BBRC
Bioscience Biotechnology
Research Communications

Deep Convolutional Neural Network for Breast Mass Classification from Mammogram

Nirmala G[1] and Suresh Kumar. P[2]
[1]Department of Electronics and Communication Engineering, Mahendra Institute of Technology (Autonomous), Namakkal, India
[2]Department of Electrical and Electronics Engineering, Mahendra Engineering College (Autonomous), Namakkal, India.

ABSTRACT

Breast cancer is the largest detection of cancer among women worldwide. Advancement in computer-aided diagnosis (CAD) makes it easy to detect and to classify benign and malignant images, henceforth to increase the life span of women. But fine-tuning of the accuracy of the existing CAD system comes to the limelight with the available resources. In recent study shows deep convolutional network provides greater accuracy. In this paper, we use deep CNN to extract the features with AlexNet. Then we Fine-tuned the various parameters to improve the accuracy with various optimizers and learning rates to classify the malignant and benign masses with CBIS-DDSM (Curated Breast Imaging Subset of DDSM) dataset. The two classifiers used the Support vector machine (SVM) and the Extreme Learning Machine (ELM) which provides an accuracy of 97.36% and 100% respectively.

KEY WORDS: DCNN, ALEXNET, ADAM, MAMMOGRAM, MASS CLASSIFICATION.

INTRODUCTION

Breast cancer is one of the leading causes of death for women globally.Tthe World Health Organization reported, the amount of cancer occurance expected in 2025 is going to be 19.3 million cases. At present, Mammography is one of the important methods to spot Breast cancer early in Deep Learning in terms of training and testing (Li Y et al., 2016 and Schmidhuber J et al., 2015). There are three types of occurrence of cancer in the lesions of the breast. They are mass, Microcalcification, and Architectural distortion. The biggest challenge in using CAD for abnormality detection in Mammograms is the high false-positive rates (FPR). False positives lead to patient anxiety, additional radiation exposure, unnecessary biopsies, high callback

rates, increased health care costs (Parameshachari et al., 2020), and extra assessment (Wang J et al.,2017).

Machine learning techniques provides significant performance in diverse healthcare applications over the traditional CAD systems for disease diagnosis (Hossain MS 2019). Nowadays (Dina et al., 2019) CNNs are used in mammography for lesion localization and recognition, risk estimation, image recovery, and classification tasks. The data needed to train a DL network is massive compared to the traditional data. CNN's also help radiologists providing more accurate diagnoses by providing precise quantitative analysis of suspicious lesions and also reduce the human error rate by 85% (Wang at al., 2017).

Recently, various algorithms have been used to build up computer-aided diagnosis (CAD) systems to enhance the diagnostic capabilities of breast cancer in medical images (Alkhaleefah et al., 2018) with supervised, semi-supervised, and Unsupervised algorithms. These algorithms are mainly supported traditional classifiers that believe had-crafted features so as to resolve a specific machine learning task. Therefore, these sorts of methods are considered to be monotonous, prolonged, and require

specialist in the field, especially in the feature extraction and selection tasks. (Schmidhuber 2017).

MATERIAL AND METHODS

CBIS-DDSM mass database is used for testing and training the benign and malignant classes. The dataset is split int0 70 % for training and 30% for testing. Many 'CAD' systems start with the image statistics being preprocessed. It may be appropriate to calibrate the scanned files, resample the data with a well-known static resolution of pre-processing steps like rescaling, contrast enhancement, etc., with our dataset performance.

Charkraborty et al., (2018) presented a new system with high-to-low intensity thresholding with multi-resolution analysis to classify masses in DDSM dataset. They achieved 85% sensitivity for mass detection an AUC of 0.92 for masses diagnosis and an accuracy of 83.30%. Ragab et al. (2019) presented two segmentation approaches with the manual determination of the region of interest (ROI) and region-based thresholding. DCNN is used for the feature extraction and AlexNet is used for fine-tuning. With SVM classifier obtained an accuracy of 87.2%.

Jain et al.,(2016) used AlexNet to classify the breast masses in mammograms of the DDSM dataset (Heath et al., 2001) and the accuracy achieved was 66%. Suzuki et al., (2016) used DCNN using transfer learning in the DCNN. The sensitivity achieved when differentiating between mass and normal lesions was 89.9% using the digital database for screening mammography (DDSM).

Lévy et al.,(2016) presented the research on pre-trained networks on AlexNet and GoogLeNet with the Digital database for screening mammography(DDSM). The experiments performed were lacking in findings of problems to find breast masses classification.it provides an accuracy of 89% and 92.9%. Huynh et al. (2016) presented a transferred learning-based CNN approach for mammogram images of 219 breast lesions. CNN extracts the features and improved classification accuracy. Due to the small number of samples used in the research, the overfitting problem might have occurred.

Hasan nasir khan et al.,(2019) proposed a trained system with four views of mammograms, after data augmentation. The datasets used are the CBIS-DDSM and mini-MIAS databases of mammograms. MVFF produced good performance over single view area under ROC curve (AUC) of 0.932 for mass and calcification. Emmanuel et al., (2019) proposed the research popular DCNN based imagenet, AlexNet. It was modified to categorize the tumors into benign and malignant (abnormal) in mammogram images. It provides improved accuracy of 95.70%. Luqman Ahmed et al., (2020) proposed two types of deep learning-based segment instances DeepLab and Masked RNN with the dataset MIAS and CBIS-DDSM. The work increased the AUC for the transfer learning method with fine-tuning. The AUC is 0.98 for mask RCNN and 0.95 for the deep lab. Also, the mean average precision

for the segmentation task is 0.80 and 0.75. The accuracy ranges from 0.80 to 0.88.

Deep Convolution Neural Networks: The Deep Convolutional Neural Networks (DCNN) is pre-trained firstly using the ImageNet dataset, which contains millions of natural images for classification and regression shown in fig.1. Nowadays, Transfer learning is widely used in many research problems in deep learning which concentrates on preserving knowledge to feed as models for research problems.

The fine-tuning process at the final layer may be substituted with any of the three layers mentioned (i) a fully connected layer (ii) a Softmax activation layer, and (iii) a custom classifier.

Convolution layer: The first layer in the DCNN is convolution layer. It performs convolution 5x5 .Each neuron in the layer computed the dot product of weignt and local region of the given input.

Activation Layer: The activation Layer used to increase the non-linearity without affecting approachable nodes of the convolution layer. ReLu is the commonly used activation layer. Sigmoid is A Non-linear function, gives a smooth gradient.

Figure 1: General Architecture of CNN
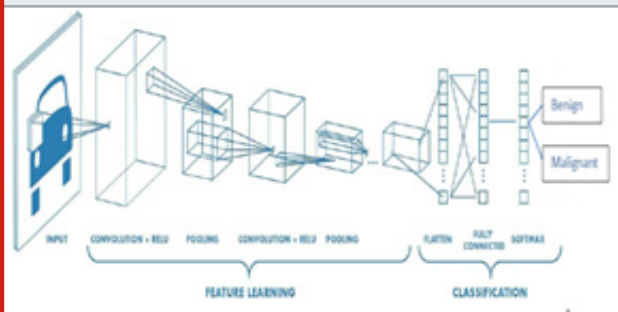
It provides good results in classification. The value normally lies between o and 1. This is normally used in binary classification in the Logistic regression model. The function is given in eqn.(1).

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{1}$$

Tanh is a Non-linear function, squashes a real-valued number to the range [-1, 1]. The output is zero centered. The value lies in (-1,1)The function is given in the eqn (2).

$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{2}$$

Softmax function calculates the probabilities distribution of the event over 'n' different events. This function will calculate the probabilities of every target class over all possible target classes. Later the calculated probabilities are going to be helpful for determining the target class

for the given inputs. It is used in Multiclass classification in the Logistic Regression model. The mathematical expression is given as eqn(3).

$$\sigma(z)j = \frac{e^{zj}}{\sum_{k=1}^{K} e^{zk}} \text{ for } j=1,2,3....k \tag{3}$$

Pooling: Pooling is a two-dimensional filter, assigned for the sample-based discretization process to reduce the dimensionality of the input. It provides down-sampling feature maps by reduces the presence of features in the feature map. Also summarizes the features present in a region of the feature map generated by a convolution layer. Thus makes the robust model. In order to introduce a translation invariance to small shifts and distortions, and decrease total number of successive learnable parameters. Commonly using pooling methods are (1) max pooling and (2) average pooling.

Regularization: Regularization plays a vital role in order to obtain accurate results in deep learning; a large number of images are required for good training. Training a model with a lesser dataset leads to overfitting. To minimize that the following regularization techniques are used.

(1)L2 & L1 regularization: The regularization term is added with the cost function. So weight metrics value is decreased. Therefore, It reduces overfitting .L2 regularization weight is decayed towards zero. The hyperparameter value is optimized to obtain good results. In L1 the weight decays to zero. The L1 and L2 equations are given in eqn (4) and eqn.(5).

$$\text{Cost function} = \text{Loss} + \frac{\lambda}{2m} * \sum ||w|| \quad\quad ..... (4)$$
$$\text{Cost function} = \text{Loss} + \frac{\lambda}{2m} * \sum ||w||2 \quad\quad .....(5)$$

Where, λ-regularization parameter.

(2)Dropout: It is used in the networks, where a large number of hidden nodes are used. At every iteration, it randomly selects some nodes and removes them together with all of their incoming and outgoing connections randomly. Hence produce a different set of outputs. This probability of selecting node percentage should be dropped in the hyperparameter

(3)Data Augmentation: Large amount of data used ,will automatically reduce overfitting is to increase the size of the training data. We can increase the training data by rotating, flipping, scaling, shifting, Noise injection, and color space transformation.

(4)Early stopping: Cross-validation is used where we keep one part of the training set as the validation set. If the validation set is poor, immediately end the training.

RESULTS AND DISCUSSION

The proposed work of multiclassifiers uses AlexNet, contains 8 deep layer that can classify into 1000 object classes with an input size of [227 227 3] is shown in

fig.2. The main significant property of AlexNet is the dropout technique which widely employed in reducing overfitting significantly [11] (Li et al., 2016). Vector parameters assigned for training leads to reduce the loss at every iteration. Generally, AlexNet uses a rectified linear activation function (ReLU) as an activation function which is simple and makes the training so easier with various parameter initializations. The ReLU is a piecewise linear function, produces one for the positive values and zero for negative output. The gradient is constant which does not require any computation in the backpropagation. The computational cost will be cheaper than the sigmoid function is given by the eqn(6).

$$R(z) = \max(0,z) \rightarrow \frac{\partial f}{\partial x} = 1 \text{ if } z \geq 0 \text{ else } 0 \tag{6}$$

Figure 2: Proposed Block Diagram of modified DCNN architecture
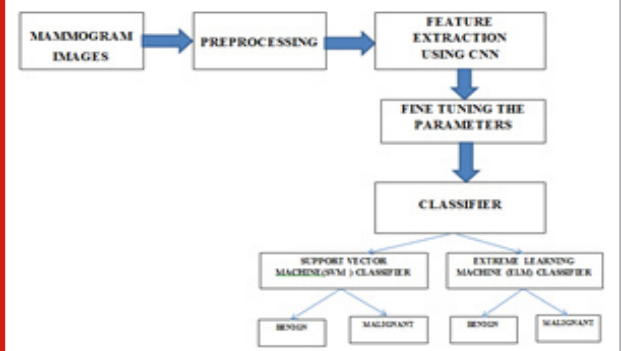
In our specific mass, the features are extracted classification there we have only two types widely known as Benign and Malignant. So the fully connected layer 'fc6' and 'fc7'are modified with the bias parameters in the typical architecture of Alexnet from 1000 classes into two classes. The features are extracted from the fully connected layer 'fc7'.Also fine-tune the parameters and optimizer given in Table.1. The final layer of Alexnet is connected classifications with the classifier with the two classes Benign and Malignant.

Table 1. Training parameter of the proposed model	
Training options	Value
Initial learn rate	10e-04 and 10e-3
Learn rate	Constant or piece wise
Mini-batch size	64
Shuffle	Every epoch
Cross validation	5
Max epochs	200

CLASSIFIER: Multiple classifiers are generally investigated by the researcher for the best performance. The entire mathematical model falls into neural networks, Support Vector Machine and Bayesian functions are often used for the effective classification with the training set. Here

using two types of a classifier to extract the performance. (1)Support Vector Machine (SVM) classifier (2) Extreme Learning (ELM) Machine.

(1)Support Vector Machine Classifier: A support vector machine is an efficient supervised pattern classifier widely used for pattern recognition and classification problems. It is a binary classifier built by constructing a hyperplane to separate non-members of the given input data. In this way, it merges the data in the high-dimensional feature space. In a given kernel space linear model is constructed and the decision is taken between two datasets.

SVM can also be suited for non-linear classification problems. For the given input data set xi, a group of training classes was constructed and the decision made based on the decision function. The Hypothesis function is given in eqn(7). The soft margin classifier is given as in eqn(8).

$$h(xi) = \begin{cases} +1 & if\ w.x + b \geq 0 \\ -1 & if\ w.x + b \leq 0 \end{cases} \quad ......(7)$$

$$\left[\frac{1}{n}\sum_{i=1}^{n} max\left(0, 1 - yi(w.xi - b)\right)\right] + \lambda||w||2 \quad ....(8)$$

Extreme Learning Machine (ELM): ELM utilizes the single hidden layer to obtain great accuracy. L neurons in the hidden layer are required to work with an activation function that is vastly differentiable without any iteration. The ELM single layer is given in eqn(9).

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x)\beta \quad (9)$$

Where x - input

β- output weight vector

H(x)→hidden layer output, it is given eqn (10)

$$h(x) = [h_1(x), h_2(x), ... h_L(x)] \quad (10)$$

To target vector is in the hidden layers is given in eqn. (11).

$$\begin{bmatrix} h(x_1) \\ h(x_2) \\ \vdots \\ h(x_N) \end{bmatrix} \quad (11)$$

The ELM uses the minimal non-linear least square methods which is given in eqn(12).

$$\beta' = H^*O = H^T(HH^T)^{-1}O \quad (12)$$

Where H*→Moore–Penrose generalized inverse. Above eqn(12) can be written as,

$$\beta' = H^T\left(\frac{1}{C}HH^T\right)^{-1}O \quad (13)$$

The output function is derived from eqn(13).

$$f_L(x) = h(x)\beta = h(x)H^T\left(\frac{1}{C}HH^T\right)^{-1}O \quad (14)$$

ELM yield good accuracy with minimal training error. The performance of the transfer learning model, AlexNet with the fine-tuned parameters is evaluated with different optimizers such as sigmoid, rmsprop adadelta, and adam with two learnig rates 1e-3 and 1e-4. Here the Alexnet with adam outperforms well with training and testing Accuracy shown in Table.2.

Table 2. Comparison of Accuracy and loss for various optimizers

Learning rate	Optimizer	Accuracy		Loss
		Training	Testing	
0.0001	adam	1.0	1.0	0.0074
	Sigmoid	0.9956	0.9736	0.0965
	rmsprop	0.9976	0.9561	0.1116
	adadelta	0.9234	0.7719	0.3384
0.001	adam	0.9922	0.9856	0.0942
	Sigmoid	0.9872	0.9649	0.1595
	rmsprop	0.997	0.9824	0.0747
	adadelta	0.9142	0.614	1.1117

Table 3. Comparison with various deep learning Methods

Authors	Method	Accuracy (%)	AUC
Luqman Ahmed et al.,(2020)	Mask R-CNN	80	0.95
Dina A Ragab et al.,(2019)	DCNN+SVM	87	0.94
Ridhi Arora et al.,(2020)	Modified DCNN	88	0.88
Gokhan Altan (2020)	CNN+Deep AE	95	0.97
Proposed Model	DCNN+ADAM+SVM	97.36	0.99
	DCNN+ADAM+ELM	100	100

Figure 3: Comparison Results of proposed work with various DCNN models
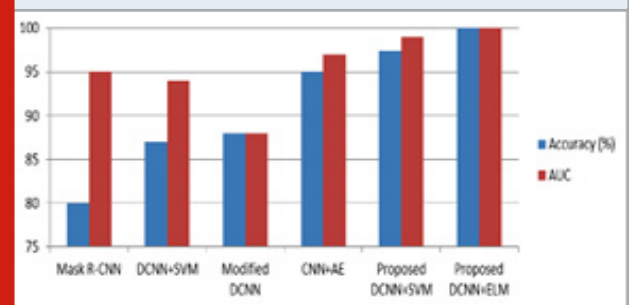
The proposed result is compared with various deep neural network models with the proposed classifier shown in fig.3. The proposed model is trained for 200 epochs whereas the loss is reduced further and obtained as 2.1287e-08 for the ELM and 0.0256 for the SVM. Thus

results even the epoch increased there are no changes in the training accuracy. The classification is results for AlexNet with SVM is 97.36 % and the ELM is 100% with a learning rate of 0.0001 shown in Table.3.

CONCLUSION

In breast cancer detection, mammography plays a significant role, but in certain situations, the radiologists cannot identify the tumors even though they have a lot of experience. The proposed work is compared with the other learning algorithms and various optimizers with the Transferred learning AlexNet. The traditional Adam layers are modified at the fully connected layers for the customized classification of mass detection. The customized transfer learning network, AlexNet with Adam optimizer provides the accuracy of 97.36% and 100% with SVM and ELM classifier respectively. The ELM is performing better than the SVM by 2.64 %. Better accuracy level leads to a better prediction of Benign and malignant masses. So the diagnosis lead is improved and hence the mortality rate of women may be increased. Deep learning needs more samples for learning and time consumption is a little high. Better augmentation techniques and good feature selection will always have room for discussion with minimum samples and minimum iterations.

REFERENCES

Akselrod-Ballin A, Karlinsky L, Alpert S, Hasoul S, Ben-Ari R, and Barkan E. (2016) A region based convolutional network for tumor detection and classification in breast mammography. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, Springer: 197–205.

Alkhaleefah, M.; Wu and C.C. (2018) A Hybrid CNN and RBF-Based SVM Approach for Breast Cancer Classification in Mammograms. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 7–10:894–899.

Dina Abdelhafiz, Clifford Yang, Reda Ammar and Sheida Nabavi1(2019) Open Access Deep convolutional neural networks for mammography: advances, challenges and applications, BMC Bioinformatics, 20(11):281. https://doi.org/10.1186/s12859-019-2823-4

Emmanuel Lawrence Omonigho, Micheal David, Achonu Adejo, and Saliyu Aliyu, (2019) Breast Cancer : Tumor Detection in Mammogram Images Using Modified AlexNet Deep Convolution Neural Network, 2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS),DOI: 10.1109/ICMCECS47690.2020.240870

Gokhan Altan (2020) A Deep Learning Architecture for Identification of Breast Cancer on Mammography by Learning Various Representations of Cancerous Mass,Deep Learning for Cancer Diagnosis:169-187

Hasan nasir khan , Ahmad raza shahid , Basit raza ,

Amir hanif dar , and Hani alquhayz , (2019) Multi-View Feature Fusion Based Four Views Model for Mammogram Classification Using Convolutional Neural Network,IEEE Access 7:165724- 165733

Heath M, Bowyer K, Kopans D, Moore R, and Kegelmeyer P. (2001). The digital database for screening mammography. In: Yaffe MJ, ed. Proceedings of the fifth international workshop on digital mammography, Medical Physics Publishing. :212_218.

Hossain MS (2017) Cloud-supported cyber-physical localization framework for patients monitoring. IEEE System Journal, 11(1):118–127.

Huynh, B.Q, Li, H,and Giger, M.L. (2016) Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. Journal of Medical Imaging, 3(3): 034501. doi: 10.1117/1.JMI.3.3.034501.

J. Chakraborty, A. Midya, and R. Rabidas,(2018) Computer-aided detection and diagnosis of mammographic masses using multi-resolution analysis of oriented tissue patterns, Expert System Applications, 99:168-179.

Jain A and Levy D (2016) Breast mass classification using deep convolutional neural networks. In: 30th conference on neural information processing systems (NIPS 2016). Barcelona, Spain:1-6.

Juan Wang, Huanjun Ding, Fatemeh Azamian Bidgoli, Brian Zhou, Carlos Iribarren, Sabee Molloi, and Pierre Baldi (2017) Detecting cardiovascular disease from mammograms with deep learning. IEEE Transaction on Medical Imaging,36(5):1172-1181.

Lévy, D. and Jain, A. (2016) Breast mass classification from mammograms using deep convolutional neural networks. arXiv 2016, arXiv:1612.00542.

Li Y, Chen H, Cao L, and Ma J(2016) A survey of computer-aided detection of breast cancer with mammography. Journal of Health Medical Informatics,4(7):1-6.

Luqman Ahmed, Muhammad Munwar Iqbal, Hamza Aldabbas, Shehzad Khalid, Yasir aleem and Saqib Saeed (2020) Images data practices for Semantic Segmentation of Breast Cancer using Deep Neural Network, Springer, Journal of Ambient Intelligence and Humanized Computing. https://doi.org /10.1007/s12652-020-01680-1.

Mehedi Masud, Amr E. Eldin Rashed, and M. Shamim Hossain(2020) Convolutional neural network-based models for diagnosis of breast cancer, Neural Computing and Applications, https://doi.org/10.1007/s00521-020-05394-5.

Nahid AA, Mehrabi MA and Kong Y (2018) Histopathological breast cancer image classification by deep neural network techniques guided by local clustering. BioMed Research International, Article ID 2362108,

2018:1-20 https://doi.org/10.1155/2018/2362108

Parameshachari, B. D., Panduranga, H. T., & liberata Ullo, S. (2020, September). Analysis and Computation of Encryption Technique to Enhance Security of Medical Images. In IOP Conference Series: Materials Science and Engineering (Vol. 925, No. 1, p. 012028). IOP Publishing.

Ragab DA, Sharkas M, Marshall S, and Ren J. (2019) Breast cancer detection using deep convolutional neural networks and support vector machines. Bioinformatics and Genomics,PeerJ7:e6201 http://doi.org/10.7717/peerj.6201

Ridhi Arora , Prateek Kumar Rai , and Balasubramanian Raman (2020) Deep feature-based automatic classification of mammograms, Medical & Biological Engineering & Computing 58(6):1199-1211

Schmidhuber J(2015). Deep learning in neural networks: An overview. Neural Networks,61:85–117.

Suckling, J.(1994) The Mammographic Image Analysis Society Digital Mammogram Database. In 2nd International Workshop on Digital Mammography; Elsevier Science: Amsterdam, The Netherlands:375–378.

Suzuki S, Zhang X, Homma N, Ichiji K, Sugita N, Kawasumi Y, Ishibashi T, and Yoshizawa M. (2016). Mass detection using deep convolutional neural network for mammographic computer-aided diagnosis. In: Proceedings of the SICE annual conference Tsukuba, Japan:1382-1386.