Designing of Ground-Truth-Annotated DBT-TU-JU Breast Thermogram Database Toward Early Abnormality Prediction

Mrinal Kanti Bhowmik ⬩, *Senior Member, IEEE*, Usha Rani Gogoi, Gautam Majumdar,
Debotosh Bhattacharjee, *Senior Member, IEEE*, Dhritiman Datta,
and Anjan Kumar Ghosh, *Senior Member, IEEE*

Abstract—The advancement of research in a specific area of clinical diagnosis crucially depends on the availability and quality of the radiology and other test related databases accompanied by ground truth and additional necessary medical findings. This paper describes the creation of the Department of Biotechnology-Tripura University-Jadavpur University (DBT-TU-JU) breast thermogram database. The objective of creating the DBT-TU-JU database is to provide a breast thermogram database that is annotated with the ground-truth images of the suspicious regions. Along with the result of breast thermography, the database comprises of the results of other breast imaging methodologies. A standard breast thermogram acquisition protocol suite comprising of several critical factors has been designed for the collection of breast thermograms. Currently, the DBT-TU-JU database contains 1100 breast thermograms of 100 subjects. Due to the necessity of evaluating any breast abnormality detection system, this study emphasizes the generation of the ground-truth images of the hotspot areas, whose presence in a breast thermogram signifies the presence of breast abnormality. With the generated ground-truth images, we compared the results of six state-of-the-art image segmentation methods using five supervised evaluation metrics to identify the proficient segmentation methods for hotspot extraction. Based on the evaluation results, the fractional-order Darwinian particle swarm optimization, region growing, mean shift, and fuzzy c-means clustering are found to be more efficient in comparison to k-means clustering and threshold-based segmentation methods.

M. K. Bhowmik, U. R. Gogoi, and A. K. Ghosh are with the Department of Computer Science and Engineering, Tripura University, Tripura 799022, India (e-mail: mrinalkantibhowmik@tripurauniv.in; ushagogoi. cse@gmail.com; anjn@ieee.org).

G. Majumdar and D. Datta are with the Agartala Government Medical College (AGMC) & Regional Cancer Centre, Tripura 799006, India (e-mail: drgmajumdar@yahoo.in; datta_dhritiman@yahoo.in).

D. Bhattacharjee is with the Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032, India (e-mail: debotosh@ieee.org).

Index Terms—Breast cancer, DBT-TU-JU breast thermogram database, evaluation metrics, ground truth, standard acquisition protocol, suspicious abnormal region.

I. INTRODUCTION

DESIGNING of a medical image database annotated with ground truth images and findings of additional medical tests may improve the quality and throughput of the research in the medical domain. In the women health issues, the breast diseases become the most common and life-threatening issue. However, if discovered early, breast cancer is a highly curable disease with 97% chances of survival [1]. In this respect, Infrared Breast Thermography (IBT) has distinguished itself as the earliest breast cancer detection methodology by having the potential to detect breast cancer ten years earlier than the traditional gold standard method, X-ray mammography [2]. The painless and radiation-free nature of thermography makes the breast thermography more reliable and feasible in early breast abnormality detection [2], [3]. In literature [4]–[6], it is reported that the thermograms those possess asymmetric thermal patterns signify the physiological dysfunction in patient's breasts most of the time. Compared to these thermograms, the thermograms having increased nipple temperature, hotspots, and vascular changes may be more suspicious and indicate more severe breast problems [7]. Hence, for identifying the presence of any abnormality in the breast thermograms, the extraction and analysis of hotspots may play a significant role. However, for extracting the hotspot regions from a breast thermogram, one must have to rely on any image segmentation method. However, in the absence of the ground truth image of the hotspot region, picking a particular method for hotspot segmentation becomes more challenging and inaccurate. Moreover, due to the unavailability of sufficient medical cases and amorphous nature of infrared thermograms, the ground truth generation is also very tedious. Unavailability of ground truth images of the breast thermograms results in (a) very limited work on hotspot extraction and hot region based breast abnormality detection in the literature and (b) difficulty in performing the quantitative evaluation of various medical image segmentation algorithms. Therefore, the availability of a ground truth image for each thermogram is a necessity for evaluation of the performance of the techniques developed for hotspot segmentation.

TABLE I
DETAILS OF VARIOUS PRIVATE/PUBLIC BREAST THERMOGRAM DATABASES USED IN RESEARCH WORKS, INCLUDING OUR DATABASE

Database Name/Designing Institute Name, year		Protocol Suite Used	Thermal Camera Used/Image Resolution (pixels)/Thermal Sensitivity	No. of Subjects	Breast thermogram Views
A Breast Cancer Screening Clinic at Moncton Hospital, 2002	[9]	PP, PA, CEC	Thermovision 680 (Agatronics)/128 × 128/NA	86	NA
Perlmutter Cancer Center at NYU Langone, 2003	[10]	PA, CEC, Distance: 5 ft	NA/320 x 240/0.05 °C.	517 (343 H, 110 M, 63- Treated)	NA
Singapore general Hospital, 2007	[11]	PP, PA, CEC	Avio TVS-2000 MkII ST/NA/NA	90	3 views: F, Lt. & Rt. lateral
Fujian Normal University, 2007	[12]	PA, CEC	Thermal Texture Mapping (TTM, TSI-21)/NA/0.05 °C at 30 °C.	44 (40H, 3M, 1Hy)	NA
Not Provided, 2009	[13]	PP, PA, Distance: 2.2 Meter	AIM256Q camera/256 × 256/17.3 mK at 300 K.	20 (20 H)	NA
Center for the Study and Prevention of Cancer (CEPREC), 2010	[14]	PA, CEC	IR camera DL-700 / 320 × 240/0.08 °C at 30 °C	911	5 views: Bilateral F, Lt. & Rt. oblique, Lt. & Rt. close-up
National Taiwan University Hospital, 2010	[15]	PP, PA, CEC, Distance: 2.5 Meter	ATIR-M301 Thermal Imaging System /320 × 240/less than 0.1 °C	276 (165 M, 111 B)	5 Views: F, Lt. & Rt. lateral, Lt. & Right oblique
'Hospital Dr. Ignacio Morones Prieto', San Luis Potosi, Mexico, 2011	[16]	PP, PA, CEC, Distance: 1.5 Meter	FLIR T400 Infrared Camera/320 × 240/0.05 K at 30 °C.	20 (20 M)	5 views: F, Lt. & Rt. Oblique, Frontal close-up of Lt. & Rt.
Vijaya Scans & Paterson Cancer Center, Chennai, India, 2011	[17]	PA, CEC, UEC	2 Cameras: Ti40FT from M/s Fluke Corp/160 × 120/90 mK. Varioscan 3021 ST/360 × 240/30 mK	50	3 views: F, Lt. & Rt. lateral
Not Provided, 2011	[18]	PA, CEC, Distance: 100 cm	Meditherm med2000 Thermal Imaging System/NA/NA	63	3 views: F, Lt. & Rt. oblique
Hospital das Clínicas of the Federal University of Pernambuco, 2013	[19]	PP, PA, CEC, Distance: 1 Meter	ThermaCAM^{TM} S45/320 × 240/NA	NA	2 image series: D1 & D2 D1: 5 views: F, Lt. & Rt. inner side, Lt. & Rt. outer side. D2: 5 views: F, Close views of Lt. & Rt., Lt. & Rt. outer side.
DMR Database, Federal University Fluminense, 2014	[8]	PP, CEC, PA, Distance: 1 Meter	FLIR SC620/640 × 480/less than 0.04 °C	287	5 views: F, Rt. lateral 90°, Lt. Lateral 90°, Lt. Lateral 45°, Rt. Lateral 45°
Breast Clinic of the Cancer Institute of Tehran, Iran, 2016	[20]	PA, CEC	FMG-MED IR camera/640 × 480/0.08 °C.	132	NA
Hospital at Manipal University & A unit of Central Diagnostic Research Foundation, 2016	[21]	PA	3 Cameras: Meditherm IRIS2000/320 × 240 FLIR E60 camera/320 × 240 FLIR T650sc camera/640 x 480	265 (120 H, 53 B, 8 non-M, 6 lacting mothers, 78 M)	5 Views: F, Lt. & Rt. lateral, Lt. & Rt. oblique
Fanavaran Madoon Ghermez (FMG) Co. Ltd & Imam Khomeini Hospital, 2016	[22]	PA, CEC, Distance: 1.5 Meter	Thermoteknix VisIR 640/640 × 480/NA	67	5 views: F, Lt. & Rt. lateral, Lt. & Rt. oblique
Breast Clinic, Vijaya Health Centre, Chennai, 2017	[23]	PP, PA, CEC	ICI7320P/480 × 640/NA	72 (36 N, 18B, 18M)	12 Views at an interval of 30°
DBT-TU-JU Database		PP, PA, CEC, Distance: 1 meter/variable distances	FLIR T650sc/640 × 480s/0.02 °C at 30 °C	100** (45 H, 36 B, 13 M, 6 Mild breast problems)	6 views: Supine, F, Lt. & Rt. lateral, Lt. & Rt. oblique

PP - Patient Preparation, PA - Patient Acclimation, CEC - Controlled Environmental Condition, NA - Not Available, H - Healthy, B - Benign, M - Malignant, F - Frontal, Lt - Left, Rt. - Right, UEC - Uncontrolled Environmental Condition, Hy – Hyperplasia, **As of 2016, December, 12th (database still growing).

A number of research works on IBT-based breast abnormality detection has been found, where most of the works were performed on the breast thermograms of some private databases that are available only for the internal purposes and can be accessed by the patients and their physicians only. However, till now, only one publicly available breast thermogram database, the DMR (Database for Mastology Research), has been reported in [8]. This database comprises of breast thermograms of 287 subjects, out of which breast thermograms of almost 47 subjects are labeled as 'Sick' and remaining are labeled as 'Healthy'. Although DMR database provides sufficient data for validation of a method, the number of pathological cases is very less compared to the healthy cases, and the results of mammography or other screening modalities are not available for all the cases. Moreover, in the absence of ground truth images of the suspicious regions in breast thermograms, evaluation of any segmentation method for hotspot extraction is also not possible. A brief summary of most of the private and public breast thermogram databases [8]–[23] used for various research purposes are provided in Table I. While providing the list of the
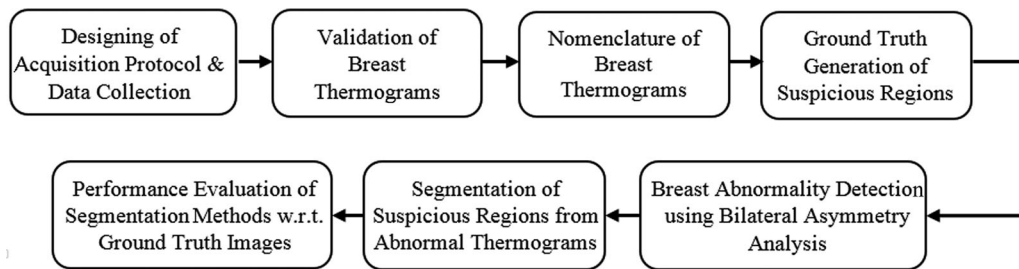
Fig. 1.   Flowchart outlining various sections of this paper.

breast thermogram datasets used by various researchers, it is worth mentioning that due to the limitation of space, we had not covered the details of all of the used breast thermogram datasets, but Table I contains the details of most of the used databases.

Considering all these factors and mainly the necessity of a ground truth associated breast thermogram database, this paper emphasizes on the designing of a breast thermogram database that comprises of the thermograms of various breast diseases. Moreover, the database is also annotated with the findings of the additional medical tests and the ground truth images of the suspicious regions. The flow of this paper is demonstrated in Fig. 1.

In the process of designing an infrared breast thermogram database, it is worth mentioning that the acquisition of breast thermograms should be performed under a strict protocol suite due to the sensitivity of IBT to the environmental conditions. However, in the absence of any universal breast thermogram acquisition protocol, a rigorous study has been made on the breast thermogram acquisition protocols. Based on the study, it has been found that there are some common factors considered by several research works during the acquisition of the breast thermograms. Considering all these critical parameters, we have designed a standard breast thermogram acquisition protocol suite for the collection of breast thermograms. In our previous work [24], a detailed description of the designed protocol suite has been illustrated. Here in this paper, due to the limitation of space, a brief description of the designed protocol suite is provided in Section II.

Designing of the acquisition protocol suite and collection of the breast thermograms are followed by the validation of the findings of IBT with the findings of other breast imaging modalities. Validation of the thermograms is followed by the ground truth image generation of the suspicious regions from the abnormal thermograms. In the absence of a standard method for ground truth preparation, we have adopted a method for making ground truth images of the hotspot areas where a ground truth image is formed based on the joint agreement of two or more medical experts' manual segmentation results. The detailed description of the ground truth image generation is provided in Section V. Although this manual segmentation may not be as perfect as the ground truth, it may provide a way to evaluate a proposed hotspot segmentation method in a quantitative way. The generation of the ground truth images is followed by the computer assisted analysis of the breast thermograms to predict the presence of breast abnormality by doing the bilateral

asymmetry analysis. Then, from the predicted abnormal thermograms, the suspicious hot regions are extracted by using six state-of-the-art image segmentation methods. A comparison of the segmentation results of all the segmentation methods is made against the ground truth images based on five widely used evaluation metrics to find out the efficient segmentation methods for abnormal hot region extraction. Extraction of these hotspot areas allows detecting and locating the potentially suspicious regions in a breast thermogram and this suspicious region localizing, in turn, can guide the subsequent diagnosis modalities like Fine Needle Aspiration Cytology (FNAC), when other medical reports appear to be normal.

Main contributive features of the designed DBT-TU-JU Breast Thermogram Database are as follows:
1) The database comprises of the breast thermograms of subjects with a variety of breast diseases.
2) It contains 6 breast thermograms, captured in supine view as well as in five different views: frontal, left lateral, right lateral, left oblique and right oblique (one for each view).
3) Except supine view, additional 5 images (in five different views) are given in zoomed view to show the area of interest in a better way.
4) The database contains 1100 images from 100 subjects, including 600 original images and 500 zoomed view images altogether.
5) For each subject's breast thermogram, the patient history, outcome of X-ray mammography, clinical diagnosis, and Fine Needle Aspiration Cytology (if available) have been collected and stored for validation of the findings of IBT.
6) For all the breast thermograms with suspicious hotspot regions, the ground truth images are also annotated and given along with this database.

The contributions of this paper are:
1) This paper illustrates the designing of a standard protocol suite for the acquisition of breast thermograms so that one can utilize the maximum potentiality of IBT in early breast abnormality detection.
2) This paper provides a procedure of generating the ground truth images of the hotspot regions of abnormal breast thermograms.
3) It provides a comparison of six most widely used state-of-the-art image segmentation methods based on five supervised evaluation metrics and thus, helps to select the most effective segmentation methods for extracting the hotspot areas from breast thermograms.
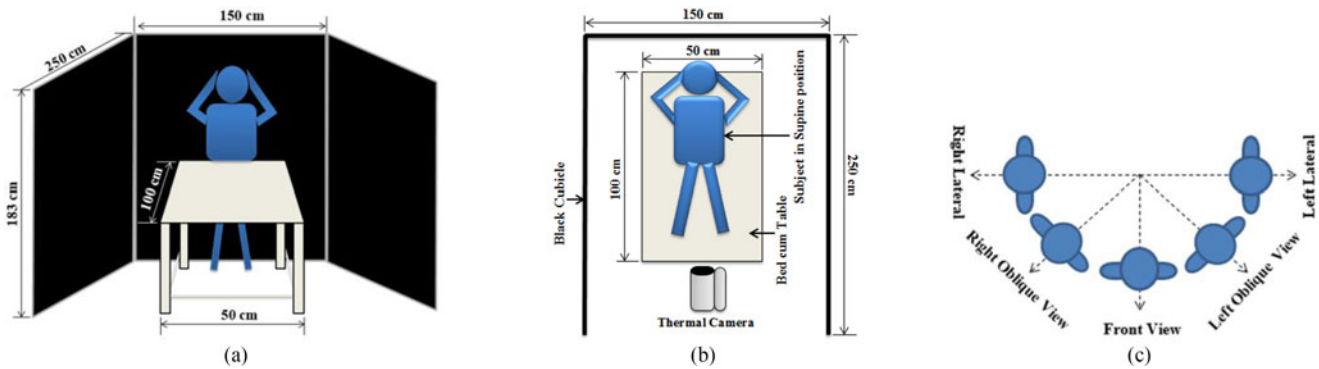
Fig. 2. (a) The Overall Acquisition setup with a patient standing inside the black cubicle, (b) the top-view of the setup with the patient in Supine position, and (c) the five top-views of the standing posture of a subject during the acquisition of breast thermograms in all views.

This paper is organized as follows. Section II elaborately describes the designing of the breast thermogram acquisition protocol suite. The validation and statistics of the developed database are provided in Section III. The validation is done by comparing the findings of the IBT with the findings of other breast screening and diagnosis methodologies. Section IV provides the nomenclature of the thermograms that is followed by the generation of ground truth images of the suspicious hotspot regions in Section V. Section VI explains the bilateral asymmetry analysis of the breast thermograms using (a) temperature features and (b) intensity value based statistical and texture features. In Section VII, six popular and widely used state-of-the-art image segmentation techniques are implemented for extraction of the hotspot regions from breast thermograms. A comparison and validation of the obtained segmented results against the generated ground truth images using five standard evaluation parameters are provided in Section VIII. Finally, Section IX concludes this paper.

II. DESIGNING OF STANDARD BREAST THERMOGRAM ACQUISITION PROTOCOL SUITE

To maintain uniformity and to collect the valuable information for medical research work, it is necessary to maintain a standard and universal acquisition protocol during the acquisition. However, due to the lack of a certain standard breast thermogram acquisition protocol, we have developed an infrared breast thermogram acquisition protocol suite by considering most of the important factors like patient preparation, examination room condition, capturing rules, etc. based on the review work made so far. Each of these factors has their individual influence on the IBT. Along with the acquisition protocol suite, an acquisition setup has also been established at Regional Cancer Centre (RCC), Agartala Government Medical College (AGMC), Govt. of Tripura, India for the collection of breast thermograms. A pictorial illustration of the acquisition setup is provided in Fig. 2(a). A detailed description of each component of the acquisition protocol suite is provided below.

A. Patient Intake Form

On the day of IBT examination, an intake form is given to each patient to fill up. The patient intake form is a kind of form in which a patient gives his/her all personal information including name, age, sex, height, weight, etc. and disease related background information like symptoms (if any), duration etc. The patient also provides his/her family history of breast cancer, previous medical tests, diagnoses, surgeries, physical therapies (if any), etc. A written consent is also taken from the patients on this intake form for using their breast thermograms for the research purposes like image analysis and printing. It takes 5 minutes to collect information from a patient. Besides, an individual 'Patient Record' is maintained for each patient for keeping all the patient related medical information.

B. Examination Room Condition

The collection of the breast thermograms is performed in a room with the controlled environmental condition. The size of the room is adequate to maintain a uniform temperature, and it is free from ventilators and windows. During the time of patient acclimation, the door of the room is kept closed to prevent the external infrared radiation or airflow inside the room. For maintaining the temperature of the room at 20 °C, an air conditioner is placed inside the room. For accurate monitoring of the ambient temperature and the humidity of the examination room, a Thermo-Hygrometer has also been utilized for the database design. In the examination room, instead of incandescent light, fluorescent lighting is used since incandescent lighting produces a high amount of infrared radiation that may interfere the process of breast thermography.

C. Patient Preparation

In the process of IBT, the patient preparation is done in two phases: Once before to the examination and the next is at the time of acquisition.

1) Pre-Exam Patient Preparation: In order to minimize the thermal artifacts generated due to improper management of patients, the subjects are asked to avoid: a) prolonged sun exposure, b) application of lotion or talcum powder on the breasts, c) performing any physical activity, d) consumption of pain medication (if any), e) wearing tight fitting clothes, f) smoking, intaking of coffee, tea, soda and g) taking of heavy meal before examination. As advised by a physician, it is also maintained

that the subject coming for breast thermography should be in her 5th–12th day and 21st day of the menstrual cycle.

2) Patient Acclimation: On the day of the examination, for acclimation of the patient, the subject is brought to a secured place inside the examination room and asked to disrobe her waist up. The patient is also asked to remove jewelry like neckpieces, chain, etc. (if any). Then she is instructed to lie down on a bed-cum- table (specially designed for our work) for 15 minutes in the supine position as shown in Fig. 2(b) and this allows the patient's body to equilibrate with the ambient temperature of the examination room. During the time of acclimation, the patient is requested to keep their both hands over their head so that the underarm area gets exposed to the air, and the contact artifact does not arise in the thermograms. For comfortability and privacy of the patient, the room is kept dark during the time of acclimation and only female persons are allowed inside the room and that is strictly maintained.

D. Breast Thermogram Acquisition Setup

The breast thermogram acquisition setup comprises of three components: the black cubicle for homogeneous background, the bed cum table and the thermal imaging camera system for breast thermogram acquisition. To reduce the impact of the radiation emitted by the background objects, a cubicle with black background has been designed for having a homogeneous background while capturing the thermograms. The second component in the acquisition setup is the bed-cum-table which is specially designed for the patient acclimation and also to capture the supine view of the patients. The third component of the acquisition setup, i.e., the thermal imaging camera system is the core of the entire breast thermogram acquisition protocol. Here, in this work, the FLIR T650sc thermal camera with thermal sensitivity of <20 mK @ 30 °C and image resolution of 640×480 pixels has been used for the acquisition of thermograms. For mounting the thermal camera, a vertically height adjustable tripod stand with a sturdy base is used so that the camera does not move or vibrate about the patient being imaged. During the acquisition, the position of the camera is kept fixed, and all the necessary movements are made by the patient.

E. Patient Positioning

During the time of acquisition, the standard positioning of the patient or the thermal camera is very critical to obtain accurate imaging. However, depending on the anatomy or body structure of the patients, maintaining a fixed camera distance for all the patients may not produce accurate thermograms in all the cases. Considering this, a distance of 1 m is maintained between the thermal camera and the patient, who has typical body structure. However, for patients whose body anatomy is larger than a typical body structure, maintaining of 1 m distance results in occluded images. Hence, for those patients, a distance in which the breast area is best fitted to the thermal camera is considered, and a measuring tape is used to record the distance of the patient from the camera.
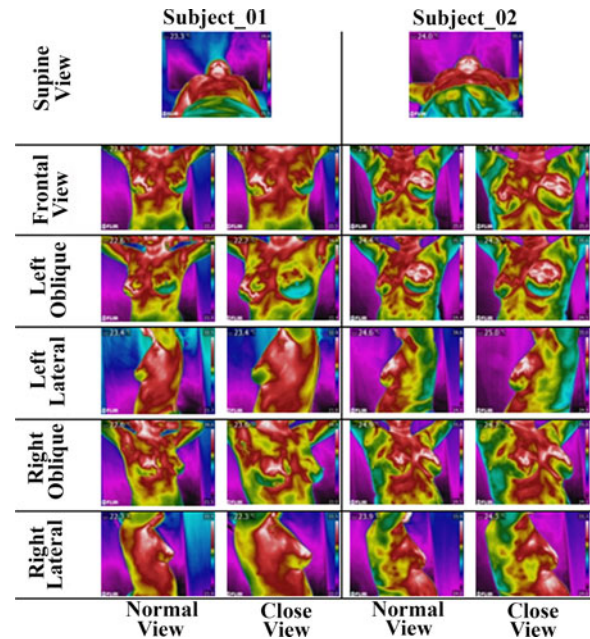

Fig. 3.   Some sample breast thermograms captured in all directions.

F. Breast Thermogram Views

The breast imaging in a limited number of views may reduce the potentiality of thermograms in abnormality detection, for which it is necessary to perform the complete imaging of each breast. Also, the formation of a breast tumor is not confined to a fixed region, for which it is necessary to take the breast images from all possible directions. Capturing of breast images from only 3 or 5 views as done in most of the research works sometimes cannot reveal the abnormality that grows underneath the breast areas. Considering this, in our work the capturing starts with the supine view of the breast, as shown in Fig. 2(b), which is followed by the capturing of frontal view, left lateral view, right lateral view, left oblique view and right oblique view. Moreover, for better visualization of the suspicious regions in breast thermograms, the zoomed or close view images of all views except the supine view are also provided in the database. The posture of a subject during the capturing of thermograms in all views except the supine view is shown in Fig. 2(c). Some sample breast thermograms acquired by using the designed protocol suite is illustrated in Fig. 3.

Using the designed protocol suite, the collection of the breast thermograms of each subject requires a duration of about 30 minutes that includes the process of collecting the patient history details also.

III. VALIDATION OF DBT-TU-JU BREAST THERMOGRAMS BASED ON OTHER MEDICAL FINDINGS AND ITS STATISTICS

Currently, the DBT-TU-JU database includes the breast thermograms of 100 females with ages between 21 and 80 years (average 44 years). To measure the efficacy of IBT in breast abnormality detection, it is crucial to validate the results of IBT with the findings of all other medical tests. While the

TABLE II
STATISTICS OF DBT-TU-JU DATABASE

Category of Thermograms	Medical Status	No. of Thermograms
Normal	Healthy	45
Abnormal	Benign Disease	36
	Malignant	13
Unknown Cases	Not Known	6

TABLE III
CODES OF EACH COMPONENT OF A NAME

Subject_ Category		Capturing_View		Imaging_Type	
IBT Type	Codes	Views	Codes	Mode	Code
Healthy	HL	Supine	S	Normal	N
Benign	BG	Frontal	F	Close	CL
Malignant	MG	Left Lateral	LL		
		Left Oblique	LQ		
		Right Lateral	RL		
		Right Oblique	RQ		

result of IBT are compared with the results of other available medical tests, it has been found that for almost all the cases, the results of all the examinations (self-examination, clinical observation, mammography, FNAC) get matched with the findings of IBT. However, for a few cases, it has been found that even though some subjects are suffering from severe breast problems, their mammography or ultrasound results come to be normal; but for those cases, the IBT shows abnormal results. It may be due to the inability of mammography to detect non-tumorous breast diseases. Moreover, for patients with dense breast, sometimes the mammography or ultrasound cannot identify any breast anomaly and thus, results in false negative rates. Due to these normal mammography reports, the patients are no more concerned about their breasts' health, and no further screening of their breasts was performed. However, in those cases, due to the high sensitivity of the IBT, the breast thermogram results get matched with FNAC report (if available), clinical observations and patient history. Moreover, among all the cases, where mammography was not able to show the presence of any abnormality, 20% of cases with abnormal thermograms are also proved to be unhealthy due to the presence of the report of FNAC. Hence, even though most of the research works have considered the mammography report as the ground truth for evaluation of the IBT based breast abnormality detection system, but due to the inefficacy and high false rate of mammography, in this work results of more than one modality are considered for labeling a breast thermogram as either 'Normal' or 'Abnormal'. The categorization of collected breast thermograms based on the findings of imaging modalities is provided in Table II.

As illustrated in Table II, among all 49 (36 benign and 13 malignant) abnormal breast thermograms, 40 thermograms have hotspots in one or both breasts and in the remaining 9 cases the thermal patterns of left and right breasts are noticeably asymmetric.

IV. NAMING CONVENTION OF BREAST THERMOGRAMS

For each person, there are 11 breast thermograms in 6 different views. Except for the supine view, for all other views, the close view thermograms are also available in the database. For better understating and convenience of the research work, the nomenclature of the breast thermograms has been done so that each image in the database gets a distinctive identity. To make the naming convention meaningful, different codes have been assigned for different capturing views, for various subjects' categories, and also for the type of each thermogram. With all

these codes, the name of a breast thermogram is like **Subject-ID _ Subject-Category _ Capturing-View _ Imaging-Type.jpg**. All the assigned codes for each component of the name are illustrated in Table III. Based on the codes provided in Table III, the image name "**010_HL_F_CL.jpg**," indicates that the thermogram is a subject with Subject_ID 10 who is found to be healthy from all medical reports and the thermogram is the close view of the frontal position.

V. GROUND TRUTH GENERATION

Generation of ground truth images for hotspot areas will allow measuring the efficiency of different segmentation algorithms in hot region segmentation. However, generation of an accurate ground truth image which contains the exact area of the hot region is one of the most tedious and challenging tasks. The most common method for generating ground truth for the medical image is to use the experts' manual segmentations [25]. However, it results in considerable spatial variations among the experts' segmentations. In such situation, consideration of a single experts' segmentation will make the system to be biased to that particular expert. The most intuitive way of solving such problems is to combine the manual segmentation results of all the experts in such a way that the resultant segmentation result is most probable to each of the expert's manual segmentation. Thus, the main intention is to maximize the joint agreement of all experts. However, the process of generating manual segmentation takes lots of time from each expert and also, these results are not exactly reproducible by the same experts. Moreover, in breast thermograms, the temperature distribution of the hotspot region does not bear a sharp boundary with other regions, and there is a smooth transition of temperature radiations to the adjacent areas. It makes the manual segmentation of the hotspot areas very tough and erroneous since discerning of the boundary between two regions is very difficult. Considering all these facts, here in this work, four popular and widely used software tools: a) Safexa segmentation tool [26], b) Photoshop tool, c) GNU Image Manipulation Program Tool [27] and d) FLIR Research IR tool are being used by four medical experts for generation of the segmented hotspot areas. While using these software, the parameters of each software tool are kept same for all the breast thermogram images. Each medical expert performs segmentation by using only one software tool. Once the software based hotspot areas are obtained, the experts
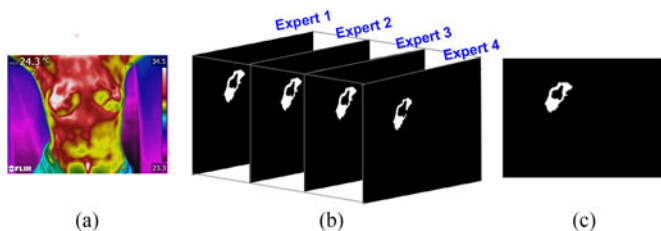
Fig. 4.   (a) The original breast thermogram, (b) four experts' reference images, and (c) the voting policy based ground truth image of the hotspot area.

further perform manual processing of the segmented images to make the segmented result more accurate. If any of the suspicious regions get discarded in software based segmentation, then these discarded areas have further been added to the reference images during manual processing. Thus, for each breast thermogram having hotspots, there are four experts' segmented reference images. After having the four segmentation images of each breast thermogram, a voting policy, the majority voting rule [25] is used to combine the result of these four segmentations to generate the ground truth image. This voting technique builds a ground truth image as maximum likelihood estimates from the set of reference images. In the method of voting policy, a threshold value T is set which depends on the number of segmented images.

Let, the number of segmented image be K, then the threshold value T is-

$$T = \left\lceil \frac{K+1}{2} \right\rceil \qquad (1)$$

and the binary labeled ground truth, GT is defined as [28]

$$GT_{ij} = \begin{cases} 1, & \text{if } \sum_{k=1}^{K} \text{Reference}_k \ (i,j) \geq T \\ 0, & \text{Otherwise.} \end{cases} \qquad (2)$$

where, Reference_k is the kth segmented image of a breast thermogram image of size $M \times N$ and $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, N$.

The generation of a ground truth image of a sample breast thermogram by using the voting policy method is shown in Fig. 4.

The ground truth generation procedure is validated by generating the ground truth images of the BRATS MRI database [29] which is already annotated with ground truth images. Since the FLIR tool is not applicable to the brain MRI images; we have used only three experts' segmentation images for ground truth generation. The similarity of the generated ground truth images is measured against the annotated ground truth images of the database by computing the DICE Similarity Coefficient (DSC) [30] score between them. A set of 25 brain MRI images with a brain tumor is randomly selected for validation of the method. By computing the DSC score between the generated and annotated ground truths, an average DSC score of 92.94 has been obtained which signifies the efficacy of the procedure for generating the ground truth images of the hot spot region of abnormal breast thermograms.

VI. Bilateral Asymmetry Analysis of Breast Thermograms

In this section, we analyze the collected breast thermograms by using some image processing methodologies to detect the presence of breast abnormality. The bilateral asymmetry analysis of the breast thermograms is an important and well-proved procedure for finding out the presence of an abnormality in breasts [4]. The bilateral asymmetry analysis is based on the hypothesis that the development of a cancer tumor or any benign tumor increases the circulations to their cells to supply nutrients by opening existing blood vessels, new blood vessels and inactive blood vessels and this increases the regional temperature of the breast in contrast to the adjacent normal areas [31]. This resulted in the significant difference between the healthy and unhealthy tissues.

Here, the bilateral asymmetry analysis of the breast thermograms has been done in two ways. First, the temperature values inherent to a breast thermogram are used for detecting the asymmetry between the two breasts of a breast thermogram. Secondly, based on the intensity values of a breast thermogram, the asymmetry analysis has been made. However, in both approaches, the nipple temperature has also been included for asymmetry analysis. The details of both ways of bilateral asymmetry analysis are provided in following subsections.

For performing the bilateral asymmetry analysis, the acquired thermograms are converted to 'Gray Palette' 24 bits image by using the FLIR software tool and then, it is converted to 8 bits grayscale image for processing. Since a breast thermogram used to contain some unnecessary regions like the background, the neck region, etc. along with the breast region, hence before performing the bilateral asymmetry analysis, extraction of the breast region is very crucial. For segmentation of the breast regions, initially, each breast thermogram and its corresponding temperature matrix are cropped to discard the unnecessary regions. Then, a semi-automatic segmentation method is used to create a binary breast mask by manually selecting the lower parabolic curves of both breasts. The binary mask is then convolved with the corresponding cropped image to obtain the intensity values of the breast region only. The same breast mask is also applied to the temperature matrix, having the extension of '.CSV' (Comma Separated Values) to obtain the temperature values of the breast region. After extracting the breast region from a breast thermogram, the left and right breasts are separated to perform the bilateral asymmetry analysis.

A. Temperature Value Based Bilateral Asymmetry Analysis

The temperature analysis may be one of the essential steps where a slight temperature difference between the left and right breast may indicate any breast pathology. Freitas stated that the temperature difference between symmetrical areas should not exceed 0.5 °C [32]. Temperature based abnormality detection is based on the premise that in an abnormal breast thermogram, the temperature distribution of one breast significantly varies from the temperature distribution of the other breast. This temperature

TABLE IV
THE AVERAGE OF MEAN, MAXIMUM AND MODAL TEMPERATURE (IN °C) OF THE NORMAL & ABNORMAL GROUP

	Avg. value of Normal Group		Avg. value of Abnormal Group		Normal Group Difference	Abnormal Group Difference	P-value
	Left Breast	Right Breast	Left Breast	Right Breast			
Mean Temperature	34.283 ± 1.0116	34.319 ± 1.1351	32.923 ± 1.3530	32.649 ± 1.2972	**0.262 ± 0.2184**	**0.501 ± 0.5059**	0.0051
Max Temperature	34.583 ± 0.7195	34.587 ± 0.8559	35.209 ± 0.9026	35.347 ± 0.9034	**0.232 ± 0.1827**	**0.573 ± 0.5076**	0.0000
Modal Temperature	32.856 ± 1.2305	32.794 ± 1.0386	32.988 ± 1.2457	32.794 ± 1.5074	**0.169 ± 0.1503**	**0.572 ± 0.5543**	0.0000

variation may be due to the fact that the surface temperature of the skin above the anomalous breast region is considerably higher than the area further away from that anomalous region [33]. However, based on the location of the tumor within the breast, the amount of heat radiation may vary.

To do temperature analysis, three temperature features: mean, mode and maximum temperature (in °C) are computed from the temperature values of both left and right breasts. Then, the difference of the temperature features of left and right breasts are computed for all breast thermograms to show the requirement of temperature analysis in breast abnormality prediction. The average and standard deviation of mean, mode and maximum temperatures of the normal and the abnormal group are listed in Table IV. From the average temperature values listed in Table IV, it has been seen that the average temperature difference of all the three features: mean, maximum and mode of abnormal group are significantly higher than the average temperature difference of the normal group. In the abnormal group, the average of mean, maximum and mode temperature differences between the left and right breasts are found to be greater than or equal to 0.5 °C. On the other hand, averages of these temperature differences are found to be less than 0.3 °C in the case of the normal group. Moreover, the statistical significance of these three temperature features is computed by using the non-parametric Mann-Whitney-Wilcoxon (MWW) test with significance level of 1% [34]. As illustrated in Table IV, it can be concluded that all three: mean, mode, and maximum temperature features are statistically significant (p < 0.01, MWW test) in signifying the presence of breast abnormality.

B. Intensity Value Based Bilateral Asymmetry Analysis

Like the temperature features, a set of statistical and texture features is also extracted from the intensity values of the breast thermograms to pinpoint the presence of abnormalities. A set of 16 statistical and texture features are extracted from both the left and right breasts. The difference between the left and right breast feature values are then computed to find the average difference that can appear in both breasts of normal and abnormal breast thermograms. The average difference of all the feature values is tabularized in Table V. As demonstrated in Table V, it can be concluded that there is a significant difference between the two breasts of an abnormal thermogram in comparison to the difference between the two breasts of a normal thermogram. Moreover, the statistical significance of these extracted features is also tested by using the non-parametric

TABLE V
AVERAGE OF FEATURE VALUE DIFFERENCE BETWEEN TWO BREASTS

Features	Average of Left & Right breast feature differences		
	Normal Group	Abnormal Group	P-value
Mean	4.535 ± 2.4067	11.969 ± 1.4936	**0.0000819**
Max	6.381 ± 5.6787	15.6 ± 13.654	0.0151
Entropy	0.046 ± 0.0323	0.1905 ± 0.1306	**0.0000011**
Skewness	0.176 ± 0.1154	0.3115 ± 0.256	0.0334
Kurtosis	0.2045 ± 0.234	0.4686 ± 0.3651	**0.0002018**
Variance	0.003 ± 0.0025	0.0083 ± 0.0056	**0.0000035**
Standard Deviation	0.008 ± 0.0076	0.0231 ± 0.0143	**0.0000038**
Correlation	0.004 ± 0.0038	0.0085 ± 0.0073	**0.0002539**
Contrast	0.018 ± 0.0130	0.0238 ± 0.018	0.1220
Energy	0.014 ± 0.0014	0.0241 ± 0.0218	0.04270
Sum entropy	0.050 ± 0.0357	0.1093 ± 0.0714	**0.0002398**
Sum Average	0.204 ± 0.1747	0.547 ± 0.4144	**0.0003552**
Sum of squares	0.722 ± 0.6045	2.2872 ± 1.7501	**0.0000117**
Sum variance	1.928 ± 1.4936	6.2879 ± 4.8756	**0.0000051**
Autocorrelation	0.719 ± 0.6073	2.2927 ± 1.7535	**0.0000126**
Inf2	0.003 ± 0.0023	0.0074 ± 4.8756	**0.0000986**

Inf2 – Information measure of Correlation2.
P-value in bold face represents Statistically Significant Features.

Mann-Whitney-Wilcoxon test [34] with significance level of 1% and the corresponding p-values are listed in Table V. Out of the initially considered 16 features, 12 features are found to be statistically significant (p < 0.01) in discriminating abnormal thermograms from the normal. However, 3 of the insignificant features: max, skewness, and energy are also found to be statistically significant against significance level of 5% in breast abnormality detection. Thus, both the temperature and intensity based bilateral asymmetry analysis can signify the presence of an abnormality.

VII. SEGMENTATION OF SUSPICIOUS HOTSPOT REGION BY USING STATE-OF-THE-ART IMAGE SEGMENTATION METHODS

This section involves the segmentation of suspicious hot areas from abnormal thermograms using some state-of-the-art and other relevant image segmentation methods. Although there are many image segmentation techniques in computer vision, not all of those work well for medical image segmentation, where low contrast, noise and image ambiguity are often the serious challenges [25], [35]. Most of the segmentation techniques are image and application dependent for which their results are highly dependent on image characteristics [35]. Moreover, due to the variation in size, shape and intensity overlapping of the hotspot areas with the surrounding normal areas, the
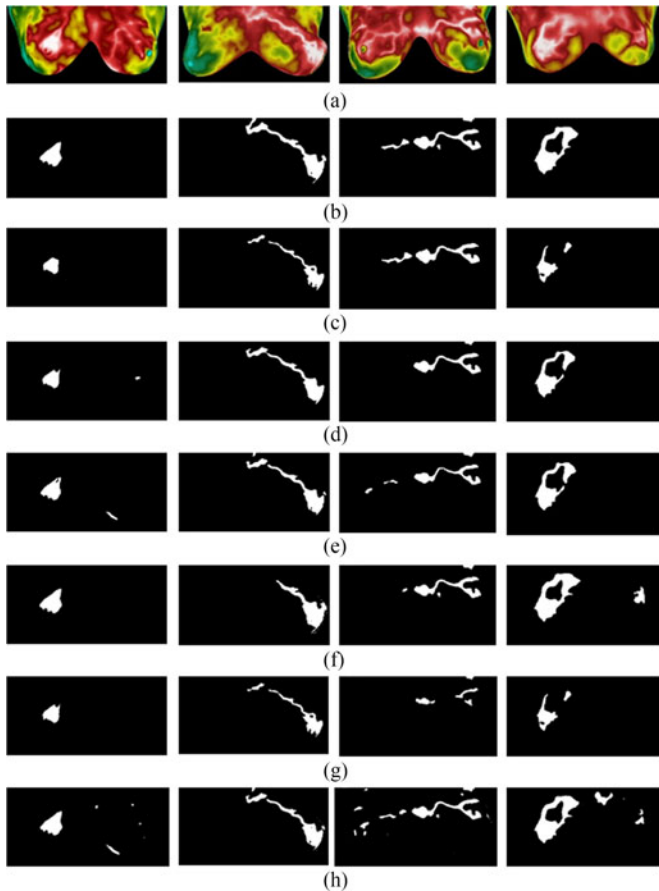
Fig. 5. (a) Infrared breast thermograms having hotspot, (b) ground truth images, (c) segmentation results of FCM, (d) segmentation results of region growing, (e) segmentation results of k-means clustering, (f) segmentation results of mean shift clustering, (g) segmentation results of thresholding, and (h) segmentation results of Fractional Order Darwinian Particle Swam Optimization.
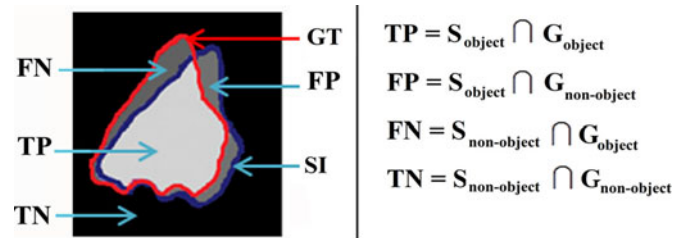

Fig. 6. The ground truth (GT) and the segmented image (SI) of the hotspot and the computation of four basic cardinalities: TP, TN, FP, FN based on the ground truth image and the segmented image.

accurate extraction of hotspot area is a very challenging task. Besides, due to the lack of clear and prominent edges, the edge-based segmentation methods are also not suitable for extraction of hotspot region. However, in literature authors had used k-means [36]–[38], fuzzy c-means [36]–[38], mean shift [37], level set [38] segmentation methods for extraction of the highest temperature regions from breast thermograms. Here, six well-known image segmentation techniques are implemented to obtain the most efficient image segmentation methods for extraction of hotspot from thermograms. These methods are Threshold based segmentation, Seeded region growing, K-means clustering (KM), Fuzzy c-means clustering (FCM), spatial information based Mean Shift Clustering (MS) [39], Particle swarm optimization method (PSO) [40]. Some samples of the segmentation result for each of the segmentation methods are illustrated in Fig. 5. A brief overview of each of these segmentation techniques is provided below.

1) Threshold based Segmentation: In a breast thermogram, the intensity value of the hotspot region is apparently different and higher than the region having normal temperature values. In almost all the abnormal thermograms, the hotspot area bears the maximum intensity value. Here, a total of 12 thresholds (T) from $T = 200$ to $T = 255$, at an interval of 5 has been used to perform the segmentation. However, for almost all the images the best result is obtained with the threshold value, $T = 200$.

2) Seeded Region Growing: Being an interactive segmentation method, it requires to select one or more seed points to start the process. The similarity criteria used here is based on the gray value difference between any neighboring pixel P and the mean intensity value of the region. Since the appearance of a hotspot is not confined to a certain area and it may appear in more than one location in the thermograms, hence in this interactive segmentation method, more than one seed points have been initialized to start the process.

3) K-means Clustering or Hard Clustering: Even though, K-means clustering is one of the simplest image segmentation method, one major drawback of the clustering based segmentation method is the selection of the optimal cluster to obtain best segmentation result. To cope with the problem of choosing an optimal number of clusters (k) for an image, we have used a cluster validity index known as I-index [41]. The value of I-index is computed for different k (k = 2 to 20) for an image and the k, against which the value of I-index is found to be the maximum is considered as the best k for segmentation of that image. Then, the segmentation result of each breast thermogram against best k is considered for doing a performance comparison.

4) Fuzzy c-means Clustering (FCM) or Soft Clustering: Unlike K-means Clustering, FCM allows a data point to belong to more than one cluster with a different degree of membership and then, the data point is assigned to a cluster to which its degree of membership is highest compared to other clusters. However, like the K-means, it also has the limitation of choosing the optimal cluster (c) for segmentation. Therefore, for FCM also, we are using the I-index for obtaining the best number of clusters (c) by computing the value of I-index [41] for different values of c (c = 2 to 20). Then, the segmentation result of each thermogram with optimal c has been considered for performance evaluation.

5) Spatial Information based Mean Shift Clustering: Mean shift is an advanced and adaptable technique for clustering based segmentation. It is a nonparametric iterative algorithm that uses a generalized kernel function. The

TABLE VI
COMPARISON OF SEGMENTATION METHODS IN BOTH DBT-TU-JU & DMR DATABASES

Database	Evaluation Metrics	Segmentation Methods					
		Region Growing	KM	FCM	Mean Shift	Thresholding (T = 200)	FODPSO (Level = 3)
DBT-TU-JU Databases	DSC	*0.774 ± 0.107*	0.469 ± 0.192	0.514 ± 0.169	**0.746 ± 0.232**	0.478 ± 0.227	**0.805 ± 0.154**
	Jaccard Index	*0.644 ± 0.146*	0.325 ± 0.163	0.359 ± 0.145	**0.635 ± 0.227**	0.342 ± 0.194	**0.695 ± 0.178**
	*Precision	*0.701 ± 0.165*	0.358 ± 0.202	0.401 ± 0.148	**0.746 ± 0.268**	0.362 ± 0.214	**0.778 ± 0.157**
	Recall	*0.919 ± 0.144*	0.857 ± 0.263	0.821 ± 0.301	**0.799 ± 0.265**	0.888 ± 0.192	**0.856 ± 0.191**
	MSE	**0.164 ± 0.070**	0.217 ± 0.076	0.207 ± 0.073	*0.162 ± 0.060*	0.207 ± 0.071	**0.129 ± 0.064**
DMR Database	Evaluation Metrics	Region Growing	KM	FCM	Mean Shift	Thresholding (T = 205)	FODPSO (Level = 2)
	DSC	**0.667 ± 0.223**	0.346 ± 0.189	0.379 ± 0.208	*0.705 ± 0.150*	0.414 ± 0.207	**0.744 ± 0.125**
	Jaccard Index	**0.537 ± 0.232**	0.224 ± 0.138	0.253 ± 0.159	*0.562 ± 0.168*	0.281 ± 0.164	**0.609 ± 0.165**
	*Precision	**0.663 ± 0.266**	0.248 ± 0.144	0.347 ± 0.168	*0.725 ± 0.269*	0.308 ± 0.189	**0.916 ± 0.137**
	Recall	**0.745 ± 0.249**	0.840 ± 0.280	0.621 ± 0.390	*0.803 ± 0.165*	0.788 ± 0.207	**0.671 ± 0.209**
	MSE	**0.139 ± 0.085**	0.190 ± 0.095	0.190 ± 0.092	**0.139 ± 0.086**	0.174 ± 0.109	*0.143 ± 0.086*

*For comparision and ranking, instead of considering the Precision and Recall value individually, their average value is considered.

performance of mean-shift clustering is very sensitive to the window size, which indicates how many data points contribute to estimate the mean. Here, to perform the mean shift clustering the bandwidth is set as 0.2 for all the images.

6) Particle Swarm Optimization (PSO): The particle swarm optimization algorithm is inspired by the behavior of the animals in an animal flock with no leader. Although it is a very simple method, it has the limitation of getting trapped in a local optimum. Two extended versions of PSO named as Darwinian Particle Swarm Optimization (DPSO) [42] and Fractional-Order Darwinian Particle Swarm Optimization (FODPSO) [43] had been used by Pedram Ghamisi *et al.* [40] for image segmentation. Here, we used the codes publically released by the authors Pedram Ghamisi *et al.* [40]. To segment the thermograms, all these three methods PSO, DPSO and FODPSO with three different levels 2, 3 and 4 have been used, while the best result is obtained with FODPSO with level 3.

VIII. EVALUATION OF THE STATE-OF-THE-ART SEGMENTATION METHODS AGAINST GROUND TRUTHS

This section evaluates the performance of the state-of-the-art and some other relevant image segmentation algorithms in hot region extraction from breast thermograms. Since none of the existing segmentation algorithms are applicable to all images, and different algorithms are not equally suitable for a particular application, the performance evaluation of segmentation algorithms is indispensable. However, a limited number of performance measures are not sufficient to describe the difference and efficiency of segmentation algorithms. Hence, in this work for evaluating the performance of different segmentation algorithms, a total of five most widely used supervised or reference image based evaluation metrics have been computed. A higher value of disparity between the reference image and the segmented image means a bigger error in the segmented image relative to the ground truth image and this, in turn, signifies the lower performance of applied segmentation algorithms.

These evaluation metrics are precision & recall [44], DICE Similarity Coefficient (DSC) [30], [45], Jaccard Index [45], and Mean Square Error (MSE). The first four evaluation metrics are commonly known as Spatial Overlap-Based Metrics as these measures are based on four basic cardinalities: TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative) where each of these four cardinalities measure the amount of overlapping or missed area between the segmentation result and the binary ground truth. The graphical representation of each of these four terms is shown in Fig. 6.

If S is the segmented image and G is the ground truth image, then $S = S_{object} \cup S_{non-object}$ and $G = G_{object} \cup G_{non-object}$, where, the object is the foreground and non-object is the background area in an image.

From a dataset of 40 segmented images, the values of these five evaluation metrics are computed for comparing the efficiency of the segmentation methods. Now to make an unbiased decision on efficient hot spot segmentation methods, all these metrics are measured against each segmentation method and tabulated in Table VI. The 3 best performance values i.e., the 1st, 2nd and 3rd best performance values of each metric are presented in *boldface*, *italics boldface* and *boldface with underline* respectively. Based on the degree of correctness of all evaluation metrics, FODPSO, region growing, and mean-shift show their potentiality in hotspot extraction. Among the remaining segmentation methods, the FCM is also found to be an effective segmentation method than the K-means and the threshold based method. Thus, we can conclude that among these six segmentation methods, FODPSO with level = 3, region growing, mean-shift and fuzzy c-means show their best performances in hot spot segmentation.

This entire procedure of ground truth generation, bilateral asymmetry analysis, segmentation of hotspot areas and evaluation of segmentation methods has also been done for the breast thermograms of the DMR database to find out the efficiency and limitations of our database in the suspicious region based breast abnormality prediction. As mentioned above, the DMR database comprises of about 47 abnormal thermograms, out of which 22 breast thermograms having hotspot regions, are ran-

domly chosen for the experimental purpose. The segmented results of each segmentation method are compared against the generated ground truth images by using all these evaluation metrics, and the result of evaluation of each segmentation method is also presented in Table VI. The experimental result shows that among all segmentation methods, the FODPSO with level = 2, mean shift and region growing provide better segmentation results since these segmentation results are closer to the ground truths. Moreover, even though FCM shows better segmentation efficacy than the k-means clustering, the segmentation result of the threshold based segmentation with T = 205 is better than both of them. However, the values of the evaluation parameters for DBT-TU-JU database is far better than the evaluation parameter values of the DMR database, which in turn illustrates that from the segmentation point of view, the breast thermograms of DBT-TU-JU database are more precise than that of the breast thermograms of the DMR database.

IX. CONCLUSION

We have presented the designing of a ground truth annotated infrared breast thermogram database by maintaining a standard acquisition protocol suite. Designing of the ground truth annotated breast thermogram database will allow the researchers to use the feature information extracted from the hotspot areas in breast abnormality detection. The bilateral asymmetry analysis of the thermograms reveals that temperature and texture feature differences of bilateral breasts of an abnormal breast thermogram are higher than the temperature and texture feature differences in a normal breast thermogram. Moreover, this paper investigates the potentiality of the state-of-the-art and some relevant image segmentation methods in hotspot extraction based on five widely used supervised evaluation metrics. The evaluation result demonstrates that with respect to all of these evaluation metrics, the FODPSO, region growing, mean-shift and fuzzy c-means clustering are more efficient than the k-means clustering and threshold based segmentation methods. Thus, the generation of ground truth image and the evaluation of existing methods for hotspot region extraction provide a new way of research in IBT based early breast abnormality detection. In future, our aim is to design a robust segmentation method for extracting the hotspot regions so that the performance of each evaluation metric gets improved i.e., the segmented result is very much close to the ground truth image. Designing of such a hotspot extraction method may also eliminate the effort of generating the ground truth image manually.

REFERENCES

[1] E. Y. K. Ng, "A review of thermography as promising non-invasive detection modality for breast tumor," *Int. J. Therm. Sci.*, vol. 48, no. 5, pp. 849–859, 2009.

[2] D. A. Kennedy, T. Lee, and D. Seely, "A comparative review of thermography as a breast cancer screening technique," *Integr. Cancer Therapies*, vol. 8, no. 1, pp. 9–16, 2009.

[3] M. Etehadtavakol, E. Y. K. Ng, V. Chandran, and H. Rabbani, "Separable and non-separable discrete wavelet transform based texture features and image classification of breast thermograms," *Infrared Phys. Technol.*, vol. 61, pp. 274–286, 2013.

[4] U. R. Gogoi, G. Majumdar, M. K. Bhowmik, A. K. Ghosh, and D. Bhattacharjee, "Breast abnormality detection through statistical feature analysis using infrared thermograms," in *Proc. IEEE Int. Symp. Adv. Comput. Commun.*, 2015, pp. 258–265.

[5] B. B. Lahiri, S. Bagavathiappan, T. Jayakumar, and J. Philip, "Medical applications of infrared thermography: A review," *Infrared Phys. Technol.*, vol. 55, no. 4, pp. 221–235, 2012.

[6] M. Etehadtavakol and E. Y. K. Ng, "Breast thermography as a potential non-contact method in the early detection of cancer: A review," *J. Mech. Med. Biol.*, vol. 13, no. 2, 2013, Art. no. 1330001.

[7] S. V. Francis and M. Sasikala, "Automatic detection of abnormal breast thermograms using asymmetry analysis of texture features," *J. Med. Eng. Technol.*, vol. 37, no. 1, pp. 17–21, 2013.

[8] L. F. Silva, D. C. M. Saade, G. O. Sequeiros-Olivera Silva, A. C. Paiva, R. S. Bravo, and A. Conci, "A new database for breast research with infrared image," *J. Med. Imaging Health Informat.*, vol. 4, no. 1, pp. 92–100, 2014.

[9] M. Frize, C. Herry, and R. Roberge, "Processing of thermal images to detect breast cancer: Comparison with previous work," in *Proc. 24th IEEE Annu. Conf. Annu. Fall Meeting Biomed. Eng. Soc. EMBS/BMES Conf.*, vol. 2, 2002, pp. 1159–1160.

[10] F. Arena, C. Barone, and T. DiCicco, "Use of digital infrared imaging in enhanced breast cancer detection and monitoring of the clinical response to treatment," in *Proc. 25th Annu. Int. Conf. Eng. Med. Biol. Soc.*, vol. 2, 2003, pp. 1129–1132.

[11] E. Y. K. Ng and E. C. Kee, "Integrative computer-aided diagnostic with breast thermogram," *J. Mech. Med. Biol.*, vol. 7, no. 1, pp. 1–10, 2007.

[12] H. Q. Yang, S. S. Xie, Q. Y. Lin, Z. Ye, S. Q. Chen, and H. Li, "A new infrared thermal imaging and its preliminary investigation of breast disease assessment," in *Proc. IEEE Int. Conf. Complex Med. Eng.*, 2007, pp. 1071–1074.

[13] V. Agostini, M. Knaflitz, and F. Molinari, "Motion artifact reduction in breast dynamic infrared imaging," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 903–906, Mar. 2009.

[14] F. G. Delgado, "Feasibility of new-generation infrared imaging screening for breast cancer in rural communities," *US Oncol. Rev.*, vol. 6, no. 1, pp. 60–64, 2010.

[15] J. Wang *et al.*, "Evaluation of the diagnostic performance of infrared imaging of the breast: A preliminary study," *Biomed. Eng. Online*, vol. 9, no. 3, pp. 1–10, 2010.

[16] F. J. González, "Non-invasive estimation of the metabolic heat production of breast tumors using digital infrared imaging," *Quant. Infrared Thermogr. J.*, vol. 8, no. 2, pp. 139–148, 2011.

[17] V. Umadevi, S. V. Raghavan, and S. Jaipurkar, "Framework for estimating tumor parameters using thermal imaging," *Indian J. Med. Res.*, vol. 134, no. 5, pp. 725–731, 2011.

[18] M. Kontos, R. Wilson, and T. Fentiman, "Digital infrared thermal imaging (DITI) of breast lesions: Sensitivity and specificity of detection of primary breast cancers," *Clin. Radiol.* vol. 66, pp. 536–539, 2011.

[19] L. A. Bezerra *et al.*, "Estimation of breast tumor thermal properties using infrared images," *Signal Process.*, vol. 93, no. 10, pp. 2851–2863, 2013.

[20] R. Omranipour *et al.*, "Comparison of the accuracy of thermography and mammography in the detection of breast cancer," *Breast Care*, vol. 11, no. 4, pp. 260–264, 2016.

[21] H. Madhu, S. T. Kakileti, K. Venkataramani, and S. Jabbireddy, "Extraction of medically interpretable features for classification of malignancy in breast thermography," in *Proc. 38th Int. Conf. Eng. Med. Biol. Soc.*, 2016, pp. 1062–1065.

[22] A. Lashkari, F. Pak, and M. Firouzmand, "Full intelligent cancer classification of thermal breast images to assist physician in clinical diagnostic applications," *J. Med. Signals Sens.*, vol. 6, no. 1, pp. 12–24, 2016.

[23] S. V. Francis, M. Sasikala, and S. D. Jaipurkar, "Detection of breast abnormality using rotational thermography," in *Application of Infrared to Biomedical Sciences*. Singapore: Springer, 2017, pp. 133–158.

[24] M. K. Bhowmik, U. R. Gogoi, K. Das, A. K. Ghosh, D. Bhattacharjee, and G. Majumdar, "Standardization of infrared breast thermogram acquisition protocols and abnormality analysis of breast thermograms," *Proc. SPIE*, vol. 9861, pp. 986115-1–986115-18, 2016.

[25] X. Li, B. Aldridge, J. Rees, and R. Fishe, "Estimating the ground truth from multiple individual segmentations with application to skin lesion segmentation," in *Proc. Int. Conf. Med. Image Understanding Anal.*, 2010, pp. 101–106.

[26] *Sefexa Image Segmentation Tool*, 2014. [Online]. Available at: http://www.fexovi.com/sefexa.html

[27] E. Oliver, J. Ruiz, S. She, and J. Wang, *The Software Architecture of the GIMP*, 2006.

[28] J. K. Kamarainen, L. Lensu, and T. Kauppi, "Combining multiple image segmentations by maximizing expert agreement," in *Proc. Int. Workshop Mach. Learn. Med. Imaging*, 2012, pp. 193–200.

[29] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[30] T. Brosch, L. Y. W. Tang, Y. Yoo, D. K. B. Li, A. Traboulsee, and R. Tam, "Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1229–1239, May 2016.

[31] U. R. Gogoi, M. K. Bhowmik, D. Bhattacharjee, A. K. Ghosh, and G. Majumdar, "A study and analysis of hybrid intelligent techniques for breast cancer detection using breast thermograms," in *Hybrid Soft Computing Approaches*. New Delhi, India: Springer, 2016, pp. 329–359.

[32] R. Freitas, *Nanomedicine Volume I: Basic Capabilities*. Washington, DC, USA: Landes Biosci., 1999.

[33] J. Kwok and J. Krzyspiak, *Thermal Imaging and Analysis for Breast Tumor Detection*, 2007.

[34] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, vol. 1, no. 6, pp. 80–83, 1945.

[35] N. Sharma and L. M. Aggarwal, "Automated medical image segmentation techniques," *J. Med. Phys.*, vol. 35, no. 1, pp. 3–14, 2010.

[36] M. EtehadTavakol, S. Sadri, and E. Y. K. Ng, "Application of K- and fuzzy c-means for color segmentation of thermal infrared breast images," *J. Med. Syst.*, vol. 34, no. 1, pp. 35–42, 2010.

[37] M. Etehadtavakol and E. Y. K. Ng, "Color segmentation of breast thermograms: A comparative study," in *Application of Infrared to Biomedical Sciences*. Singapore: Springer, 2017, pp. 69–77.

[38] N. Golestani, M. EtehadTavakol, and E. Y. K. Ng, "Level set method for segmentation of infrared breast thermograms," *EXCLI J.*, vol. 13, pp. 241–251, 2014.

[39] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[40] P. Ghamisi, M. S. Couceiro, J. A. Benediktsson, and N. M. Ferreira, "An efficient method for segmentation of images based on fractional calculus and natural selection," *Exp. Syst. Appl.*, vol. 39, no. 16, pp. 12407–12417, 2012.

[41] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, pp. 1650–1654, Dec. 2002.

[42] J. Tillett, T. M. Rao, F. Sahin, R. Rao, and S. Brockport, "Darwinian particle swarm optimization," in *Proc. 2nd Indian Int. Conf. Artif. Intell.*, 2005, pp. 1474–1487.

[43] M. S. Couceiro, N. M. F. Ferreira, and J. A. T. Machado, "Fractional order Darwinian particle swarm optimization," in *Proc. Symp. Fract. Signal Syst.*, Coimbra, Portugal, Nov. 4/5, 2011, pp. 127–136.

[44] J. Xu *et al.*, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med. Imaging*, vol. 35, no. 1, pp. 119–130, Jan. 2016.

[45] J. Pont-Tuset and F. Marques, "Supervised evaluation of image segmentation and object proposal technique," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1465–1478, Jul. 2016.