International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1467-1472 www.ijaem.net ISSN: 2395-5252

Breast Cancer Detection Using Deep Learning

K.S.DemelAbhisha M.E, K Roopini, Roshni T,P Swathimithra
Dhanalakshmi college of Engineering, Tambaram, Chennai-601301, Tamil Nadu India

Date of Submission: 20-04-2023 Date of Acceptance: 30-04-2023

ABSTRACT
A computer-aided diagnosis (CAD) system based on mammograms enables early breast cancer detection, diagnosis, and treatment. However, the accuracy of the existing CAD systems remains unsatisfactory. This paper explores a breast CAD method based on features fusion with Deep convolutional Neural Network (DCNN)deep features. First, we proposed a mass detection method based on CNN deep features and unsupervised extreme learning machine (ELM) clustering. Second, we build a feature set fusing deep features, morphological features, texture feature, and density feature. Third an ELM classifier is developed using fused features set to classify Benign and Malignant Breast masses. Extensive experiments demonstrate the accuracy and efficiency of our proposed mass detection and Breast cancer classification method.

INDEX TERMS: Mass Detection, Computer Aided Diagnosis,Deep Learning, Fusion Features, VGGNET, Deep convolutional Neural Network.

I. INTRODUCTION:

Breast cancer is a serious threat to women life and health, and the morbidity and mortality of breast cancer ranked first and second out of all female diseases. Earlydetection of lumps can effectively reduce the mortality rate of breast cancer. The mammogram is widely used in early screening of breast cancer due to its relatively low expense and high sensitivity to minor lesions. In the actual diagnosis process,however, the accuracy can be negatively affected by many factors, such as radiologist fatigue and distraction, the complexity of the breast structure, and the subtle characteristics of the early-stage disease [4], [5]. The computer-aided diagnosis (CAD) for breast cancer can help address thisissue.

The classical CAD for breast cancer contains three steps:
(a) finding the Region of Interest in the preprocessed mammogram, and hence locating the region of the tumor. (b) then, extracting features of the tumor based on expert knowledge, such as shape, texture, and density, to manually generate feature vectors. (c) finally, diagnosing benign and malignant tumors by classifying these feature vectors [6], [7]. Although the classical diagnosis method has been commonly used, its accuracy still needs to be improved [8]. The quality of the handcrafted feature set directly affects the diagnostic accuracy, and hence an experienced doctor plays a very important role in the process of manual feature extraction. The commonly used features, including morphology, texture, density and other characteristics are manual set, which are obtained based on doctors ‘experience, that is, subjective features. In recent years, deep learning methods, such as the convolutional neural network (CNN), that can extract hierarchical features from image data without the manual selection, which is also called objective features, have been successfully applied with a great improvement on accuracies in many applications, such as image recognition, speech recognition, and natural language processing [9], [10]. There are some shortcomings in either subjective or objective features Subjective features ignore the essential attributes of images, while objective features ignore artificial experience. Therefore, the subjective and objective features are fused so that these features can reflect the essential properties of the image as well as the artificial experience. Meantime, Extreme Learning Machine (ELM) has better classification effect on multi-dimensional features than other classifiers including SVM, decision tree, etc., based on our previous research.

II. RELATED WORKS:

The research efforts related to breast cancer CAD mainly focus on the detection and diagnosis of breast tumor. This section briefly summarizes existing works related to these two aspects. In the aspect of breast tumor detection, Sun et al. [11] proposed a mass detection method, where an adaptive fuzzy C-means algorithm for segmentation is employed on each mammogram of the same breast. A super- vised artificial neural network is used as a classifier to judge whether the segmented area is a tumor. Sai din et al. [12] employed pixels as an alternative feature and used a region growing method to segment breast tumor in the mammogram. Xu et al. [13] proposed an improved watershed algorithm. They first make a coarse segmentation of breast tumor, fol- lowed by the image edge detection via combining regions that has similar gray-scale mean values. Hu et al. [14] pro- posed a novel algorithm to detect suspicious masses in the mammogram, in which they utilized an adaptive global and local thresholding segmentation method on the original mammogram. Yap et al. [15] used three different deep learning methods to detect lesion in breast ultrasound images based on a Patch-based LeNet, a U-Net, and a transfer learning approach with a pretrained FCN-AlexNet, respectively.

In the aspect of differentiating benign and malignant breast tumors, Kahn, Jr., et al. [16] created a Bayesian network that utilizes 2 physical features and 15 manually marked probabilistic characteristics to conduct the computer-aided diagnosis for breast cancer. Wang et al. [17] utilized ELM to classify features of breast tumors and compare results with SVM classifier. Qiu et al. [18] applied CNN to the risk prediction of breast cancer by training CNN with a large amount of time series data. Sun et al. [19] also used deep neural networks to predict the risk of breast cancer in the near term based on 420 time series records of mammography. Jiao et al. [20] proposed a deep feature-based framework for breast masse classification, in which CNN and a decision tree process are utilized. Arevalo et al. [21] used CNN to abstract representations of breast tumor and then classified the tumor as either benign or malignant. Carneiro et al. [22] proposed an automated mammogram analysis method based on deep learning to estimate the risk of patients of developing breast cancer. Kumar et al. [23] presented an image retrieval system using Zernike moments (ZMs) for extracting features since the features can affect the effectiveness and efficiency of a breast CAD system. Aličković and Subasi [24] proposed a breast CAD method, in which genetic algorithms are used for extraction of informative and significant features, and the rotation forest is used to make a decision for two different categories of subjects with or without breast cancer.

METHODS:

In this paper, we consider the following five steps in breast cancer detection: breast image preprocessing, mass detection, feature extraction, training data generation, and classifier training. In the breast image preprocessing, denoising and enhancing contrast processes on the original mammogram have been utilized to increase the contrast between the masses and the surrounding tissues.

1. Image processing
2. Pre-processing
3. Feature Extraction
4. Segmentation
5. Prediction

IMAGE PROCESSING:

There are several preprocessing methods. The adaptive mean filter algorithm is selected to eliminate noise on the original mammograms in order to avoid the impact of noise on subsequent auxiliary diagnosis. The main idea is to use a fixed size window sliding in the line direction age, calculating the mean, variance, and spatial correlation values of each sliding window to determine whether the window contains noise. If the noise is detected, replace the pixel values of the selected window with the mean value. In this paper, a contrast enhancement algorithm [30] has been used to increase the contrast between the suspected masses and surrounding tissues. The main idea is to trans- form the histogram of the original image into uniformly distributed. After this process, the gray scale of the image is enlarged, thus the contrast has been enhanced and the image details become clearer.

Computerized Tomography images is used as input.
It takes image from mammographic image Analysis society mammograms dataset.
It determines the size, Quality and nature of the image.

PREPROCESSING:

The aim of preprocessing is an improvement of the image data that suppresses unwanted distortion.
It is the most important step in mammogram analysis.
It increase quality of the image for further steps.
There are many filtering Techniques available for preprocessing.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1467-1472 www.ijaem.net ISSN: 2395-5252

Techniques used in preprocessing:
This filter used to improve quality remove noise, preserves the edge within an image.
The filters used is Average Filter, Adaptive median filter, Wiener filter.

FEATURE EXTRACTION:
Feature Extraction is a part of reduction process in which an initial set of raw data is divided and reduced to more manageable groups.
Feature extraction uses group of pixels
Each pixel is used to classify image.
Extract the features of image here features refers to number of columns in the image. The image will be in the map format.

SEGMENTATION:
Image segmentation is the process by which a digital image is partitioned into various subgroups. Objects, which can reduce the complexity of image and thus analysis the image becomes simpler.

In this dynamic threshold technique is used to optimize the number of features which can achieve the maximum classification of accuracy rate.

PREDICTION:
In prediction stage the stages of breast cancer is predicted. In this stage whether the breast cancer is in Normal or Benign, or Insitu or Invasive.

EXISTING SYSTEM
The existing model uses mammogram images as input. The first step in the existing system is the detection of masses from the given image. The techniques that have been employed are based on CNN deep features and extreme learning clustering. Mass detection includes the following steps.

- Extract the Region of Interest
- Segment the region in to several sub-regions
- Traverse through the sub region and extract the deep features
- Cluster the sub region
- Find the boundary.

The next step is the creation of feature set. The feature set is built by fusing the following feature.

- Deep features
- Morphological features
- Texture features
- Density features.

The features are extracted with the CNN. The CNN model contains 7 layers.

- 3 convolutional layers
- 3 Max Pooling Layer

1 Fully connected Layer

After the features are built, an ELM classifier is used for classifying the image in to benign or malignant. The ELM algorithm is a single hidden layer feed-forward neural network.

There are 400 mammograms in the image dataset used in the existing model, which contains 200 malignant mass images and 200 benign mass images. These images are generated using the Stenograph 2000D all-digital mammography camera from 32 to 74 years old female patients. The locations of the masses of all images have been marked by the professional doctor, and the diagnosis results are also confirmed by the pathologist. It has been observed that the model used produces good results in each of the processes when compared with the baseline models.

III. PROPOSED SYSTEM

Before implementing the proposed convolutional neural network models, a set of base line models is chosen and the breast cancer classification is tested with those models. the models that has been chosen as the base line models are random forest, K-Nearest Neighbour and Gradient boosting. When the implementation of the proposed model is made comparison will also be extended to the convolutional neural networks.

The proposed model is a kind of deep convolutional networks. The dataset used in the proposed model is MIAS Mammogram dataset. It contains 323 images in the portable gray map format.

The following details could be extracted from the image set and could be used to build the classifier.

1. The characteristics of the background tissue can be any one of the following.

- F Fatty
- G Fatty-glandular
- D Dense-glandular

2. The class of abnormality present could be any one of the following .

- CALC- Calcification
- CIRC -Well-defined/circumscribed masses
- SPIC - Spiculated masses
- MISC- Other, ill-defined masses
- ARCH - Architectural distortion
- ASYM - Asymmetry
- NORM -Normal

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1467-1472 www.ijaem.net ISSN: 2395-5252

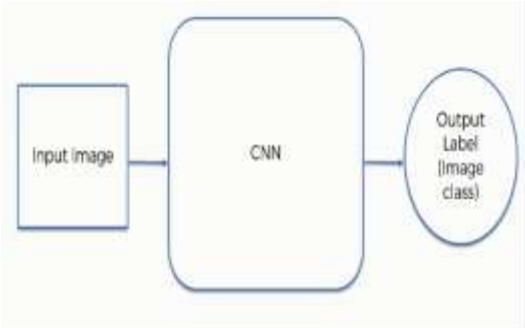
3. Severity of abnormality;
- B Benign
- M Malignant
4. x,y image-coordinates of centre of abnormality.
5. Approximate radius (in pixels) of a circle enclosing the abnormality.

The ideology of the proposed model is to use VGGNET and separable convolutional neural networks for classifying the breast cancer. The architecture of the proposed model is given in the following section.

CNN ALGORITHM:

The algorithm called Convolutional Neural Network Improvement for Breast Cancer Classification (CNNI-BCC) is presented by the authors to assist medical experts in breast cancer diagnosis in timely manner. The CNNI-BCC uses a convolutional neural network that improves the breast cancerlesion classification in order to help experts for breast cancer diagnosis. CNNI-BCC can classify incoming breast cancer medical images into the malignant, benign, and healthy patients. datasetthat has been used in this work is MIAS mammogram images. Convolutional neural network consists of 1 input layer, 28 hidden layers and 1 output layer. The input layer is used original size abnormally cancer images in RGB channels. A convolutional layer is located at first hidden layer for feature map extraction. A dropout rate of 0.5, a learning rate of 0.002, fully connected hidden layer of 1024 neurons, and Rectified Linear Units (ReLUs). Data augmentation is performed on the image patches to overcome the over fitting issue faced by other researchers. ReLUs is applied as nonlinear layers in the fully connected layer section. The process is initialized with annotations and over patches. The pre-processing module of the work includes the minimization of the size of the input mammogram to reduce the computational burden during segmentation phase. The air–skin interface and nipple of the breast are detected using fuzzy c-means clustering technique. Single view approach detects and characterizes suspicious lesions on craniocaudal (CC) and mediolateral oblique (MLO) view separately using geometric and textural features. Lesions detected on each view are paired with potential lesions on another view.

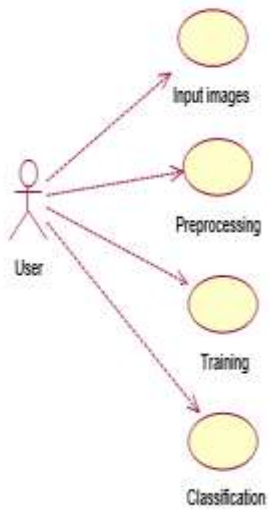
The algorithm proposed by the authors computes the correspondence score of each lesion pair. Single view information is fused with two views correspondence score to discriminate malignant tumours from benign masses using the SVM classifier.



Given an input image the convolutional neural network would classify the image and outputs the label of the given image. The system can be trained with any set of images. it can be either animals, traffic signals, humans etc.

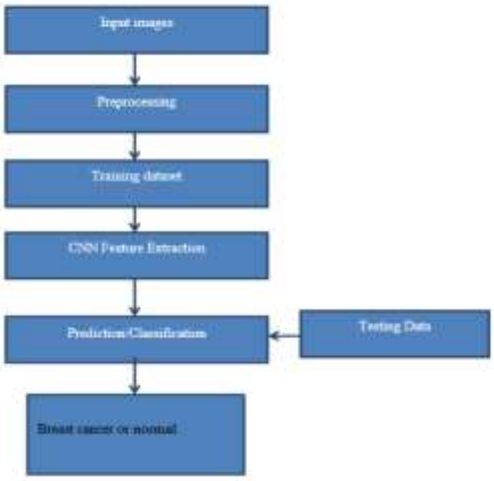
As an initial step, the black and white and colour images are represented in different formats by the convolutional neural networks. Each pixel of the black and white image is represented as a value between 0 to 255. Where "0" represents black and "1" represents white. The values in between are grey shades. While in case of the colour images, each pixel will be represented in RGB Format. With this representation, the steps followed in the convolutional neural networks are explained below.

USE CASE DIAGRAM:

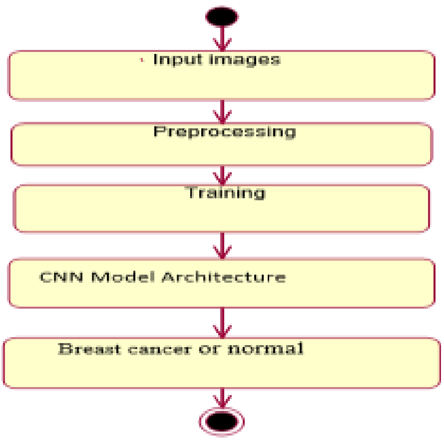


International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1467-1472 www.ijaem.net ISSN: 2395-5252

DATA FLOW DIAGRAM:



CTIVITY DIAGRAM:



CLASS DIAGRAM:

SOFTWARE REQUIREMENTS:

- Integrated Development Environment is Anaconda
- Working Environment is Jupiter and spider:
- Language is Python 3.6.5 (64 bit)
- Operating system is windows

HARDWARE REQUIREMENTS:

- Processor Intel core i3 2.40 GB
- RAM 4 GB

IV. CONCLUSION:

This paper proposes a breast CAD method based on deep Convolution Neural Network. Its main idea is to apply deep features extracted from CNN to the two stages of mass detection and mass diagnosis. In the stage of mass detection, a method based on sub-domain CNN deep features and VGGNET clustering is developed. In the stage of mass diagnosis, an VGGNet classifier is utilized to classify the benign and malignant breast masses using a fused feature set, fusing deep features, morphological features, texture features, and density features. In the process of breast CAD, the choice of features is the key in deter- mining the accuracy of diagnosis. In previous studies, either traditional subjective features or objective features are used, in which traditional subjective features include morphology, texture, density, etc., and objective features include features extracted from CNN or DBN. These features are flawed to some extent. In this paper we combine subjective and objective features, taking the doctor's experience and the essential attributes of the mammogram into account at the same time. After extracting the features, the classifier is used to classify the benign and malignant of the breast mass. In this paper, ELM, which has a better effect on multi-dimensional feature classification, is selected as the classifier. Through the experiments using breast CAD of 400 cases of female mammograms in the northeastern China, it demonstrates that, in mass detection and mass diagnosis, our proposed methods outperform other existing methods.

REFERENCE:

[1]. R. L. Siegel et al., ''Colorectal cancer statistics,'' CA, Cancer J. Clinicians, vol. 64, no. 2, pp. 104–117, Mar. 2014

[2]. J. B. Harford, ''Breast-cancer early detection in low-income and middle-income countries: Do what you can versus one size fits all,'' Lancet Oncol., vol. 12, no. 3, pp. 306–312, Mar. 2011.

[3]. C. Lerman et al., ''Mammography adherence and psychological distress

among women at risk for breast cancer,'' J. Nat. Cancer Inst., vol. 85, no. 13, pp. 1074–1080, Jul. 1993.

[4]. P. T. Huynh, A. M. Jarolimek, and S. Daye, ''The false-negative mammo- gram,'' Radiographics, vol. 18, no. 5, pp. 1137–1154, Sep. 1998.

[5]. M. G. Ertosun and D. L. Rubin, ''Probabilistic visual search for masses within mammography images using deep learning,'' in Proc. IEEE Int. Conf. BioInform. Biomed. (BIBM), Nov. 2015, pp. 1310–1315.

[6]. S. D. Tzikopoulos, M. E. Mavroforakis, H. V. Georgiou, N. Dimitropoulos, and S. Theodoridis, ''A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry,'' Comput. Methods Programs Biomed., vol. 102, no. 1, pp. 47–63, 2011.

[7]. D. C. Pereira, R. P. Ramos, and M. Z. do Nascimento, ''Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm,'' Comput. Methods Programs Biomed., vol. 114, no. 1, pp. 88–101, Apr. 2014.

[8]. S. A. Taghanaki, J. Kawahara, B. Miles, and G. Hamarneh, ''Pareto- optimal multi-objective dimensionality reduction deep auto-encoder for mammography classification,'' Comput. Methods Programs Biomed., vol. 145, pp. 85–93, Jul. 2017.

[9]. X.-W. Chen and X. Lin, ''Big data deep learning: Challenges and perspec- tives,'' IEEE Access, vol. 2, pp. 514–525, 2014.

[10]. K. Ganesan, U. R. Acharya, C. K. Chua, L. C. Min, K. T. Abraham, and K. H. Ng, ''Computer-aided breast cancer detection using mammograms: A review,'' IEEE Rev. Biomed. Eng., vol. 6, pp. 77–98, 2012.

[11]. X. Sun, W. Qian, and D. Song, ''Ipsilateral-mammogram computer-aided detection of breast cancer,'' Comput. Med. Imag. Graph., vol. 28, no. 3, pp. 151–158, Apr. 2004.

[12]. N. Saidin, U. K. Ngah, H. A. M. Sakim, N. S. Ding, M. K. Hoe, and I. L. Shuaib, ''Density based breast segmentation for mammograms using graph cut and seed based region growing techniques,'' in Proc. 22nd Int. Conf. Comput. Res. Develop., 2010.

[13]. S. Xu, H. Liu, and E. Song, ''Marker-controlled watershed for lesion seg- mentation in mammograms,'' J. Digit. Imag., vol. 24, no. 5, pp. 754–763, Oct. 2011.

[14]. K. Hu, X. Gao, and F. Li, ''Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms,'' IEEE Trans. Instrum. Meas., vol. 60, no. 2, pp. 462–472, Feb. 2011.

[15]. M. H. Yap et al., ''Automated breast ultrasound lesions detection using convolutional neural networks,'' IEEE J. Biomed. Health Inform., vol. 22, no. 4, pp. 1218–1226, Jul. 2022