



Breast Cancer Detection Based DenseNet with Attention Model in Mammogram Images

Tawfik Ezat Mousa^(✉), Ramzi Zouari, and Mouna Baklouti

National School of Engineering of Sfax, University of Sfax, Sfax, Tunisia
tawfikezat@yahoo.com, mouna.baklouti@enis.tn

Abstract. Breast cancer has become a very interesting topic due to the massive number of deaths among women across the world. Radiologists can diagnose breast cancer faster and more accurately because of advances in the computer-aided diagnosis (CAD) system. In this paper, we presented a new breast cancer detection system based on the integration of self attention model in the pre-trained deep neural networks DenseNet. First, we extracted automatic high-level features from breast images using DenseNet extraction layers, and thereafter attention model was applied to focus the treatment on the relevant parts of the region of interest. The experiments were conducted on a multi-class Mammographic Image Analysis Society (MIAS) database, including three classes of breast cancer images. We achieved the accuracy of 0.9939 when applying both transfer learning, data augmentation, and self attention mechanism.

Keywords: Breast cancer · DenseNet · Self Attention · Detection · Mammogram

1 Introduction

Cancer is a disease that affects cells and its detection at an early stage increases the chances of recovery. In fact, breast cancer is one of the most common types among women, and breast cancer has always shown a very high incidence and mortality rate of about 10% of women at some point in their lives. It is the second-largest cause of death among females after lung cancer [1]. The World Health Organization's International Agency for Research on Cancer (IARC) reported an anticipated increase in the number of breast cancer cases to 1.1 million by 2030, with the gap between developed and developing nations expected to widen [2]. Cancer can be described as the uncontrolled proliferation of abnormal cells that form masses. These tumors can be benign or malignant. Benign tumors remain localized and grow slowly. Malignant tumors invade nearby structures and may destroy other parts of the body [3]. Therefore, mammography is an effective imaging technique used in the detection of breast cancer. It is the

most used imaging technique in screening programs [4]. It helps in detecting suspicious lesions like masses and micro calcification. However, mammography in a 2D image results in tissue overlap, which can mask the lesion, or create a false lesion, thus producing false-positive and false-negative results. In addition, mammography is known to be less sensitive to breast density (30–64%) compared to fatty breasts (76–98%), as it has been shown that women with dense breasts are more susceptible to breast cancer.

In the last decade, several works based artificial intelligence tools were developed to enhance computer-assisted breast cancer (CAD) diagnosis. These approaches have shown their ability to treat the problem of abnormal lumps and calcification in the breast and predict their growth. They help radiologists and oncologists diagnose breast cancer by providing a second opinion. In this work, we propose a new system of breast cancer detection using a pre-trained DenseNet, with the integration of attention mechanism. The combination of these two models has demonstrated their effectiveness in improving the detection performance. In fact, attention mechanism allows to increase the weights of the relevant features of the model and decreasing the others, to make a better decision. Furthermore, we applied data augmentation technique to increase the number of training images and to improve the model generalization.

The rest of this paper is organized as follows. Section 2 presents some related works in breast cancer detection field. Section 3 illustrates the different parts of the proposed methodology. Section 4 presents the experimental results and a comparison with similar works. We finish this paper by a conclusion and some prospects.

2 Related Works

In the literature, numerous approaches were proposed for breast cancer detection based on mammography images. Samee et al. [5] proposed a breast cancer detection system based on several deep learning architectures. In fact, they extracted automatic features from both AlexNet, VGG, and GoogleNet models. These features are fused to make the prediction task. This system was evaluated on INbreast database and achieved the accuracy of 98.50%. In other work, Jiang et al. [6] integrated a new dataset of breast mammograms named Film Mammography dataset number 3 (BCDR-F03). They applied both GoogLeNet and AlexNet models to classify segmented tumors found on mammograms, and obtained the accuracy of 88% and 83% for GoogleNet and AlexNet respectively. Ribli et al. [7] used Regional based CNN(R-CNN) model to detect and classify breast lesions using mammograms. They obtained the accuracy of 95% on INbreast dataset. Alruwaili et al. [8] used ResNet50 model to distinguish between malignant and benign breast cancer. Data augmentation technique was applied to increase the number of training images and prevent the model from over-fitting. The proposed model was assessed on MIAS dataset and achieved the accuracy of 89.5%. Kaur et al. [9] proposed an hybrid model where both deep learning neural networks (DNN) and Support Vector Machines (SVM) were used. SVM was implemented after the DNN classification part instead of regular dense layers. The results

showed that SVM allows improving the recognition rate from 70% to 96.9% on multiclass MIAS dataset. Mohapatra et al. [10] evaluated several pre-trained deep learning models such as AlexNet, VGG16, and ResNet50 on mammogram images. Due to the limited number of training images, they applied data augmentation to address the problem of over-fitting. The experiments were done on Mini-DDSM dataset and reached the accuracy of 65% when using ResNet50. Muduli et al. [11] proposed a deep convolution neural network (CNN) model for breast cancer classification using mammograms and ultrasound images. This model facilitates the extraction of prominent features from the images with only few tune parameters. They applied data augmentation to increase the number of training images and prevent the model from over-fitting. The proposed model was evaluated on MIAS and INbreast datasets, and achieved the accuracy of 90.68% and 91.28% respectively. Rouhi et al. [12] proposed a new model of primary breast cancer detection using region growing method. Their model is based on the hybridization of cellular neural network with genetic algorithm, and achieved the accuracy of 96.47% and 95.13% on MIAS and DDSM databases respectively. In [13], transfer learning technique was applied with the pre-trained deep neural networks Inception V3, ResNet50, VGG16, and Inception-ResNet. The best result was obtained using VGG16 model which achieved the accuracy of 98.96% on MIAS database. Punithavathi et al. [14] proposed an hybrid model based on SVM and KNN classifiers. They introduced multiple categories of images to the SVM, and the final decision was done by KNN algorithm. This model produces higher diagnostic accuracy on MIAS dataset and achieved the accuracy of 99.34%. Pillai et al. [15] evaluated several pre-trained deep learning models such as EfficientNet, AlexNet, VGG16, and GoogleNet on MIAS database. They applied data augmentation to increase the number of training images and prevent the model from over-fitting. The best performance was obtained with VGG16 model and achieved the average accuracy of 75.46%. Chougrad et al. [16] applied fine-tuned Inception-v3 model on MIAS database to classify breast lesions and obtained the accuracy of 98.23%. Selvathi et al. [17] proposed a new system for breast cancer detection. Their approach consists of using a stack autoencoder architectures with softmax classifier. Moreover, they applied some processing on MIAS database images to remove noise, background, and pectoral muscle, and obtained the accuracy of 98.5%.

3 Proposed Methodology

In this section, we present our system for multi-class breast cancer detection based on mammogram images. The proposed methodology employs the pre-trained DenseNet121 model truncated at the feature extraction part, followed by an attention model to give more importance to relevant features of the Region of Interest (ROI). Thereafter, convolutions and attention modules are combined to fuse both the high-level information and the interesting semantic information. The obtained features are fed into a Global Average Pooling (GAP) to reduce the feature maps dimensions and preserve pertinent features for the classification part.

3.1 DenseNet121 Architecture

Dense Convolution Network (DenseNet) is a modern CNN architecture designed for visual object recognition with only few parameters [18]. It achieved the state-of-the-art results on several image classification datasets, such as CIFAR-10, SVHN and ImageNet [19]. The basic structure of the network mainly includes two-component modules: Dense and Transition blocks (Fig. 1). In DenseNet-121, there are a total of 4 dense blocks and 3 transition blocks. Each layer in the Dense Block is connected to all subsequent layers in a densely manner [20]. Moreover, each dense block is composed of a stack of two convolution layers with a kernel size of (1×1) and (3×3) respectively. In each transition block, (1×1) convolution and (2×2) max pooling operations are done. Table 1 shows the overall architecture of DenseNet121 model. We notice that DenseNet121 alternates dense and transition blocks. At each pass, the convolution layers of the dense block are reproduced 6, 12, 24, and 16 times respectively.

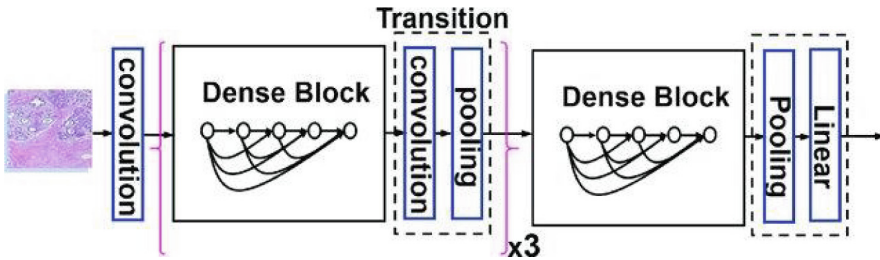


Fig. 1. DenseNet121 model concept [21]

Table 1. DenseNet121 structure

DenseNet121 parts	Layers	Parameters
Input	–	$(224 \times 224 \times 3)$
Extraction	convolution	kernel size = (7×7)
	Max Pooling	pool size = (2×2)
Dense Block ($\times 4$)	convolution	kernel size = (1×1)
	convolution	kernel size = (3×3)
Transition Block ($\times 3$)	convolution	kernel size = (1×1)
	Average pooling	pool size = (2×2)
Classification	Global Average Pooling	pool size = (7×7)

3.2 Self Attention Model

After the global average pooling layer, we implemented a Multi-Head Self Attention (MHSA) model to improve the model effectiveness (Fig. 2). In fact, MHSA

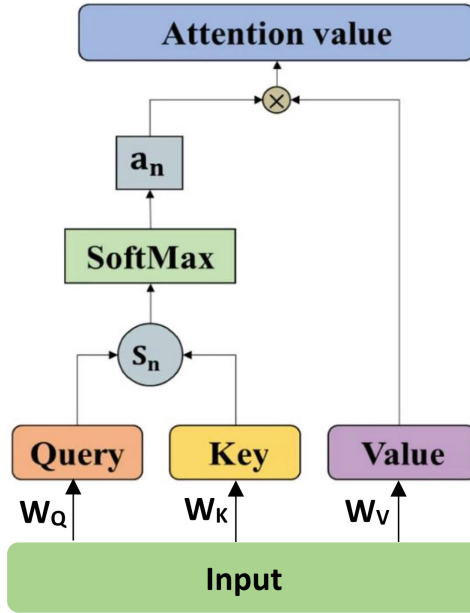


Fig. 2. Attention model architecture

is a mechanism used to provide an additional focus on a specific component in the data. It enables the network to concentrate on a few aspects at a time and ignore the rest [22]. MHSA consists of several attention layers running in parallel, instead of performing one single attention function. In particular, the input consists of queries and keys of dimension d_k (Q and K respectively), and values of dimension d_v (V). The output of the attention model is done by computing the scaled dot product of the queries with all keys and applying a SoftMax function to obtain the weights on the values V (Eq. 1). The attention mechanism is linearly projected h times with different learned weights (W_Q , W_K , W_V). These different representation sub spaces are concatenated into one single attention head to form the final output result (Eq. 2). We applied a particular version of attention model called self-attention, in which query, key and value inputs are the same. The calculation process follows these steps: First, we made the dot product (MatMul) of query and keys tensors and scale the obtained scores. Next, we apply a SoftMax function on these scores to obtain attention probabilities. Finally, we take a linear combination of these distributions with the value input tensors and concatenate them into one channel.

$$Attention(Q, K, V) = Softmax\left(\frac{Q \times K^T}{\sqrt{d_k}}\right) \times V \quad (1)$$

$$\begin{cases} \text{MHA}(Q, K, V) = \text{concat}(\text{head}_1, \dots, \text{head}_h) \\ \text{head}_{i=1..h} = \text{Attention}(QW_Q, KW_k, VW_V) \end{cases} \quad (2)$$

The proposed methodology consists of making the dot product of DenseNet121 and Self Attention models outputs. Thereafter, we applied Global Average Pooling on both attention model output and the resulting dot product tensors. The classification part is composed of two dense layers with dropout function to prevent the model from over-fitting. Figure 3 illustrates the different parts of the proposed breast cancer detection system.

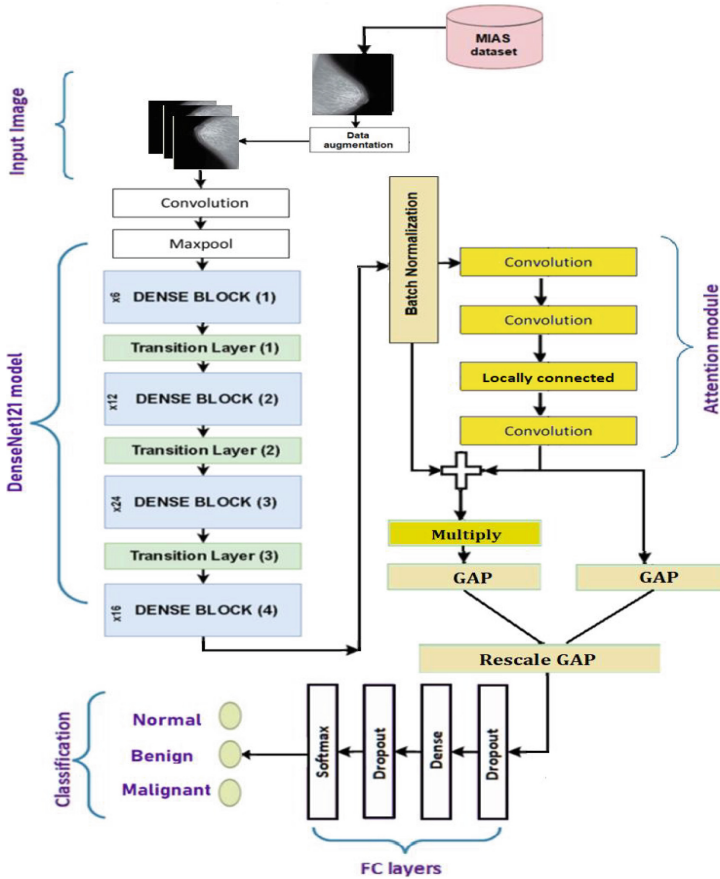


Fig. 3. Proposed methodology architecture

4 Experimentation and Results

4.1 Database Description

The proposed methodology was evaluated on MIAS multi-class database containing images of normal, benign, and malign breast cancer [23]. This database consists of 322 mammogram images of size (1024×1024) pixels and stored according to Portable Gray Map (PGM) format. These images belong to three types: glandular dense, fatty, and fatty glandular. Each type is divided into three categories: normal, benign, and malignant. The dataset also contains radiologists' actual estimations of the location of abnormalities (benign, malignant), with an approximate determination of the radius surrounding the center of the anomaly. In this work, we use all the images in the dataset, which consists of 207 normal images, 64 benign images, and 51 malignant images. Figure 4 shows three images from MIAS database representing three categories (Normal, Benign, and Malign).

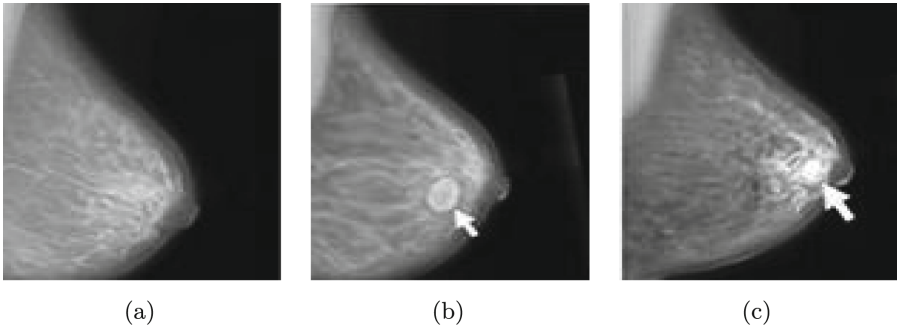


Fig. 4. MIAS database samples. (a) Normal (b) Benign (c) Malign

4.2 Data Augmentation

Since MIAS dataset contains only 322 images, the proposed model may not be generalized. For this purpose, we applied data augmentation operation to increase the number of training samples in each class and prevent the model from overfitting. In this work, data augmentation is mainly based on geometric transformations including rotation, flipping, and shifting. Thus, we obtained a new database of 1836 breast cancer images evenly distributed over the three classes (612 images per class). Figure 5 shows an example of data augmentation where vertical and horizontal flip were applied on the original image.

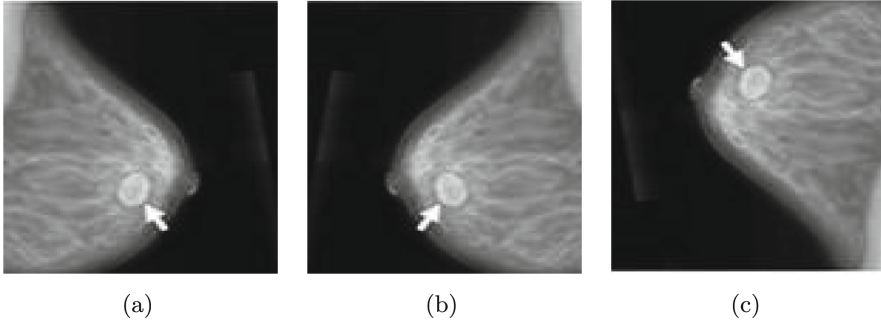


Fig. 5. Data augmentation samples. (a) original (b) horizontal flip (c) vertical flip

4.3 Experimental Setup

During the experiments, the training database was divided into batches of size 32, with shuffling option to make different min-batch samples in each epoch. Moreover, in each iteration categorical cross entropy method was used to compute the loss between desired and calculated outputs. The model was trained using Adam (Adaptive Moment Estimation) optimizer with an initial learning rate of 0,001. This value can be reduced by a factor of 0.5 once learning stagnates. Moreover, the early stopping approach is applied as a regularization method. It consists of stopping the training process early before it has over-fit the training database. In the multi-head self attention model, we employed 8 parallel attention layers or heads. For each of these, we used 64 units in the linear projector of both query, key, and value matrices (Table 2).

Table 2. Hyperparameters setting

Hyper parameter	Value
learning rate (lr)	0.001
lr decrease rate	0.5
optimizer	Adam
batch size	32
epochs	100
loss function	cross entropy

4.4 Evaluation Metrics

To illustrate the performance of the proposed model, the confusion matrix and other metrics were calculated like Accuracy, Recall, Precision, and F1-score (Eq.

3–6). They are all based on the calculation of True positive (TP), False positive (FP), False negative (FN) and True negative (TN) values. TP denotes images predicted with breast cancer when they were. TN relates to normal images predicted as healthy. FP concerns normal images which are predicted as breast cancer, and FN refers to images predicted as normal, but they were not.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$F1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

4.5 Experimental Results

In the experiments, the images shape was fixed to $(256 \times 256 \times 3)$. Moreover, several models were studied with different values of splitting and optimizers. All of these models have been used with pretrained weights. First, we evaluated the model’s performance without the use of self attention mechanism. The best result was obtained with DenseNet-121 (Table 3). When applying multi-head self attention mechanism, the DenseNet-121 accuracy was improved by 6%, and we reached the accuracy of 0.9939 for 90% of database split. On the other hand, several other metrics were evaluated such recall, precision, and AUC (Table 4). In all of these metrics, the best results have been obtained using DenseNet-121 model with Adam optimizer. Figures 6 and 7 represent the confusion matrices related to the classification report for different split ratios. We notice that the model performances was improved when using multi-head self attention mechanism. Moreover, the proposed model allows a good discrimination between benign and malign image samples, but it confuses between normal and benign classes (Table 5).

Table 3. Models accuracies without and with attention

Model	Accuracy (split ratio 90:10)	
	without Attention	with Attention
VGG-16	0.8712	0.9387
MobileNet	0.9027	0.9675
Xception	0.9029	0.9600
InceptionResNetv2	0.9363	0.9714
DenseNet-121	0.9405	0.9939

Table 4. DenseNet model performance with different split ratio

Split ratio (Train:Test)	Without Attention			With Attention		
	(70:30)	(80:20)	(90:10)	(70:30)	(80:20)	(90:10)
Accuracy	0.6363	0.8727	0.9405	0.7490	0.9264	0.9939
Recall	0.6363	0.8545	0.9297	0.7436	0.9202	0.9939
Precision	0.6375	0.8935	0.9398	0.7490	0.9259	0.9939
F1-score	0.6368	0.8735	0.9347	0.7462	0.9230	0.9939
AUC	0.7807	0.9539	0.9749	0.8964	0.9882	0.9978

Table 5. Performance results with optimizers

Optimizer	Accuracy	Recall	Precision	F1-score	AUC
Rmsprop	0.9693	0.9632	0.9691	0.9661	0.9983
Adam	0.9939	0.9939	0.9939	0.9939	0.9978
SGD	0.9387	0.9773	0.9597	0.9166	0.9857

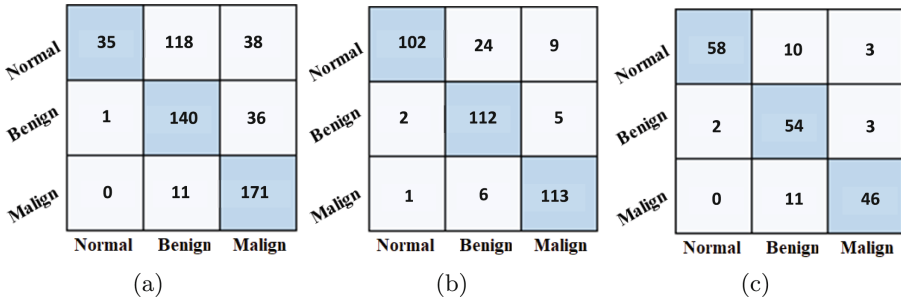


Fig. 6. Confusion matrices without self attention. (a) split ratio (70:30) (b) split ratio (80:20) (c) split ratio (90:10)

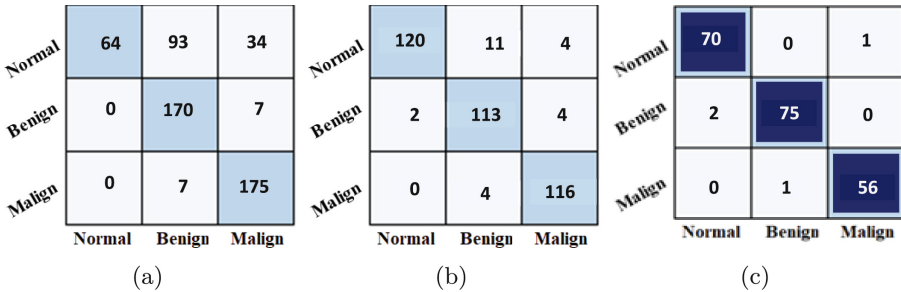


Fig. 7. Confusion matrices with self attention. (a) split ratio (70:30) (b) split ratio (80:20) (c) split ratio (90:10)

4.6 Comparative Study and Discussion

Table 6 summarizes several works evaluated on multi-class MIAS dataset. When applying the split ratio of 90% and multi-head self attention mechanism, the proposed model achieves the state-of-the-art performances on MIAS dataset, and outperforms the models based on ADL-BCD and ResNet50. However, in the case of split ratio of 80%, the proposed approach is better than DenseNet-201 model, and it is slightly less efficient than VGG16 and OMLTS-DLCN approaches. Furthermore, the proposed work is the only one to combine multi-head self attention mechanism with the pre-trained deep neural networks DenseNet-121. This combination has led to significant improvement in the classification rates. In fact, the attention model was frequently applied to sequential data. In this work, we turned it to image classification task to associate high attention weights to the parts of images with relevant features.

Table 6. Results comparison on MIAS database

Authors	Model	Split ratio	Accuracy
Alruwaili et al. [8]	ResNet50	90	89.5
Saber et al. [13]	VGG16	80	98.96
Zeng et al. [24]	DenseNet-201	80	92.73
Kvitha et al. [25]	OMLTS-DLCN	80	98.50
Maqsood et al. [26]	TTCNN	60	96.57
Gutierrez et al. [27]	ADL-BCD	90	96.07
Jebarani et al. [28]	GMM + K-means	70	95.5
Proposed work	DenseNet121 + Attention	70	74.90
		80	92.64
		90	99.39

5 Conclusion

In this paper, we proposed a deep architecture for breast cancer classification based on mammographic images to help medical doctors in breast cancer detection and diagnosis. The approach provides the breast image classification into normal, benign, and malignant. The virtue of our method is to combine pre-trained deep convolution neural networks DenseNet121 with a self-attention model. Moreover, data augmentation was applied to increase the number of images and prevent the model from overfitting. During the experiments, several hyper-parameters were tuned such as optimizer and learning rate to boost the diagnostic efficiency. The proposed methodology achieved the accuracy of 92.64% and 99.39% for a split ratio of 80% and 90% respectively. Finally, it

can be concluded that by integrating the CNN using learning transfer with the attention mechanism, a clear improvement was achieved compared with other existing approaches. The results presented in this study open new windows for the use of self-attention-based architectures with vision transformer technology for breast cancer classification to obtain high-performance CAD schemes with better results.

References

1. Ara, S., Das, A., Dey, A.: Malignant and benign breast cancer classification using machine learning algorithms. In: 2021 International Conference on Artificial Intelligence (ICAI), pp. 97–101. IEEE (2021)
2. Krithiga, R., Geetha, P.: Deep learning based breast cancer detection and classification using fuzzy merging techniques. *Mach. Vision Appl.* **31**, 1–18 (2020)
3. Elmannai, H., Hamdi, M., AlGarni, A.: Deep learning models combining for breast cancer histopathology image classification. *Int. J. Comput. Intell. Syst.* **14**(1), 1003 (2021)
4. Mendes, J., Domingues, J., Aidos, H., Garcia, N., Matela, N.: AI in breast cancer imaging: a survey of different applications. *J. Imaging* **8**(9), 228 (2022)
5. Samee, N.A., Atteia, G., Meshoul, S., Al-antari, M.A., Kadah, Y.M.: Deep learning cascaded feature selection framework for breast cancer classification: hybrid CNN with univariate-based approach. *Mathematics* **10**(19), 3631 (2022)
6. Jiang, F., Liu, H., Yu, S., Xie, Y.: Breast mass lesion classification in mammograms by transfer learning. In: Proceedings of the 5th International Conference on Bioinformatics and Computational Biology, pp. 59–62 (2017)
7. Ribli, D., Horváth, A., Unger, Z., Pollner, P., Csabai, I.: Detecting and classifying lesions in mammograms with deep learning. *Sci. Rep.* **8**(1), 4165 (2018)
8. Alruwaili, M., Gouda, W.: Automated breast cancer detection models based on transfer learning. *Sensors* **22**(3), 876 (2022)
9. Kaur, P., Singh, G., Kaur, P.: Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification. *Inform. Med. Unlock.* **16**, 100151 (2019)
10. Mohapatra, S., Muduly, S., Mohanty, S., Ravindra, J.V.R., Mohanty, S.N.: Evaluation of deep learning models for detecting breast cancer using histopathological mammograms images. *Sustain. Oper. Comput.* **3**, 296–302 (2022)
11. Muduli, D., Dash, R., Majhi, B.: Automated diagnosis of breast cancer using multi-modal datasets: a deep convolution neural network based approach. *Biomed. Signal Process. Control* **71**, 102825 (2022)
12. Rouhi, R., Jafari, M., Kasaei, S., Keshavarzian, P.: Benign and malignant breast tumors classification based on region growing and CNN segmentation. *Expert Syst. Appl.* **42**(3), 990–1002 (2015)
13. Saber, A., Sakr, M., Abo-Seida, O.M., Keshk, A., Chen, H.: A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique. *IEEE Access* **9**, 71194–71209 (2021)
14. Punithavathi, V., Devakumari, D.: A hybrid algorithm with modified SVM and KNN for classification of mammogram images using medical image processing with data mining techniques. *Eur. J. Mol. Clin. Med.* **7**(10), 2956–2965 (2020)

15. Pillai, A., Nizam, A., Joshee, M., Pinto, A., Chavan, S.: Breast cancer detection in mammograms using deep learning. In: Iyer, B., Ghosh, D., Balas, V.E. (eds.) *Applied Information Processing Systems*. AISC, vol. 1354, pp. 121–127. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-2008-9_11
16. Chougrad, H., Zouaki, H., Alheyane, O.: Deep convolutional neural networks for breast cancer screening. *Comput. Methods Programs Biomed.* **157**, 19–30 (2018)
17. Selvathi, D., Aarthi Poornila, A.: Breast cancer detection in mammogram images using deep learning technique. *Middle-East J. Sci. Res.* **25**(2), 417–426 (2017)
18. Hasan, N., Bao, Y., Shawon, A., Huang, Y.: DenseNet convolutional neural networks application for predicting COVID-19 using CT image. *SN Comput. Sci.* **2**(5), 389 (2021)
19. Albelwi, S.A.: Deep architecture based on DenseNet-121 model for weather image recognition. *Int. J. Adv. Comput. Sci. Appl.* **13**(10), 2022
20. Kim, T.-H.: Electricity theft detection using fusion DenseNet-RF model (2021)
21. Zeng, L., Lang, J.: Classification of breast cancer histopathological image based on lightweight network. In: *CIBDA 2022; 3rd International Conference on Computer Information and Big Data Applications*, pp. 1–6. VDE (2022)
22. Vaswani, A., et al.: Attention is all you need. *Advances in Neural Information Processing Systems*, vol. 30 (2017)
23. Yoon, W.B., Oh, J.E., Chae, E.Y., Kim, H.H., Lee, S.Y., Kim, K.G.: Automatic detection of pectoral muscle region for computer-aided diagnosis using MIAS mammograms. *Biomed. Res. Int.* (2016)
24. Xiang, Yu., Zeng, N., Liu, S., Zhang, Y.-D.: Utilization of DenseNet201 for diagnosis of breast abnormality. *Mach. Vis. Appl.* **30**, 1135–1144 (2019)
25. Kavitha, T., et al.: Deep learning based capsule neural network model for breast cancer diagnosis using mammogram images. *Interdisc. Sci.: Comput. Life Sci.* 1–17 (2021)
26. Maqsood, S., Damaševičius, R., Maskeliūnas, R.: TTCNN: a breast cancer detection and classification towards computer-aided diagnosis using digital mammography in early stages. *Appl. Sci.* **12**(7), 3273 (2022)
27. Escorcía-Gutierrez, J., et al.: Automated deep learning empowered breast cancer diagnosis using biomedical mammogram images. *Comput. Mater. Continua* **71**, 3–4221 (2022)
28. Jebarani, P.E., Umadevi, N., Dang, H., Pomplun, M.: A novel hybrid k-means and GMM machine learning model for breast cancer detection. *IEEE Access* **9**, 146153–146162 (2021)