Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 225 (2023) 3938–3947

Procedia Computer Science

www.elsevier.com/locate/procedia

27th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2023)

Enhanced Computer-Aided Diagnosis Model on Ultrasound Images through Transfer Learning and Data Augmentation Techniques for an Accurate Breast Tumors Classification

Ikram BEN AHMED[a,c,*], Wael OUARDA[b], Chokri BEN AMAR[c]

[a]*Higher Institute of Computer Science and Communication Technology (IsitCom), 4011, Hammam Sousse, Sousse University, Tunisia*
[b]*Digital Research Center of Sfax (CRNS), Tunisia*
[c]*ReGIM-Lab- REsearch Groups in Intelligent Machines, National School of Engineers (ENIS), Sfax University, 3038, Sfax, Tunisia*

Abstract

Cancer is a critical global public health problem with meager median survival. It is therefore quite essential to detect this disease at an early stage to improve diagnostic results and consequently avoid serious complications. For this purpose, various researchers have implemented automated methods with the use of different medical imaging modalities. Accordingly, the expansion of deep learning techniques grants opportunities to enhance diagnosis, cure, and prevention. In this study, a diagnostic system for accurate classification of ultrasound breast abnormalities based on the powerful ResNet-50 CNN is proposed with the aim of providing early detection of breast cancer decease. The contribution of this work lies in the novel approach taken to improve the performance of the ResNet50 model in the classification of ultrasound breast cancer images. Transfer learning allows for the model to leverage pre-existing knowledge, while the application of data augmentation techniques enhances the diversity and quality of the training data. Additionally, the optimization of the batch size as a hyperparameter ensures that the model is able to effectively learn from the training data, leading to improved accuracy and efficiency in the classification process. This approach is crucial in the early detection and treatment of breast cancer. Quantitative and qualitative evaluations have been detailed in this study using Breast Ultrasound Dataset BUSI. Our presented work shows interesting results in terms of accuracy, specificity, sensitivity, and AUC which exceed the performance of other compared works. Moreover, the proposed method helps boost the clinical diagnosis of breast cancer. It may integrate a radiologist network, allowing them to constantly follow up on the patient's medical history.

Keywords: Breast Cancer; Deep Learning; Transfer Learning; Data augmentation; Classification; Ultrasound.

* Corresponding author. Tel.: +21695906125 ; fax: +73498079.
E-mail address: benahmedikram@gmail.com

10.1016/j.procs.2023.10.389

1. Introduction

Breast cancer is known to be the leading cause of death among women. Based on 2018 statistics, in the U.S.A, 266,120 new malignant cases as well as 63,960 new benign cases are estimated to be detected in women [1]. Due to its high mortality rate, early diagnosis of this disease is very essential to avoid heavy consequences. As a very efficient diagnostic tool, ultrasound imaging offers high sensitivity in detecting this tumor [2]. Computer-Aided Diagnosis (CAD) systems in ultrasound images have been receiving recently growing attention [4]. They have presented a high percentage of accuracy in breast cancer detection and classification. Indeed, in some cases, this system may surpass the human performance and therefore may decrease the percentage of deaths among women and men diagnosed with breast cancer [5]. Deep learning is considered today a powerful technique in classifying tumors in medical images. Thus, it has shown that it is very relevant in solving biomedical imaging problems [6]. Since transfer learning is considered to be the most powerful technique in deep learning methods [8], it represents the perfect solution to overcome the problem of having a small database [9]. Therefore, in order to reuse the parameters already learned by the transfer learning algorithms, several works use this technique which has proven its effectiveness in solving new problems [10]. The major contributions of the study are as follows:

- We present various geometric data augmentation techniques applied to the dataset to enhance the diversity and quality of the training data by artificially increasing the dataset size.
- We investigate the principle of the transfer learning technique by implementing it on the ResNet50 model for breast cancer classification.
- We focus on post-processing by optimizing the batch size as a hyperparameter.

This paper is organized as follows: Section 2 clarifies state-of-the-art approaches. Subsequently, Section 3 describes our developed solutions. Then Section 4 thoroughly explains the experiments. Section 5 tackles the discussion of the obtained results. Last by not least, section 6 presents the conclusion.

2. Related work

The main interest in medical image processing is the extraction of useful information that could help radiologists detect cancerous tumors that are difficult to reveal with the naked eye. In literature, numerous researchers tried to elaborate trustworthy methods in order to detect breast cancer from ultrasound images. Several papers have been written on breast cancer detection and many approaches have been proposed. We review briefly Some popular ones. Some researchers have used various machine learning algorithms such as decision trees, random forests, support vector machines (SVMs), and convolutional neural networks (CNNs) [12] for the classification of ultrasound breast images [13]. Other researchers have combined image processing techniques with machine learning algorithms to improve classification performance. Lately, works based on deep learning approaches have gained an important place. This technique has been widely used for medical image classification, including ultrasound breast cancer classification, such as AlexNet [14], VGG [15], DenseNet [16], and Residual network [17], and they all have excellent performance in numerous areas. But the Fully Convolutional Networks (FCN) [18] provide superior performance when compared with other deep-learning models with respect to semantic medical image segmentation. Convolutional neural networks (CNNs) are the most commonly used deep learning models for this task.

Transfer learning is a technique that allows a model trained on one task to be fine-tuned for another task, which has been applied to improve the performance of ultrasound breast cancer classification. Several studies have demonstrated the benefits of this technique for the detection of tumors using ultrasound images. For instance, we summarize some data augmentation techniques and classification methods for breast cancer detection.

2.1. Discussion and motivation

There are numerous research studies in the literature dedicated to detecting and classifying breast cancer, but only a few focus on using ultrasound scan images for classification. In fact, there are several factors that make the detection of breast cancer in ultrasound images challenging such as image quality, heterogeneity of lesions, overlapping tissue,

Table 1: Data augmentation techniques and classification methods for breast cancer detection in the literature.

Author/Method/year	Data augmentation Used	Transfer Learning Used
Ayana et.al[20]. 2022	scaling, rotation, and translation	AlexNet, DenseNet201
idossov et.al [21].2023	Rotation, scaling, translation, and horizontal flip	MobileNet
Ogundokun et.al[22]. 2023	Rotation, Width shift, Height shift range, Horizontal Flip, Vertical Flip.	MobileNet-SVM
Jabeen et.al [23].2022	horizontal flip, vertical flip, and rotate 90	DarkNet53
Hijab et .al [1].2019	rotation (90º, 180º, 270º), width shift, height shift	VGG16
Al-Dhabyani et.al [24].2019	Horizontal flip, brightness, scaling and zooming	VGG16, ResNet, Inception

small lesions, and subjectivity of interpretation. All these specificities of ultrasound breast images can make the detection of breast cancer challenging. ResNet50 uses residual connections, which allow it to mitigate the vanishing gradient problem, a common issue in deep CNNs. ResNet50 has shown high performance in a number of image classification tasks, making it a good choice for the classification of ultrasound images for the detection of breast cancer. It is a deep network with 50 layers, which allows it to learn complex representations of the input data. This is important for tasks like breast cancer classification, where the images may contain subtle differences that are indicative of cancer.

The added value of using ResNet50 is that it can leverage its pre-trained weights to learn relevant features from the ultrasound images, reducing the amount of data and computational resources required for training. thus, ResNet50 could be a good choice for the classification of ultrasound images for the early and accurate detection of breast cancer due to its ability to leverage transfer learning, its high performance on image classification tasks, and its deep architecture.

3. Proposed approach

In this study, we mainly aim to improve the performance of the breast cancer classification task using ultrasound images by utilizing transfer learning and data augmentation techniques. By combining these two approaches, our goal is to achieve improved accuracy and robustness in the classification of breast cancer. Then we focus on post-processing by optimizing the batch size as a hyperparameter. So, in this section, the methodology of our work is detailed into three parts. The Data augmentation module will be clarified in the first part. In the second part, we explain the features of learning. Finally, we illustrate the data classification part and the post-processing. We illustrate all components of the full system architecture in Figure 1. Furthermore, in the upcoming subsections, we will explain all subsystems.

3.1. Data Augmentation

For better performance and generalization of the models, the size of the dataset is enlarged by using different data augmentation methods. The model training data are small samples, thus they do not represent enough information for relevant training. Therefore, we adopt the technique of data augmentation in order to solve the problem of overfitting that could be caused. In order to establish that the learning of the network is well performed, employed data augmentation on images by applying some geometric transformations which were rotation, horizontal flip, vertical flip, and zoom range. We illustrate these techniques in Fig. 2. Upon applying the aforementioned transformations to the initial dataset, comprising 780 images, a total of 3,120 images were generated, effectively augmenting the dataset to a final size of 3,900 images. This expanded dataset was then utilized for subsequent analysis and model training purposes.

3.2. Feature Learning

Feature learning is a sequence of steps capable of automatically extracting the necessary data in order to detect, classify or predict features from the preprocessed dataset. Thus the machine improves its performance in the classification and segmentation of suspicious regions. A recent and personalized technique that could be used in deep learning is transfer learning. It is represented by the fact of transferring the capacities of the model acquired during the training on a database to a new images-set. And that was mentioned in Figure 1.

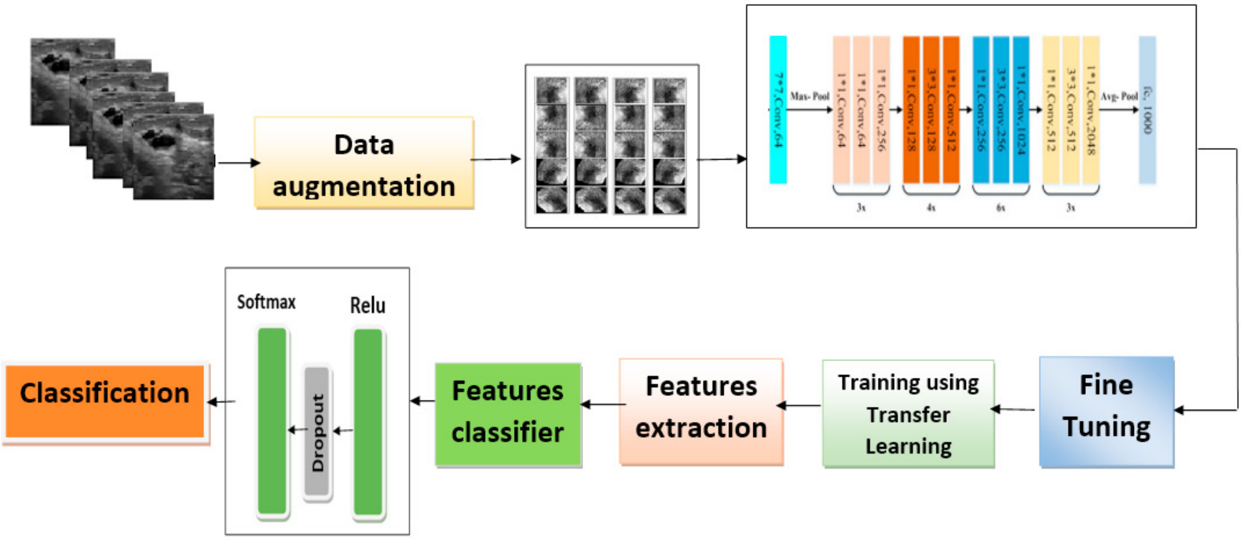


Fig. 1: Proposed System Architecture: Breast Ultrasound images classification using Transfer Learning and data augmentation.

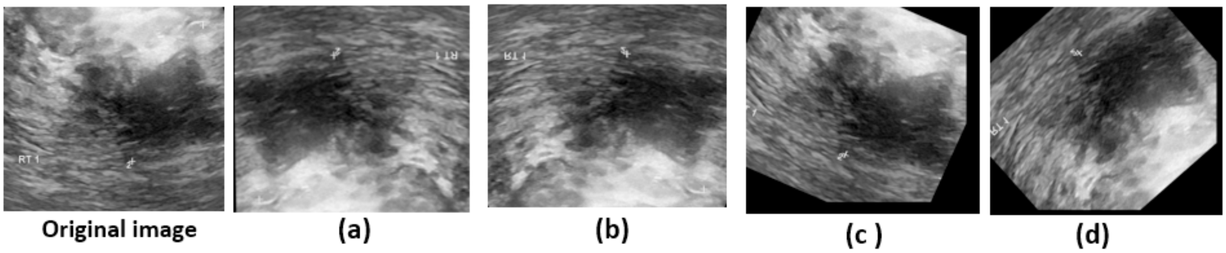


Fig. 2: Data augmentation for one image from the 'BUSI' Dataset with our Geometric transformations pipeline: rotation (a), zoom range (b), horizontal flip (c), and vertical flip (d).

3.3. Data classification

Data classification is the final step that systematically leads to decision-making. It is based on the classification layer which interprets the results deduced from the fully connected layers and then calculates the cross entropy loss in case of multi-class classification problems. Thus the module recovers the functionalities of Resnet50 and generates 1024 neurons as the output of the fully connected layer. Subsequently, these last units will be activated by the Relu activation function [26]. The main motivation for using the Relu function is that experiments in other works have shown that this method helps the model to converge faster than the sigmoid function. Finally, the output layer will generate the probabilities for each class.

4. Experiementations

4.1. Dataset

The developed model is tested on the 'BUSI' Dataset [25]. The Breast Ultrasound Images Dataset (or BUSI Dataset) is a public database made up of 780 ultrasound images. these images were taken from 600 patients aged between 25 and 75 years. The average image size is 500*500 pixels. Among the total number of images, 133 are

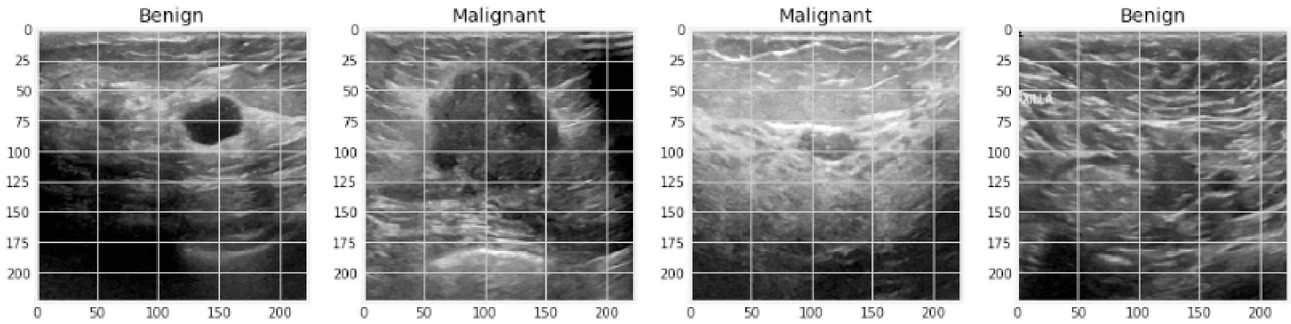


Fig. 3: Various ultrasound images and their corresponding class From BUSI database.

normal, they do not contain cancerous tumors while 437 contain cancerous masses and 210 were masses of benign category. Various ultrasound images are showen in Fig. 3.

4.2. Evaluation metrics

We opt to figure out the performance of our approach with certain evaluation metrics which are the most used in the classification of medical images [29]. Thus, We briefly explained the evaluation indicators used in this part which are specificity, sensitivity, confusion matrix [27], recall, and accuracy and precision [28].

- True Positive (TP): When it is indeed a patient with breast cancer and the model asserts that the suspicious region is positive.
- False Positive (FP): When it is indeed a healthy patient whereas the model predicts that there is a tumor.
- True Negative (TN): when the system classifies the case as negative and the patient is also not suffering from cancer.
- False Negative (FN): when the system predicts that there is not a tumor however the patient has cancer.
- Sensitivity: known as True Positive rate. It presents the ability that a malignant cancer tumor to be correctly diagnosed as malignant. Formula (1) indicates how sensitivity would be calculated.

$$Sensitivity = \frac{TP}{TP + FN} \tag{1}$$

- Specificity: Or True Negative rate. This metric calculates the probability that a test is correctly predicted as negative and effectively there are no tumors. The specificity measurement is detailed in Formula (2).

$$Specificity = \frac{TN}{FP + TN} \tag{2}$$

- Accuracy: It specifies how many tests were predicted correctly over all cases. Formula (3) presents how accuracy should be measured.

$$Accuracy = \frac{TN + TP}{FP + FN + TN + TP} \tag{3}$$

- Recall: It is nominated also as True-Positive Rate. This metric presents the probability of successfully predicting the presence of a pathology when it actually exists.

$$Recall = \frac{TP}{FN + TP} \tag{4}$$

- Precision: It specifies the fraction of instances that have been correctly estimated among all the retrieved instances of a distinct class. It can be expressed as

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

- The F1 score: It is the most common metric used on imbalanced classification problems. It can be defined as the harmonic mean of the model's recall and precision. Formula (6) shows how The F1-score measure may be calculated.

$$F1 - score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{6}$$

- Confusion Matrix: It illustrates the performance evaluations of the methods used after the classification task. It represents the correctly classified true positive cases, false positive cases affected to a specific class while it should be in another class. Also, it indicates the false negative cases predicted as another class when they should be relevant to another one. Furthermore, it precisely the true negative cases correctly classified in other classes.

4.3. Post-processing: optimizing the batch size

For a good study of the specificity and behavior of our model, we resort to a qualitative and quantitative evaluation to better visualize and understand the behavior of the proposed method throughout the training and then deduce the results found.

4.3.1. Quantitative evaluation

To accomplish the proposed classification task, breast ultrasound images were tested and trained using the resources provided by Google Collaboratory. We used the free GPU provided in the cloud during processing as an accelerator. In order to find the right combination of parameters that leads systematically to a good classification of images, we tried to change each time the batch size and at another time the number of epochs. So, as shown in Table 2. we notice that each time we decrease the batch size, we observe that the value of the loss function decreases, consequently the accuracy measure increases. Afterward, we plot the ROC curve, also called the Receiver Operating Characteristic. This curve allows us to visualize the variation of the rate of true positives in the function of the false positives.

As a result, it can be noted that each point on the curve refers to a specific decision threshold represented by the sensitivity/specificity pair. We can then deduce that when the ROC curve crosses the upper left corner, in this case, we have good discrimination. thus, we obtain a higher global precision value, when the curve comes closest to the upper left corner [17]. In our Experimentation, the best value found for the accuracy under the curve is 0.905. Table 2 illustrates the classification report. It provides a very good visualization of the details of the behavior of our trained model by specifying the different values of the studied evaluation metrics which are F1 score, precision, recall, and support.

We take the example of 8 for batch size since it represents the best performance. For the precision value: of all lesions that should be detected, 94.12% are predicted correctly.
For the Recall: Out of all the lesions that actually did get predicted, the model only detected this outcome correctly for 76.19% of those lesions.
For the F1 score value, it was equal to 84.21%, Since this value is close to 1, it tells us that the model does a good job of predicting whether or not lesions will get predicted. We present the best results for loss, accuracy, recall, and precision with an indication of the best epoch in Figure 4 below.

Evenly, we curried out of a Confusion Matrix which is a graphical representation of the Actual VS Predicted values to measure the effectiveness of our Machine Learning classification approach. That is illustrated in Figure 5.

4.3.2. Qualitative evaluation

Under qualitative classification, data are classified on the basis of some properties like similarity. When the data are classified according to geographical location or region it is known to be a geographical classification and when data

Table 2: Variation of Precision, recall, and F1 score during training for 600 epochs with values of batch size=16 and 8.

		Precision	Recall	F1 Score	Support
Batch size= 16 , Nb epochs=600	Benign	0.8913	0.9762	0.9318	84
	Malignant	0.9412	0.7619	0.8421	42
	Accuracy			0.9048	126
	Macro avg	0.9162	0.8690	0.8870	126
	Weighed avg	0.9070	0.9048	0.9019	126
Batch size= 8 , Nb epochs=600	Benign	0.9222	0.9881	0.9540	84
	Malignant	0.9722	0.8333	0.9257	42
	Accuracy			0.9365	126
	Macro avg	0.9472	0.9107	0.9257	126
	Weighed avg	0.9389	0.9365	0.9352	126

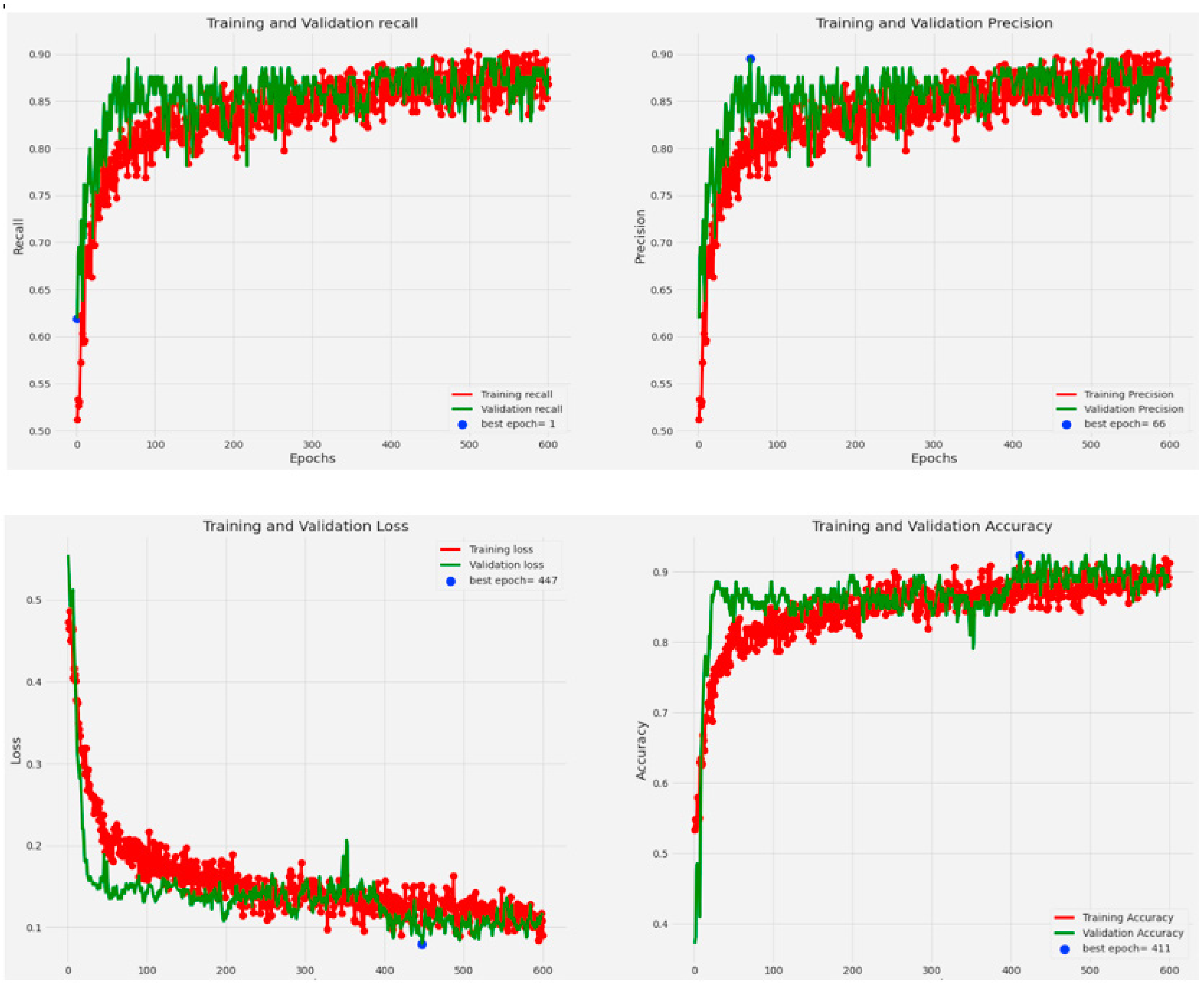
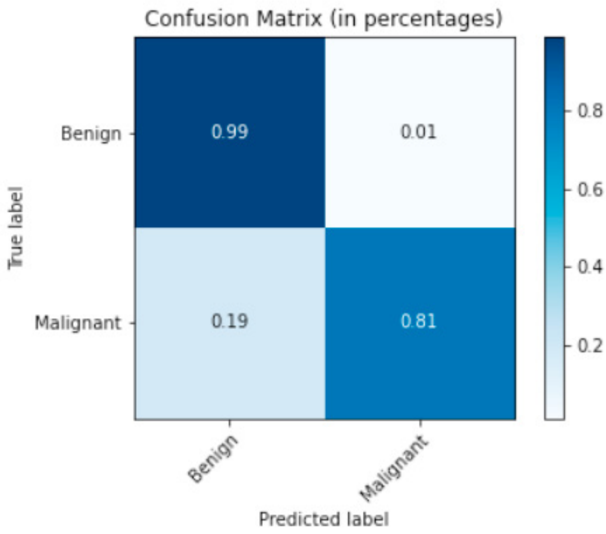


Fig. 4: Best results for loss, accuracy, precision, and recall with an indication of the best epoch.

are observed over a period of time the type of classification is referred to as chronological classification. Therefore, it is quite clear that qualitative classification can only be found whether it is present or absent in the units of study. Figure 7 details the results of our model classification by specifying each image, the predicted result, and the ground truth result.

Fig. 5: Confusion matrix for the binary ultrasound breast cancer classification.



5. Results and discussion

Lately, the transfer learning technique has been integrated into various application fields including the axis of medical image processing in the particular case of ultrasound images, although there still exists other grounds for improvement. Even though the existing related works use different programming techniques and network configurations to classify the breast ultrasound images dataset used. Our model proves its robustness by showing comparable accuracy metric values during the training and testing phase. Therefore, table 3 reveals that our proposed model has reported a promising result in terms of accuracy, specificity, sensitivity, and AUC and proves that it surpasses the results of other compared works.

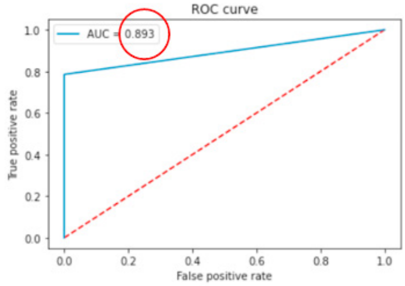
Table 3: Comparison of the proposed method with state of art breast cancer classification methods.

Model	ACC(%)	TPR(%)	TNR(%)	AUC
ResNet-50 [18]	84.94	77.39	88.74	0.91
ResNet-50 [1]	82.72	80.24	83.31	0.832
ResNet-50 [19]	80.27	79.61	82.40	0.803
Our model	93.65	**83.33**	**97.22**	0.905

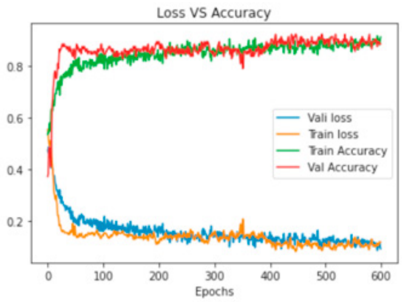
6. Conclusion

In this paper, we resorted to the transfer learning technique using a robust CNN which is ResNet50. A crucial step was used which was data augmentation, fine-tuning of the learned model, and post-processing optimization. We used this efficient classification approach in order to classify BUSI ultrasound breast cancer pathology. The obtained results showed that this fine-tuned approach improved the quality of the medical data classification in terms of accuracy compared to other research works. For the coming works, we aim to expand the ultrasound database and design other neural networks for accurate cancerous tumor detection and classification. Besides, we have already published a work that generates the detection of anomalies on ultrasound images, thus merging the two approaches and creating an end-to-end model, that ensures the classification and segmentation of tumors will be our next goal.

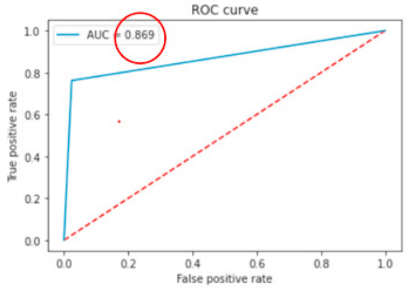
(a) Nb Epochs=400 ; Batch=16



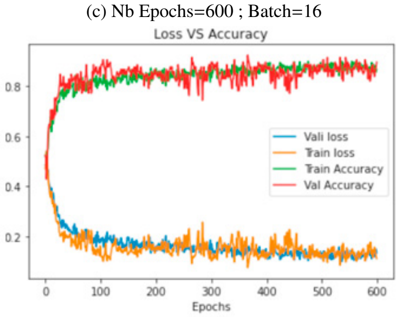
(b) Nb Epochs=400 ;Batch=16



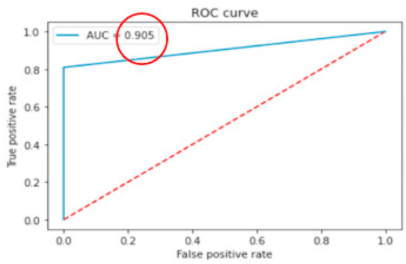
(c) Nb Epochs=600 ; Batch=16



(d) Nb Epochs=600 ; Batch=16



(e) Nb Epochs=600 ; Batch=8



(f) Nb Epochs=600 ; Batch=8

Fig. 6: Qualitative training experiments aspects of Loss vs Accuracy and curve ROC for the study of the model behavior by varying the epoch number and the batch size

Acknowledgements

The research leading to these results has received funding from the Ministry of Higher Education and Scientific Research of Tunisia under the grant agreement number LR11ES48.

References

[1] Hijab, Ahmed, et al. "Breast cancer classification in ultrasound images using transfer learning." 2019 Fifth international conference on advances in biomedical engineering (ICABME). IEEE, 2019.
[2] Ben Ahmed, Ikram, Wael Ouarda, and Chokri Ben Amar. "Hybrid UNET Model Segmentation for an Early Breast Cancer Detection Using Ultrasound Images." Computational Collective Intelligence: 14th International Conference, ICCCI 2022, Hammamet, Tunisia, September 28–30, 2022, Proceedings. Cham: Springer International Publishing, 2022.
[3] Houssein, Essam H., et al. "Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review." Expert Systems with Applications 167 (2021): 114161.
[4] Misra, Sampa, et al. "Bi-modal transfer learning for classifying breast cancers via combined B-mode and ultrasound strain imaging." IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 69.1 (2021): 222-232.

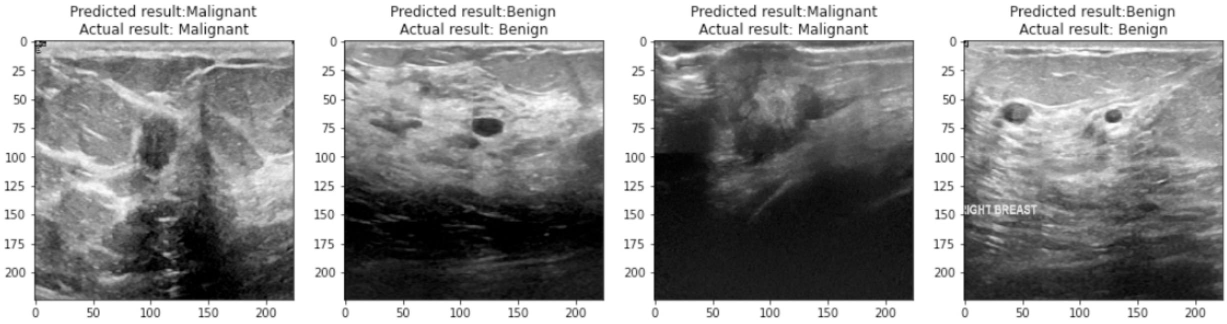


Fig. 7: Qualitative classification results: test image with the predicted result and the actual result.

[5] Murtaza, Ghulam, et al. "Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges." Artificial Intelligence Review 53.3 (2020): 1655-1720.

[6] Masud, Mehedi, et al. "Pre-trained convolutional neural networks for breast cancer detection using ultrasound images." ACM Transactions on Internet Technology (TOIT) 21.4 (2021): 1-17.or classification in ultrasound images using combined deep and handcrafted features." Sensors 20.23 (2020): 6838.

[7] Al-Antari, Mugahed A., et al. "An automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network." Journal of Medical and Biological Engineering 38.3 (2018): 443-456.

[8] Daoud, Mohammad I., et al. "Breast tumor classification in ultrasound images using combined deep and handcrafted features." Sensors 20.23 (2020): 6838.

[9] Yan, Rui, et al. "Richer fusion network for breast cancer classification based on multimodal data." BMC Medical Informatics and Decision Making 21.1 (2021): 1-15.

[10] Prasanna, M., M. Thamarai, and S. P. Malarvizhi. "Breast Cancer Prediction using ResNet50." Breast Cancer 7.9 (2020).

[11] Al-Haija, Qasem Abu, and Adeola Adebanjo. "Breast cancer diagnosis in histopathological images using ResNet-50 convolutional neural network." 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS). IEEE, 2020.

[12] Shen, Li, et al. "Deep learning to improve breast cancer detection on screening mammography." Scientific reports 9.1 (2019): 1-12.

[13] Gheflati, Behnaz, and Hassan Rivaz. "Vision transformers for classification of breast ultrasound images." 2022 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2022.

[14] Hepsağ, Pınar Uskaner, Selma Ayşe Özel, and Adnan Yazıcı. "Using deep learning for mammography classification." 2017 International Conference on Computer Science and Engineering (UBMK). IEEE, 2017.

[15] Sun, Dongdong, Minghui Wang, and Ao Li. "A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data." IEEE/ACM transactions on computational biology and bioinformatics 16.3 (2018): 841-850.

[16] Ibrahim, Norhayati, et al. "Automated detection of clustered microcalcifications on mammograms: CAD system application to MIAS database." Physics in Medicine and Biology 42.12 (1997): 2577.

[17] Lopez, Miguel G., et al. "BCDR: a breast cancer digital repository." 15th International conference on experimental mechanics. Vol. 1215. 2012.

[18] Alkhaleefah, Mohammad, and Chao-Cheng Wu. "A hybrid CNN and RBF-based SVM approach for breast cancer classification in mammograms." 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2018.

[19] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[20] Ayana, Gelan, et al. "A novel multistage transfer learning for ultrasound breast cancer image classification." Diagnostics 12.1 (2022): 135.

[21] Aidossov, N., et al. "Evaluation of Integrated CNN, Transfer Learning, and BN with Thermography for Breast Cancer Detection." Applied Sciences 13.1 (2023): 600.

[22] Ogundokun, Roseline Oluwaseun, et al. "MobileNet-SVM: A Lightweight Deep Transfer Learning Model to Diagnose BCH Scans for IoMT-Based Imaging Sensors." Sensors 23.2 (2023): 656.

[23] Jabeen, Kiran, et al. "Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion." Sensors 22.3 (2022): 807.

[24] Al-Dhabyani, Walid, et al. "Deep learning approaches for data augmentation and classification of breast masses using ultrasound images." Int. J. Adv. Comput. Sci. Appl 10.5 (2019): 1-11.

[25] Al-Dhabyani, Walid, et al. "Dataset of breast ultrasound images. Data Brief 28, 104863 (2020)." (2019).

[26] Prananda, Alifia Revan, and Eka Legya Frannita. "Toward Better Analysis of Breast Cancer Diagnosis: Interpretable AI for Breast Cancer Classification." IT Journal Research and Development 7.2 (2023): 220-227.

[27] Bai, Yujie, et al. "Recognizing breast tumors based on mammograms combined with pre-trained neural networks." Multimedia Tools and Applications (2023): 1-20.

[28] Wiacek, Alycen, et al. "Coherence Metrics for Reader Independent Differentiation of Cystic From Solid Breast Masses in Ultrasound Images." Ultrasound in medicine and biology 49.1 (2023): 256-268.

[29] Brunetti, Nicole, et al. "Artificial Intelligence in Breast Ultrasound: From Diagnosis to Prognosis—A Rapid Review." Diagnostics 13.1 (2023).