Breast Cancer Classification Using Convolutional Neural Network

Eman Alshanbari, Hanaa Alamri, Walaa Alzahrani, and Dr.Manal Alghamdi
maalghamdi@uqu.edu.sa
Umm Al-Qura University, Department of Computer Science, Makkah, Saudi Arabia

Summary
Breast cancer is the number one cause of deaths from cancer in women, knowing the type of breast cancer in the early stages can help us to prevent the dangers of the next stage. The performance of the deep learning depends on large number of labeled data, this paper presented convolutional neural network for classification breast cancer from images to benign or malignant. our network contains 11 layers and ends with softmax for the output, the experiments result using public BreakHis dataset, and the proposed methods outperformed the state-of-the-art methods.
Key words:
Breast cancer; Convolutional Neural Network; medical image; classification.

1. Introduction

Cancer is one of the most common chronic diseases, as the number of cases infected with this disease in 2008 reached 12 million new cases, and it is expected that cases will reach 27 million cases of cancer [2]. In the current year 2020 cases of breast cancer have reached more than 2.2 million cases, nearly 1 in 12 women will develop breast cancer in their lifetime and Approximately 685 000 women died from the disease [1]. Breast cancer is the leading cause of cancer-related deaths in women, according to the World Health Organization [1]. Breast cancer classification is an important task in medical image processing; there are several methods for diagnosing breast cancer, including computerized tomography (CT) scans, positron emission tomography (PET) scans, and breast magnetic resonance imaging (MRI) [3]. Diagnostic mammograms (x-rays), ultrasound (sonography) and magnetic resonance imaging [4].

Image classification is a fundamental topic in computer vision. It is described as the effort of categorizing images into one of several specified groups. Convolutional Neural Network have been successfully applied for multiple image tasks such as classification, detection and many other. CNNs are feedforward networks in the sense that information flows only in one direction, from inputs to outputs [5]. The convolutional neural network (CNN) model is the most successful, as it automatically learns and extracts the required information for medical picture understanding. The CNN model is made up of convolutional filters, the major function of which is to learn and extract necessary features for effective medical picture understanding [6].

This paper presented Breast Cancer images classify within two main classes: benign or malignant using CNN. In the beginning, we applied five learning methods in Neural Network models, first step we applied three first-order methods and three second- order methods. second, we compared the two higher accuracy between them and improved the result by generalizing the result using weight decay and Early stooping methods to avoid overfitting. The experiments result show that the presented model outperform the state of the state-of-the-art methods.

2. Related Works

For breast cancer classification, there are numerous papers available. However, the majority of the works are less accurate than the presidents. Phu et al. [7] They used the CNN method to classify breast cancer photos in an experiment on the BreakHis dataset. The accuracy from a construction model is 73.68%. Teresa et al. [8] For feature extraction, CNN is employed, while support vector machines are used for classification. They obtained accuracy around 77.8%.

Alexander et al. [9] For breast cancer classification, deep neural networks and gradient boosted trees classifiers were used and obtained accuracy around 87.2%. Tomas et al. [10] They used capsule net architecture to solve the challenge of breast cancer classification. and obtained accuracy around 87%. The photos are classified using a Back Propagation Neural Network (BPNN). The performance improved by using radial basis neural networks (RBFN) [11]. A convolutional neural network with a convolutional layer and a small SE-ResNet module was used. The results accuracy 93.81%[12].

Support vector machines were used to train two coding models: bag of words and locality limited linear coding, also convolutional neural networks have been used.where we achieved accuracy 88.23% [13].

Manuscript received June 5, 2021
Manuscript revised June 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.6.14

3. Materials and Method

3.1 BreakHis Database

Microscopic biopsy images of benign and malignant breast tumors can be found in the BreaKHis database. Images were collected through a clinical study from January 2014 to December 2014. During this time period, all patients referred to the P&D Lab in Brazil with a clinical indication of BC were invited to participate in the study. The study was approved by the institutional review board, and all patients supplied written informed permission. Anonymization was applied to all of the data. [14].

Surgical (open) biopsy (SOB) is used to collect the samples, which are then prepared for histological analysis and labeled by pathologists at the P&D Lab. The approach employed in this study was the standard paraffin process, which is commonly used in clinical practice [15].

BreakHis dataset contains 7909 photos of breast cancer divided into two categories: benign and malignant. There are 2440 photos in the benign class and 5429 images in the malignant class [14].

Microscopic biopsy images of benign and malignant breast cancers can be found in the BreaKHis database. From January to December 2014, images were collected as part of a clinical study. During this time span, all patients sent to the P&D Lab in Brazil with a clinical indication of BC were invited to participate in the study. The study was approved by the institutional review board, and all patients supplied written informed permission. Anonymization was applied to all the data [14].

Surgical (open) biopsy (SOB) is used to collect the samples, which are then prepared for histological analysis and labeled by pathologists at the P&D Lab. The approach employed in this study was the standard paraffin process, which is commonly used in clinical practice [15].

3.2 Convolutional Neural Network architecture

The layers and parameters that make up our CNN architecture are as follows:

1. Input layer:
   A Matrix of pixel values in the shape of [WIDTH, HEIGHT, CHANNELS]. Our input is [32x32x3].
2. Convolutional layers:
   The input image is combined with a set of nuclei, each of which creates a unique map in the output image.
3. Pooling layers:
   A down sampling operation will be performed by the pooling layer. as a result of the spatial dimensions (width, height), in outputs like [16x16x12] for pooling size = (2, 2).

Table 1: Model Summary

Layer (type)	*Output shape*	*Param #*
densenet201 (Model)	(None,7,7,1920)	18321984
global average pooling2d 1	(None,1920)	0
dropout 1 (Dropout)	(None,1920)	0
batch normalization 1	(Batch (None, 1920)	7680
dense 1 (Dense)	(None , 2)	3842
Total params: 18,333,506		
Trainable params: 18,100, 610		
Non trainable params: 232,896		

In this model we add two layers on the last of pretrained model DenseNet201. First layer GlopalAveragePooling2D and the Dropout layer with ratio 0.5. We also use Batch Normalization. The last two layers for classifier. We use Tanh activation Function in All layers. The last layer is SoftMax and the number of Trainable Parameters is 18,333506 .

In addition , we used the six learning methods, the learning methods in the neural network models can be divided into first-order and second-order derivatives learning algorithms.

3.3 First Order learning algorithms

The first-order derivatives method constructs the next training using gradient information. Only the gradient information is used in the first-order technique, and it does not perform as well as the second-order method [15].

3.3.1 Backpropagations

Backpropagation is a form of supervised learning. It's used to teach a multilayer neural network how to map the relationship between the target and actual outputs. The input pattern is sent through the network using network connection weights and biases of the activation or transfer functions throughout the training period.when we apply backpropagations we get accuracy of the model 92%.

3.3.2 Stochastic gradient descent with momentum

Stochastic gradient descent is an optimization algorithm that work iteratively it is well suited for neural network optimization, the term momentum is brought from physic we can think of the parameter value as being ball traveling through space when we introduce momentum it is similar to achieve roll raise, the ball keep rolling roughly in the same direction, this prevent unnecessary oscillation this well plot in neural network training (Momentum updates the weights of neural network). when we apply Stochastic gradient descent with momentum (0.01) we get accuracy of the model 75.35 %[17].

3.3.3 Steepest Descent

One of the oldest known approaches for reducing a broad nonlinear function is the steepest descent method. Most minimization algorithms work on the principle of computing a step along a particular search direction. When we apply Steepest descent, we get accuracy of the model 69.53 % [18].

3.4 Second Order learning algorithms

Hessian is used to compute the iteration based on the optimization trajectory in second-order derivatives. Due to the modification of curvature information, the second-order technique can produce superior outcomes. As a result, using second-order derivatives in neural network training allows for adaptive step-length adjustment during the artificial neural network training process [19].

3.4.1 QuickProp

We aim to find the best solution in the quickest amount of time, and QuickProp accomplishes this by merging two traditional approaches: The first is dynamic learning rate regulation. second, the error's second derivative with regard to each weight is calculated. when we apply QuickProp we get accuracy of the model 85% [20].

3.4.2 Follow The Regularized Leader

The goal of this FTRL optimize the regret bound, the difference between FTRL and proximal gradient descent (PGD) and other online learning methods i.e. (FOBOS and RDA is that it can get very sparse solutions.) .when we apply Follow The Regularized Leader we get accuracy of the model 69% [21].

3.4.3 Levenberg–Marquardt

The most extensively used optimization algorithm is the Levenberg-Marquardt (LM) algorithm. In a wide range of situations, it outperforms simple gradient descent and other conjugate gradient approaches. Nonlinear Least Squares Minimization is the name of the problem that the LM method solves. When we apply LM, we get accuracy of the model 86% [22].

4. Improvement

We improve the result of our model to get the high accuracy by using two regularization techniques such as Weight Decay and Early Stopping.

4.1 Weight Decay

One of the most common strategies in the neural network toolbox is weight decay. Because it can be calculated from the gradient of the L2 norm of the weights in the gradient descent setting, it is commonly viewed as a type of L2 regularization. We can build more efficient and robust neural network designs if we have a better understanding of the impact of weight decay [23].

$$L_{new}(w) = L_{Original}(w) + \lambda w^T w$$

The value of lambda parameter determining the strength of the penalty

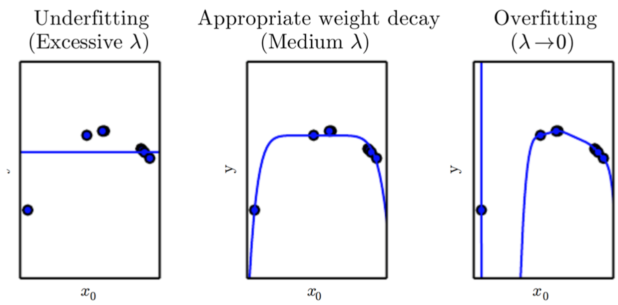


Fig.1 Weight Decay [24]

4.2 Early Stopping

When using gradient-based optimization to train an over-expressive model, early halting is a common strategy to avoid poor generalization performance [24]. If the weights are allowed to grow enough during training and then stop training at this point, it is possible to restrain the network from overfitting [25].

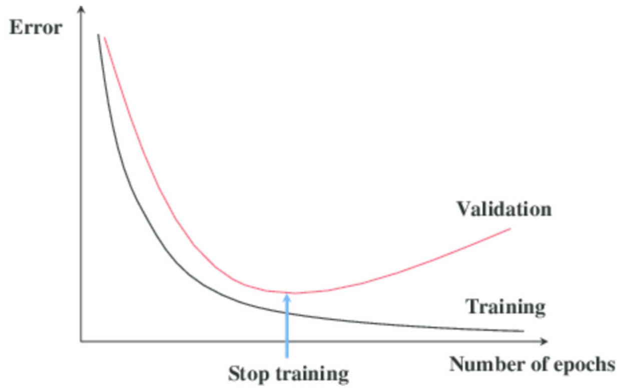


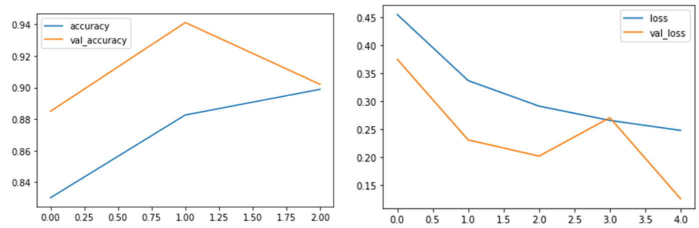
Fig.2 Early Stopping [26]

5. Experiment

Our model is based on [27] model, we used many libraries to build the CNN model such as keras Neural Network library [28] and tensorflow open-source library [29] with python programming language. It contains 5 epochs and 16 Batch Size, Dense layer with 'Softmax' Activation Function, Adam optimizer and different first and second order optimization methods.
We divided the data into two sets 80%train and 20% test.
Finally, we improve the results of the model by using weight decay and early stopping techniques.
We picked one method from three first-order methods which we applied in our CNN model and another one method from three second-order methods which is also applied in our CNN model which all these two methods have highest accuracy between all these six methods, so this is the reason we have picked them.

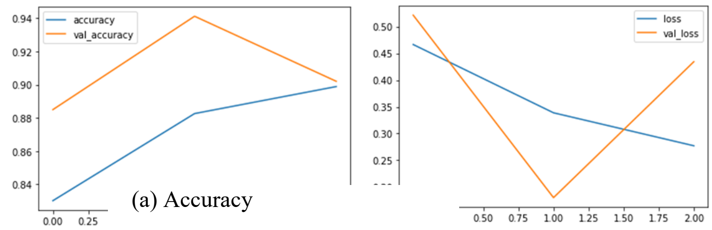
The first method which is Back Propagation give us 92% accuracy and 0.22 loss value in our model and then after improvement with Weight Decay the accuracy became 96% , and 93% with Early stopping . Fig.3 shows the impact of Wight Decay and Early Stopping techniques on Back Propagation method.

Levenberg Marquardt is the second method we have in improvement stage which the accuracy was 86% and after optimize it by using weight decay the result of accuracy became 90% and with early stopping also 90% Fig.4 shows the impact of Wight Decay and Early Stopping techniques on Levenberg–Marquardt method.



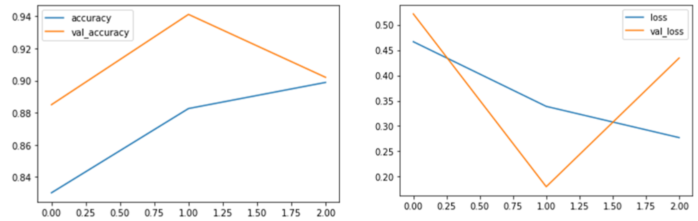
(a) Accuracy (b) Loss

Fig.3 Impact of Wight Decay on Back Propagation method



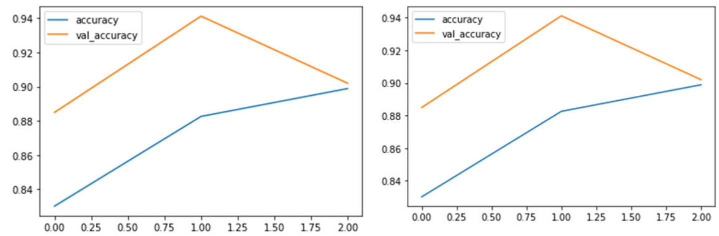
(a) Accuracy (b) Loss

Fig.4 Impact of Early Stopping on Back Propagation method



(a) Accuracy (b) Loss

Fig.5 Impact of Wight Decay on Levenberg–Marquardt method



(a) Accuracy (b) Loss

Fig.6 Impact of Early Stopping on Levenberg–Marquardt method

Table 2 presents the difference between test accuracy (%) before and after using of various improvement methods.

Table 2 The difference between test accuracy (%) before and after using of various improvement methods.

Method	***Before using improvement method (%)***	***After using improvement method (%)***
First Order (Back Propagation)	92%	Weight Decay (96%) Early Stopping (93%)
Second Order (Levenberg–Marquardt)	86%	Weight Decay (90%) Early Stopping (90%)

Comparing with many other models using the same dataset (BreakHis Database) with different approaches we get the high number of accuracy, The model used in (Nguyen.,2019) [7] was based on the classification of the data set into 8 subcategories by using convolutional neural networks,the dataset was divided into 90% train set and 10% test and they reached an accuracy of 73.68% Which is considered a result much less than the result that we reached through our experiments .

Table 3 Presents the results of of various models using BreaKHis dataset.

Paper	***Model***	***Accuracy (%)***
(Nguyen.,2019) [7]	CNN	73.68%
(Araújo.,2017) [8]	CNN and SVM	77.8%
(Rakhlin.,2018) [9]	Deep CNN	87.2%
(Iesmantas.,2018) [10]	Capsule Network	87%
(Jiang, Y,2019) [11]	ANN	70.49%
(Bardou, D ., 2018) [12]	CNN	93.81%
(Spanhol.,2015) [13]	SVM and CNN	88.23%
Our Model	CNN	96%

6. Conclusion

We conducted an experiment using the CNN method to classify breast cancer images on the BreakHis dataset in this paper. In the BreakHis dataset, there are two subclasses of breast cancer: benign and malignant. For the precise properties of the input image, the accuracy obtained from a building model is adequate at 96%. We were able to get a high accuracy due to our use of Weight Decay weight decay regularization, weight Decay can help with generalization in two ways: first, it will choose the smallest vector to solve the learning problem, and second, if the size value is chosen correctly, it will be able to suppress some of the static noise effect on the targets [30].

In future work, we will expand our selection of methods that will help us achieve greater accuracy than what we have obtained.

References

[1] "Breast Cancer." World Health Organization. World Health Organization. Accessed April 24, 2021. https://www.who.int/news-room/fact-sheets/detail/breast-cancer.

[2] Boyle, Peter, and Bernard Levin. World cancer report 2008. IARC Press, International Agency for Research on Cancer, 2008.

[3] Jaglan, Poonam, Rajeshwar Dass, and Manoj Duhan. "Breast cancer detection techniques: issues and challenges." Journal of The Institution of Engineers (India): Series B 100, no. 4 (2019): 379-386.

[4]Iranmakani, Sepideh, Tohid Mortezazadeh, Fakhrossadat Sajadian, Mona Fazel Ghaziani, Ali Ghafari, Davood Khezerloo, and Ahmed Eleojo Musa. "A review of various modalities in breast imaging: technical aspects and clinical outcomes." Egyptian Journal of Radiology and Nuclear Medicine 51 (2020): 1-22.

[5]Rawat, Waseem, and Zenghui Wang. "Deep convolutional neural networks for image classification: A comprehensive review." Neural computation 29, no. 9 (2017): 2352-2449.

[6] Sarvamangala, D. R., and Raghavendra V. Kulkarni. "Convolutional neural networks in medical image understanding: a survey." Evolutionary intelligence (2021): 1-22.

[7] Nguyen, Phu T., Tuan T. Nguyen, Ngoc C. Nguyen, and Thuong T. Le. "Multiclass breast cancer classification using convolutional neural network." In 2019 International symposium on electrical and electronics engineering (ISEE), pp. 130-134. IEEE, 2019.

[8] Araújo, Teresa, Guilherme Aresta, Eduardo Castro, José Rouco, Paulo Aguiar, Catarina Eloy, António Polónia,

and Aurélio Campilho. "Classification of breast cancer histology images using convolutional neural networks." PloS one 12, no. 6 (2017): e0177544.
[9] Rakhlin, Alexander, Alexey Shvets, Vladimir Iglovikov, and Alexandr A. Kalinin. "Deep convolutional neural networks for breast cancer histology image analysis." In international conference image analysis and recognition, pp. 737-744. Springer, Cham, 2018.
[10] Iesmantas T, Alzbutas R. Convolutional capsule network for classification of breast cancer histology images. InInternational Conference Image Analysis and Recognition 2018 Jun 27 (pp. 853-860). Springer, Cham.
[11] Spanhol, Fabio Alexandre, Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. "Breast cancer histopathological image classification using convolutional neural networks." In 2016 international joint conference on neural networks (IJCNN), pp. 2560-2567. IEEE, 2016.
[12] Bardou, Dalal, Kun Zhang, and Sayed Mohammad Ahmad. "Classification of breast cancer based on histology images using convolutional neural networks." Ieee Access 6 (2018): 24680-24693.
[13] Spanhol, Fabio A., Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. "A dataset for breast cancer histopathological image classification." Ieee transactions on biomedical engineering 63, no. 7 (2015): 1455-1462.
[14] Junqueira, Luis C., and Anthony L. Mescher. "Junqueira's basic histology: text & atlas/Anthony L. Mescher." (2013).
[15] Battiti, Roberto. "First-and second-order methods for learning: between steepest descent and Newton's method." Neural computation 4, no. 2 (1992): 141-166.
[16] Hossain, Md, Md Rahman, Uzzal Kumar Prodhan, and Md Khan. "Implementation of back-propagation neural network for isolated Bangla speech recognition." arXiv preprint arXiv:1308.3785 (2013).
[17] Hu, Yuzheng, Licong Lin, and Shange Tang. "Second-order Information in First-order Optimization Methods." arXiv preprint arXiv:1912.09926 (2019).
[18] Meza, Juan C. "Steepest descent." Wiley Interdisciplinary Reviews: Computational Statistics 2, no. 6 (2010): 719-722.
[19] Tan, Hong Hui, and King Hann Lim. "Review of second-order optimization techniques in artificial neural networks backpropagation." In IOP Conference Series: Materials Science and Engineering, vol. 495, no. 1, p. 012003. IOP Publishing, 2019.
[20] Goh, Anthony TC. "Back-propagation neural networks for modeling complex systems." Artificial Intelligence in Engineering 9, no. 3 (1995): 143-151.
[21] Hoi, Steven CH, Doyen Sahoo, Jing Lu, and Peilin Zhao. "Online learning: A comprehensive survey." arXiv preprint arXiv:1802.02871 (2018).
[22] Ranganathan, Ananth. "The levenberg-marquardt algorithm." Tutoral on LM algorithm 11, no. 1 (2004): 101-110.
[23] Zhang, Guodong, Chaoqi Wang, Bowen Xu, and Roger Grosse. "Three mechanisms of weight decay regularization." arXiv preprint arXiv:1810.12281 (2018).
[24] Paperswithcode.com. 2021. Papers with Code - Weight Decay Explained. [online] Available at: https://paperswithcode.com/method/weight-decay [Accessed 24 April 2021].
[25] Mahsereci, Maren, Lukas Balles, Christoph Lassner, and Philipp Hennig. "Early stopping without a validation set." arXiv preprint arXiv:1703.09580 (2017).
[26] LaptrinhX. 2021. A Practical Introduction to Early Stopping in Machine Learning. [online] Available at: https://laptrinhx.com/a-practical-introduction-to-early-stopping-in-machine-learning-2456645136 [Accessed 24 April 2021].
[27] Medium. 2021. Convolutional Neural Network for Breast Cancer Classification. [online] Available at: https://towardsdatascience.com/convolutional-neural-network-for-breast-cancer-classification-52f1213dcc9 [Accessed 24 April 2021].
[28] Chollet F. keras.
[29] TensorFlow. 2021. TensorFlow. [online] Available at: https://www.tensorflow.org [Accessed 24 April 2021].

Hanaa Alamri received her bachelor's degree in Computer Science with sconed class honours from Umm Al-Qura University, in the Kingdom of Saudi Arabia in 2018. She is currently studying a master's degree in Artificial Intelligence at Umm Al-Qura University, in the Kingdom of Saudi Arabia. She is interested in scientific research in data visualization, neural networks, and various fields of artificial intelligence.

Eman Alshanbari have a bachelor's degree in Computer Science from Taif University, in the Kingdom of Saudi Arabia in 2017. She is currently studying a Master's degree in Artificial Intelligence at Umm Al-Qura University, in the Kingdom of Saudi Arabia. She is interested in scientific research in various fields of artificial intelligence.

Walaa Alzhrani have a bachelor's degree in computer science from PNU. Saudi Arabia in 2018. Her study involved Java programing language,structure of data and computer. Currently she is magister student at artificial intelligence specialization in UQU Saudi Arabia. Her research's interest in learning algorithms and its application in different fields.

Manal Algamdi received her Ph.D. degree in Computer Vision from University of Sheffield, United Kingdom in 2015. Her study involved video representation and video similarity measurements. Currently, she is an Associate Professor at the Department of Computer Science, UQU, Saudi Arabia. Manal's research interests include deep learning and its application in different areas of computer vision including HealthCare applications.