CNN Model with MDWT2 and different parameters for Early Breast Cancer diagnosis

Sema Nizam Abdulghani[1,*], Maha Mohammed Attieya[2], Ahmed Freidoon Fadhil[3], and Seyfettin Sinan Gültekin[4]

[1] Department of Physics, University of Kirkuk

semanizam@uokirkuk.edu.iq

[2] Department of Information Technology, University of Kirkuk

mahamohammed@uokirkuk.edu.iq

[3] Electrical Engineering Department, University of Kirkuk

ahmedfadhil83@uokirkuk.edu.iq

[4] Electrical and Electronics Engineering Department, Konya Technical University

ssgultekin@ktun.edu.tr

https://doi.org/10.31972/iceit2024.049

Abstract

Breast Cancer is the most common cancer among women worldwide. The most widely used method for diagnosis of this types of cancer is the Histopathological analysis. Many researchers focused on developing computer-aided diagnosis system to support pathologist experience for correct diagnoses. This paper proposes an improved model for CNN using the publically available (BreakHis dataset) to categorize breast cancer histopathological images. The proposed model uses a 2-dimensional discrete wavelet transform (DWT) for feature extraction. Then, the approximate sub-band will be used for training and testing of the CNN instead of the original raw images. It is observed that the use of DWT features attain better results than the use of raw images itself. For instance, proposed MDWT2 CNN method has showed a better performance compared to the previous published of binary classification task with a performance range between 88.9% and 89.9% at image levels. A further investigation was implemented using the step scheduler with increasing mini-batch size for updating Stochastic Gradient Descent (SGD) network parameters at the early stages of the training phase. The results show that the proposed CNN based DWT model achieves the highest accuracy of 90.8% at the patient level and 89.1% at image level, thus outperforms all the previous published approaches. Therefore, the

researchers could suggest using the proposed MDWT2 CNN model for future investigation of breast cancer diseases.

Keywords: Breast Cancer; Convolutional Neural Networks; Discrete Wavelet Transform; and Stochastic Gradient Descent

Introduction

Currently, breast cancer is a substantial public health problem and the second cause of death for women around the world (Boyle, P. et al, 2008). Among many molecular markers, the most widely used method for diagnosis of this types of cancer is the Histopathological analysis. Pathologists use visual inspection of the samples under a microscope for grading and staging the Breast Cancer diagnosis (Lakhani, S. R., 2012). Histopathological analysis done by the pathologists is a time-consuming task with an accuracy rate depends on the experience of the pathologist. For this reason, Computer-Aided Detection/Diagnosis (CAD) systems were needed to assist the pathologists in the diagnosis process. Recently, the development in machine learning and digital image processing fields allow researchers to construct an advanced CAD system that supports pathologists in their diagnosis. Building an automated image processing system for breast cancer diagnosis has been considered a challenging task for researchers in the last years due to the complexity of the images (Das, A. et al, 2020). A well-designed CAD system for histopathological image classification can improve the diagnosis accuracy and make the pathologists more productive, objective and consistent in diagnosis (Araújo, T. et al, 2017).

Filipczuk P. et al, (2013), analyzed images of fine needle biopsies using a BC diagnosis system with a four different classifiers and reported performance of 98% on 737 images. Similarly, George Y. M.et al, (2013), presented a BC diagnosis system based on the nuclei segmentation of cytological images using neural networks and support vector machines reporting accuracy ranging from 76% to 94% on 92 images. These recent works focused on a small data set and Whole-Slide Imaging (WSI) imaging which still has many drawbacks, including the complexity and many unsolved technologies related concerns (Evans, A. J, 2015).

Spanhol F. A. et al, (2016), introduced an extensive histopathological breast data set of 7,909 images obtained from 82 patients. The authors evaluated six different textural descriptors

and different classifiers and achieved accuracy ranging from 80% to 85%, depending on the image magnification factor. Four different classifiers 1-nearest neighbor (1-NN), quadratic linear analysis (QDA), support vector machines (SVM), and random forests (RF) were used. The higher recognition rate (85.1%) was achieved for (200X factor) by the SVM classifier trained with the Parameter-Free Threshold Adjacency Statistics (PFTAS) descriptors.

The Convolutional Neural Network (CNN) is widely used in many pattern recognition problems (Niu, X. X. et al, 2012). In Spanhol F. A. et al, (2016), assessed the use of deep learning approach merging different CNNs using simple fusion rules and reported an improvement in classification accuracy of 6% when compared to the experiments reported in (Spanhol F. A. et al, 2016). Besides, the authors in Spanhol F. A. et al, (2017), used DeCAF features with previously trained CNN achieving an accuracy of 84% on breast cancer images. Another deep classifier is represented by the supervised intra-embedding of Fisher vectors with a multilayer neural network model, followed by a CNN in (Song Y. et al, 2017, Mewada, H., 2024).

Even though Development in machine learning in the classification of the histopathological images for breast cancer, the over-fitting of the system due to increase of CNN parameters is still a challenging problem. Although increasing the data set of images overcomes the over-fitting problem, it increases the complexity of the system and consumes time. The main aim of this paper is to reduce the complexity of an existing CNN architecture model by using the approximation sub-band of the wavelet domain instead of the image itself. This sub-band reduces the size of input images which reduces the number of image patches from the dataset.

Methodology

Microscopic images of histopathologic sections is usually classified visually. Thus, image classification is a challenging problem due to presence of a large amount of geometrical structures and complex textures. In deep learning, high-level features can be represented by using multiple levels of representation of the data (LeCun Y. et al, 2015). Among other deep learning techniques, Convolutional Neural Networks (CNNs) have achieved improved results in image classification problems in analyzing of medical images (Yamashita R. et al, 2018). The

CNN includes multi-layers which has the ability to be trained, arranged on each other. Then, sets of arrays named feature maps administered as a classifier which represent both input and output for every stage (Sakkari, M. et al, 2020). Mainly, CNN architectures consist of three types of layer: convolutional layer, pooling layer, and fully connected layer. The design of CNN could be set depending on the types and numbers of layers included, depending on the application or data (Zainel, Q.M., et.al, 2021). Since the data set consisting of complex colored images, a CNN with multiple convolutional and fully connected layers was required. The CNN layers architecture presented in this paper is shown in detail in Table 1.

Table 1. The CNN Layers

CNN Layer	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
Layer Type	**CONV + POOL**	**CONV + POOL**	**CONV + POOL**	**FC**	**FC**
Conv Feature maps	**32**	**32**	**64**	**-**	**-**
Conv Filter Size	**5 x 5**	**5 x 5**	**5 x 5**	**-**	**-**
Conv Output Size	**60 x 60 x 32**	**27 x 27 x 32**	**11 x 11 x 64**	**-**	**-**
Conv Stride	**1 x 1**	**1 x 1**	**1 x 1**	**-**	**-**
Pooling Type	**Max**	**Average**	**Average**	**-**	**-**
Pooling Size	**3 x 3**	**3 x 3**	**3 x 3**	**-**	**-**
Pooling Stride	**2 x 2**	**2 x 2**	**2 x 2**	**-**	**-**
Padding Size	**2 x 2**	**2 x 2**	**2 x 2**	**-**	**-**
Pooling Output Size	**31 x 31 x 32**	**15 x 15 x 32**	**7 x 7 x 64**	**-**	**-**
FC Input Size	**-**	**-**	**-**	**3136**	**64**
FC Output Size	**-**	**-**	**-**	**64**	**2**

Convolution with a collection of filters improves the representation of the features. At the first layer, simple primitive features computed across all the input channels. Then, a pooling layer will be place in between two succeeding convolutional layers to reduce the size, which

helps to prevent over-fitting. Although using max function for pooling layer has been shown good results, the average function can also be used. Neurons of fully connected layer will be fully connected to all other activations from the previous layer. The last fully connected layer will hold the probability distribution of the classes (Nannia, L. et al, 2018).

The Discrete Wavelet Transform (DWT), for a signal x, which is mathematically expressed in equation below, is calculated by passing it through a low pass and high pass filters separately.

$$y_{low}[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k].g[n-k] \quad (1)$$

$$y_{high}[n] = (x * h)[n] = \sum_{k=-\infty}^{\infty} x[k].h[n-k] \quad (2)$$

Where, the first step represent passing the signal x to a low pass filter g, while, the second step is passing the signal with a high pass filter h. These two filters should be connected to each other in a method named a quadrature mirror filter. According to Nyquist's rule, the outputs of the quadrature mirror filter will be halved (Fadhil, A. F., 2014). Usually, the 1-dimensional signal will passed twice to generate a 2-dimensional signals, such as images. The image is passed through a low pass and high pass filters to acquire two sub-bands, namely, approximate and detail sub-bands. The detail sub-band include of vertical, horizontal, and diagonal sub-bands.

The proposed method deals with small image patches instead of large images to reduce the complexity and the extensive set of parameters of the system. Also, the time that is necessary for training the parameters was reduced. The image sizes reduced to half first then patches were extracted by a sliding window with 25% of overlapping between patches resulting of 28 patches by each image. Also, the proposed model implements data normalization. This is achieved by subtracting the mean image of the training set from every input image. The patch images were passed through 2-dimensional DWT, and only the approximate sub-bands were selected. The approximate sub-band contains most of the information from the image and can be used instead of the original image pixels in many applications. The approximate sub-band will be used instead of the raw pixel values of the original image to reduce the size of the images, which reduces the time consumption of the system. On the other hand, Stochastic Gradient Descent (SGD) method

(Bottou, L., 2012), with back-propagation to compute gradients and different values for learning rate and mini-batch size was used to change the network's parameters every time (Abdulghani, S., et.al., 2020).

Finally, in the classification stage, results for each image can be conducted by combining the patch results. The grid patches of the images were extracted and allowed 25% of overlapping. This is the same strategy used for the training which in practice proved a sensible balance between classification performance and computational cost.

Experimental Results and Discussions

About 7,909 microscopic biopsy images from the BreaKHis database (Spanhol, F. A. et al, 2016) collected from 82 patients containing both benign and malignant breast tumors. The benign class consists of 4 different types: adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenoma (TA); and the malignant class also consists of 4 different types: ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC). The main aim in this work is classifying images into benign or malignant cases at patch, image, and patient levels. Besides, five-fold cross-validation method at the patient level were used to evaluate results. Furthermore, five splits sets of dataset has been initially separated for cross- validation. Then, each split set divided into 80% of images for training and 20% of images for testing. Finally, the original dataset split patient wise were used to ensure that the classifier covered all patients including unseen patients. The researchers investigated the effect of changing mini-batch size and the learning rate parameters of the Stochastic Gradient Descent network.

The results of fixed, decreasing and increasing the mini-batch numbers are presented in Tables 2, 3 and 4, respectively. In Table 2, the mini-batch size was fixed at a value of (100) every 5 epochs with a learning rate of (10−2). Then in Table 3, the mini-batch size was started with a value of (100) then decreased by (25) each 5 epochs with a learning rate of (10−2). While in Table 4, the mini-batch size is increased by (50) every 5 epochs.

Table 2: The Average Recognition Accuracy presented for MDWT2 method (fixed mini-batch size)

MF	Max Epoch	Mini-batch	Learning Rate	Recognition Rate		
				Patch	Image	Patient
40 X	5	100	10^{-2}	81.54	82.68	84.06
	10	100	10^{-2}	84.27	85.52	87.99
	15	100	10^{-2}	82.58	83.52	86.03
	20	100	10^{-2}	**85.73**	**87.19**	**87.99**

Table 3: The Average Recognition Accuracy presented for MDWT2 method (decreased mini-batch size)

MF	Max Epoch	Mini-batch	Learning Rate	Recognition Rate		
				Patch	Image	Patient
40 X	5	100	10^{-2}	81.54	82.69	84.07
	10	75	10^{-2}	83.56	84.17	86.03
	15	50	10^{-2}	83.23	84.71	86.03
	20	25	10^{-2}	**84.06**	**85.26**	**85.91**

Table 4: The Average Recognition Accuracy presented for MDWT2 method (increased mini-batch size)

MF	Max Epoch	Mini-batch	Learning Rate	Recognition Rate		
				Patch	Image	Patient
40 X	5	100	10^{-2}	81.54	82.69	84.07
	10	150	10^{-2}	84.39	86.12	87.99
	15	200	10^{-2}	83.72	85.79	87.87
	20	250	10^{-2}	**84.27**	**85.64**	**87.99**

From Tables 2, 3 and 4, fixed mini-batch size overcome the results of both increasing and decreasing mini-batch size. Therefore, we could suggest using the fixed mini-batch size for further investigation. Furthermore, the impact of the learning rate on model performance is investigated. A dynamic learning rate strategy was suggested and conducted in this research. The step scheduler is the most used for scheduling the SGD learning rates. In the step scheduler, four hyper-parameters need to be fine-tuned carefully: initial learning rate, training epochs, decay stages, and decay rate. It updates the learning rate every specified number of epochs by multiplying with a specific factor. In Table 5, the step scheduler starts the learning rate from (10−2), then drops by a factor (10−1) every 5 epochs with mini-batch size = 100.

Table 5: The Average Recognition Accuracy presented for MDWT2 method (dynamic learning rate)

MF	Max Epoch	Mini-batch	Learning Rate	Recognition Rate		
				Patch	Image	Patient

40 X	5	100	10^{-2}	81.54	82.69	84.07
	10	100	10^{-3}	84.47	85.50	85.91
	15	100	10^{-4}	84.44	85.20	87.99
	20	100	10^{-5}	84.33	85.13	87.99

In Table 5, the accuracy rates decrease after 10 or 15 epochs when decreasing the learning rate. These results state that using a smaller learning rate every 5 epochs is not enough for the model to learn effectively with a fixed mini-batch number. These results lead to an assumption that changing the learning rate does not work correctly alone while fixing the mini-batch value. Therefore, a different scenario was used in Table 6 by increasing the mini-batch size with a factor of (100) and using the same previous step scheduler every 10 epochs.

Table 6: The Average Recognition Accuracy presented for MDWT2 method (increasing mini-batch size with dynamic learning rate)

MF	Max Epoch	Mini-batch	Learning Rate	Recognition Rate		
				Patch	Image	Patient
40 X	5	100	10^{-2}	81.54	82.69	84.07
	10	100	10^{-2}	84.27	85.52	87.99
	15	200	10^{-3}	84.91	86.72	89.95
	20	200	10^{-3}	**85.36**	**89.10**	**90.81**

Results from Table 6 show that when mini-batch size changes in conjunction with learning rate, the combination yields better results. The accuracy ranges from approximately 85%, 89% and 91% for Patch, Image and Patient recognition rates, respectively.

We Also compared the results presented in this paper, Table 6 , with the results of previous works (Spanhol, F. A., et al 2016 A), (Spanhol, F. A. et al , 2016 B), (Spanhol, F. A. et al 2017), and (Song Y., 2017) at both patient and image levels. Table 7 shows the average accuracy of the proposed methods at patient and image levels for five trails using five-fold cross-validation. The proposed MDWT2 with step error scheduler achieves the highest accuracy of 90.8% at the patient level and 89.1% at image level, thus outperforms all the previous published approaches.

Table 7: Mean Accuracy rates for different strategies

Level	Strategy	MF of 40X

Patient Level	PFTAS (1-NN) [8]	80.9
	PFTAS (QDA) [8]	83.8
	PFTAS (RF) [8]	81.8
	PFTAS (SVM) [8]	81.6
	CNN (Sum) [10]	88.4
	CNN (Product) [10]	89.2
	CNN (Max) [10]	90.0
	CNN + DeCAF [11]	88.5
	Intra-embedding Fisher Vectors [12]	90.2
	Proposed (MDWT2)	**90.8**
Image Level	CNN (Sum) [10]	85.4
	CNN (Product) [10]	85.5
	CNN (Max) [10]	85.6
	CNN + DeCAF [11]	88.0
	Intra-embedding Fisher Vectors [12]	87.7
	Proposed (MDWT2)	**89.1**

Conclusions

In this paper, a new CNN based 2-dimensional DWT model (MDWT2) was proposed and evaluated on BreaKHis dataset. The main observation was using the DWT features can attain better results than the use of raw images on the same CNN architecture. The performance of the suggested method was assessed in patient, image, and patch based levels. Furthermore, The SGD network parameters were changed to investigate the performance of the system at the early stages of the training phase. The step scheduler for fine-tuning the learning rate parameter for SGD algorithm with increasing mini-batch size presented the best accuracy results with the proposed algorithm.

The experimental results presented in patient levels using five-fold cross-validation shows improvement of average recognition accuracy on the BreakHis dataset against the other machine learning models tested on the same dataset. Also, the proposed MDWT2 CNN method has showed a better performance compared to the previous published of binary classification task

with a performance range between 88.9% and 89.9% at image levels. Therefore, we could suggest using the proposed MDWT2 CNN model for future investigation of breast cancer diseases.

Acknowledgements

The authors are grateful to Spanhol, Fabio A., et al. for establishing the BreakHis dataset and making data publically available for researchers.

Conflict of interests

The authors declare that there is no conflict of interests or financial interests or personal relationships presented in this paper.

References

Abdulghani, S., Fadhil, A., & Gültekin, S. S. (2020). Transfer Learning using Alexnet with Support Vector Machine for Breast Cancer Detection. Avrupa Bilim Ve Teknoloji Dergisi423-430. https://doi.org/10.31590/ejosat.806679

Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., & Campilho, A. (2017). Classification of breast cancer histology images using convolutional neural networks. PloS one, 12 (6).

Bottou, L. (2012). Stochastic gradient descent tricks. In Neural networks: Tricks of the trade (pp. 421-436).Springer, Berlin, Heidelberg.

Boyle, P. Levin B, editor. (eds). World Cancer Report 2008. Lyon, France: World Health Organization. International Agency for Research on Cancer.

Das, A., Nair, M. S., & Peter, S. D. (2020). Computer-Aided Histopathological Image Analysis Techniques for Automated Nuclear Atypia Scoring of Breast Cancer: a Review. Journal of Digital Imaging, 1-31.

Evans, A. J., Krupinski, E. A., Weinstein, R. S., & Pantanowitz, L. (2015). 2014 American Telemedicine Association clinical guidelines for telepathology: another important step in support of increased adoption of telepathology for patient care. Journal of pathology informatics, 6.

Fadhil, A. F. (2014). Formulation of detection strategies in images. Thesis, Southern Illinois University atCarbondale.

Filipczuk, P., Fevens, T., Krzyżak, A., & Monczak, R. (2013). Computer-aided breast cancer diagnosis basedon the analysis of cytological images of fine needle biopsies. IEEE transactions on medical imaging, 32 (12), 2169-2178.

George, Y. M., Zayed, H. H., Roushdy, M. I., & Elbagoury, B. M. (2013). Remote computer-aided breast cancer detection and diagnosis system based on cytological images. IEEE Systems Journal, 8 (3), 949-964.

Lakhani, S. R. (Ed.). (2012). WHO Classification of Tumors of the Breast. International Agency for Research on Cancer.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

Mewada H. Extended Deep-Learning Network for Histopathological Image-Based Multiclass Breast Cancer Classification Using Residual Features. Symmetry. 2024; 16(5):507. https://doi.org/10.3390/sym16050507.

Nannia, L., Ghidoni, S., & Brahnam, S. (2021). Ensemble of convolutional neural networks for bioimageclassification. Applied Computing and Informatics, 17(1), 19-35.

Niu, X. X., & Suen, C. Y. (2012). A novel hybrid CNN–SVM classifier for recognizing handwritten digits. Pattern Recognition, 45 (4), 1318-1325.

Sakkari, M., & Zaied, M. (2020). A Convolutional Deep Self-Organizing Map Feature extraction for machinelearning. Multimedia Tools and Applications, 79(27), 19451-19470.

Song Y., Chang H., Huang H., Cai W. (2017) Supervised Intra-embedding of Fisher Vectors for Histopathology Image Classification. In: Descoteaux M., Maier-Hein L., Franz A., Jannin P., Collins D., Duchesne S. (eds) Medical Image Computing and Computer Assisted Intervention − MICCAI 2017. MICCAI 2017. Lecture Notes in Computer Science, vol 10435. Springer, Cham.

Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016 A). A dataset for breast cancer histopathological image classification. IEEE Transactions on Biomedical Engineering, 63 (7), 1455-1462.

Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016 B). Breast cancer histopathological image classification using convolutional neural networks. In 2016 international joint conference on neural networks (IJCNN) (pp. 2560-2567). IEEE.

Spanhol, F. A., Oliveira, L. S., Cavalin, P. R., Petitjean, C., & Heutte, L. (2017, October). Deep features for breast cancer histopathological image classification. In 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 1868-1873). IEEE.

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overviewand application in radiology. Insights into imaging, 9(4), 611-629.

Zainel, Q.M., Khorsheed, M.B., Darwish, S., Ahmed, A.A. (2021). An optimized convolutional neural

network architecture based on evolutionary ensemble learning. Computers, Materials & Continua, 69(3), 3813-3828. https://doi.org/10.32604/cmc.2021.014759.