Intelligent Diagnosis of Breast Cancer with Thermograms using Convolutional Neural Networks

Nurduman Aidossov, Aigerim Mashekova, Yong Zhao, Vasilios Zarikas, Eddie Yin Kwee Ng and Olzhas Mukhmetov

1School of Engineering and Digital Sciences, Nazarbayev University, 53 Kabanbai batyr street, Nur-Sultan, Kazakhstan
2School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore

Keywords: Breast Cancer, Thermography, Convolutional Neural Network, Intelligent Diagnosis.

Abstract: Breast cancer is a serious public health issue among women all over the world. The main methods of breast cancer diagnosis include ultrasound, mammography and Magnetic Resonance Imaging (MRI). However, the existing methods of diagnosis are not appropriate for regular mass screening in short intervals. On the other hand, there is one non-invasive and low-cost method for mass and regular screening which is the so-called thermography. Recent studies show rapid quality improvement of thermal cameras as well as distinct development of machine learning techniques that can be combined together to enhance the technology of breast cancer detection. Machine learning technologies can potentially be used to support the interpretation of thermal images and help physicians to automatically determine the locations and sizes of tumors, blood perfusion, and other patient-specific properties of breast tissues. In this study, we aim to develop CNN techniques for intelligent precision breast tumor diagnosis. The main innovation of our work is the use of breast thermograms from a multicenter database without preprocessing for binary classification. The results presented in this paper highlight the usefulness and efficiency of deep learning for standardized analysis of thermograms. It is found that the model developed can have an accuracy of 80.77%, sensitivity of 44.44% and the specificity of 100%.

1 INTRODUCTION

Breast cancer is one of the most serious health problems with possible fatal consequences for women in modern times. The risk factors and causes could include changes in the cell genome, hormonal dysfunction, family history, hormone therapy, lifestyle features and undesirable life habits (Francis, 2017; Singh, 2020; NCI, 2014; WHO, 2014).

Currently the mainstream methods of breast cancer diagnosis include ultrasound, mammography and MRI. The mammograms also known as gold standard method of breast cancer diagnosis. However, it is mainly recommended for women over 40 years old as an invasive method which uses ionizing radiation (X-rays). Ultrasound is recommended as a first test to detect whether a lump is a cyst filled with liquid or a solid tumor. Ultrasound examination uses high-frequency acoustic waves. It is useful for younger patients as their mammary glands have denser structure. However the success of ultrasound diagnosis critically depends on the experience of the specialist conducting the test. MRI is the most effective and accurate method to diagnose breast cancer or tumor. But up to now this is the most expensive method, which could only be found in large and well equipped hospitals.

The existing methods of diagnosis are not appropriate for regular mass screening in short intervals.
Intelligent Diagnosis of Breast Cancer with Thermograms using Convolutional Neural Networks

One of the non-invasive and low cost method for mass and regular screening is thermography. It is well known that body temperature is an accurate indicator of some disorder in the human body. The body temperature distribution depends on such factors as: blood perfusion, metabolic rate and ambient temperature. Any abnormality in the body such as tumor could be detected by thermography (Melal, 2016; Kandlikar, 2017; Ito, 2019).

Recent studies (Sing, 2019; Kandlikar, 2017; Jiang, 2010; Bezerra, 2020; Saniei, 2016; Omranipour, 2016; Sarigoz, 2020; Zeng, 2020) discovered fast quality improvement of thermal cameras as well as distinct development of machine learning techniques that can be used to enhance the technology of breast cancer detection. Machine learning algorithms in principle, can be used to support the interpretation of thermal images and help physicians to automatically make diagnosis and even to determine the locations and sizes of tumors, blood perfusion, and other patient-specific properties of breast tissues (Khan, 2018).

One recently developed image recognition method is the so called convolutional neural network (CNN), which is a deep-learning neural network system that processes input images by extracting specific training weights and biases to determine meaningful characteristics that distinguish one input image from another (Khan, 2018). Thereby, a diagnosis tool can be built using CNNs in order to classify “healthy” and “with-tumor” types of thermograms without any human experts’ intervention.

In early studies of CNN the segmentation was an important part of image recognition. There are a lot of studies dedicated to different methods of image recognition among them studies (Antonini, 2015; Dayakshini, 2015; Kermani, 2015; Mahmoudzadeh, 2015; Diaz-Cortes, 2018; Etehad, 2010; Golestani, 2014). Study by Antonini et al. estimated the ability of thermogram diagnose multicentric or multifocal breast carcinomas (Antonini, 2017). Study by Dayakshini segmented thermograms by using projection profile method and by asymmetry analysis, comparing the left and right breasts (Dayakshini, 2015). Study by Kermani used Gaussian mixture model segmentation method (Kermani, 2015). Study by Mahmoudzadeh used the novel method of Hidden Markov Model to optimize the segmentation (Mahmoudzadeh, 2015).

2 MATERIALS AND METHODS

The public Visual Lab database (Visual Lab, 2021), which contains about 287 thermal images, was used to extract thermal images as input for our diagnosis tool. However, for the present study only 76 thermal images were selected as the most appropriate. These thermal images were accompanied with doctors’ diagnosis and also had three views: frontal, left and right.

In addition a second database was used which consists of thermograms of patients obtained in the "Multifunctional Medical Center" of the Nur-Sultan city of Kazakhstan by the authors. 38 thermal images most suitable for this work were selected. The database currently includes breast thermograms for women between the ages of 18 and 80. To protect the privacy of patients, the nomenclature of breast thermograms has been designed so that every image in the database is given a distinguished name.

Temperature distributions on the breast skin surfaces were recorded by the thermal camera IRTIS-2000 ME, which is used for medical research and the diagnosis of a wide range of illnesses, including
oncological diseases. Its temperature resolution for the entire field of view is 0.02 °C and its temperature measurement accuracy is 0.1 °C (see Figure 1).

An experimental procedure was developed, together with the instructions for conducting the experiment with which the doctor-oncologist was familiarized. When conducting an experiment, patients underwent a familiarization procedure with the research being conducted and gave their consent to participate in the research, since participation was voluntary. The study was approved by the institutional ethics committee of Nazarbayev University AEO (identification number is 294/17062020).

A typical breast thermogram used has three RGB channels with a square size of 224 × 224 × 3 by numbers of pixels. A breast thermogram should include half the armpit to analyze the entire breast tissue and nearby ganglion groups. The area of interest of the breast thermogram shows a significant increase in temperature compared to the temperature in the adjacent area for a patient with breast tumor. Examples of thermograms used are shown in Figure 2.

Two databases were combined to increase the size of the dataset, since the format of the thermograms and their images were similar. After they were mixed, the dataset was classified into two sets: Training and Validation ones, containing 88 and 26 images, respectively.

Image classification is the process of classifying images according to their visual contents. The learning process for neural network involves recognizing breast thermograms with a predetermined label, for example healthy and sick. This problem is known as supervised learning (Simeone, 2018). Thus, in the current study, the image set was divided into “Healthy” and “Sick” (as shown in Figure 2). The sigmoid function is a non-linear continuous function. Classification in CNN is based on inference, which implies that its output can be the entire range of $x$ to the domain $[0,1]$ of $f(x)$.

Mathematically, the sigmoid function is defined by (Sanjeev, 2017):

$$f(x) = \frac{1}{1 + e^{-\frac{x-\alpha}{\beta}}} \quad (1)$$

The parameters $\alpha$ and $\beta$ define the center and width of the sigmoid function, respectively.

In the current study, the CNN architecture consists of 5 layers of convolution and pooling. This is followed by flattening and 2 fully connected layers with the latter to obtain a binary output of probability (Figure 3).

CNN is a computational model that typically consists of three types of neural layers: convolution, pooling and fully connected ones. The convolution and pooling layers extract features from the input images, while the fully connected layer converts the extracted features into final output, such as binary classification. In the convolution layers have a small grid filled with parameters called kernel as a feature extractor, applied at each image position. One layer
feeds its output into the next layer, thus extracted features may progressively grow more complex as a forward process. Then the parameters in kernels can be optimized through gradient-based optimization scheme in a backward propagation process called training, which is performed so as to minimize the difference between outputs and given labels.

A batch size of 16 was used as a standard number of training examples utilized in one iteration of forward/backward pass. Activation function Rectified Linear Unit (ReLU) was used in the CNN architecture in the convolution layer. The size of the kernel is $3 \times 3$ pixels. First we began with a filter value of 32 (32 filters) in the first 3 layers, and the following two layers had the filter size of 64. After the convolution, flattening the input after CNN layers is standard procedure to go with and also adding the ANN layer as well.

First callback list was defined as follows. For the model to learn effectively it was necessary to define the EarlyStopping function. It was used to halt the epochs on metric of “loss” value and “patience” value. This function was used to avoid overfitting. In this CNN model, “loss” value is tracked and “patience” of 3 epochs is defined. What it means is that once the loss value reaches the minimum, and in the next 3 iterations the value of loss increases, then training will stop at that epoch. Another adherence is reducing the learning rate. So, once the metric stagnates, the learning rate reduces. Patience is 2 for this callback, and if no improvement is detected, then the learning rate reduces by a factor of 0.3, because in that way loss value will decrease gradually and finally arrive to the lowest value.

Another important parameter to define was class weight. Since the dataset consisted mostly of patients who had breast cancer, then it was necessary to assign higher class weight to minority classes, so it could learn in a balanced way from all classes.

CNN learning took 23 iterations to reach the stopping point as mentioned above. Each iteration took 10-11 seconds on a computer whose technical specifications are as follows, in Table 1:

<table>
<thead>
<tr>
<th>Total Cores</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Threads</td>
<td>20</td>
</tr>
<tr>
<td>Max Turbo Frequency</td>
<td>3.20 GHz</td>
</tr>
<tr>
<td>Processor Base Frequency</td>
<td>2.20 GHz</td>
</tr>
<tr>
<td>Cache</td>
<td>13.75 MB</td>
</tr>
<tr>
<td>RAM size</td>
<td>64GB</td>
</tr>
<tr>
<td>Maximum Memory Speed</td>
<td>2400 MHz</td>
</tr>
</tbody>
</table>

In summary, the overall flow diagram of the proposed study is presented on Figure 4.

**Figure 4:** The overall flow diagram of the proposed study.

### 3 RESULTS AND DISCUSSION

During training with stochastic gradient descent iterations the loss gradually decreased to 0.2151, while the accuracy of the training data increases to 80.77%. The learning rate remains relatively low from 30.0e-05 to 9.0e-05 as seen in Figure 5.

**Figure 5:** Graph of iteration versus loss, accuracy and learning rate values.
The accuracy of prediction of the CNN model is 80.77%, as shown in Figure 6.

![Accuracy](image)

Figure 6: Displaying the accuracy of the CNN model.

To analyze the inaccuracies, a confusion matrix was plotted.

![Confusion Matrix](image)

Figure 7: Confusion matrix showing the correctly and incorrectly predicted test data for each class.

In the proposed CNN study, the network was trained with thermograms with doctor’s diagnoses so that it could later be used to give a direct diagnostic response with a normal or abnormal decision when fed with a thermogram without the need for manual preprocessing and training of the features. Thus, it is easier to analyze thermograms in a standardized way, as opposed to studies with human judgment.

One limitation of this study is that the number of cases for analysis is less than the amount of data typically collected for deep learning. Although data augmentation can partially solve this problem, this disadvantage can be eliminated by using data exchange structures such as those implemented in neuroimaging (Yamashita, 2018). It can also be solved if it is combined with physics-driven and PINN diagnosis (Mukhmetov, 2021; Karniadakis, 2021), which we are working on.

Since thermography is intended as an adjunct to mammography in breast cancer screening, its most important value is sensitivity to detect the possible presence of an abnormality (Lalkhen, 2008). The study showed a sensitivity of 44.44%. Along with this, the performance of the classifier was assessed in terms of sensitivity, specificity, and accuracy. In general, the accuracy value was 80.77%, which can be considered quite high, since it allows detecting breast cancer with almost 81% confidence in thermograms. The specificity of the classifier was 100%.

Another metric that was used is ROC (receiver operating characteristic) curve and AUC value (Figure 8).

A negative predicted value (NPV) indicates the likelihood that the patient is not sick if the test is negative. Figure 7 shows that 5 out of 26 were incorrectly identified as healthy when in fact they are sick. The authors acknowledge that the positive predictive value (PPV) is low and is a metric that we would like to improve.

![ROC Curve](image)

Figure 8: ROC curve and AUC value.

Main innovation of our work is the use of breast thermograms from a multicenter database without preprocessing the images for efficient and automatic binary classification. The results presented in this paper highlight the usefulness of deep learning for standardized analysis of thermograms with efficiency. Future work will apply these algorithms in a longitudinal study without tagged data and evaluate their effectiveness in comparison with experts. Furthermore, an integrated methodology that will combine Bayesian Networks (Zarikas, 2015; Zholdasbayeva, 2020) and CNNs will be developed in order not only to improve the diagnosis but also to dig out the key factors that determine a successful diagnosis. It is important to remember that the current deep learning methods cannot yet replace the clinician when making a clinical diagnosis, but it can help the clinician make more accurate diagnoses and treatment recommendations. Furthermore, before widespread adoption in clinical practice, deep learning models should be tested on representative datasets of different communities in order to solve generalization problems for new populations.

4 CONCLUSIONS

Early detection of breast cancer remains an important part of the fight against breast cancer. World Health
Organization recommends regular self-examination to detect the breast cancer at early stages. The review of the previous studies shows that thermography is a promising supplementary tool for breast cancer detection at early stages. The combination of thermography and computer technology can considerably enhance breast cancer detection at early stages. Modern models of neural networks have led to an increase in the accuracy of classification of breast cancer thermograms, especially in distinguishing between healthy and deceased cases.

In the present study, a successful diagnosis tool is presented using convolutional neural network (CNN) to implement and validate the deep learning model. This algorithm could accurately classify breast cancer thermograms as “Healthy” and “Sick” using two databases and utilizing multi-view images. Moreover, our results were calculated automatically without any image pre-processing to obtain perspective sensitivity values, thus reducing human error and bias and improving efficiency. Reason of that is usage of Data Augmentation technique that is artificially enlarging the dataset size that helps for CNN to better learn and distinguish in binary classification. The limitation of the present study is that the patients’ data available for the analysis were less than the amount of data typically collected for deep learning. In addition, the positive predictive value (PPV) is still considered low, which can be further improved via physics-informed Neural Network (PINN) models in the future which are being developed by us.

ACKNOWLEDGEMENTS

The authors are grateful to the Ministry of Education and Science of the Republic of Kazakhstan for financing this work through the grant for the “Application of artificial intelligence to complement thermography for breast cancer prediction” (AP08857347) and Nazarbayev University for managing the research project.

REFERENCES


