A New Mammography Image Classification System by Deep Learning and Feature Selection

1Baljeet Kumar, 2Sumit Chopra
1baljit.sapra@gmail.com, 2sumitonhunt@gmail.com;
1, 2 Department of computer science KCCEIT SBSNAGAR

Abstract: The classification of breast masses from mammograms into benign or malignant has been commonly addressed with machine learning classifiers that use as input a large set of hand-crafted features, usually based on general geometrical and texture information. In this paper, we propose a novel deep learning method that automatically learns features based directly on the optimization of breast mass classification from mammograms, where we target an improved classification performance compared to the approach described above. The novelty of our approach lies in two step features optimization particle swarm optimization (PSO) and learning with convolution neural network (CNN). In proposed approach improve by accuracy 98.45%.

Keywords: PSO, CNN, optimization, texture features

I. INTRODUCTION

Breast cancer is the second leading cause of death among the women in all over the world. The proper diagnosis of the breast cancer reduces the risk of death many times. It helps the patient to relieve the physical pain and mental stress. For the diagnosis of cancer mainly four methods are used that are the surgical biopsy, magnetic resonance imaging (MRI), mammography and fine needle cytology [4]. It can help decrease the quantity of passing from breast cancer among ladies ages 40 to 70. In any case, it can likewise have downsides. Mammograms can now and again discover something that looks strange however cancer isn’t. This prompts additionally testing and can cause you anxiety. Now and again mammograms can miss cancer when it is there. It likewise exposes you to radiation. In this paper, the author discussed about the mammography approach. There are two types of the mammography: Mammography screening and diagnostics. The screening method is used when there is no sign or symptom of the tumor in the women. The diagnostic method is used when it is not possible to diagnose in the screening process. The detection process is mainly based on the shape, age, margin, the density of the tissues in the mammograms [2].

Digital Mammography is also known as full-field digital mammography (FFDM). In this method of mammography X-ray film is replaced by the electronics and it provides the images of breasts. These images are also look like the normal images. These images are captured under low amount of radiations for providing the effective image and tissues seen clearly [6] [8]. The experience of the patient during a digital mammogram is very similar to having a conventional film mammogram. Computer-aided detection (CAD) systems are used to search the abnormal tissue parts from the mammographic images. It searches the on the basis on mass, density and calcification. CAD images highlight the area which has the large amount then the normal. This method distinguishes between the malignant and benign tissues in the mammography. Some methods are based on the deep convolution neural network which is used for the classification of the features in the mammographic images [9]. In [3] textural features and fractal features are combined to overcome the problem to the classification of the mammograms. In this paper, malignant features and benign features are extracted separately for better classification result.

II. LITERATURE SURVEY

Jiao [1], proposed the deep convolution neural network for proper identification of the breast tissues in the mammograms. Performance is improved by using the layered approach of CNN. In this method, the author used the parasitic method which defines the clear relation between the tissues and it is easy to identify the mammograms. The performance of the explained approach over the existing approach is better. All the experiments are performed by CNN model and on medical images.

Khan [2], Local binary Pattern method is proposed to identify the abnormal mass classification in the breast mammography. In the proposed approach features are classified by using Support vector machine classification method. Mammograms are taken from different angles and view for the examination. It classifies the ‘cranial-caudal’ view and ‘mediolateral-oblique’ view separately. The proposed method reduces the classification error and provides the high accuracy rate in recognition.

Shirazi [3], proposed a method of tumor detection in mammography images using support vector machine and gravitational search algorithm. Images are segmented by using template matching method and extract the region of interest. GSA is used for the optimization of the parameters of the classifier. The main goal is to combine the two techniques to reduce the number of features and improves the accuracy of the classifier. The results of the paper show that the proposed method is able to optimize the features and tumor detection.

Swapnil [4], in this paper the author introduced an analysis of region marking and Grid Based. The proposed method reduced the human error by using objective analysis. In this process region of interest is marked by the radiologist to extract the features using classifiers. In the next step, a random image is taken from the database which is processed to remove x-ray annotation and pectoral muscle. This image is divided into small blocks and
features are extracted from each block. These features are either malignant or benign. The proposed method provided the results with high accuracy.

Magna, Gabriele [5], in this paper the author worked on the Adaptive artificial Immune Network. This network is applied to the Mammography image indicators. In this classification models were trained by using the set of the descriptor and it measures the degree of similarity in the left and right breast. The classification result of the proposed approach achieved high accuracy. The training and testing of the model are done on the different data set and provides the good rates of accuracy, sensitivity, and specificity.

Dehache [6], Proposed an artificial immune recognition system for mammographic mass classification. In this experiment recognition process is performed on the malignant and benign mass of the breast. In this method 3 classifiers are used for the classification of the masses. The results of the paper show the proposed method gives very effective results.

DeSampaio[7], In this paper, the methodology is divided into two main part according to the objectives of the difficulties comes into the detection of masses. The first method is used to detect the density of breast is either dense or non-dense. In this part segmentation is done according to the structure of masses. In this paper micro genetic algorithmic process is used to select the region which containing lesions. The second part of this methodology used for the reduction of false positive. False positive is based on the DBSCAN and proximity ranking of the textures of the ROIs.

Oliver, Arnau,[8] In this paper, the author proposed an approach for the density segmentation of the mammographic images this method is totally based on the pixel based classification. This approach result checked by comparing manual annotation with automatically obtained estimations. Transversal analysis of the breast density analysis is based on the two techniques craniocaudal (CC) and mediolateral oblique (MLO).

Cardona [9], the author proposed the recognition method of the mammographic image by using the Fractal Texture Analysis. In this research, a medical database is constructed by using mammographic images under the expert labeling. In this Support Vector Machine, the classifier is used to classify the features. The proposed methodology is compared with the local binary patterns (LBP) method which is mostly used in the digital image processing. The results of the proposed approach and the existing approach show that the SFTA performs better.

Bessa [10], proposed a new method of identification or normal breasts. In this experiment, the author takes the samples of the breasts which consist of the malignant part. It divides the area or the breast into the small number of blocks and compared the blocks according to the pairs. If the entire block shows the normal properties then the breast is normal otherwise it is considered suspicious. For finding the similarity between the blocks intra-block pixel distribution method is used. This method worked on the machine learning techniques. The implementation results of this method show that it works very effective than the normal mammograms.

Wang [11], in this paper the author proposed the support vector machine method for cancer detection of breast in women. This method classifies the images in two parts cancerous and normal. The problem of the optimization is also solved by the proposed method. The proposed experiment is performed on the dataset of the mammography screening. It also extracts the features based on the texture of the tissues. The results of the proposed method show that it provides the better detection rate of mammographic images.

III. PROPOSED METHODOLOGY

In this section, we discussed the proposed approach and the methodology used to achieve the results. In this methodology, we use the Prewitt filter for the edge detection of the mammographic images. It works on the 3*3 convolution mask. It detects the horizontal as well as vertical edge. It is also called as discrete differentiation operator. It is used to compute the gradient of the image intensity function. In this filter it reduces the noise from the image then sharpens the edges of the object in the image. After this, it detects the features which have to discard and which have to maintain.

![Flow chart of proposed methodology.](image)

PSO stands for particle swarm optimization. PSO is a stochastic optimization algorithm which is based on the behavior of birds. It works similar to the genetic algorithm.
In PSO, it is initialized with a group of random particles. In every iteration, each particle is updated by the two “best” values. The first best solution shows the fitness of the particles and this called as pbest. The second best value is tracked by the optimizer is the best value. This value is called as global best (gbest). When a particle takes part of the population as its topological neighbors; the best value is a local best and is called lbest.

**IV. RESULTS**

**Detection Results**

**Classification Results**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Proposed</th>
<th>SVM</th>
<th>N.N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>96.26</td>
<td>88.26</td>
<td>84.23</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>92.45</td>
<td>79.92</td>
<td>75.62</td>
</tr>
<tr>
<td>Specificity</td>
<td>98.45</td>
<td>92.26</td>
<td>90.23</td>
</tr>
</tbody>
</table>

![Proposed method results](image-url)
V. CONCLUSION

The ultimate goal of this paper is to show an efficient classification technique to detect the presence of the tumor cells in breasts and to give an early prediction of breast cancer so that many women's lives could be saved as it was a major public problem. In this approach, the MIAS dataset is used to apply the technique whether the given data is benign or malignant. This prediction gives the maximum accuracy of 96.45%. But SVM does not perform well as well as neural network (NN).

REFERENCES