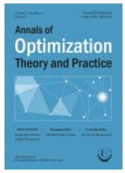
Annals of Optimization Theory and Practice

Volume 3, Number 2, 71-79
November 2020

DOI: 10.22121/aotp.2020.251387.1046



Automatic breast thermography images classification based on deep neural networks

Mahmoud Azza

Department of Basic Science, Faculty of Engineering, Pharos University, Alexandria, Egypt.

✉ azza.elbagoury@gmail.com

(Received: October 5, 2020 / Accepted: November 8, 2020)

Abstract Breast thermography is a screening tool which is capable of detecting cancer at an early stage. The main objective of this work is using the full power of deep neural network (DNN) and exploring its ability to learn the discriminative features of input data. The transfer learning and data augmentation are performed to solve the problem of lack of labled data. To improve the accuracy, the support vector machine (SVM) classifier will hybrid with the convolutional neural network (CNN) instead of using the deep model as end-to-end. The performance is verified by the k-fold cross-validation. The proposed techniques are trained and evaluated on DMR-IR dataset to classify the thermographic images to normal and abnormal groups. The proposed technique of employing AlexNet hybrid with SVM achieves the best performance, producing 92.55% accuracy, 95.56% sensitivity, 89.80% precision, 92.63% F1 score.

Keywords Breast cancer; Breast thermography; Deep neural network; Convolutional neural network; AlexNet; The support vector machine

1. Introduction

Breast cancer is the most common kind of cancer in women worldwide, including 16% among women's cancers. Although breast cancer is to be a disease of the developed world, a majority of 69% of all breast cancer mortality happens in developing countries (Siegel *et al.*2019). In recent years, the occurrence rate of breast cancer has increased. The breast cancer survival rate has also grown over the past few years; with the advancement of more effective diagnostic tools and treatment methodologies. The low survival rates in less developed countries may be due to a lack of early detection methodologies; thus, a large number of cases is detected in late-stage.

There are many techniques available to detect the presence of cancer: ultrasound, thermography, magnetic resonance imaging (MRI), microwave, X-ray, etc. Mammography is known as the gold standard imaging tool for detecting breast cancer

Published by Firouzabad Institute of Higher Education, Firouzabad, Fars, Iran

(Lee and Yang 2010). This exam needs X-ray radiation doses to identify the suspicious tissue in breast volume. However, this method has some constraints (Woloshin and Schwartz 2010). It is uncomfortable for women because of breast compression. Also, the X-ray radiation may damage the cancerous masses. Breast density also influences the diagnosis of mammography. Finally, the searchers realize that mammogram is less efficient for women less than 50 years old (Kennedy *et al*. 2009).

Several screening tools were evaluated to overcome the limitations of mammography (Brekelmans *et al*. 1996). One of the most popular technique is the infrared thermography (Ng 2009). Breast thermography technique is forming an image of a temperature distribution of the breast surface. It has been used as a technique or medical imaging since 1960s. The thermal infrared camera has a sensor to detect and capture the thermal radiation emits from the human body. A processing system measures and displays the heat pattern of the human skin (Keyserlingk *et al*. 2000).

Breast thermography has some significant advantages over mammography; such as its capability to deal with dense breast tissues and efficient in all age groups (Ng 2009). Thermography is also harmless, fast method, and helps in early detection of breast cancer (Kennedy *et al*. 2009). Finally, the ionization, high pressure, and compression of the breast are not required in thermography (Hossam *et al*. 2018), (Mazhar and Tayel 2020).

Thermography is based on the principle that metabolic and vascular circulation in cancerous tissue are always higher than in normal breast tissue. Thus, the temperature of the regional skin surface will be relatively higher than the normal breast temperature. For non-cancer patients, the temperature distribution of the thermogram is generally symmetrical across the mid-line of the body. Therefore, in breast cancer, thermography detects disease by identifying areas of asymmetric temperature distribution on the breasts' surface. The analysis of the thermograms should be carried out by a trained physician in comparing the temperature distribution of the left and right breasts (Debi and Tina 2012). This comparison takes scrupulous training which is dependent on the experience and watchfulness of the physician. Thus, computer-aided diagnostic (CAD) in thermograms can show great promise in being used as an adjunct tool for early breast cancer detection.

The SVM is an algorithm for the machine learning which analyzes the classification data. It is a supervised method of learning, which separates data into categories. SVM aims to formulate a computationally efficient way of learning by separating hyperplanes in a high dimensional characteristics space. There are many hyperplanes which could distinguish two sets of data. The best hyperplane to choose from is the one with the highest range. The margin is defined as the width by which, before hitting a data point, the boundary could increase. Support vectors are the data points pushed up by the margin. Thus, the goal of the SVM is to find the optimum hyperplane that separates clusters of target vectors on the opposing sides of the plane.

In this work, a CAD technique is proposed for effective, accurate and objective classification of normal and abnormal breast tissues using infrared thermographic images. The main contribution is that the deep convolutional neural network (CNN); i.e. AlexNet; is hybrid with SVM to accomplish high rates of breast cancer classification task. Data augmentation technique is employed on DMR-IR dataset to overcome the deficiency of dataset; also transfer leaning followed by the fine-tuning are performed to reduce consuming computational power. To evaluate the performance of these classifier, we used accuracy, sensitivity, specificity and F1 score value as the evaluation. The organization of the paper is as follows: Section 2 illustrates the literature review; Section 3 details the methods and materials used in the current study for the classification of thermograms as normal and abnormal. The obtained results are displayed and discussed in Sections 4. Finally, the paper concludes in Section 5.

2. Literature Review

Several supervised learning methods have been reported for breast cancer detection based on thermography such as ANN (Artificial Neural Network), K-nearest neighbors, fuzzy and SVM, as illustrated in (Karim *et al*. 2018). The SVM is the most powerful and widely used in the classification of health diseases, especially in oncology, due to its high accuracy and ability to deal with high-dimensional data.

Pramanik *et al*.(2016) have presented a local feature extraction mod¬ule called block variance to classify malignant and benign of breast thermographic images. They have used a feed-forward neural network for classification. Also, the proposed method identifies abnormality using asymmetric tempera¬ture distribution between right and left breast images. The proposed system has achieved a classification accuracy of 90%.

Sathish *et al.* (2017) have proposed an automated method for seg¬menting right and left breast images. The segmentation technique depends on the breast's shape features as well as the polynomial curve fitting. The proposed system extracts histogram and GLCM features from the segmented image. Then, an SVM classifier with RBF kernel function is employed to differentiate between healthy and non-healthy images. The proposed method has been assessed using some images from the DMR-IR database that consists of 40 healthy and 40 non-healthy thermo¬graphic images. The obtained experimental results have shown an accuracy of 90%.

In 2018 Hossam *et al.* (2018) have introduced an automated segmentation method to determine the ROI in breast thermographic using the statistics of the images. Then, the detected boundaries were enhanced using the Hough transform algorithm. The proposed system extracted statistical features from the ROI images and employed two classifiers to differ¬entiate between healthy and non-healthy images, namely SVM classifier and ANN classifier. The proposed system has been evaluated using the DMR-IR and achieved 96.67% and 96.07% for the SVM classifier and the ANN classi¬fier, respectively.

Published by Firouzabad Institute of Higher Education, Firouzabad, Fars, Iran

3. Methods and Materials

In this paper, breast thermograms are obtained from an open on-line database ("DMR, Visual Lab. Database for Mastology Research. Http://Visual.Ic.Uff.Br/ Dmi/, 2017.," n.d.) which called DMR-IR-IR database (Silva *et al*. 2014); example of images are shown in Figure 1. The images are taken in the University Hospital of UFPE (Brazil) and stored in the database with several other information such as age, exam date, patient family history, and patient preparation before the exam. Ethical Committee of UFPE approved the acquisition procedure and the storage of images registered at the Brazilian Ministry of Health. The images resolution are around 320 x 240 pixels. Total of 48 normal and 47 abnormal thermal images was considered for this work.

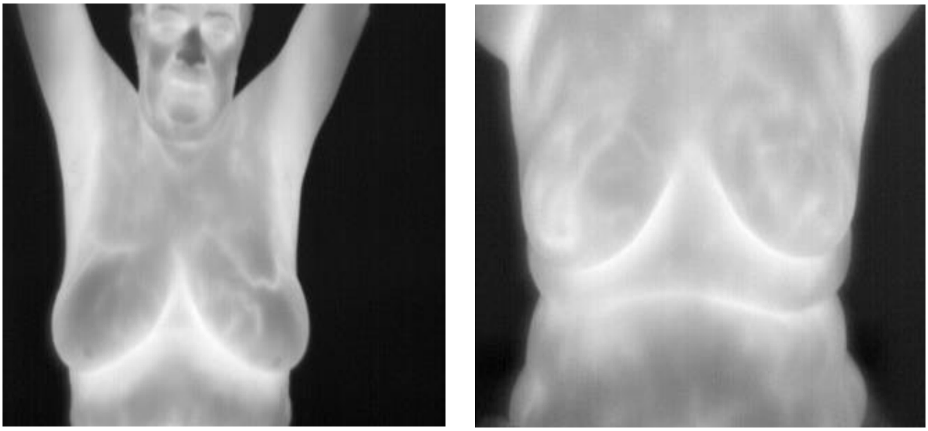


Figure 1. Example of breast thermographic images

This work proposes automatic techniques to classify the thermography. This work depends on taking the full power of pre-trained AlexNet; as starting the training from scratch will lead to over-fitting, consume time and need high computational resources. Also, to overcome the thermographic images deficiency, this work utilizes the data augmentation and fine-tuning the transferred parameters of pre-trained AlexNet. Eventually, the testing performance will be introduced to assess the proposed technique.

AlexNet is the first and the most successful deep convolutional neural network for classification in the computer vision field. AlexNet (Krizhevsky *et al*. 2012) was developed by Krizhevsky *et al.* 2012. It was the first time a model performed so well on ImageNet dataset. AlexNet illustrated the power and benefits of CNNs and its performance has backed CNNs up with record-breaking in the competition of 2012 ILSVRC (ImageNet Large-Scale Visual Recognition Challenge), which achieved a top-5 error of 15.4% (Top-5 error is the rate at which, given an image, the model does not output the correct answer with its top-5 predictions). AlexNet composed of 5 convolutional layers. The architecture of AlexNet is shown in Figure 2. Note that, each

Published by Firouzabad Institute of Higher Education, Firouzabad, Fars, Iran

of these layers is learnable and followed by a non-linearity (non-learnable) which counted as one layer. Finally, three fully connected layers are added. Moreover, the techniques utilized by AlexNet such as data augmentation and dropout are still used until today.

AlexNet is pre-trained on the ImageNet dataset (1.3 million natural images) with 1000 classes. Furthermore, the new classification layer for two classes; normal and abnormal, is utilized as the last layer instead of 1000 classes.

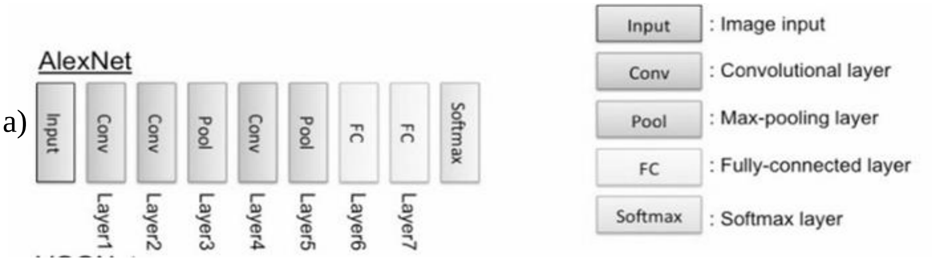


Figure 2.Architecture of AlexNet.

Unfortunately, there are many obstacles to apply state-of-the-art DL to medical problems and take advantage of the deep models. Medical images suffer from the problems of data deficiency (not as large as the size of classification datasets such as ImageNet Deng *et al.* (2009). To overcome this challenge, many strategies are introduced: the Data Augmentation, transfer learning and fine-tuning Mazhar ans Tayel (2020), (Tayel and Elbagoury (2020) are discussed below.

Data augmentation Sun *et al.* (2017) enlarges the training dataset as it generates new samples by applying a series of random transformations to the already existing data. It has many benefits, among which, the speeding up of the convergence process, preventing over-fitting and increasing the capabilities of generalization Wong *et al.* (2016). The simplest approach is to apply affine transformations (translation, zooming, flipping, mirroring, rotation, etc.) on small datasets Flusser and Suk (1993).

Training a DNN from the scratch with random initializing weights is often not feasible. It is often helpful to start the training with pre-trained weights instead of randomly initialized ones. Transfer learning technique is to transfer pre-trained parameters of a DNN which is trained on a very large, broad domain dataset to learn general features of images. This is possible because the pre-trained filters in the convolutional layers contain specific shape representations of an image that improve the task. Then, fine-tuning the pre-trained model to learn more features of a specific domain is undertaken, which could also be done for a different task Yosinski et al. (2018). Fine-tuning is a training process that differs slightly from the training from the scratch.

Then, some parameters must be set to start the fine-tuning process. Firstly, the learning rate is set to 10^{-3} in the AlexNet model. The number of epoch is 40. It is observed that the momentum should be 0.9 and the weight decay should be 5×10^{-4}. Gaussian distribution with a zero mean and variance of 10^{-2} is used to initialize the new classification final layer since only that layer will be trained from scratch. These parameters are the best fine-tuned for the diagnosis of medical breast cancer. The other parameters are set to default values.

There are several metrics for assessing a classifier, including the accuracy, the sensitivity, the precision, and the F1 score.

The accuracy

Accuracy is the indicator of the classifier making the correct prediction. It gives the ability of performance of the whole classifier. The following equation describe the accuracy.

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

Where: True positives (TP), True negatives (TN), False positives (FP), False negatives (FN).

The sensitivity

The sensitivity is the proportion of abnormal cases correctly classified. It is particularly useful when finding real positives (abnormal) is more important than finding real negative (normal). Moreover, it measures the ability of the the classifier to identify the real positives (abnormal).

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

The specificity

The specificity is the proportion of normal cases correctly classified. Also, it measures the ability of the classifier to identify the real negatives (normal). It is particularly useful when finding real negatives is more necessary than finding real positives (non-healthy).

$$\text{Specificity} = \frac{TN}{TN + FP}$$

The Precision

The precision is the ratio of positive observations correctly predicted to total positive observations expected.

$$\text{Precision} = \frac{TP}{TP + FP}$$

F1 score

The weighted average of precision and recall are described by F1 score. It is used as a statistical measure to score the classifier's efficiency. Therefore, this score takes both false positives and false negatives into account.

Published by Firouzabad Institute of Higher Education, Firouzabad, Fars, Iran

$$\mathrm{F-Score} = \frac{2 * \mathrm{Precision} * \mathrm{Sensitivity}}{\mathrm{Precision} + \mathrm{Sensitivity}}$$

The k-fold cross-validation technique is employed to validate the classification results Stone (1974). In the 5-fold cross-validation, data is divided into five groups randomly, every time one group is selected for testing and the other four groups for training. The 5-fold average classification accuracy, sensitivity and precision are calculated by averaging the classification results of the 5-fold. The final classification results are obtained by averaging the results of all k-folds cross-validation (from k = 1 to 5). The average classification results and standard deviation are illustrated in Tables 1 for all k-folds cross-validation of the AlexNet hybrid with SVM using DMR-IR databases. The proposed framework achieves an accuracy of 92.55%. Moreover, the sensitivity, specificity, precision, and F1 score reach 95.56%, 89.80%, 89.80% and 92.63%.

A comparison between the proposed classification technique and other classification techniques in the literature is described in Table 2. The comparison reveals the superior performance of the proposed AlexNet hybrid with SVM over other existing systems.

Table 1. Average k-fold cross-validation classification results using AlexNet hybrid with the SVM.

Accuracy	92.55%
Sensitivity	95.56%
Specificity	89.80%
precision	89.80%
F1 Score	92.63%

Table 2. Comparison the proposed classifier against some of the recent work which used the same dataset.

Parameter Model	Accuracy %	Sensitivity %	Specificity %
Pramanik *et al*. 2016)	90.00%	95.00%	85.00%
Sathish et al. (2017)	90.00%	87.50%	92.50%
The proposed framework	92.55%	95.56%	89.80%

4. Conclusion

The effectiveness of the proposed techniques is examined on real thermographic images from the DMR-IR database. The classification achievement is evaluated in terms of the accuracy, sensitivity, precision, F1 score using k-fold cross-validation. Transfer learning technique is applied since the training is initialized by weights of a pre-trained network. The data augmentation technique is utilized to defeat data scarcity and add diversity to the dataset. Also, data augmentation technique strengthens the generalization ability of the pre-trained network and further mitigates over-fitting. This work is very beneficial and shows that there is no need for human interface with pre- or post-processing or hand-crafted features. The AlexNet hybrid with SVM attains the best performance: 92.55% accuracy, 95.56% sensitivity, 89.80% precision and 92.63% F1 score. The robustness of the proposed techniques is reviewed by evaluating and comparing with other recent systems, and the obtained performance shows the superior accomplishment of the proposed technique**.**

References

1. Brekelmans, C. T., Westers, P., Faber, J. A., Peeters, P. H., & Collette, H. J. (1996). Age specific sensitivity and sojourn time in a breast cancer screening programme (DOM) in The Netherlands: a comparison of different methods. *Journal of Epidemiology & Community Health*, *50*(1), 68-71.
2. Debi, W. N. D., & Tina, K. (2012). Breast thermography: History, theory, and use. Nat. Med. J, 4.
3. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.
4. DMR, Visual Lab. Database for Mastology Research. Http://Visual.Ic.Uff.Br/ Dmi/, 2017.n.d.
5. Flusser, J., & Suk, T. (1993). Pattern recognition by affine moment invariants. Pattern recognition, 26(1), 167-174.
6. Hossam, A., Harb, H. M., & Abd El Kader, H. M. (2018). Automatic Image Segmentation Method for Breast Cancer Analysis Using Thermography. Journal of Engineering Sciences, 46(1), 12-32.
7. http://www.globalscientificjournal.com/researchpaper/An_Efficient_and_Reliable_Method_for_Regional_Analysis_of_Breast_Thermographic_Images.pdf
8. Karim, C. N., Mohamed, O., & Ryad, T. (2018). A new approach for breast abnormality detection based on thermography. Medical Technologies Journal, 2(3), 245-254.
9. Kennedy, D. A., Lee, T., & Seely, D. (2009). A comparative review of thermography as a breast cancer screening technique. Integrative cancer therapies, 8(1), 9-16.
10. Keyserlingk, J. R., Ahlgren, P. D., Yu, E., Belliveau, N., & Yassa, M. (2000). Functional infrared imaging of the breast. IEEE Engineering in Medicine and Biology Magazine, 19(3), 30-41.

11. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.
12. Lee, M. Y., & Yang, C. S. (2010). Entropy-based feature extraction and decision tree induction for breast cancer diagnosis with standardized thermograph images. Computer methods and programs in biomedicine, 100(3), 269-282.
13. Mazhar B.Tayel, Azza M.Elbagoury. 2020. "An Efficient and Reliable Method for Regional Analysis of Breast Thermographic Images." Global Scientific Journal 8 (9): 1508–18.
14. Ng, E. K. (2009). A review of thermography as promising non-invasive detection modality for breast tumor. International Journal of Thermal Sciences, 48(5), 849-859.
15. Pramanik, S., Bhattacharjee, D., & Nasipuri, M. (2016, September). Texture analysis of breast thermogram for differentiation of malignant and benign breast. In 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 8-14). IEEE.
16. Sathish, D., Kamath, S., Prasad, K., Kadavigere, R., & Martis, R. J. (2017). Asymmetry analysis of breast thermograms using automated segmentation and texture features. Signal, Image and Video Processing, 11(4), 745-752.
17. Siegel, Rebecca L, Kimberly D Miller, and Ahmedin Jemal. 2019. "Cancer Statistics, 2019." CA: A Cancer Journal for Clinicians 69 (1): 7–34
18. Silva, L. F., Saade, D. C. M., Sequeiros, G. O., Silva, A. C., Paiva, A. C., Bravo, R. S., & Conci, A. (2014). A new database for breast research with infrared image. Journal of Medical Imaging and Health Informatics, 4(1), 92-100.
19. Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society: Series B (Methodological), 36(2), 111-133.
20. Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. In Proceedings of the IEEE international conference on computer vision (pp. 843-852.
21. Tayel, M. B., & Elbagoury, A. M. (2020). Breast Infrared Thermography Segmentation Based on Adaptive Tuning of a Fully Convolutional Network. Current Medical Imaging, 16(5), 611-621.
22. Woloshin, S., & Schwartz, L. M. (2010). The benefits and harms of mammography screening: understanding the trade-offs. Jama, 303(2), 164-165.
23. Wong, S. C., Gatt, A., Stamatescu, V., & McDonnell, M. D. (2016, November). Understanding data augmentation for classification: when to warp?. In 2016 international conference on digital image computing: techniques and applications (DICTA)(pp. 1-6). IEEE.
24. Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks?. In Advances in neural information processing systems (pp. 3320-3328).

Published by Firouzabad Institute of Higher Education, Firouzabad, Fars, Iran