Classification of Breast Thermography using Machine Learning for Early Detection of Malignancy

Krithika Sureshkumar[1], K.Suriya[2], Karthika.R[2], Priyadharshini.S[3]

[2] Assistant Professor, Easwari Engineering College, Chennai, India
[1,3,4] UG Scholar, Easwari Engineering College, Chennai, India

Abstract— Uncontrolled growth of cells in the breast tissue causes Breast Cancer and it has been the major cause of fatality in the developing countries. Although the existing methods of detection have been efficient, they largely focus on women above the age of 40. Early detection in a stage 1 mutation would be helpful in better diagnosis and treatment protocol thus ensuring longer lifespans. Therefore the thermography method proposes to take thermal images of the breast, highlighting the areas with extra growth. A pattern recognition technique that has been most effective in classifying tumours as benign or malignant – Fast support vector machine (FSVM) is used on the Thermo-gram. It is a three-step approach. In the first step, thermal image pre-processing and segmentation is done. The second step focuses on Textural Features Extraction using Curvelet transform. The percentage of area occupied by pixels with the higher temperatures of the image is calculated. The third step is used to extract classification of the tumour and highlight a boundary with the affected cells. Using this screening technique will pave way for a non-invasive, highly efficient, early detection method for women of all age groups.

Index Terms— Breast Cancer, Curvelet Transform, K-means, MEB, SVM, Thermogram

1 Introduction

Breast Cancer is one of the leading causes of death in the world right now and detection and treatment protocols are being devised every day to help women fight and survive it. Early detection is the key to faster and more efficient treatment. Unfortunately awareness of detection methods is considerably less in developing countries compared to high income countries. This, coupled with minimal access to detection techniques, prolonged treatment processes has curbed the survival rates to a bare minimum. The 5-year survival rates for breast cancer are much worse for low- and low-middle income countries such as Gambia (12%), Algeria (38.8%), India (52%) and Brazil (58.4%) than the developed countries. These staggering numbers show that though breast cancer is a threatening problem, proper methods of detection and treatment are not prevalent for people from different walks of life and ages. This is where the breast thermogram plays a vital role. Research has proved that thermography can be used as a screening tool and it has the ability to diagnose breast cancer at least ten years in advance. Since thermogram does not depend upon the structural aspects of the breast, it gives more conclusive answers for all women irrespective of their age, size of breast, tissue density, pregnancy or lactation period. Moreover infrared imaging does not require radiation, venous chemical inputs or any other invasive methods thereby causing no harm to the patient. Following this, feature extraction and boundary detection is done to sequester the infected area from the normal tissue. A region of interest is taken into account and transformation methods are used to collect features representing each part of the thermogram. [1] Further, these features, cumulatively provide an outline of the location of the infected tissue. Once the segregation is complete, machine learning algorithm is used to train the computer to classify malignant/ benign cells Extracted features are compared to a training database with threshold levels corresponding to cancerous cells. After multiple iterations, a definitive result of classification is given. This result along with the calculation of the affected area, its boundaries and manifestations through various layers of tissue equips the physicians with better and faster method of treatment.

2 Screening Process

2.1 Thermogram Acquisition

Thermogram is basically a depiction of infrared radiations from objects with temperatures above absolute zero. The intensity of radiation increases with temperature and is portrayed as various temperature levels on a thermal image. To obtain a thermogram, the body is cooled down below the normal temperature range for humans. A Thermo-cam is used to capture the radiations from various parts of the body, in this case the breasts. [2] The presence of a cancerous growth is associated with the excessive formation of blood vessels and inflammation in the breast tissue. These show up on the thermal image as areas with a higher skin temperature. Investigation of the skin temperature provides information about the extent of infected area which needs to be treated. Due to the high probability of real time imaging, thermograms are used to monitor patients even after surgery for removal of infected tissue. Though it is effective and 100 % harmless, obtaining high contrast images with less signal to noise ratio is still proving to be difficult. Despite these, Thermography is considered the best choice for screening.

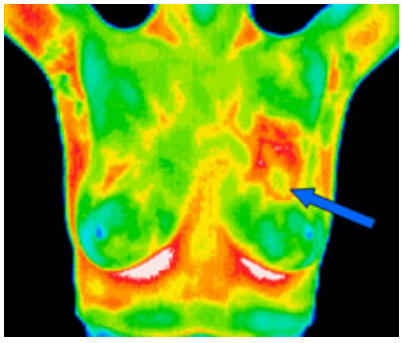

Fig 1: Thermogram showing a high temperature region with tumour growth

2.2 Drawbacks of Older Methods

Breast Self-Exam is very instrumental in discerning a problem at the early stage. Confirming the existence of tumour needs further screening and processing. For many years, various methods have been implemented, each having their own merits and demerits. It is vital to take into account the patient's age, body structure, tissue density, medical history etc before finalizing on a detection method. The most prevalent methods are as follows,

1. 3D Mammogram
2. Sonogram
3. MRI Imaging
4. Molecular Imaging

These methods, though effective in their own way, need to be followed up by a biopsy to conclude if the malignant cells are present or not. Generally, an uncomplicated, straightforward result takes only about 2-3 days. If the tissue has to be extracted from the patient and analyzed, it may take upto a week or more, during which time the growth of cells may spread to neighboring regions causing the formation of contra lateral breast cancer. Metachronous contralateral breast cancer is defined as a tumour in the opposite breast which was diagnosed more than 6 months following the detection of the first cancer. [3] The second cancer could have occurred due to the initial cancer in the same breast or it could be a new cancer in the same breast or a new growth in contralateral breast. 50% of all secondary cancers diagnosed are among patients with primary breast cancer. Therefore, early, non-invasive detection will be instrumental in diagnosing the cancer and avoiding stimulation of further growth of cells. Taking the other methods into consideration, various drawbacks can be deduced.

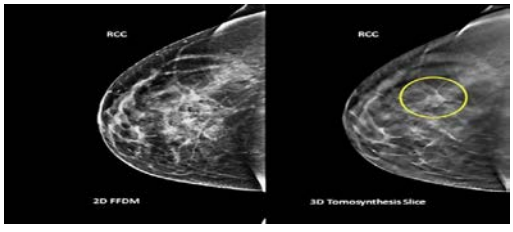

Fig 2: Comparison of 2D and 3D Mammogram.

The image depicts a 3D mammograms ability to discern a cancer cell from within layers of tissue as supposed to the 2D mammogram that only shows the growth of a mass. The mammogram will be able to give a definite answer only after multiple screenings. This exposes the patient to constant and recurrent radiation and may trigger the growth of more cells. Moreover recall screening is done in most cases because of an initial false result. From the medical standpoint, a screening mammogram (subsequent round) has a false-positive rate of up to 4% which is considerably less, but keeping in mind the emotional distress the patient has to go through and the fact that nearly one in seven women is diagnosed with breast cancer, the number does seem large.

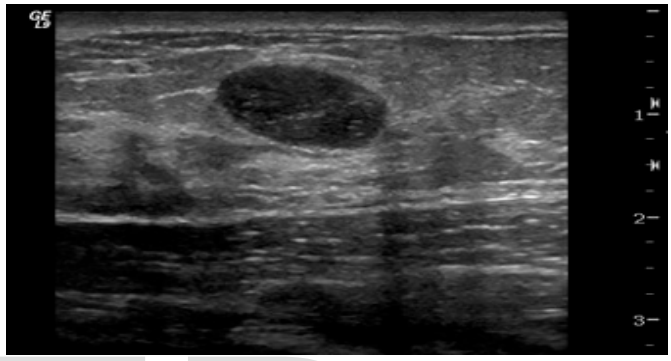

Fig 3: Sonogram of a breast having a fluid filled cyst

A breast ultrasound or sonogram is used to see whether a breast lump is filled with fluid (a cyst) or if it is a solid lump. An ultrasound acts an as axillary to the mammogram and not as an independent screening method. The use of excessive ultrasonic waves may also cause issues since the breast tissue isn't identical to the lining of the stomach over which fetal ultrasounds are taken. The possibility of false positives is also comparatively high to that of mammogram.

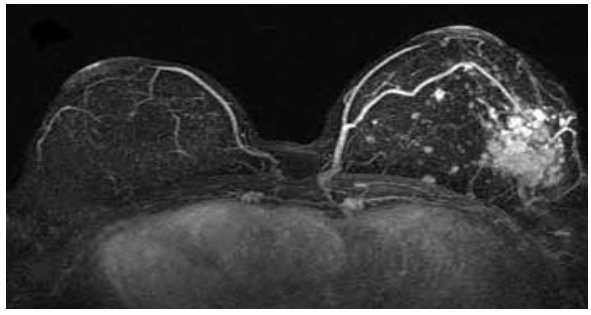

Fig 4: MRI Imaging of cancer in the right breast

Breast magnetic resonance imaging (MRI) is more invasive than mammography because a contrast agent is given through an IV before the procedure. It uses magnetic fields to create an image of the breast and is mostly used in breast cancer staging rather than detection. Though accuracy levels have improved in recent times, this method was not opted for a very long period due to its complexity and prolonged time of operation.

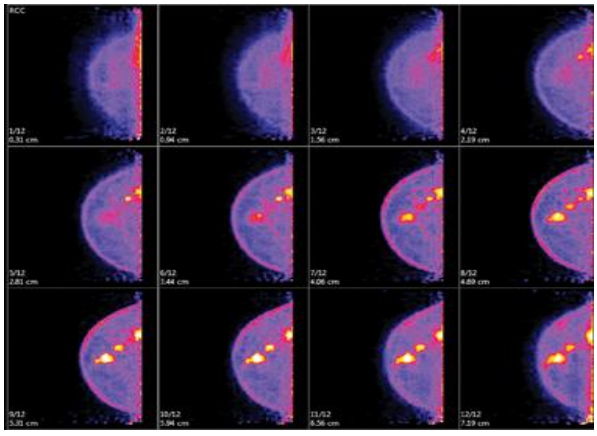

Fig 5: Molecular detection of tumour using a nuclear tracer

Molecular Imaging is a new and upcoming method. Though it may give the required results with the advent of new technology and safer equipments, this procedure is still uncommon in deveopling countries. In addition to that, a nuclear tracer is injected into the veins to clearly illuminate the structure of the tumour. [4] This may cause long term defects in the patient. Since, this mechanism depends on the breast structure and tissue, it's advisable only for older women with denser breasts.

2.3 Processing and Segmentation

Image acquisition is generally followed by a set of processing steps that carry out denoising and enhancement operations on the image making it better suitable for further processing and calculation. The infrared image that is collected is an RGB image unlike those from the other screening methods. Therefore, the techniques used must cater to the RGB model and primitive noise filters will not be sufficient. Contrast enhancement is done on the image to sharpen the edges and intensify the difference in the background and foreground colors giving the thermal image, a much needed meaning. Generally areas depicted with red in the infrared image may suggest the presence of tumour cells because that region exhibits maximum heat. On the contrary blue regions are cooler and mostly will not contain any abnormality.

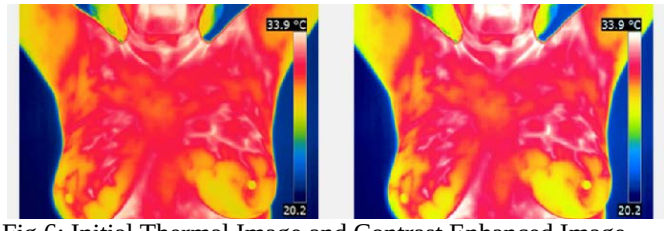

Fig 6: Initial Thermal Image and Contrast Enhanced Image

After pre-processing, segmentation is required to be performed. Segmentation and deduction of the region of interest (ROI) is vital in understanding the extent of the growth and to make targeted treatment plans. Clustering is the segmentation technique used. It separates the initial image into three separate regions based on the primary colors. This assists in narrowing down the region where the tumour cells may be present. Separate iterations are done on each cluster to find the presence or absence or partial presence of the abnormal cells.

K-means clustering algorithm (Hartigan and Wang, 1979) (Lloyd, 1957) (MacQueen, 1967) is the most common investigatory algorithms in data analysis. [5] It subtracts the interest area from the background and divides the image into 2 layers which separates the hotter region from the background image. This will ease the hot spot detection on thermal infrared images. The algorithm locates K - means value throughout a data set and highlights its representation. It is an iterative algorithm in which K-Means values are spread throughout the set in accordance with its Euclidean mean value.

Each cluster is separately iterated and checked for abnormalities. This process ensures that the region of interest calculated is a true and acceptable value.

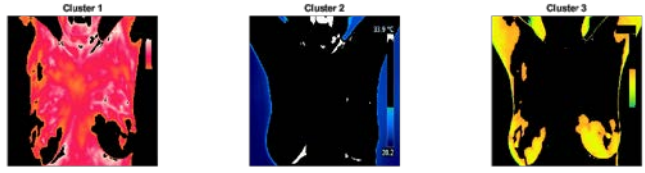

Fig 7: 3 way clustering of thermogram using K-means algorithm based on RGB model

3 Feature Extraction

A thermal image has a plethora of features that can be extracted and analyzed to come to a definite conclusion about the nature of the tumour. These features, namely Mean, variance, standard deviation, energy density, entropy, contrast, skewness, kurtosis, smoothness and many others are extracted using Curvelet or Ridgelet Transform. Generally, wavelet transform is the much sought after method, but in recent years, curvelet transform has surpassed its predecessor on many accounts.

3.1 Curvelet Transform

Curvelet transform is a multi-scale geometric wavelet transforms, and it is more efficient in representing edges, curves and singularities than traditional wavelet. Curvelet combines multi-scale analysis and geometrical ideas to achieve the optimal rate of convergence by simple thresholding. Multi-scale decomposition slices the image into innumerable segments thereby capturing point discontinuities in the linear structures Curvelets in addition to a variable width have a variable length and so a variable anisotropy making it suitable for extracting features that are harder to notice in plain sight. [6] It gives the possibility to analyse an image with different block sizes, but with a single transform. Curvelets can be interpreted as numerous groupings of nearby wavelets into linear structures so that they can capture the smooth discontinuity of curve more precisely.

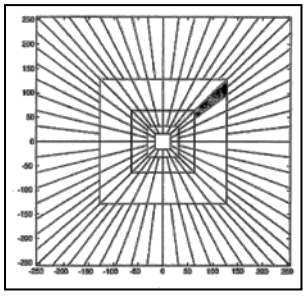

Fig 8: Depiction of anisotropic decomposition in curvelet transform

De-noising of the thermal image is also carried out by the transform and here too, the curvelet performs better than than wavelet transform.

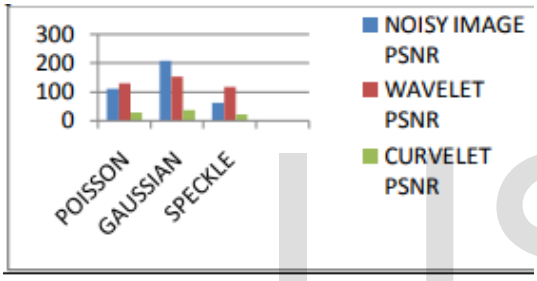

Fig 9: Comparison of Curvelet and Wavelet De-noising capabilities

3.2 Features

As mentioned above, various quantities can be deduced from the image and they are instrumental in the SVM classification process. Some of the important features are:

1. THERMAL DENSITY: Thermal density is the measure of difference between the textural densities in the thermal image and the complementary function image.
   BW1= binary image area for the thermal threshold image
   BW2= binary image area for the complementary function image
   Total1 = bwarea(BW1)
   Total2 = bwarea(BW2)
   Thermal density=total2-total1

2. ENERGY DENSITY: The energy is expected to be high if the occurrence of repeated pixel pairs is high. Where, P(i,j) are the pixel values at the (i,j) coordinates of the image.

3. ENTROPY: Entropy measures the randomness of a gray-level distribution. The entropy is expected to be high if the gray levels are distributed randomly throughout the image

4. HOMOGENITY: The distribution of the infected cells and it's spread is calculated
5. SKEWNESS: Pixel differences in the same layer of inspection
6. KURTOSIS: Colour and thereby heat density difference between two layers of image segmented according to the RGB model
7. MEAN AND VARIANCE: Mean temperature value and it's difference with neighbouring cells
8. CONTRAST: Contrast is a measure of colour variations between different texture areas in an image. Image with poor edges has low contrast. Image with sharp edges has high contrast.

Once the values are obtained using the formulas and respective functions, these values are grouped as an array of similar descriptors. (i.e) [T1,T2,T3…..] where T1,T2,T3…Tn refer to the various groups of descriptors. This array is further compared to another existing array to finish the classification process.

4 CLASSIFICATION

Classification techniques usually consist of taking a biopsy and analyzing it for abnormalities that suggest the presence of cancer or benign cells. Since this procedure is invasive and sometimes takes a long time to give a conclusive answer. Therefore, the fastest classification machine learning method called Fast/ Core Support Vector Machine is trained with a medical database of values pertaining to cancer patients. [7] These existing values are compared and correlated with the features and values extracted from the curvelet transform. An SVM train is added to the existing image and the computer calculates the probability of malignant cells present. The arrays of similar descriptors are grouped according to the probability of features in a specific region being high, medium or low. [8] High probability of deviant values in one place is a manifestation of malignant cells.

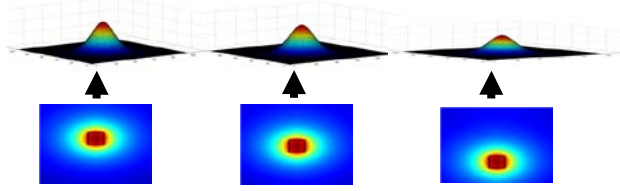

Fig 10: Grouping of feature descriptors into high, medium and low probability

The various classification outputs that can be derived are: [9]

1. True Positive: This result confirms the cells as malignant and the SVM computes the affected area and accuracy of calculation by multiple iterations.
2. True Negative: This result confirms the tumour cells as benign and prompts the computer to analyze a different cluster in the odd chance of the cells being present in other layers and regions of tissue.
3. False Positive: This a unacceptable result that points out the presence of malignant cells when they are benign in reality.

4. False Negative: This result is when the SVM algorithm fails to detect the presence of cancer cells irrespective of their presence because their manifestation isn't very prominent and the features don't match with the exact threshold values of the training database.

The last two results are unsatisfactory and usually occur in biopsies. The use of SVM is to eliminate these false cases and provide a high level of accuracy and faster results. Moreover FSVM is used to make sure the number of training data can be reduced for easier implementation. After the classification result finalizes the presence of cancer, identification of the exact location of the cells is carried out. This is performed by Minimum Enclosing Ball Algorithm (MEB). The malignant cells may not only be present over the surface tissue layer, but it may have also penetrated into the interior walls. MEB is used to detect a boundary of infected cells and sequester them in the shape of spheres. [10] This spherical area is decomposed and further layers are inspected for abnormality and a graphical output is presented depicting the exact cell locations in the entire thermogram.

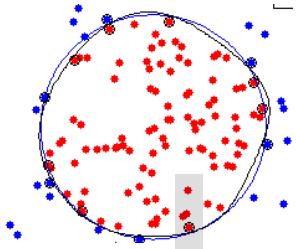

Fig 11: MEB boundary detection. Red dots represent malignant cells, blue dots are benign cells.

5 RESULT

After the preprocessing, segmentation, feature extraction and classification procedures, the thermogram image is further processed for boundary detection of the malignant cells. Once a cumulative answer with the nature of the tumour, its potential expansion in the tissue, its texture and density are calculated, an accurate report can be made for the patient. This method not only provides a non-invasive screening method, but it also utilizes the fastest classification method ensuring the provision of highly accurate results when compared to earlier methods. The ultimate aim is providing a platform for women of all ages to be able to diagnose the cancer at an early stage and get them treated so as to ensure higher survival rates. It is a seamless and easily operable mechanism which provides instant, efficient, automated results thus revolutionizing and entangling the fields of Medicine, Image Processing and Machine Learning.

REFERENCES

[1] R.C.Gonzalez, R.E. Woods, Digital Image Processing 3rd Edition, Pearson Education

[2] Elham Mahmoudzadeh, Maryam Zekri, Mohammad Ali Montazeri, Saeid Sadri, Sima Taghizadeh Dabbaggh,—Directional SUSAN image boundary detection of breast thermogram, IET Image Process ,2016, pp: 552-560

[3] Tomi F. Akinyemiju, Risk of Asynchronous Contralateral Breast Cancer: Multiple Approaches for a Complex Issue, Gland Surg. 2013 May; 2(2): 110–113.

[4] HumayunIrshad, Student Member, IEEE, Antoine Veillard, Ludovic Roux, and Daniel Racoceanu, Member, IEEE 2015, Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review—Current Status and Future Potential, IEEE Reviews in Biomedical Engineering (Vol 7) , 2013,pp. 97 – 114

[5] Faten Abu Shmmala, Wesam Ashour, Color Based Image Segmentation using DifferentVersions of K-Means in two Spaces, Global Advanced Research Journal of Engineering, Technology and Innovation (ISSN: 2315-5124) Vol. 1(9), January, 2013, pp. 030-041,

[6] Monika Shukla, Dr.Soni Changlani,—A Comparative Study of Wavelet and Curvelet Transform for Image Denoising, (IOSR-JECE) Vol 7(4), 2013,pp: 63-68.

[7] Bryan Christopher Catanzaro, Narayanan Sundaram, Kurt Keutzer, —Fast Support Vector Machine Training and Classification on Graphic Processors, UCB/EECS, 2008

[8] J. Phan, R. Moffitt, J. Dale, J. Petros, A. Young, M. Wang; "Improvement of SVM Algorithm for Microarray Analysis Using Intelligent Parameter Selection"; Engineering in Medicine and Biology Society, 2005, IEEE-EMBS 2005; 27th Annual International Conference of the IEEE; vol., no., pp.4838-4841.

[9] Pragati Shrivastava, Piyush Singh, Gaurav Shrivastava,—Image Classification using SOM and SVM Feature Extraction,(IJCSIT) Vol 5(1), 2014,pp:264-271

[10] Jair Cervantes, XiaoouLi, WenYu, KangLi, ,—Support Vector Machine Classification of Large Data Sets via Minimum Enclosing Ball Clustering, Neurocomputing, (Vol 71) Issues 4-6, Jan 2008, pp 611-619