Cancer detection Using CNN for Mammogram Imaging System

2.1 Mammogram

A mammogram is an X-ray of the breast. For many women, mammograms are the best way to find breast cancer early, when it is easier to treat and before it is big enough to feel or cause symptoms. Having regular mammograms can lower the risk of dying from breast cancer.

2.2 Image Pre-processing

The next stage is data pre-processing that contains removing redundant and irrelevant data that can improve hugely from the performance of ML algorithms. The raw data pixel size is too large. Huge segment of time is spent to adjust the input data for the classifier. Accordingly, most breast cancer tissue size is less than 20% percent from a 1024 x1024 pixel raw data, so it will affect the accuracy of the classifier performance and computation time. For a large pixels data as input of CNN classifier, it will spend a lot of computation time for every not necessary information or not important noise. At the same time, it also will affect the last result for the classifier

2.3 Segmentation

As mentioned in the introduction, different imaging modalities are used by radiologists. However, the most techniques that are used for classifying breast cancer are ultrasound imaging and mammograms. The former imaging technique is low contrast and blurry boundaries that impact automatic segmentation. In the contrast, the later imaging technique is high resolution, low energy x-ray and it aids in discovering abnormalities in tissue cell.

For mammography images, segmentation is applied to extract the interest region from the image and remove any noise such as lesions of the breast tumor, pectoral muscles, and any region that does not belong to the breast.

2.4 Features extraction

Features extraction refers to a technique that minimizes the number of features by generating a new set of attributes with the same information as the old ones. Working with a huge dataset with hundreds or thousands of features without extracting the most ones that represent the actual observations about given variables could lead to overfitting in a model of ML. Therefore, applying this technique reduces the risk of overfitting and increases the performance of the machine learning model. In other words, the purpose of features extraction is to discard the original features by proposing new ones that summarize most of the information of the old characteristics in the dataset. In addition, this technique increases the speed of training, accuracy, and enables visualization.

2.5 Features selection

Effective classification scheme also depends highly on selecting features technique, which reduces in a number of features and also ranks them from most important one to least important ones. This reduction possibly gives many benefits through statistical analysis (e.g., improving accuracy, reducing the risk of overfitting, increasing training

speed, the possibility of data visualization). There are several methods that can be utilized to apply feature selection (i.e., filter method, wrapper method, and embedded method). The former includes selecting a subset from the dataset that contains only relevant features by using filtering methods such as Pearson Correlation. The second method is more accurate than the previous one as it uses machine learning algorithms for evaluating features, but it needs more processing time. The technique involves adding and removing features based on the performance of the model.

2.6 Dataset

At this stage, a dataset is selected to train a model firstly and then evaluate it secondly. There are various datasets for breast cancer tumors that are available globally, but the most popular ones are Wisconsin Breast Cancer dataset (WBCD), Wisconsin Diagnostic Breast Cancer (WDBC), Digital Database for Screening Mammography (DDSM), and Mammographic Image Analysis Society (MAIS). Each dataset include necessary information for processing and modelling, for instance, the second dataset contains the ID for a patient, features, and diagnosis. The ID refers to the identification number for the patient, the number of features are 10 that are computed from a digital image of a breast mass, and the diagnosis for the patient (i.e., positive or negative). The number of patients are 569, 357 of them are diagnosed as benign and 212 are classified as malignant.

2.7 Software Requirement

Python: - Python is a computer programming language often used to build websites and software, automate tasks, and conduct data analysis. Python is a general-purpose language, meaning it can be used to create a variety of different programs and isn't specialized for any specific problems.

Python includes a modular machine learning library known as PyBrain, which provides easy-to-use algorithms for use in machine learning tasks. The best and most reliable coding solutions require a proper structure and tested environment, which is available in the Python frameworks and libraries.

2.7.1 Libraries Required

OpenCV-Python:- Is a library of Python bindings designed to solve computer vision problems. cv2.imwrite() method is used to save an image to any storage device. This will save the image according to the specified format in current working directory.cv2 (old interface in old OpenCV versions was named as cv) is the name that OpenCV developers chose when they created the binding generators It is used in various applications such as face detection, video capturing, tracking moving objects, object disclosure, nowadays in Covid applications such as face mask detection, social distancing.

PIL :- It is stand for Python Imaging Library it is a free and open-source additional library for the Python programming language that adds support for opening, manipulating, and saving many different image file formats. PIL can perform tasks on an image such as reading, rescaling, saving in different image formats. PIL can be used for Image archives, Image processing, Image display.

Tkinter :- Tkinter is the standard GUI library for Python. Python when combined with Tkinter provides a fast and easy way to create GUI applications. Tkinter provides a powerful object-oriented interface to the Tk GUI toolkit. Creating a GUI application using Tkinter is an easy task. All you need to do is perform the following steps −

1. Import the *Tkinter* module.
2. Create the GUI application main window.
3. Add one or more of the above-mentioned widgets to the GUI application.
4. Enter the main event loop to take action against each event triggered by the user.

Tensorflow :-

TensorFlow is a software tool . It is an artificial intelligence library that allows developers to create large-scale multi-layered neural networks. It is used in Classification, Recognition, Perception, Discovering, Prediction, and Creation, etc. Some of the primary use cases are Sound Recognition, Image recognition, etc.

3. RESULT AND DISCUSSION

Proposed system is implented using Tenserflow with convolutional neural network(CNN) architecture. Fig 2,3 and 4 shows the result of proposed system .

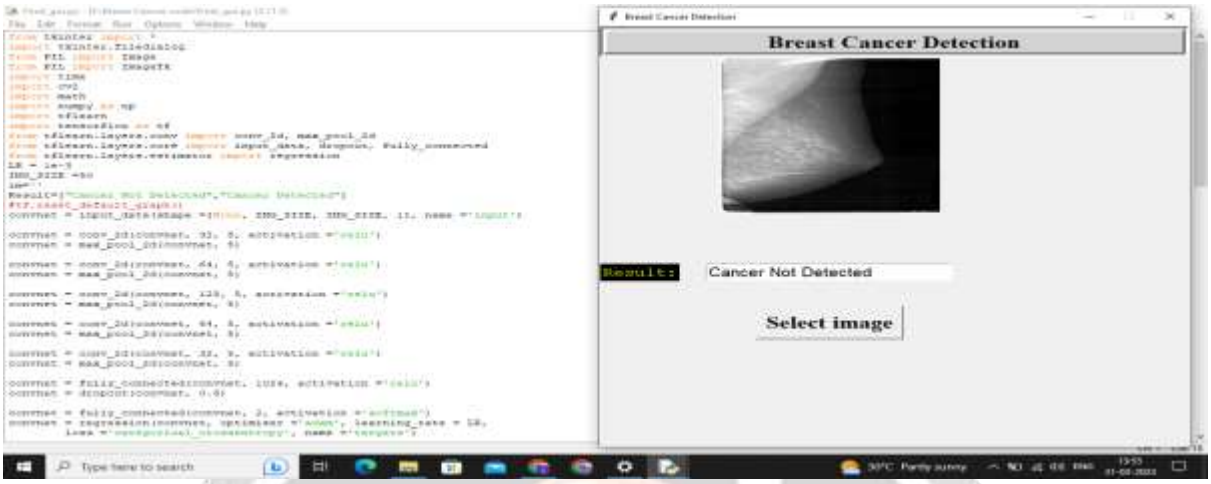

Fig no.2 :- Cancer is not detected

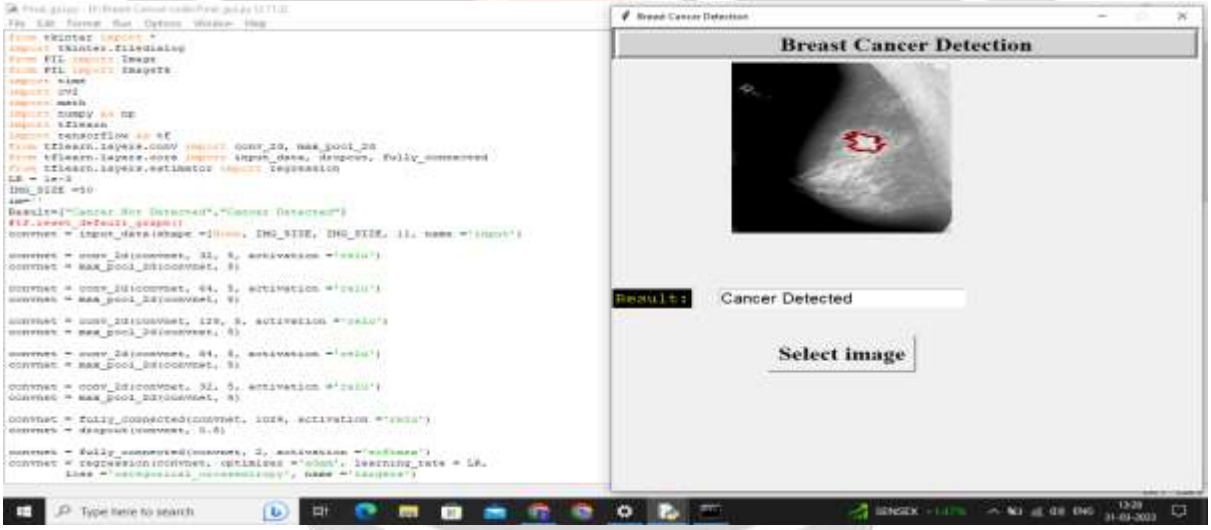

Fig no.3:- Cancer is detected

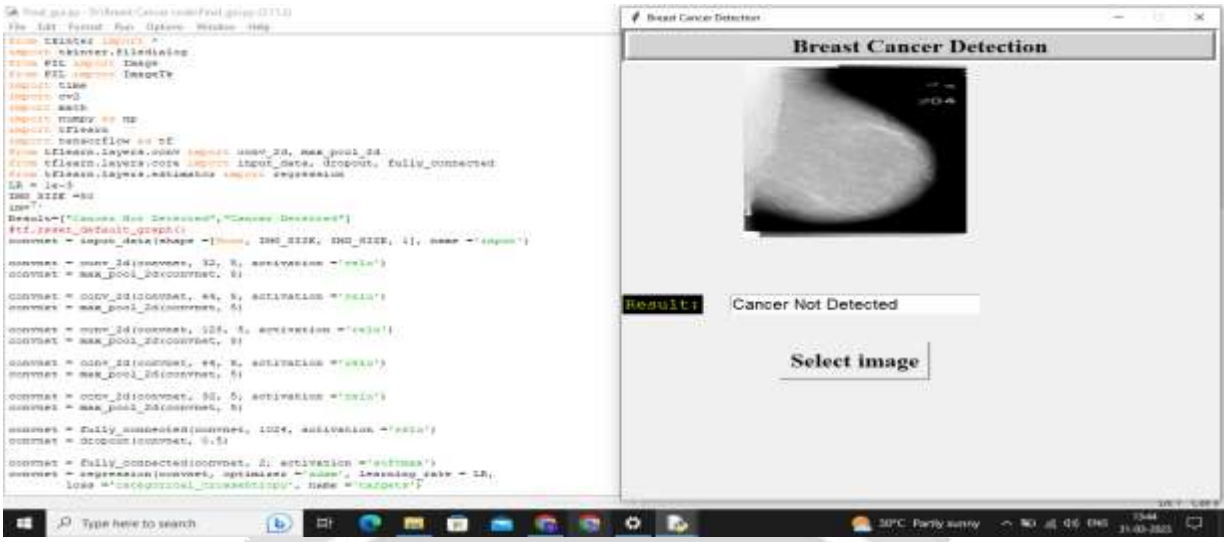

Fig no.4:-Cancer not detected

Python language is used for Breast cancer detection. We have used Tensorflow library for recognition, also CNN Architecture for training data. A system that uses the mammogram image dataset for training process and then we use it for cancer detection.

4. PERFORMANCE PARAMETER

Accuracy:

The accuracy of the system's cancer detection and classification, as measured by the percentage of correct predictions made by the model.

The model was trained on a dataset containing a total of 270 images, consisting of both cancerous and non-cancerous images. Of the 270 images, 97 were cancerous images, and the remaining 89 were non-cancerous images. During the evaluation process, the model achieved a detection accuracy of 85% for cancer images and 90% for non-cancer images.

For cancer images, out of the 10 images tested, 8 were detected accurately, while 2 was misclassified as a non-cancer, indicating a potential issue with the model's classification capabilities..

For non-cancerous images, out of the 10 images tested, 7 were detected accurately, and 3 were not detected. This indicates a relatively high detection accuracy rate for cancer images, but there is still room for improvement, particularly in detecting the 2 missed images.

5. CONCLUSIONS

Breast cancer is one of the major causes of female deaths worldwide. Cancer is a disease raised by the changes occurred in cells and spread uncontrollable. When the early symptoms of breast cancer are ignored, the patient might die with major effect in her health and can lead to death. Breast cancer can be in under control when it will be detected early. Many studies focus mainly on the application of classification techniques to breast cancer prediction. In conclusion, the breast cancer detection by using Convolutional Neural Network had been successfully developed and tested with mammogram images. This system will provide a fast diagnosis time and high accuracy system.

6. REFERENCES

[1] Siyabend Turgut et al. (IEEE 2018)“Microarray Breast Cancer Data Classification Using Machine Learning Methods”

[2] Saria Eltalhi1 , Huda Kutrani 2 (April 2019) “Breast Cancer Diagnosis and Prediction Using Machine Learning and Data Mining Techniques”

[3]Sanjana Balasubramanian(April 2021) “Breast cancer prediction using machine learning”

[4] Srwa Hasan Abdulla 1, a Ali Makki Sagheer 2, b , Hadi Veisi 3,c (September 2021) “Breast Cancer Classification Using Machine Learning Techniques”

[5]Manav Mangukiya1 , Anuj Vaghani2 , Meet Savani3 (February 2022)“Breast Cancer Detection with Machine Learning”

BIOGRAPHIES

	Rashmi Kishor Kolambe Student, Electronics & Telecommunication, Sinhgad College Of Engineering Pune ,Maharashtra, India
	Dhanashree Satish Dashpute Student, Electronics & Telecommunication, Sinhgad College Of Engineering Pune ,Maharashtra, India
	Chaitanya Rameshchandra Shah Student, Electronics & Telecommunication, Sinhgad College Of Engineering Pune ,Maharashtra, India
	Dr.S.R. Ganorkar HOD of Information Technology, Sinhgad College Of Engineering Pune, Maharashtra, India