

Software Engineering Model Based Early Detection Method of Breast Cancer using Deep-Learning Framework

¹Muayad S. Croock, ¹Ayad E. Korial, ²Tara F. Kareem,

³Qusay Sh. Hamad and ¹Ghaidaa M. Abdulsahab

¹Department of Computer Engineering, University of Technology, Baghdad, Iraq

²Oncology Teaching Hospital, Medical City Complex, Baghdad, Iraq

³University of Communications and Information Technology, Baghdad, Iraq

Abstract: Recently, the serious diseases that can attach people have been increased in scary way. One of these diseases is the breast cancer. This type of diseases affects the women in particular more than men. In this study, a deep-learning framework is proposed for detecting the breast cancer in early stages based on mammography images. It adopts the incremental development software process model as a part of software engineering in the designing framework for more reliability and extendibility. The presented method has been established on extracting the features of the employed images as a learning dataset for convolutional neural network inside the deep-learning strategies. Different step algorithms have been used for performing the detection of the standard benchmarks of breast cancer in the soft tissue, shown in the utilized mammography images. A real dataset, collected from Baghdad hospital is considered and it is divided into 30% test and 70% training sets. The obtained results show a high accuracy in terms of feature extraction of training set about 100% and breast cancer detection from test set as a validation accuracy about 90%.

Key words: Deep-learning, breast cancer, mammography, software engineering, employed, incremental

INTRODUCTION

It is well known that the breast cancer is considered as the most dangerous disease that attacks women for different reasons such as the life habits and inheritance. There are different ways utilized to detect the changing in the soft tissue of the breast in early stages. Mammography imaging is the most common way of breast cancer detection. It is X-ray images taken from different angles to create a complete vision of the disease. These images suffer from low contrast due to low volume X-ray in which the image processing algorithm might not work in proper way. Therefore, different methods are used for implementing the image enhancement including artificial intelligent strategies and deep learning (Duraismy and Emperumal, 2017; Kallenberg *et al.*, 2016). In the deep learning method, the preprocessing stage performs the image enhancement and increase the contrast of allocated areas in the mammography images to ease the detection of breast cancer. The convolutional neural network algorithm is adopted in the deep learning to train the detection system by extracting different features that can lead to allocating the change in the soft tissue layers to be classified as a breast cancer. The convolutional neural network is a classification algorithm that works strongly

on the images and feature extraction to avoid the high capacity size required for processing the whole images. As a result, the processing time is reduced as well in efficient way due to low size of underlying images (Kallenberg *et al.*, 2016; Zhang *et al.*, 2018). Amongst the previous research work done, so far, the convolutional neural network classifies the small patches of suspicious tumors that can be allocated as area of interest. These classifications of the underlying images are based on either clinical or automatic detections. By Duraismy and Emperumal (2017) the researchers introduced a new deep learning classifier for digital mammography images. The classifier was used to detect the tumor tissues in addition of recovering the low contrast images of mammography. The contour was performed using Chan-Vese level set method and the required features were extracted using deep learning based convolutional neural network. A complex valued relaxation was added to the classifier to reduce the false results and increase the accuracy up to 99%. By Kallenberg *et al.* (2016), a method of learning a feature hierarchy of unlabeled dataset was proposed. The data set was fed to the classifier for breast density segmentation and mammography texture scoring. Both of lifetime and population sparsity were incorporated by introducing a regularizer to control the capacity of the

presented model. The strong point of the proposed method was the easy of implementation and the obtained results proved the high accuracy. The researchers by Carneiro *et al.* (2017) addressed the risky developmen of breast cancer showed in the mammography images using Cranio-Caudal (CC) and Mediolateral Oblique (MLO). They used deep learning models for solving the issue of unregistered breast images and related segmentations. This affected the performance accuracy of breast cancer detection. By Yap *et al.* (2018) deep learning approaches were adopted to detect and investigated of breast cancer in the ultrasound lesion. A Patch-based LeNet, a U-Net and a transfer learning approach with a pertained FCN-AlexNet had been used for performing the aim of the study. The obtained results showed the high accuracy in comparison with the traditional methods.

By Zhang *et al.* (2018), the researchers proposed a tomosynthesis classification method based on deep learning for over 300 mammography images collected from University of Kentucky. Different models of deep learning based convolutional neural network method were used in building the classifier for working on 2 and 3D images. The obtained results expressed the efficiency of the proposed method. The researchers by Mehdy *et al.* (2017) presented a review paper for the utilized techniques of using of neural network in image processing used for early detection of breast cancer in mammography samples. In this review, different types of neural models were studied including the hybrid adaptation in breast cancer detection.

At the other hand, artificial neural network based breast cancer detection system was proposed by Saubhagya *et al.* (2016) Saad *et al.* (2016) and Ayer *et al.* (2013). The proposed methods enhanced the micro-calcification based on illumination and non-regularity. They also allocated using iterative selection threshold method by rebuilding the shape of images and removing the redundant pixels. In addition, the proposed methods extracted the features of these mammography images for detecting the breast cancer. The obtained results showed the high accuracy of performance of the proposed methods in comparison with the previous approaches.

In this study, an early breast cancer detection method is proposed based on deep learning convolutional neural network. This method is aimed to extract the training features from the training dataset that includes over 500 mammography images. These images are collected from Baghdad hospital to be classified into two groups of infected and non-infected. The conventional neural network classifier uses the built training dataset for detecting the breast cancer of the incoming test dataset.

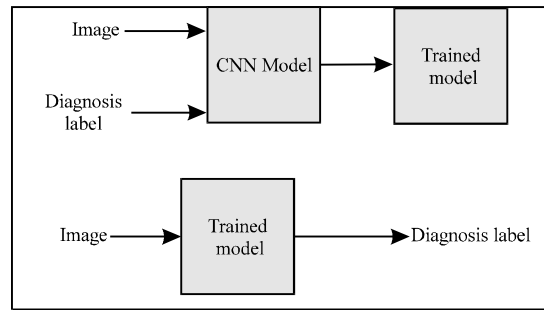


Fig. 1: The block diagram of the proposed method

It also recover the low density of construction of the utilized images to ease the disease detection in early stages. It is important to note that the proposed method adopts the lightweight soft computing algorithms to satisfy the required complexity reduction. This is to accelerate the processing speed required for building the training and test dataset. Therefore, the proposed method motivates over the previous research by Duraisamy and Emperumal (2017), Ayer *et al.* (2013) in two points: reducing the computational complexity and tackling the problem of lacking in specialist doctors by automating the cancer detection (Fig. 1).

MATERIALS AND METHODS

The proposed method: As mentioned earlier, the proposed breast cancer detection method is based on deep learning approach. It can be divided into three subsections for easing the understanding of the study reading as follows.

Block diagram: The block diagram of the proposed method is presented in Fig. 1. In this figure, the mammography images are entered to the Convolutional Neural Network (CNN) for processing. The obtained results represents the trained model that includes the training dataset. Moreover, the diagnosed labels represents the classification of the disease in terms of infected and not-infected. At the other hand, the images of the dataset, required to be diagnosed are entered to the trained model for searching over the trained dataset of the extracted features and classifying. The obtained results are the diagnosed labels which include two types of infected and non-infected. Figure 2 illustrates the steps of work flow regarding the processing inside the CNN Model.

It is shown that the enter image is decoded for obtaining a binary format. The binary image is resized for optimization the resolution under the condition of limited size. This is to reduce the processing complexity. The

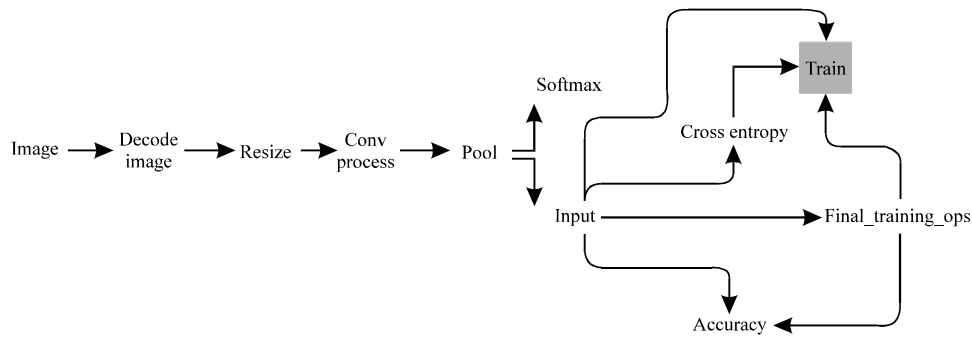


Fig. 2: CNN work flow

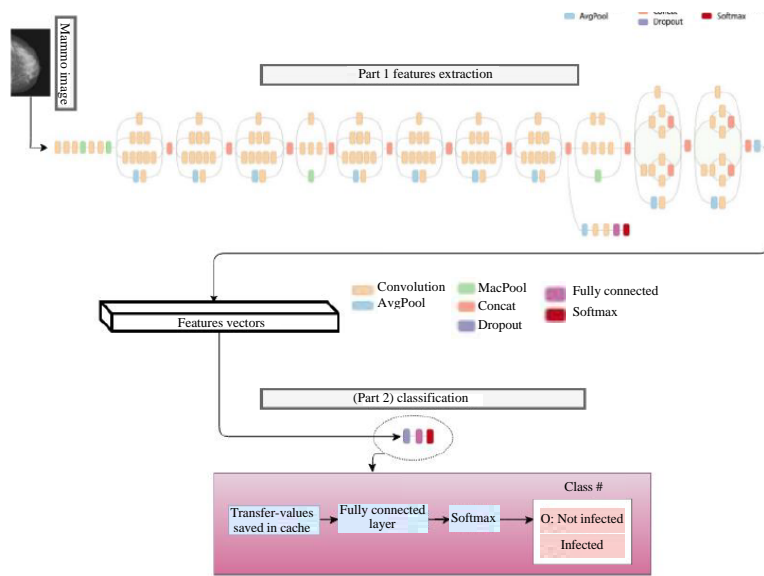


Fig. 3: Functional work flow of the proposed method

now are convolved to be ready for feature extraction. At the pool, the softmax approach is applied in the deep learning method for feature extraction and classification according to the entered diagnosed labels. At this stage, the trained model is being ready and the cross entropy in addition to the accuracy are evaluated.

Functional work flow: Figure 3 shows the functional work flow of the proposed breast cancer detection method. This figure includes group of small and colored blocks that represents different layers of functions, performed inside the CNN Model. These functions include: Convolution, Average Pool (Avg. Pool) for normalize the underlying images, maxPool for feature extraction, concat for segmentation and contour, Dropout for exporting the results, fully connected for communication reason and softmax. The research flow is divided into two main parts: feature extraction and classification. The first part extracts the features following

the mentioned functional blocks in Fig. 3 while the second part is classifying the test dataset into two parts infected and not infected.

It is shown that each image in the training dataset is passed through the mentioned functions to obtain the feature vector in a cross layer sequential way. The cross layer information exchange is continued till the required features are obtained used for classification which leads to early detection. The research flow of Fig. 3 is based on the lightweight processes of different functions. The highlighted point is regarding the use some functions employing the parallel processing of deep learning.

Software engineering model: It is well known that the concepts of software engineering can play as important factors in designing a reliable, extendable software systems. Therefore, in this research, we adopted the incremental development software process model, shown in Fig. 4. This model is basically depends on improving

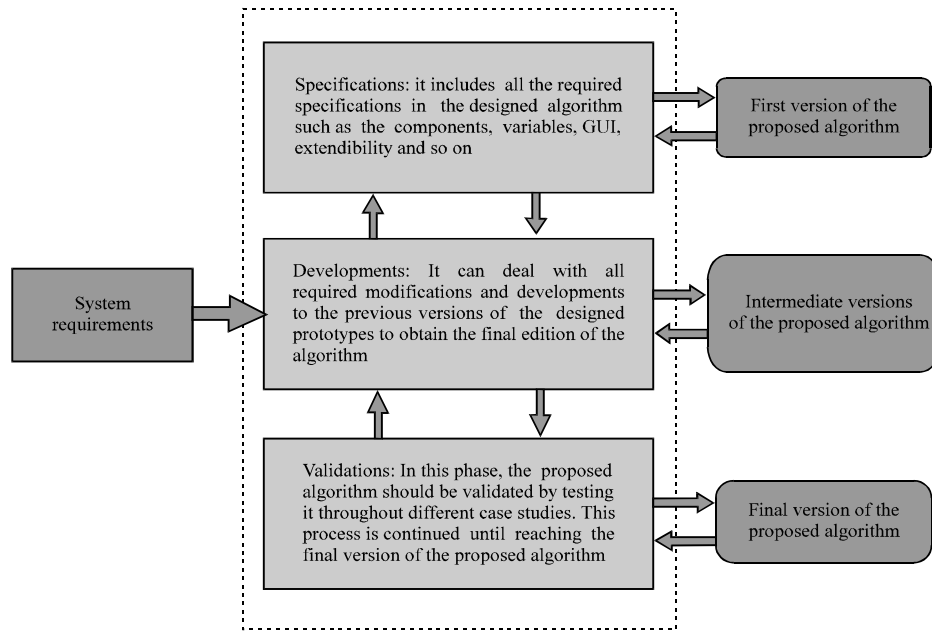


Fig. 4: The adopted incremental development software process model

the first version of the implemented prototype using the evaluation feedback of the proposed system. This process is continued in iterative way until reaching the final version. Figure 4, the system requirements is entered to the designed model to pass over three processes: specifications, developments and validation. The specifications can include all the appointed points by the proposed algorithm. The developments contains the corrections done on the previous versions of the prototypes to have the final version. While the validation takes care from testing the proposed algorithm to reach the required targets.

The proposed algorithm: The proposed method is based on two algorithms one for training model and the other for testing model. Both algorithms are working together to perform the objectives of the proposed algorithm.

Training model: The proposed algorithm for building the training model is shown in Fig. 5 as a flowchart. The steps of this algorithm can be summarized as:

- Preparing the necessary directories used as classification labels. This is employed during the training process
- Gathering the information related to the training model architecture. This model is used in the building of extracted features dataset
- Setting up the pre-trained graph. This graph is considered as the main structure of the adopted classifier of CNN method

- Looking at the labeled folder structures for any mistakes. These folders are used to create lists of all the images, classified according to the designed system. Here, they classified into infected and not-infected classes
- Checking the command-line flags for needing of applying any distortions
- Setting up the image decoding sub-graph. These changes, applied on the underlying images are necessary for increasing the accuracy and enhancing such images
- In case of detecting any distortion, the algorithm can apply distortions processes by setting up the required operations
- Calculating the ‘bottleneck’ image summaries and caching them on disk
- Adding the required and new layers as main steps in the training
- Creating the operations, needed to evaluate the accuracy of our new layer
- Merging all the image summaries and writing them out to the summaries directory
- Setting up the required weights to their initial default values
- Running the training for as many cycles as required on the command line
- Obtaining a batch of input bottleneck values, either by calculating freshly every time with distortions applied or by fetching them from the cache stored on disk

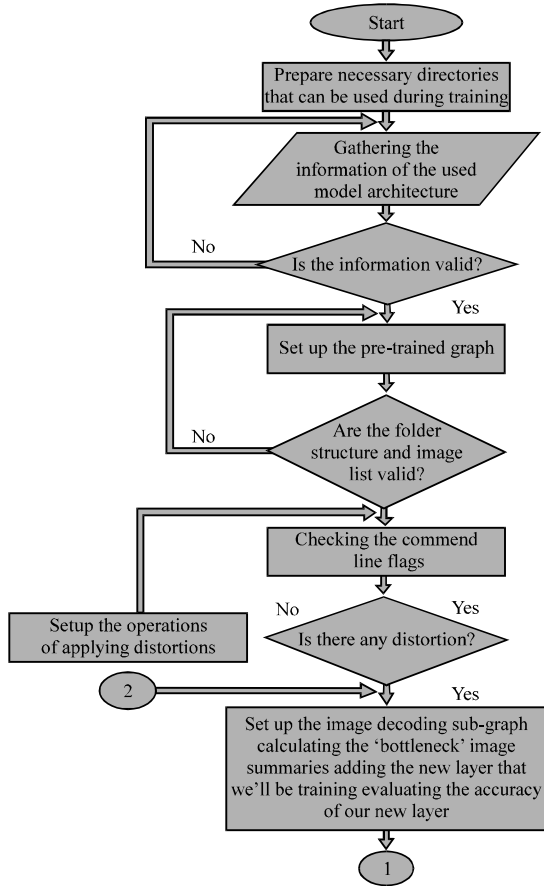


Fig. 5: The proposed training model algorithm (continue)

- Feeding the bottlenecks and ground truth into the graph and run a training step
- Capturing training summaries for TensorBoard with the 'merged' operation
- Every, so, often, print out how well the graph is training
- Running a validation step and capture training summaries and storing intermediate results
- Completing the training dataset and running a final test evaluation on some new images we have't used before
- Writing out the trained graph and labels with the weights stored as constants

Testing model: After obtaining the training model which includes the training dataset, the tested images can now be used for detecting the breast cancer if happened. Figure 6 shows the proposed algorithm as a flowchart used for detecting the disease from the mammography images. This algorithm can be summarized as:

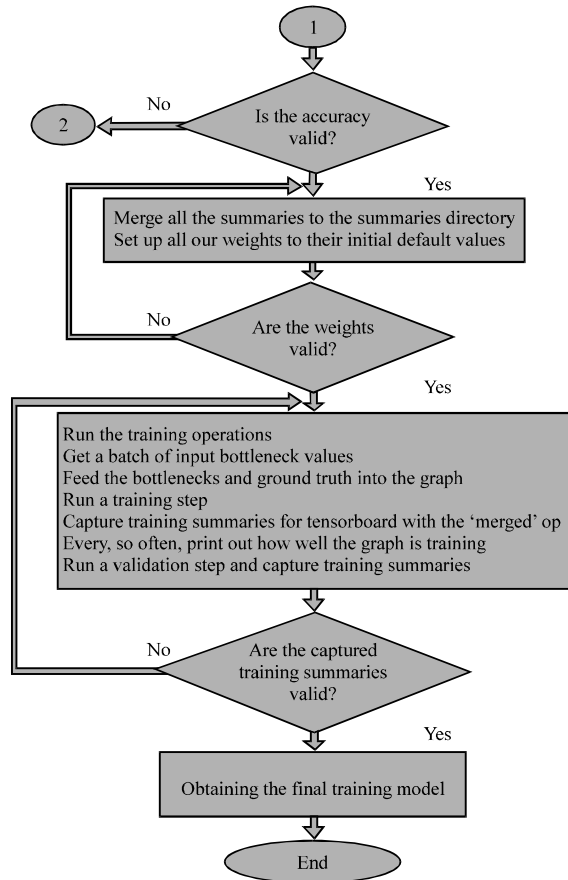


Fig. 6: The proposed training model algorithm

- Read the path of an image file
- Read image data
- Load label file
- Load the trained graph form file
- Feed image data in the step 2 into the loaded graph as input of it
- Sorting the prediction set to show labels of first prediction in order of confidence
- Classifying the entered images and evaluating the results

RESULTS AND DISCUSSION

As mentioned earlier, the proposed system is aimed to detecting the breast cancer at early stages using deep learning based CNN method. Mammography images are considered as input dataset for processing. In this research, a dataset of over 500 images are adopted and collected from Baghdad hospital, Department of Mammography. This data set is divided into training set of 70% and testing set of 30% as shown in Table 1. The testing is performed on Hp laptop with 2.4 GHz

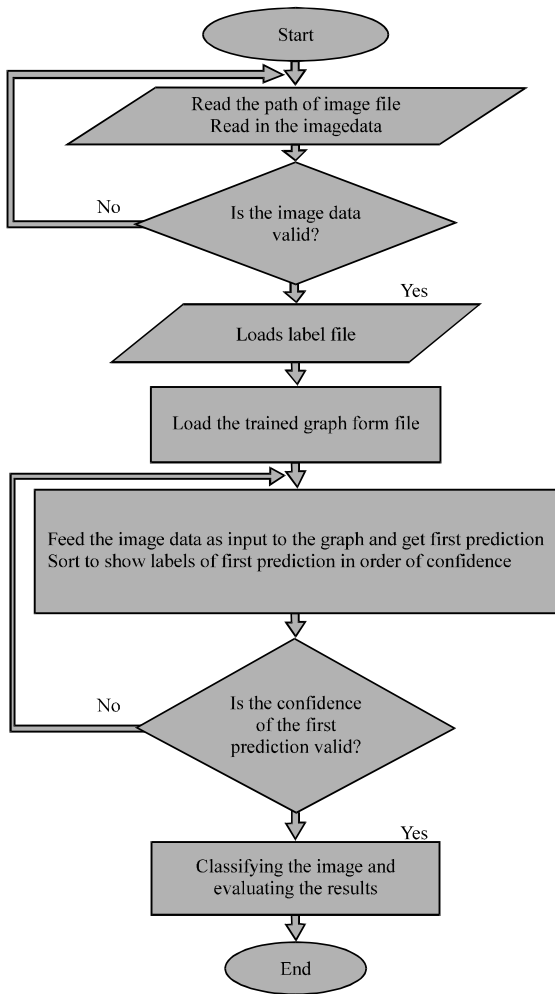


Fig. 7: The proposed testing model algorithm

Table 1: The classification of datasets

Dataset	Number	Classification		Ratio (%)
		Infected	Not-infected	
Testing set	150	70	80	30
Training set	350	200	150	70
Total dataset	500	270	230	100

processor, 4GB RAM with dedicated display adapter of (2GB) and under operating system of Windows 10 pro. With these specifications, the proposed method is run in efficient way with processing time up to half hour from the initial point. Figure 7 illustrates the evaluated accuracy of training process, performed throughout the preparation of training model, amongst the processing steps. It is shown that the accuracy is increased with advanced steps due to the residual data set of training and features from CNN. After the step of 2000, the accuracy records a ratio of 100%. From this figure, we can conclude that the training process passes the acceptable ratio of accuracy which is

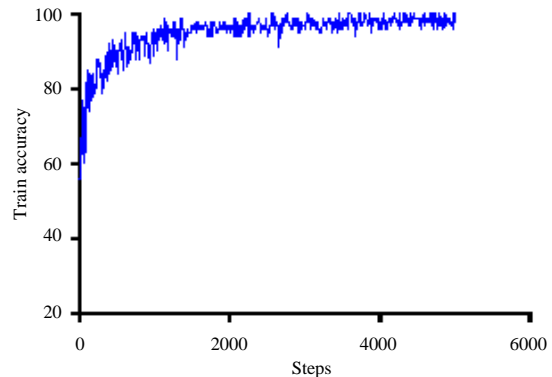


Fig. 8: The training accuracy

considered as important factor of the proposed method efficiency. In addition, the preprocessing functions, performed on the received dataset, support the increasing of accuracy ratio as a result of image enhancement. The low contrast in the mammography images can reduce the accuracy of extraction the required features in proper way without preprocessing. At the other hand, the cross entropy is measured over the processing steps for the training model as shown in Fig. 8. It is noted that the cross entropy is reduced with the advances of the processing step. This is the result of reducing the redundant features of the images running inside the CNN. In addition, the advanced steps of processing degrades the crossing between the rules of CNN used as a platform for classification. After the processing step of 2000, the cross entropy is reached the value of 0.2. This value is continued to be reduced till touch the value of 0.1 at the processing step of 5000. The optimal case is happened when the cross entropy being zero value but this is just in theory part as in real-time processing different reasons affect the entropy. In order to test the validation accuracy of the proposed method, Fig. 9 and 10 describes this validation as a result of detecting the breast cancer of the testing dataset. This figure proves the high validation of the results of the proposed method in training and testing phases. Figure 9 shows the validation accuracy of almost 90% at the processing step 2000 and over. It is highlighted from this figure that the accuracy is varied from 50% at the lower processing step and reached up to 90% over step 2000. The validation accuracy is being in the acceptable level after step 2000 for the same reasons increasing the training accuracy and reducing the cross entropy.

As a result of the testing outcome, the proposed method proves its efficiency in terms of training accuracy, cross entropy and validation accuracy. Although, the collected dataset is not prepared for

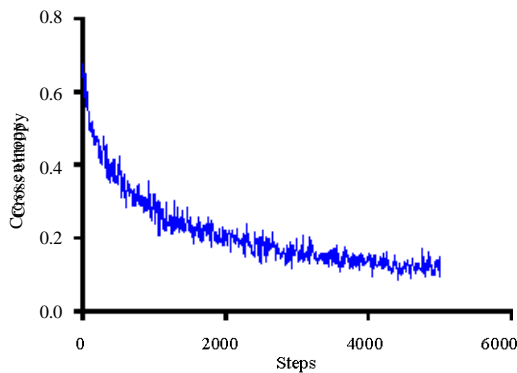


Fig. 9: The evaluated cross entropy of training model

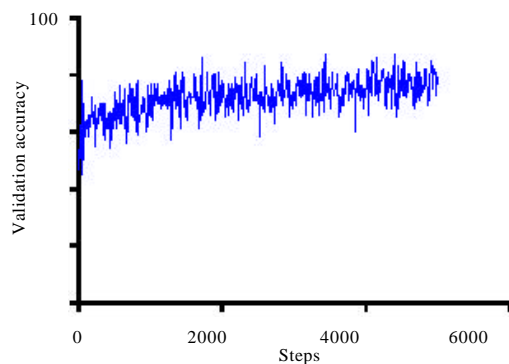


Fig. 10: The computed validation accuracy

computer programming use the preprocessing functions performed by the proposed method reduces these effects to very minimum value of error ratio.

CONCLUSION

A breast cancer early detection method based on deep learning CNN approach has been proposed. The mammography images have been adopted for this research as a dataset source, collected from Baghdad hospital. The dataset was divided into training set of 70% and testing set of 30%. Two proposed algorithms were presented in the proposed method. The first one was the training model which performed the preprocessing functions on the training set and passing the results over different processing steps to obtain the training model. The second algorithm is the testing model that applied the detection of the changing in the soft tissue which may be diagnosed as a cancer. The proposed method was tested over more than 500 images in terms of training accuracy, cross entropy and validation accuracy. The obtained results showed an acceptable accuracy and cross entropy for both training and testing phases. The proposed

method classified the testing dataset into infected and not-infected classes depending on the changing at the tissue of the breast. It is important to note that the proposed system reduces the required computational complexity by applying the preprocessing functions on the images. In addition, in Iraq it is the first time to tackle the diagnosis of breast cancer electronically (automatically) based on the proposed method.

REFERENCES

- Ayer, T., Q. Chen and E.S. Burnside, 2013. Artificial neural networks in mammography interpretation and diagnostic decision making. *Comput. Math. Methods Med.*, 2013: 1-10.
- Carneiro, G., J. Nascimento and A.P. Bradley, 2017. Automated analysis of unregistered multi-view mammograms with deep learning. *IEEE. Trans. Med. Imaging*, 36: 2355-2365.
- Duraisamy, S. and S. Emperumal, 2017. Computer-aided mammogram diagnosis system using deep learning convolutional fully complex-valued relaxation neural network classifier. *IET. Comput. Vision*, 11: 656-662.
- Kallenberg, M., K. Petersen, M. Nielsen, A.Y. Ng and P. Diao *et al.*, 2016. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. *IEEE. Trans. Med. Imaging*, 35: 1322-1331.
- Mehdy, M.M., P.Y. Ng, E.F. Shair, N.I. Saleh and C. Gomes, 2017. Artificial neural networks in image processing for early detection of Breast Cancer. *Comput. Math. Methods Med.*, 2017: 1-15.
- Saad, G., A. Khadour and Q. Kanafani, 2016. ANN and adaboost application for automatic detection of microcalcifications in Breast Cancer. *Egypt. J. Radiol. Nucl. Med.*, 47: 1803-1814.
- Saubhagya, V.K., A. Rani and V. Singh, 2016. ANN based detection of breast cancer in Mammograph images. *Proceedings of the 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, July 4-6, 2016, IEEE, Delhi, India, ISBN:978-1-4673-8588-6, pp: 1-6.
- Yap, M.H., G. Pons, J. Marti, S. Ganau and M. Sentis *et al.*, 2018. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE. J. Biomed. Health Inf.*, 22: 1218-1226.
- Zhang, X., Y. Zhang, E.Y. Han, N. Jacobs and Q. Han *et al.*, 2018. Classification of whole mammogram and tomosynthesis images using deep convolutional neural networks. *IEEE. Trans. Nanobiosci.*, 17: 237-242.