Artif Intell Rev
DOI 10.1007/s10462-011-9274-2

A review of computer assisted detection/diagnosis (CAD) in breast thermography for breast cancer detection

Mehrdad Moghbel · Syamsiah Mashohor

Abstract Breast cancer is the leading type of cancer diagnosed in women. For years human limitations in interpreting the thermograms possessed a considerable challenge, but with the introduction of computer assisted detection/diagnosis (CAD), this problem has been addressed. This review paper compares different approaches based on neural networks and fuzzy systems which have been implemented in different CAD designs. The greatest improvement in CAD systems was achieved with a combination of fuzzy logic and artificial neural networks in the form of FALCON-AART complementary learning fuzzy neural network (CLFNN). With a CAD design based on FALCON-AART, it was possible to achieve an overall accuracy of near 90%. This confirms that CAD systems are indeed a valuable addition to the efforts for the diagnosis of breast cancer. Lower cost and high performance of new infrared systems combined with accurate CAD designs can promote the use of thermography in many breast cancer centres worldwide.

Keywords Thermography · Breast cancer · Computer aided detection/diagnosis · Fuzzy logic · Neural networks

1 Introduction

In 1956 Lawson suggested that the skin temperature over a cancerous area of the breast and the Venus blood draining from the cancerous tumor were higher than the surrounding normal tissue (Lawson 1956). In 1965, infrared imaging was introduced to the United

M. Moghbel (✉)
Intelligent Systems and Robotics Laboratory, Institute of Advanced Technology, University Putra Malaysia, Serdang 43400, Selangor, Malaysia
e-mail: mehrdad2275@gmail.com

S. Mashohor
Department of Computer and Communication Systems, Faculty of Engineering, University Putra Malaysia, Serdang 43400, Selangor, Malaysia
e-mail: syamsiah@eng.upm.edu.my

Published online: 06 July 2011

States by Gershen-Cohen et al. (1965), from the Albert Einstein Medical Center. Unlike mammography, infrared imaging does not use any ionizing radiation, venous access or any invasive procedures. Therefore, the examination poses no known harm to the patient because thermography is based on the infrared radiation emitted from the human body. All objects with temperature higher than absolute zero emit infrared radiation. The imaging procedure is also simple. The patient is seated in a room with a temperature of 18∼22 degrees centigrade for a time of 5–15 minutes so an equilibrium between skin temperature and environmental temperature is reached. Then an infrared camera records the infrared radiation emitted from the patient's body.

This ease of imaging and absence of the problems related to the ionizing radiation associated with mammography brought a high initial interest in Thermography and in the early studies done by the pioneers of this field consisting of more than 60,000 patients (Gershen-Cohen et al. 1965; Hoffman et al. 1967; Stark and Way 1974; **?**), thermography achieved an average specificity and sensitivity of over 88% which were higher than mammography at the time.

This high interest in thermography convinced the NCI (National Cancer Institute) to include thermography in their large scale study to compare different breast cancer detection methods which took place from 1973 till 1979, known as the Breast Cancer Detection and Demonstration Project (BCDDP) (Baker 1982). But due to inconsistent image interpreting protocols and poor study design, thermography was dropped from BCDDP. This led to a decreased interest on thermography especially in the US and only a handful of centers continued research and publication on thermography. On the other hand, other countries considered thermography to be a valuable assistance and were more active on the research. Today, many countries are considering thermography as a first line detection system such as japan with more than 1,500 hospitals and clinics which are using thermography as a first-line detection, Korea uses more than 450 centers and there are also many centers currently operating in Germany, Austria, United Kingdom, Poland, Italy, China, United States, Canada, Russia, Norway, Australia and South America (**?**).

With the declassification of the military infrared sensing technology in the 1990s a new type of uncooled infrared camera hit the market also known as a second generation infrared camera, which almost fixed all the problems faced by the previous generation infrared cameras such as thermal drift, poor sensitivity, time required for image acquisition (Head et al. 1996; Anbar 1998).

The availability of higher resolution images coupled with the emphasis on the early detection of cancer brought back the interest in thermography. Thermography has some unique properties such as the ability to detect early warning signs of cancer up to 10 years before the appearance of cancer on other imaging modalities (Gamagami et al. 1997; Head et al. 2000) or its ability to assess the treatment efficiency (Gamagami et al. 1997; Keyserlingk et al. 2001) and the prognosis condition of the patient (Head et al. 1993, 2000; Ohsumi et al. 2002). It was shown that 44% of the patients with abnormal thermograms will develop breast cancer within 5 years, also hot cancers showed a significantly poor prognosis with a 24% survival rate at 3 years while cooler cancers showed up to 80% survival rate at 3 years (**?**).

The sensitivity of mammography decreases on younger persons or women with dense breasts but thermography does not depend on the age of the patient or the density of breasts (Head et al. 1993; Qi and Diakides 2003), In 70% of cases signs of breast cancer will be detectable by infrared imaging up to 1 year before it is diagnosable with mammography (Head et al. 2000). Average tumor size undetected by thermal imaging is 1.28 cm while the average tumor size undetected by mammography is 1.66 cm (Keyserlingk et al. 2002). Thermography has obtained an average sensitivity and specificity of 90% (Head et al. 1999).

As a future risk indicator for breast cancer a persistent abnormal thermograph carries a 22 time higher risk and is 10 times more significant than a first order family history of the disease (Head et al. 2002; **?**). In recent years with the advances in image processing techniques and Computer Assisted Detection (CAD) thermography has achieved an average sensitivity and specify higher than mammography (Ng et al. 2005; Schaefer et al. 2007; Tan et al. 2007; Schaefer et al. 2008).

2 Acceptance of CAD in medical imaging

Early studies on using computers for analyzing medical images date back to 1960s, during that period many believed that computers can replace the physicians and so, much faith was put in computer diagnosis. However, computers were at early stages of development and lacked the processing power, also the advanced image processing techniques which exist today were not available during that time and digital images were rare. This yield to failure of computers in detecting abnormalities with an acceptable performance and the idea was pushed aside. However in 1980s scientists begin using computers in aiding the physicians by highlighting the problematic areas or helping physicians by providing second opinion on the diagnosis of abnormalities. This concept known as Computer Assisted Detection/Diagnosis (CAD) became widely accepted around the world. CAD does not try to replace the physician but rather works on the principle of aiding the physician on the diagnosis by providing additional information on image abnormalities or indicating the parts of image which looks normal to the naked eye but in fact abnormalities exists on that region or by giving a diagnosis result which can serve as a second opinion in decision making. CAD is currently employed in many aspects of daily clinical practice such as breast cancer, skeletal, brain and vascular disorders.

3 Definition of sensitivity, specificity and accuracy in medical imaging

3.1 Definitions of true and false positive/negative

- True positive: Sick people correctly diagnosed as sick (TP)
- False positive: Healthy people incorrectly identified as sick (FP)
- True negative: Healthy people correctly identified as healthy (TN)
- False negative: Sick people incorrectly identified as healthy (FN)

3.2 Sensitivity

Sensitivity is the probability that a test will correctly identify a person with the disease when in fact they do have the disease or sensitivity is the system's ability in truly identifying the sick people. Sensitivity is calculated by the following formula:

$$\text{Sensitivity} = \frac{\text{number of TP}}{\text{number of TP} \quad + \quad \text{number of FN}}$$

3.3 Specificity

Specificity is the probability that a test will correctly identify a person whom does not have the disease when in fact they are disease free. Specificity is calculated by the following formula:

$$\text{Specificity} = \frac{\text{number of TN}}{\text{number of TN} \quad + \text{number of FP}}$$

3.4 Total accuracy

Total accuracy shows the performance of the diagnostic system is determined based on the combination of sensitivity and specificity and is calculated by the following formula:

$$\text{Total Accuracy} = \frac{\text{number of TP} \quad + \text{number of TN}}{\text{number of TP} \quad + \text{number of FP} \quad + \text{number of TN} \quad + \text{number of FN}}$$

3.5 Tradeoff between sensitivity and specificity

Ideally, an imaging system should have high sensitivity and high specificity, around 100% but usually there are some tradeoffs between sensitivity and specificity. An imaging system might have a high sensitivity but lower specificity and vice versa, this is due to limitations on hardware manufacturing and available materials which affect the performance of the system diversely. Usually a system is designed in a way to keep both sensitivity and specificity high but still false positives and false negatives are inevitable (Institute 1999; Ng et al. 2002;2011).

4 Computer aided diagnosis/detection (CAD)

4.1 A brief history of CAD between 1977 and 2008

One of the earliest attempts for a thermography CAD system was attempted by Negin et al. (1977). They called their system; a computerized breast Thermographic Interpreter (CBTI), although their design was based on linear discriminant classifiers and the computer used was so basic that it did not even had a display and required 5 min to interpret one case and decide whether the breast is cancerous or not. They managed to achieve an average accuracy of 79% (Negin et al. 1977) and this is quite surprising considering the technology of that time. Even in 2002 with images from a second generation camera, the backpropagation neural network struggled to achieve such accuracy (Ng et al. 2002). During 1990s with the development of the Artificial Neural Networks (ANN) and introduction of systems based on Fuzzy Logic (FL) for image classification and the availability of high quality thermograms from second generation infrared cameras, Thermographic CAD systems received a new wave of interest from researchers. The most prominent CAD designs developed between 1977 and 2008 are based on ANN and FL.

4.2 Definition of artificial neural networks and fuzzy logic

4.2.1 Artificial neural networks (ANN)

An ANN is a bio physiological model of the human brain that attempts duplicate its operational procedures. The main part of any ANN is the processing element. ANNs combine these processing elements in different architectures to address different computational requirements.

4.2.2 Back propagation artificial neural network (BP-ANN)

BP-ANN was first developed by Werbos in 1974 and Parker developed the same idea in 1982. BP-ANN was popularized by Rummelhart et al. in 1986. It is a technique used to determine the training signal used for adjusting the weights of a given neuron in a neural network (Bronzino 2006).

BP-ANN is a feed forward network consisting of at least three layers: an input layer, a hidden layer and an output layer. It is an easy system to train and is extensively used in pattern recognition problems and image classification tasks. The BP neural network uses supervised learning to train, which is the use the output error of the neural network as a training signal. This means the data set used for training contains both the inputs and the corresponding outputs thus the error of the system will be determined and used for training.

4.3 Radial basis function network (RBFN)

RBFN was introduced to ANN during the late 1980s and it is a technique used to determine the training signal used for adjusting the weights of a given neuron in a neural network. While the output of RBFN is linear, its input is non-linear and this gives them great capabilities in modeling complex mappings, learning, classification, and decision making capabilities compared to BP-ANN. They have been used in many areas such as interpolation, control engineering, image restoration and 3D modeling. RBFN is a feed-forward neural network consisting of three layers; input, output and a hidden layer between them. The RBFN neural network requires supervised training.

4.4 Fuzzy logic

Fuzzy Logic was introduced in 1965 by Professor Lotfi Zadeh from the University of California at Berkeley. It is based on a human like reasoning for solving problems and thus the programming language used is very easy to understand. Fuzzy logic incorporates a simple, rule-based IF X AND Y THEN Z approach in solving problems rather than modeling solutions which are based solely on mathematical approaches.

Although it was initially used for the problem solving related to control systems, it soon found its way in artificial intelligence and image classification fields. The problem with the fuzzy logic is its dimensionality which means the number of rules increases exponentially with respect to the number of attributes involved which can present high computational and memory requirements and steps should be taken to ensure that the number of rules stays within a reasonable criteria. The fuzzy systems are commonly very robust and have high tolerances for errors in input data.

5 Performance of CAD in breast thermography

With the availability of more powerful computers, improved infrared cameras and better image classification algorithms, researchers experimented on different CAD systems for breast thermography incorporating both ANN and FL approaches. Although the results obtained vary from one approach to another, they all confirm the effectiveness of CAD in thermography.

The worst accuracy was obtained with the help of a BP-ANN which was 61.54% which is even lower than what was achieved 22 years before that. Although the system had good sensitivity of 68.97% the specificity of 40% lets it down significantly (Ng et al. 2002, 2003) The next improvement was to switch BP with RBFN, nonlinear properties of this approach increased sensitivity and specificity of CAD systems significantly, able to achieve 81.2% sensitivity, 88.2% specificity and a total accuracy of 80.95%. The performance of the CAD system based on RBFN is now comparable with the mammography supplying an average accuracy of 80% (Ng et al. 2005).

Some researchers adapted the FL for analyzing thermograms; the high point of fuzzy system is the human like reasoning of the system and its ability to explain the reasons for a certain decision. This gives a certain degree of confidence in the eyes of the operator as he/she can easily check the system's reasons for a diagnosis made. The CAD scheme based on fuzzy systems can reach an accuracy of 79.53% which is almost the same as RBFN and mammography (Schaefer et al. 2007), but the drawback of this approach is the number of the rules required for the classification. Basically more than 2,500 rules are required, which presents high computational and memory wise requirements. In an attempt to reduce the number of required rules, a genetic algorithm was used to come up with a smaller set of rules which is more practical.

Genetic Algorithms (GA) was first introduced by John Holland in 1975. It is based on the theory of evolution proposed by Darwin. It considers the each rule as a part of a population and by combining different rules it produces new rules, a process known as mutation, then all the rules are assessed on their ability on solving the problem which the system is being designed to solve, a value is given to each rule known as the fitness. Then the process repeats and new rules are generated and their fitness levels are assessed. This process can be repeated until a criterion is met like the total number mutations, time, best solution and etc. Then a predetermined number rules with the highest fitness number is selected as the rules set for the system. This process can take a long time but it reduces the computational requirements of the designed system significantly. Using these genetic operations a new hybrid fuzzy system was designed with a set of just 100 rules (Schaefer et al. 2008) compared to 2,812 rules of the previous system (Schaefer et al. 2007), yet the new system was able to achieve an improved accuracy of 80.89%. This means an improvement of almost 1% despite a reduction of near 250% in the number of the rules which confirms the usefulness of the genetic algorithms as a tool for reduction of rule set for a fuzzy system design.

Perhaps the biggest improvement can be seen when the superior reasoning properties of FL is combined with the learning properties of ANN. The design is called Complementary Learning Fuzzy Neural Network (CLFNN), which is capable of both positive and negative learning which means the system will learn just like a human. Thus the rule set derived mimics human reasoning as closely as possible, with both positive and negative reasons for its decision.

Fuzzy Adaptive Learning Control Network-Another Adaptive Resonance Theory (FALCON-AART) is a good example of CLFNN with superior data classification capabilities and requires relatively short time for training with respect to other approaches.

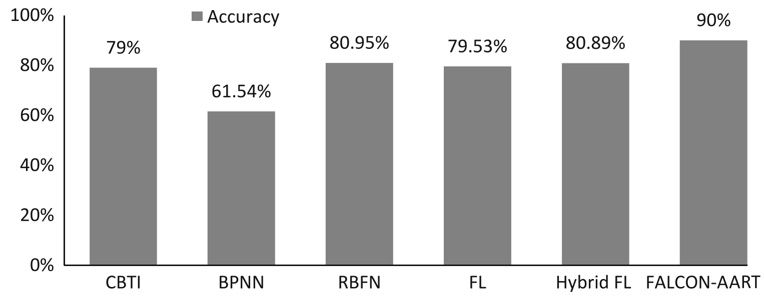

Fig. 1 Accuracy of different CAD designs for thermography image analysis

It generates a relatively small set of rules which makes it possible to incorporate the system on a wide variety of hardware designs. Also it has the ability to explain its reasoning which provides the user confidence in the system, as the operator can check the logic behind a certain decision given by the FALCON-AART. With a relatively small training data set of just 78 patients the FALCON-AART was able to achieve accuracy of almost 90% in detection of breast tumors while the data set used in the training of BPNN was 200 and in the case of FL data from 150 patients were used. The designed system based on FALCON-AART also showed an amazing accuracy of near 93% in differentiating malignant tumors from benign tumors providing a valuable advice to the physician (Tan et al. 2007), refer to Fig. 1.

6 Future of breast thermography

Although breast thermography has its limitations in sensitivity and specificity and it is dependent on examination conditions, it provides valuable information about the physiological condition of the breasts. Its ability to detect early physiological changes in breasts due to cancer formation can be used to detect patients whom require more thorough examinations, thus making the treatment more effective. With the help of computer protocols such as FALCON-AART for interpretation of thermograms, the thermography can be a very useful modality in detecting the patients whom are at a higher risk of developing cancer or on early stages of cancer which is not detectable on mammography, thus it help in reducing financial and sociological cost of cancer. But due to early problems associated when thermography when it was first introduced many physicians which were students during that time refuse to consider thermography even as a second opinion to the mammography. Blaming its high false positive rate of thermography which has been dealt with for almost two decades, also it is shown that the absence of cancer on mammography does not necessarily mean that it is not there. In fact 44% of people with an abnormal Thermogram will develop detectable breast cancer within 5 years and in 70% percent of patients, the cancer will be detectable by mammography one year after it was detected by thermography.

Although thermography's ability in risk assessment and determination of the prognostic conditions of the patient has been proved for a long time, now with the availability of highly accurate CAD systems and reduced cost of high performance infrared Thermographic systems, many cancer centers can easily afford to offer thermography alongside mammography. Also thermography can highlight areas of breast which represent some risk so the mammography of that area or breast can be examined more thoroughly or a CAD system incorporating mammography, thermography and physiological data of the patient can be utilized which can increase the detection of cancer to almost 95% (Keyserlingk et al. 2000), see Fig. 2.

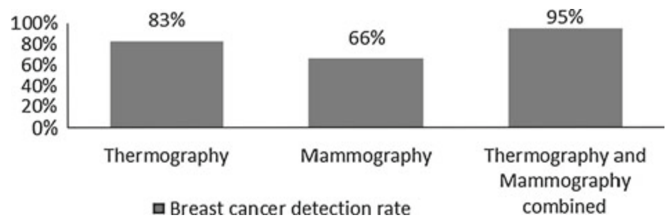

Fig. 2 Breast cancer detection rate for different imaging combinations (Keyserlingk et al. 2000 #94)

With the research in identification of the underlying causes of thermal properties of cancer such as angiogenesis, thermography will become more popular by time.

Despite the fact that there has been much improvement in thermal technology in recent years, most American physicians whom lost their interest in thermography mainly due to the breast cancer detection demonstration project (BCDD) still refuse to consider thermography as a complement to mammography, despite the fact that the FDA has approved thermography for breast cancer monitoring.

On the other hand due to simple image capturing procedures and lack of any ionizing radiation on the patient thermography continues to attract the attention of researchers from other parts of the world, with the advances in CAD systems for image analysis, the requirement for high level of training of image interpreters will decrease. Now more mammography centers can offer thermography alongside mammography to their patients and hopefully detection rate would increase dramatically. Many countries are using thermography such as Japan with 1,500 centers and South Korea with more than 450 centers and hopefully this number will increase each year.

References

(2011) Sensitivity and specificity. Retrieved 6 March 2011, from http://en.wikipedia.org/wiki/Sensitivity_and_specificity

Anbar M (1998) Clinical thermal imaging today. Mag IEEE Eng Med Biol 17(4):25–33

Baker LH (1982) Breast cancer detection demonstration project: five-year summary report. CA Cancer J Clin 32(4):194

Bronzino JD (2006) Medical devices and systems. CRC/Taylor & Francis

Gamagami P, Silverstein MJ, Waisman JR (1997) Infrared imaging in breast cancer. In: Proceedings of the 19th annual international conference of the IEEE engineering in medicine and biology society

Gershen-Cohen J, Haberman J, Brueschke (1965) Medical thermography: a summary of current status. Radiol Clin North Am 3(3): 403

Head JF, Wang F, Elliott RL (1993) Breast thermography is a noninvasive prognostic procedure that predicts tumor growth rate in breast cancer patients. Ann New York Acad Sci 698(1):153–158

Head JF, Elliot RL (2002) Infrared imaging: making progress in fulfilling its medical promise. Eng Med Biol Mag IEEE 21(6):80–85

Head JF, Lipari CA, Elliott RL (1999) Comparison of mammography and breast infrared imaging: sensitivity, specificity, false negatives, false positives, positive predictive value and negative predictive value. In: [Engineering in medicine and biology, 1999. 21st annual conference and the 1999 annual fall meeting of the biomedical engineering society] BMES/EMBS Conference, 1999. Proceedings of the first joint

Head JF, Lipari CA, Wang Fen, Davidson JE, Elliott RL (1996). Application of second generation infrared imaging with computerized image analysis to breast cancer risk assessment. Bridging disciplines for biomedicine. In: Proceedings of the 18th annual international conference of the IEEE engineering in medicine and biology society

Head JF, Wang F, Lipari CA, Elliott R (2000) The important role of infrared imaging in breast cancer. Mag IEEE Eng Med Biol 19(3):52–57

Hoffman RL (1967) Thermography in the detection of breast malignancy. Am J Obstet Gynecol 98(5):681

Institute NSL (1999) Sensitivity and specificity
Keyserlingk JA, Ahlgren P, Yassa M, Belliveau N (2002) Overview of functional infrared imaging as part of a multi-imaging strategy for breast cancer detection and therapeutic monitoring. In: [Engineering in medicine and biology. 24th annual conference and the annual fall meeting of the biomedical engineering society] EMBS/BMES Conference
Keyserlingk JR, Yassa M, Ahlgren P, Belliveau N (2001) Preliminary evaluation of preoperative chemohormonotherapy-induced reduction of the functional infrared imaging score in patients with locally advanced breast cancer. IEEE, London
Keyserlingk JRA, Ahlgren PD, Yu E, Belliveau N, Yassa M (2000) Functional infrared imaging of the breast. Eng Med Biol Mag IEEE 19(3):30–41
Lawson R (1956) Implications of surface temperatures in the diagnosis of breast cancer. Can Med Assoc J 75(4):309
Negin M, Ziskin MC, Piner C, Lapayowker MS (1977) A computerized breast thermographic interpreter. IEEE Trans Biomed Eng BME-24(4):347–352
Ng EY, Fok SC, Peh YC, Ng FC, Sim LS (2002) Computerized detection of breast cancer with artificial intelligence and thermograms. J Med Eng Technol 26(4):152–157
Ng EY, Fok SC, Peh YC, Ng FC, Sim LS (2003) A framework for early discovery of breast tumor using thermography with artificial neural network. Breast J 9(4):341–343
Ng EYKK, Kee EC, Acharya UR (2005) Advanced technique in breast thermography analysis. Proceedings of the 2005 IEEE engineering in medicine and biology 27th annual conference. IEEE, pp 710–713
Ohsumi STS, Aogi K, Usuki H (2002) Prognostic value of thermographical findings in patients with primary breast cancer. Breast Cancer Res Treat 74:213–220
Qi H, Diakides NA (2003) Thermal infrared imaging in early breast cancer detection—a survey of recent research engineering in medicine and biology society. In: Proceedings of the 25th annual international conference of the IEEE
Schaefer G, Nakashima T, et al (2007) Breast cancer classification using statistical features and fuzzy classification of thermograms. IEEE, London
Schaefer GN, Nakashima T, Zavisek M, Yokota Y, Drastich A, Ishibuchi H (2007) Breast cancer classification using statistical features and fuzzy classification of thermograms fuzzy systems conference, 2007. FUZZ-IEEE 2007. IEEE International, pp 1–5
Schaefer G, Nakashima T, Zavisek M (2008) Analysis of breast thermograms based on statistical image features and hybrid fuzzy classification. Adv Vis Comput 5358:753–762
Stark AM, Way S (1974) The screening of well women for the early detection of breast cancer using clinical examination with thermography and mammography. Cancer 33(6):1671–1679
Tan TZ, Quek C, Ng GS, Ng EYK (2007) A novel cognitive interpretation of breast cancer thermography with complementary learning fuzzy neural memory structure. Expert Syst Appl 33(3):652–666