Journal of Mechanics in Medicine and Biology
Vol. 7, No. 1 (2007) 1–10

INTEGRATIVE COMPUTER-AIDED DIAGNOSTIC WITH BREAST THERMOGRAM

E. Y. K. NG* and E. C. KEE

College of Engineering, School of Mechanical and Aerospace Engineering
Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798
**mykng@ntu.edu.sg*

Thermography is a non-invasive and non-contact imaging technique widely used in the medical arena. This paper investigates the analysis of thermograms with the use of bio-statistical methods and Artificial Neural Networks (ANN). It is desired that through these approaches, highly accurate diagnosis using thermography techniques can be established.

The proposed advanced technique is a multipronged approach comprising of Linear Regression (LR), Radial Basis Function Network (RBFN) and Receiver Operating Characteristics (ROC). It is a novel and integrative technique that can be used to analyze complicated and large numerical data. In this study, the advanced technique will be used to analyze breast cancer thermogram for diagnosis purposes.

The use of LR shows the correlation between the variables and the actual health status (healthy or cancerous) of the subject, which is decided by using mammography. This is important when selecting the variables to be used as inputs, in particular, for building the neural network.

For ANN, RBFN is applied. Based on the various inputs fed into the network, RBFN will be trained to produce the desired outcome, which is either positive for cancerous or negative for healthy cases. When this is done, the RBFN algorithm will possess the ability to predict the outcome when there are new input variables. The advantages of using RBFN include fast training, superior classification and decision making abilities as compared to other networks such as back-propagation.

Next, ROC is used to evaluate the accuracy, sensitivity and specificity of the outcome of RBFN Test files. The best results obtained are an accuracy (score) rate of 80.95%, with 81.2% sensitivity and 88.2% specificity. For breast cancer diagnosis, clinical examination by experienced doctors has an accuracy rate of approximately 60–70%. Hence, the proposed method has a higher accuracy rate than the existing practice.

Through the use of Bio-statistical methods and ANN, improvements are made in thermography application with regard to achieving a higher level of accuracy rate in diagnosis as compared to clinical examination. It has now become possible to use thermography as a powerful adjunct tool for breast cancer detection, together with mammography for diagnosis purposes.

Keywords: Breast cancer; neural network; ROC; thermography; non-invasive; bio-statistic.

1. Introduction

Thermography is a non-invasive diagnostic method that is economic, quick and does not inflict any pain on the patient. It is a relatively straightforward imaging

*Corresponding author.

method that detects the variation of temperature on the surface of the human skin.

Thermography is widely used in the medical arena. This includes the detection of breast cancer, which is the focus of many biomedical researches in recent years. Thermograms alone will not be sufficient for the medical practitioner to make a diagnosis. Analytical tools like biostatistical methods and ANN will be incorporated to analyze the thermogram.

Every day, 3 women are diagnosed with breast cancer in Singapore and it kills 5 women every week. It is the most common type of cancer in women here today. Thermography application in breast cancer thermography holds great promise in detecting early breast cancer.[1] Keyserlingk *et al.*[2] observed that the tumour diameter missed in thermogram was 12.8 mm and that in a mammogram was 16.6 mm. In fact, it has the potential to detect breast cancer ten years earlier than the traditionally golden method — mammography. However, due to inconsistencies in diagnosis from breast cancer thermograms, it is not yet used and regarded as a reliable adjunct tool to mammography in Singapore currently. This paper seeks to achieve a high level of consistency in the use of breast cancer thermography by virtue of a novel and unique approach encompassing biostatistical method and ANN.

2. Methodology

2.1. *Data acquisition*

The data collection for breast cancer was done in Singapore General hospital.[3] 90 patients for breast thermography were chosen at random. The thermal imager used is Avio TVS-2000 MkII ST. It possesses a wide range of capabilities, including image enhancement, freeze-frame mode, automated tracking of the heat pattern and recording. The venue was an indoor environment where the room temperature was between 20°C and 22°C and the humidity was about 60%. Heat sources such as sunlight or other electrical appliances were reduced to a minimum due to its effect on the ambient temperature. Prior to the screening, the patients were instructed to abstain from alcohol, cigarettes and any form of drugs that will affect the body biological system, which will result in a change in body temperature. In addition, the patient breast surface should be free from powder or ointments.

2.2. *Procedures for thermal imaging*

The patients were required to abstain from any physical activities for 20 minutes before the start of the thermal screening.[3] The rationale is to reduce the body metabolism rate so as to allow the overall body temperature to stabilize. During the thermographic examination, the patients are required to take off the top clothing and their hands will be positioned behind their heads. During the imaging, 3 thermograms are taken — 1 frontal image and 2 lateral images. Each image was then improvised digitally to enhance the resolution. For the analysis here, thermograms

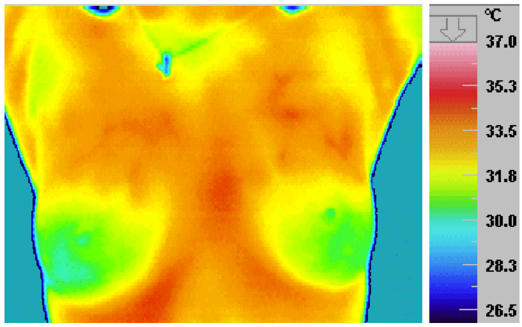

Fig. 1. Typical thermogram of an asymptomatic volunteer (age 24).

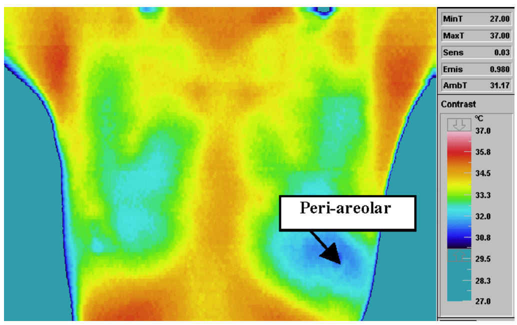

Fig. 2. Asymmetric thermogram of a 52 year old woman with left breast abnormality.

of 8 patients are excluded for analysis as 7 of the patients have history of mastectomy and 1 patient has highly distorted breast on one side. Thus, thermograms of 82 patients were used in the analysis: 30 asymptomatic patients (age $= 51 \pm 8$), 48 patients (age $= 46 \pm 10$) with benign breast abnormality on either side of the breast, and 4 patients (age $= 45 \pm 5$) with cancer on either side of the breasts.

Figures 1 and 2 show a typical thermogram of an asymptomatic volunteer (age 24) and a typical asymmetric thermogram of a 52 years-old woman with left breast abnormality. Based on free mammography conducted for 1000 Singapore women on the eve of the breast cancer awareness month (October 1998), the average size of tumor cancerous lump first detected in clinic[4] was 1.415 cm in spheroid shape.

From the breast thermograms, temperature data are extracted. The thermograms are consisted of many colored pixels, each representing a temperature. From the thermograms alone, it is also possible for an experienced medical practitioner to diagnose abnormalities such as a cyst. After every pixel's temperature is compiled, biostatistical technique can be used to treat them, such as determining the mean, median and modal temperature of the breast region.

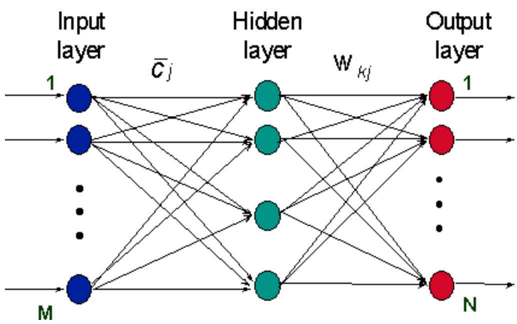

Fig. 3. Typical RBFN architecture.[5]

2.3. *Multi-pronged analysis of thermogram*

The analytical tools used in this study are biostatistical methods and ANN.[5] To enhance the performance of ANN RBF, a biostatistical method LR is incorporated to increase the accuracy of the diagnosis. This is done by selecting only useful and relevant inputs, which are used to predict the outcome.

For ANN, RBF is the main focus (Fig. 3). ANN is a pattern recognition program that has the ability to predict the outcome based on the various inputs fed into the program.[5] For breast cancer thermography applications, it has the ability to predict whether the breast is healthy or cancerous.

The final step is an evaluation test using ROC analysis. By generating the ROC curve, sensitivity, specificity and the area under the curve are obtained. These statistics are useful for the user to decide if the neural network is well built or not.

3. Design Approach

The proposed approach is a multipronged analysis comprising of LR, RBFN and ROC analysis (Fig. 4). LR and RBFN will do the prediction while ROC is used for evaluation and analysis purposes (Table 1).

3.1. *Step 1: Linear Regression (LR)*

LR reflects the correlation between the variables and the actual health status (healthy or cancerous) of the subject, which is decided by mammography. Hence, LR is used to decide if a particular variable should be used for inputs in the train file. In other words, a variable is used as input in the NN if, and only if, it has a strong correlation with the outcome (health status of the patient).

The following data are collected[3,6]:

- Temperature data from thermograms
- Mean temperature of left breast
- Mean temperature of right breast

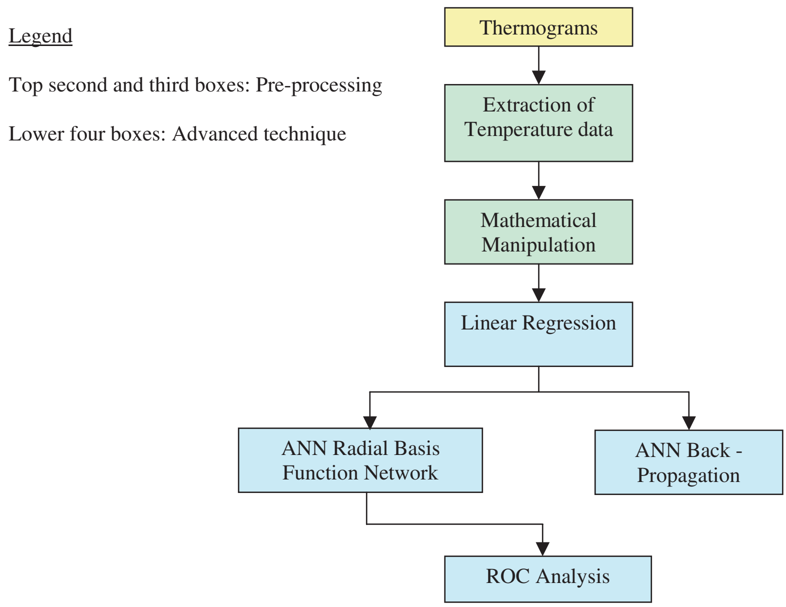

Fig. 4. Flow chart of advanced technique.

Table 1. Summary of flow chart for advanced technique in BC thermograms.

Purpose	Software
• View thermograms from thermal imager & extract temperature data	Image J
• Normalize raw temperature data	
• Perform statistical analysis (e.g. mean, median, std dev)	MS Excel Statistical Toolbox
• Determine the correlation of each variable with the output (health status)	MedCal
• Training and testing of data	NeuralWorks Pro II
• Building an algorithm for the data	
• Evaluate the effectiveness of the computed method	MedCal

- Median temperature of left breast
- Median temperature of right breast
- Modal temperature of left breast
- Modal temperature of right breast

Biodata from questionnaire include:

- Age of patient
- Family history of breast cancer
- Hormone replacement therapy
- Age of menarche

- Presence of palpable lump
- Previous breast surgery/biopsy
- Presence of nipple discharge
- Pain in the breast
- Menopause over 50 years of age
- First child over 30 years of age

3.2. *Step 2: ANN RBFN*

Based on the various inputs fed into the network, RBFN will be trained to produce the desired outcome, which is either positive (1) for cancer and benign cases or (0) for healthy cases. Different combinations of Learn Rule, Transfer Rule and Options are tested under the comprehensive umbrella of RBFN. When this is done, the RBFN algorithm possesses the ability to predict the outcome when there are new input variables. For this breast cancer study, back-propagation (BP) training and testing is also included and the results (accuracy of diagnostic = 61.5%)[7] are compared with that of RBFN.

3.3. *Step 3: ROC analysis*

ROC[8] is used as a final step to evaluate the effectiveness of the neural network by virtue of sensitivity, specificity and the area under curve. In other words, it is used to evaluate if the RBFN is well built or not.

4. Results and Discussion

4.1. *Step 1: Linear regression*

Table 2 indicates that the coefficient of determination for the temperature related data is generally higher than that of the biodata. This reinforces the fact that thermography can be used as an adjunct tool, as there is a relatively stronger correlation between the surface temperature of the breast and the health status of the patient as compared to the biodata.

The variable with the highest coefficient of determination is the modal temperature of the right breast as shown in Fig. 5. The variable with the lowest coefficient of determination is the "First child at more than 30 years old" criterion. Such observation also shows that the biodata has minimal implications on the health status of the subject.

4.2. *Selected results for Step 2: ANN RBF Single Layer Perceptron (SLP) with selected combination of learn rule and transfer rule*

The NNs that achieved an accuracy rate of 80% or more are presented in Table 3. Various combinations of Learn Rule and Transfer Rule are tested. Learn Rule

Table 2. Summarized results for step 1 in advanced technique of breast cancer thermograms.

No.	Independent X	Coeff. of determination
1	Mean Temp of Left Breast	0.03412
2	Median Temp of Left Breast	0.03110
3	Modal Temp of Left Breast	0.02850
4	Mean Temp of Right Breast	0.04740
5	Median Temp of Right Breast	0.04520
6	Modal Temp of Right Breast	0.04900
7	Age of Patient	0.00430
8	Family History	0.00500
9	Hormone Replacement Therapy	0.27313
10	Age of Menarche	0.04740
11	Presence of Lump	0.05190
12	Previous Breast Surgery/Biopsy	0.02650
13	Presence of Nipple Discharge	0.00830
14	Pain in the Breast	0.02500
15	Menopause at more than 50 yr old	0.02500
16	1st child at more than 30 yr old	0.00650

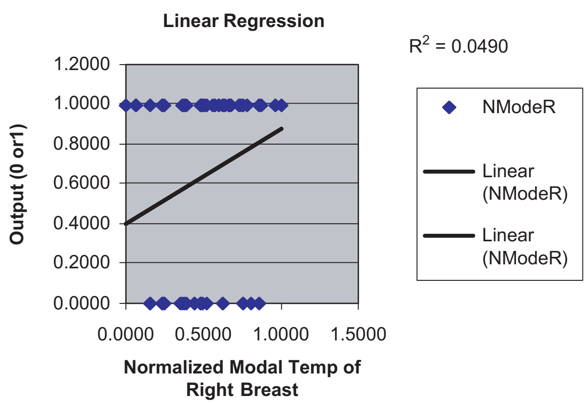

Fig. 5. Linear regression for modal temperature of right breast.

specifies how the connection weights are varied during the learning process. The more popular ones are Delta Rule and Norm-cum-delta. The former is the standard BP learning rule, whereas the latter uses the accumulation of weight changes and updates the weights at the end of the epoch. Transfer Rule is a non-linear function that transfers the internally generated sum for each processing element (PE) to a potential output value. Both linear and TanH (non-linear) are common mathematical functions. DNNA is the acronym for Digital Neural Network Architecture. It is an efficient implementation of a sigmoid-like activation function that has high densities of connections per area.

Table 3. Selected results for ANN RBF SLP with advanced technique on BC thermograms.

Learn rule	Transfer rule	Score (%)
Delta Rule	DNNA	80.95
Norm-Cum-Delta	DNNA	80.95
Ext DBD	DNNA	80.95
QuickProp	Linear	80.95
Delta Bar Delta	TanH	80.95

Table 4. Selected results for ROC analysis with ANN RBF SLP for advanced technique on BC thermograms.

Learn rule	Delta bar delta
Transfer Rule	Tanh
Area Under Curve	0.888
Sensitivity	81.2%
Specificity	88.2%

The highest level of accuracy attained by ANN RBF is 80.95% (versus BP score rate of 61.5%[7]), as shown in Table 3. This indicates that RBF is credible and effective in the prediction of the breast cancer to a huge extent. Its data are very complicated and large, as it has 10 input variables per subject (selected from LR).

4.3. *Selected results (with area* $>$ **0.85***) for Step* **3***: ROC analysis*

The score rate in Step 2 is only based on the number of correct predictions. However, it does not take into account the percentage of correct predictions out of the positive cases and the percentage of correct predictions out of the negative cases. Hence, there is a need for ROC analysis on the selected RBF NNs with high accuracy rate to further verify its effectiveness.

Evaluating the RBF using ROC shows that the NN model is well built (Table 4). The area under curve for many RBF NNs is larger than 0.85. In addition, the RBF NNs have high sensitivities (>75%) and high specificities (≈85%). This suggests that the overall diagnostic performance is competitive to that of mammography.

The best performing RBF NN (Table 4) is a Single-Layered Perceptron with Delta Bar Delta as the Learn Rule and TanH as the Transfer Rule. The area under the curve for this NN is 0.89 as illustrated in Figure 6.[a] It possesses very high sensitivity (81.2%) and high specificity (88.2%).

[a] Figure 6 illustrates the true positive (abnormal) rate in function of the false positive rate at different threshold temperature points. This nonparametric approach is free of distributional assumptions in that it depends only on the ranks of the observations in the combined sample, but the resulting empirical ROC curve is a series of horizontal and vertical steps (in the absence of ties and limited sample size), which may be quite jagged.

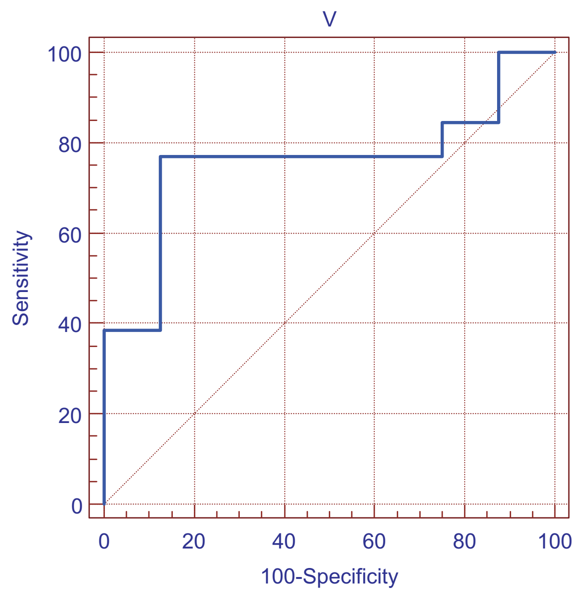

Fig. 6. A typical ROC curve for the selected RBFN.

5. Conclusion

Through the use of biostatistical methods and ANN, progress is made in thermography application with regard to achieving a higher level of consistency. This is made possible with the introduction of the advanced technique in thermogram analysis. It has become possible to use breast cancer thermography as a powerful adjunct tool, together with mammography for diagnosis. The advanced integrative technique (LR + RBF + ROC) has a high level of accuracy rate in prediction based on the temperature data extracted from the thermograms. This allows thermography to play an increasingly important role in the medical field, as it has the ability to detect cancerous cells in the early stages.

Acknowledgments

The authors would like to thank Dr. L. S. J. Sim, Consultant Radiologist of Singapore General Hospital for his valuable discussions and input to this project.

References

1. Breast Cancer Foundation: http://www.bcf.org.sg/ (Last accessed 15th July 2005).
2. Keyserlingk JR, Ahlgren PD, Yu E, Belliveau N, Infrared imaging of breast: Initial reappraisal using high-resolution digital technology in 100 successive cases of stage I and II breast cancer, *The Breast Journal* **4**(4):241–251, 1998.

3. Ung LN, Thermography as breast imaging modality, M.Eng. Thesis, School of Mechanical & Aerospace Engg, Nanyang Technological University, Singapore, 2000.
4. Wang SC, The singapore national breast screening programme: Principles and implementation, *Ann Acad Med Singapore* **32**:466–76, 2003.
5. Hopgood A, *Intelligent systems for engineers and scientists*, CRC Press, London, 2000.
6. Ng EYK, Fok SC, Ng FC, Sim LSJ, Breast tumour detection by thermography, *AcRF Report No:* NTU RG 69/98, 2001.
7. Ng EYK, Fok SC, Peh YC, Ng FC, Sim LSJ, Computerized detection of breast cancer with artificial intelligence and thermograms, *International Journal of Medical Engineering & Technology* **26**(4):152–157, 2002.
8. Receiver Operating Characteristics (ROC): http://www.medcalc.be/ (Last accessed 25th Sept. 2005).