International Journal of Medical Engineering and Informatics

ISSN online: 1755-0661 - ISSN print: 1755-0653
https://www.inderscience.com/ijmei

Harnessing the power of machine learning for breast anomaly prediction using thermograms

Aayesha Hakim, R.N. Awale

DOI: 10.1504/IJMEI.2021.10040645

Article History:

Received:	12 August 2020
Last revised:	13 January 2021
Accepted:	15 January 2021
Published online:	30 November 2022

Harnessing the power of machine learning for breast anomaly prediction using thermograms

Aayesha Hakim* and R.N. Awale

Veermata Jijabai Technological Institute,
H.R. Mahajani Marg, Matunga, Mumbai, Maharashtra 400019, India
Email: aayesha.hakim@gmail.com
Email: rnawale@el.vjti.ac.in
*Corresponding author

Abstract: Breast cancer is the most fatal cancer among women globally. Thermography provides an early sign of a developing abnormality based on the temperature changes in breasts. In this work, statistical features extracted from the segmented breast region are used for breast cancer prognosis. Machine learning algorithms like support vector machine (SVM), k-nearest neighbourhood (kNN), naïve Bayes and logistic regression without and with principal component analysis (PCA) as a pre-cursor are applied to the extracted data to classify thermograms as malignant or benign. Classification was also performed using tree-based classifiers, namely, decision tree and random forest. This work indicates that thermal imaging is capable of predicting breast pathologies coupled with machine learning algorithms. The PCA-SVM model has the best accuracy, sensitivity, specificity and AUROC of 92.74%, 77.77%, 95.83% and 0.8699 respectively. Among tree-based classifiers, random forest classifier has the best accuracy, sensitivity, specificity and AUROC of 94.4%, 97.5%, 78.72% and 0.97961 respectively with five-fold cross validation. Our study produced competitive results when compared to other studies in the literature.

Keywords: breast cancer; breast thermography; infrared imaging; thermal imaging; machine learning.

Reference to this paper should be made as follows: Hakim, A. and Awale, R.N. (2023) 'Harnessing the power of machine learning for breast anomaly prediction using thermograms', *Int. J. Medical Engineering and Informatics*, Vol. 15, No. 1, pp.1–22.

Biographical notes: Aayesha Hakim is a research scholar at Veermata Jijabai Technological Institute. Her areas of interest are biomedical image processing, machine learning, pattern recognition and its applications.

R.N. Awale is a Professor at Veermata Jijabai Technological Institute. His active research areas are cross layer design for QoS for ad hoc wireless networks, image processing, antenna design, and data security.

1 Introduction

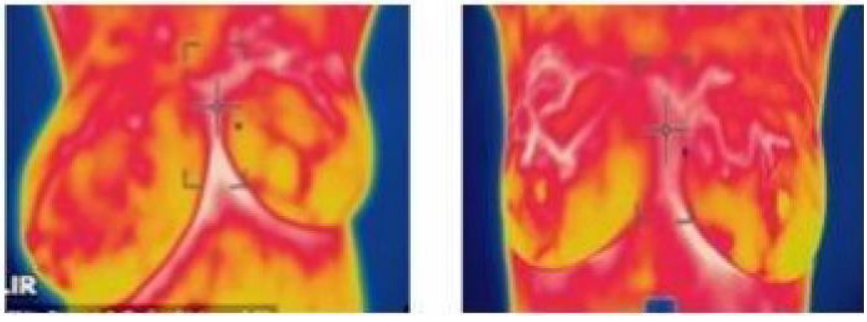
According to the World Health Organization (WHO), one of the major reasons of death among women globally is breast cancer. It impacts 2.1 million women per year worldwide (Breast Cancer India, 2020). Approximately 60% deaths are due to delay in diagnosis. For better survival of patients and lesser use of the treatments, many imaging

systems are continually being developed to diagnose this disease as early as possible. Although, the gold standard for imaging breasts is mammography, its performance is poor in younger women with dense breasts (Keyserlingk et al., 1998). Patients who repeatedly undergo mammography, for evaluation of suspected lesions are exposed to harmful X-ray radiations (Khandpur, 1994). Being a structural imaging modality, ultrasound (USG) (Prasad and Houserkova, 2007) is used to find the size, shape, texture and density of a breast lump. Its diagnostic performance is poor in fatty breasts due to poor penetration of sound waves. Magnetic resonance imaging (MRI) (Prasad and Houserkova, 2007) machine uses a large magnet and radio waves to create images of detailed cross sections of the breast. MRI is often used in conjunction with mammography. Ng and Sudharsan (2001) suggested that human skin temperature pattern is symmetric bilaterally. Neo-angiogenesis (Ng et al., 2001) is the formation of new blood vessels that develop to feed cancerous tumours. Proliferating tissues generate more infrared radiation (Jones, 1998) which leads to high vascularity and production of heat which gets transferred to the skin surface and is detected by thermal imaging (Bronzino, 2006). Cancer increases the breast temperature by 1°C–2°C which leads to asymmetrical patterns in the thermogram (Fok et al., 2002). Thermography is efficient in detecting non-palpable breast cancer early for women between 30–50 years with dense breasts. It is a private, painless, contactless process and does not expose the patient to any radiation hazard. This opens potential for thermography to be used as a safe risk marker for routine examination of breasts.

1.1 Related works

Clinical interpretation of breast thermograms is based on colour analysis (Shahari and Wakankar, 2015). Healthy breasts indicate low heat level and appear purple on a thermograph. Warmer areas emit more heat than cooler ones. Spots appearing red, orange, or yellow in a breast thermograph indicate the presence of abnormality as shown in Figure 1. US Food and Drugs Administration (FDA) (2011) in 1982 approved the usage of thermal imaging with mammography for the screening of breasts.

Figure 1 Thermograms of (a) benign case and (b) malignant case (see online version for colours)



(a) (b)

Source: Visual Lab (2014)

Head et al. (2000) used infrared index as a metric to quantify breast abnormalities for infrared images of 220 patients. They analysed the possibility of a link between family

history or hormone therapy with the results obtained and the study showed no correlation. Qi et al. (2000) and Scales et al. (2004) used Canny edge detector and Hough transform (Borchartt et al., 2013) to identify breast contours. Edge detection gave false results in case of sagging breasts, i.e., flat lower part. Hough transform is a slow operation and takes more than 96% of processing time. Studies conducted in Acharya et al. (2012, 2010), Madhu et al. (2016) and Wakankar and Suresh (2016) used an support vector machine (SVM) classifier by feeding it the extracted statistical and texture features from thermograms. The accuracy of classification was 88.1% in Acharya et al. (2012). The obtained sensitivity and specificity in Madhu et al. (2016) were 90% and 94.3% as compared to 85.71% and 90.48% from Acharya et al. (2010). A similar work (Borchartt et al., 2011) reported the use of a free LibSVM classifier. Kuruganti and Qi (2002) used the extracted features from thermograms to measure the asymmetry between left and right breasts using K-means clustering and k-nearest neighbourhood (kNN) methods. However, the dataset used was too small to confirm their results. A case study (Kirubha et al., 2018) was conducted on two patients – one, who had undergone screening mammography and the other had cancer (proven by biopsy) respectively. Segmented tumour region extracted from mammogram and thermogram was compared for the diagnosis of breast cancer. SVM with radial basis function (RBF) kernel was used in Gogoi et al. (2019) and Sathish et al. (2017) to classify breast thermograms as healthy, benign and malignant and hotspots were categorised quadrant wise in breasts. Accuracy reported was 90%. In Silva et al. (2014), SVM and ANN were used to segregate the thermal images into three classes – normal, benign and malignant. Their results were promising as many studies reported in literature classified thermograms into two classes only. Work conducted in Francis et al. (2014) used principal component analysis (PCA) to reduce the data dimensions and fed the transformed feature set to an SVM reporting a sensitivity of 83.3% which indicated that rotational thermography can be potentially used for screening breast cancer. Khan and Arora (2018) and Gaber et al. (2013) used Gabor filter to extract the texture features from the left and right breasts. SVM was used to classify images and an accuracy of 84.5% and 92.06% respectively was obtained. Sh et al. (2016) used different training-testing data partition, with an SVM classifier for thermograms and proved 80%–20% gave the best accuracy of 99.51%.

This study investigates the use of thermal imaging in decision-making for the diagnosis of breast cancer coupled with machine learning algorithms. Rest of the paper is organised as follows – Section 2 describes the methods and materials used to conduct the investigation and various machine learning algorithms used. Classification results and discussions are presented in Section 3. Section 4 concludes the paper with some discussion on scope for future work.

2 Methods and materials

The objective of this study is to determine with a high degree of certainty if a tumour is malignant or benign using thermal images and machine learning algorithms. We used a set of 287 images from the Database of Mastology Research (DMR) (Visual Lab, 2014) which were captured by FLIR SC-620 IR camera at the Hospital of the Federal University of Pernambuco, Brazil (Hakim and Awale, 2020). The demographic data of subjects is mentioned in Table 1.

Table 1 Demographic data of subjects in DMR

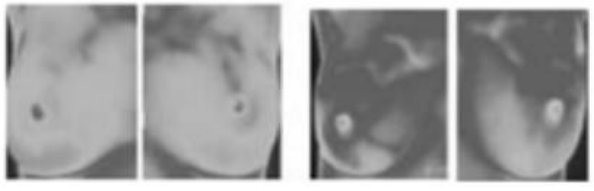
Age range	*No. of malignant cases (pathology)*	*No. of benign cases (healthy)*
29–50	26	134
51–70	18	79
71–85	03	27
Total	47	240

Source: Visual Lab (2014)

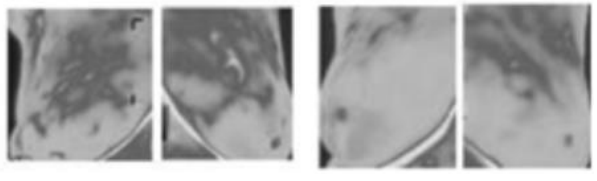
2.1 Region of interest extraction

The RGB images acquired from DMR are converted into greyscale images. We segment only the breast, i.e., region of interest (ROI) from each image in order to eliminate errors from other unnecessary warm areas like the background, arm-pits, head, neck portion, area underneath the breast (Qi and Head, 2001). Breast region is extracted manually by generating a unique breast mask (Gonçalves et al., 2019) for each thermogram from the Canny edge detected image. We fixed the size of all the segmented breast images as 256 × 256 since the original image size is large. The individual breast masks are then multiplied with their corresponding greyscale images obtained after removal of the irrelevant regions. Thus, we obtain only the breast region from each breast thermogram as shown in Figure 2.

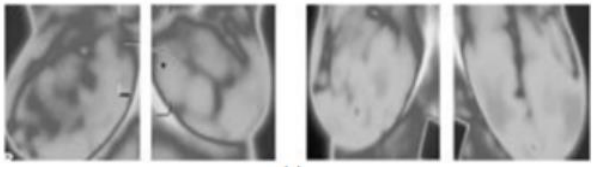
Figure 2 Segmentation of (a) small breast sizes, (b) big breast sizes and (c) asymmetric breast sizes from DMR



(a)



(b)



(c)

2.2 Feature extraction

Statistical features (Hakim and Awale, 2020; NIST, 2017; Schaefer et al., 2009) are extracted from the segmented breast region and used them for classification. Before building machine learning models, we have scaled the features to ensure that no independent variable is dominating other variables in the model. A .csv (comma separated values) file is prepared with all the feature values and the target variable, i.e., malignant (1) or benign (0).

1 First order statistics
   - *Mean:* Mean is defined as the average colour in the image.
   - *Variance:* Variance measures the deviation of grey levels from the mean value. Standard deviation is the square root of variance.

2 Second order statistics
   - *Skewness:* Skewness is a measure of asymmetry of a pixel distribution around the mean value.
   - *Kurtosis:* Kurtosis is the fourth moment and characterises the peakiness of the distribution in comparison to a normal distribution.

3 Texture features (Hakim and Awale, 2020; Haralick et al., 1973)

   Texture features help in understanding the way the intensity varies within an image and follows a pattern. The texture information is obtained in pixel domain using grey-level co-occurrence matrix (GLCM). We averaged every value obtained from the four GLC matrices corresponding to four directions ($\theta = 0°$, 45°, 90°, and 135°) keeping $d = 1$ pixel.
   - *Contrast:* Contrast is a measure of grey level variations between a pair of pixels.
   - *Correlation:* Correlation presents the linear dependency of grey level values in GLCM.
   - *Energy:* Energy measures the local uniformities of grey levels. Images with similar pixels have large energy values.
   - *Homogeneity:* Homogeneity gives the distribution of elements with respect to the diagonal of GLCM.
   - *Entropy:* Entropy is a measurement of randomness present in an image and represents the degree of disorder present in an image.

2.3 Classification using machine learning models

Classification is a powerful strategy used to categorise datasets, so that the analysis can be utilised in prognosis to make faster decisions. The extracted features from thermograms are fed into the classification algorithms to analyse the breast thermal images. Each image has nine biostatistical features. So, we have a vector of 9×287 features for the classification purpose. We apply PCA and then fit machine learning algorithms like SVM, kNN, naïve Bayes and logistic regression on the reduced dataset to classify thermograms. Tree-based classifiers like decision tree (DT) and random forest does not need feature scaling, as they are if else nested loop and cause branching of

data. For the proposed work, we used open-source tool Python and Jupyter notebook (scikit-learn.org, 2019) for machine learning and data visualisation. We divided the dataset into training (80%) and test set (20%) for all algorithms.

2.3.1 Principal component analysis

Some of the features in the datasets are more selective and decisive than other features which are redundant. Two highly correlated variables bias the output and make samples of both classes look the same. PCA reduces the overwhelming number of dimensions by constructing principal components (PCs) based on maximum variance along the axis (Lashkari et al., 2016). PCs are calculated only from the knowledge of features and not classes. Hence, PCA is an unsupervised method. PCA provides efficient visualisation of our nine-dimensional feature set and speeds up the machine learning algorithms. We use it as a pre-processing step for supervised learning tasks and fit a classifier on the PCA-transformed data. The eigenvectors give the proportion of dataset's explained variance that lies along the axis of each PC. All the nine components capture 100% variance in data. Eigenvectors and eigenvalues are ordered in descending order where eigenvector with the highest eigenvalue is the most significant (i.e., the first PC). As seen from Table 2, the first two PCs contribute to 86.46% of the total variance and can be chosen for classification. At least 80% of original dataset's information should be retained (Lashkari et al., 2016) and rest are dropped as they represent very less information.

Table 2 Eigenvalues and explained variance for various PCs

Principal component #	*Eigenvalue*	*Proportion of variance*
1	4.0	0.6295
2	2.0	0.2351
3	1.1	0.0721
4	0.8	0.0238
5	0.8	0.0231
6	0.7	0.010
7	0.6	0.0045
8	0.5	0.0010
9	0.4	0.0009

A scree plot helps to identify the number of components needed to summarise the data and captures the variation contributed by each PC. Here, first two PCs are sufficient to describe the essence of the data after which the scree plot bends quickly to flatten out as seen in Figure 3.

A loading plot helps to identify how strongly each variable influences a PC. When two vectors are close, with a small angle between them, the two variables they represent are positively correlated. If vectors form 90°, they are not likely to be correlated. When vectors diverge away from each other with 180° angle, they are negatively correlated. The loading plot for our data shown in Figure 4 shows that variables – variance, kurtosis, contrast, correlation and homogeneity have large positive loadings (values close to 1) on the first PC. Skewness and energy have large negative loadings (values close to –1) on second component.

Figure 3 Scree plot for PCs and explained variance

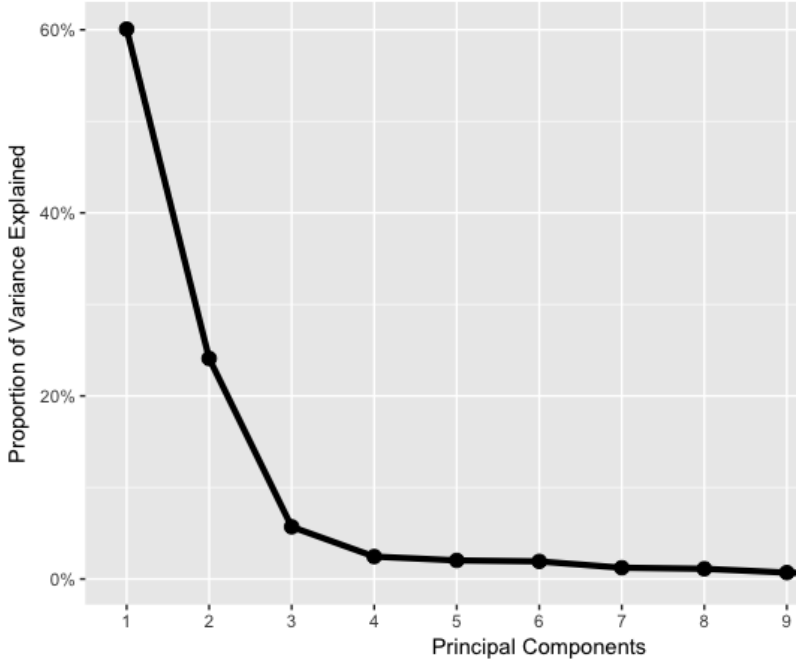
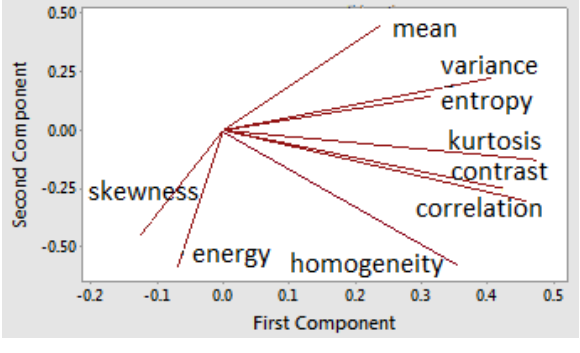


Figure 4 Loading plot for first two PCs (see online version for colours)

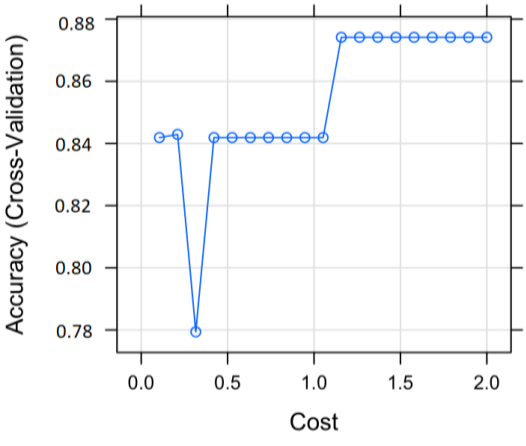


2.3.2 SVM with grid search

SVM is a supervised learning method that implements classification by constructing a hyper plane as a decision boundary (Nunes et al., 2010; Rejani and Selvi, 2009) which maximises the distance (margin) between the classes and lowers the generalisation error of the classifier. We build a two-class SVM classifier over a feature vector obtained from the features extracted from thermograms and the class of the samples. We used a linear kernel (mapping function) to make a decision boundary. C (cost) is a regularisation parameter that imposes a penalty to the model for misclassifications. Higher the penalty, it is less possible to misclassify a sample. But a very large C leads to overfitting. We

performed a ten-fold cross validation grid search and computed model accuracy for different values of C. We select the optimal C as 1.16 as it maximised the training model's ten-fold cross-validation accuracy (87.70%) and gave the kappa value (k) as 0.754 as compared to kappa of 0.68 for C = 0.4 (84% accuracy). The plot of accuracy vs. cost for training model is shown in Figure 5.

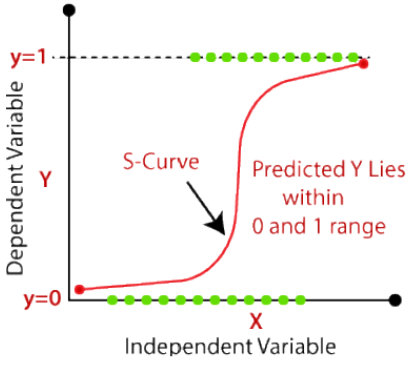
Figure 5 Plot of SVM training model's accuracy vs. cost (see online version for colours)



2.3.3 Logistic regression

Logistic regression is a supervised learning technique where labelled data is provided for the classifier to make decisions rationally for the new data. A decision boundary is converted to probabilities (Pearl, 1998) and a threshold of 0.5 is set in a sigmoid curve as shown in Figure 6. The probability greater than 0.5 is considered as 'malignant (1)' and lower than 0.5 is considered as 'benign (0)'. We performed 50 iterations of training with tolerance of 0.0001 after which convergence is reached and training stops.

Figure 6 Sigmoid curve decision boundary for logistic regression (see online version for colours)



2.3.4 k-nearest neighbours

kNN is a classification technique which assigns an unknown sample to the class to which the majority of its 'k' nearest neighbours belong to (Mejia et al., 2015). Euclidean distance metric is computed between test data point and all labelled datapoints. The top 'k' labelled datapoints are selected from the ascending order of distance and their class labels are looked at. The test data is assigned to the class label that majority of 'k' labelled datapoints belong to, as seen in Figure 7.

$$d(x_i, x_i) = \sqrt{(x_{i1} - x_{i1})^2 + (x_{i2} - x_{i2})^2 + \ldots + (x_{ip} - x_{ip})^2}$$

$$\text{A Ri} = \left\{ \text{x} \in R_p : d(x, x_i) \le d(x, xm), \forall \text{i} \ne \text{m} \right\}$$

Figure 7 Illustration of the kNN classifier (see online version for colours)

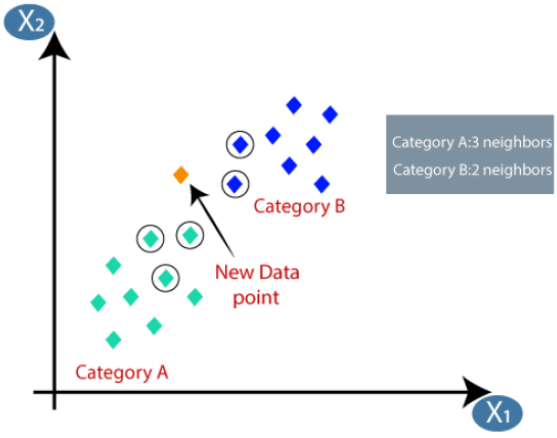
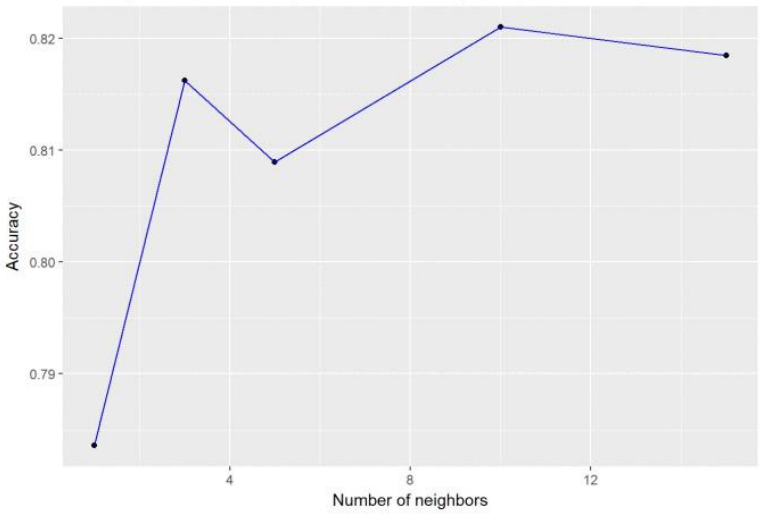


Figure 8 Plot of accuracy vs. number of neighbours (see online version for colours)



The misclassification error for each 'k' value between 1 and 15 was calculated. The number of misclassified samples was least for 'k' = 10 and the best performance with accuracy of 82.2% were achieved as seen in Figure 8.

2.3.5 Naive Bayes classifier

Naïve Bayes classifier (Zhang, 2004) is a probabilistic classifier based on Bayes' theorem. It assumes that features are conditionally independent. For binary class dataset, a sample is classified in the class which has higher probability (Pearl, 1998). To decide between two class labels L1 and L2, the ratio of the posterior probabilities for each label is computed. With Gaussian distribution model in place for each class, we computed the likelihood probability P(features | L1) and P(features | L2) for each data point. The posterior ratio is finally computed, thus determining which class label is the most probable for a given point (Mandal, 2017).

$$\frac{P(L_1 \mid \text{features})}{P(L_2 \mid \text{features})} = \frac{P(\text{features} \mid L_1)}{P(\text{features} \mid L_2)} \frac{P(L_1)}{P(L_2)}$$

2.3.6 DT classifier

A top-down tree is constructed using features and a decision is made on the first good search (Lior and Oded, 2005). Data is recursively split into branches to build a tree for breast cancer prediction by using Gini index to find a variable and corresponding threshold for splitting the input data into two or more subgroups. Gini index reduces for every passing node in the tree. To predict a response, one needs to follow the decisions/rules in the tree from the root (beginning) node down to a leaf node. Leaf nodes are classes, 0 (benign) or 1 (malignant) as shown in Figure 9. In our work, maximum number of splits allowed was 100, minimum leaf samples were 3. To protect against overfitting and biasing of tree, five-fold cross-validation was used. The partition stops when:

a all samples for a given node belong to same class

b no attributes remaining for partitioning

c no samples are left (Mandal, 2017).

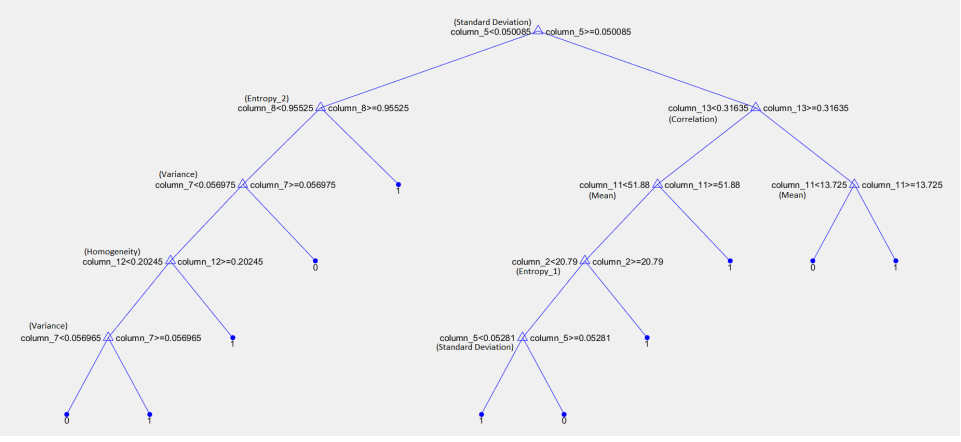
$$GI = 1 - \sum_{i=1}^{n} (p)^2$$

$$GI = 1 - \left[\left(P_{(+)} \right)^2 + \left(P_{(-)} \right)^2 \right]$$

$$H(s) = -P_{(+)} \log_2 P_{(+)} - P_{(-)} \log_2 P_{(-)}$$

$$\text{Here } P_{(+)} / P_{(-)} = \%of + ve \text{ class} / \%of - ve \text{ class}$$

Figure 9 DT model for features extracted from the DMR dataset (see online version for colours)



2.3.7 Random forest classifier

A risk of overfitting is involved in the DT classifier. The random forest classifier has the same basic structure as a DT and uses Gini index to split the tree; however, it selects a random subset of variables at each node of each tree instead of splitting a tree node using all variables, and only those variables selected are used as candidates to determine the best split for the node. Thus, it builds multiple trees (forest) from the same dataset and selects the best tree to classify the test samples (Lior, 2010; Breiman, 2001). It is an ensemble classification algorithm that reduces the risk of overfitting (following a full-grown tree like in DT) and the required training time. In our work, we have 12 features and number of learners/classifiers (DTs) are 30. We fixed the maximum amount of splits/depth of DT to 20 and learning rate to 0.1. The square root of total number of features is used as the maximum number of features in each DT. All classifiers are bagged to classify the test data based on voting method.

2.4 Performance evaluation parameters

Comparison of the performance of a classifier with other classifiers to categorise thermograms is done using some evaluation metrics (Koprowski, 2014) like accuracy, sensitivity, specificity, PPV and NPV. They are calculated using the data in the confusion matrix. The diagonal grid values in the confusion matrix show the number of cases that are correctly classified and the off-diagonal values show the falsely classified cases. Accuracy gives the percentage of correct classification. However, it is not enough alone to reveal how well the model predicted 'benign' and 'malignant' cases independently. Sensitivity is the ability of a classifier to detect malignancy while specificity is the ability to detect benign cases. PPV reflects the malignant possibility of positive result while NPV reflects the benign possibility of negative result. F1-score is the harmonic mean of precision and recall. It ranges between 0 (worst value) and 1 (best value).

$$\text{Precision} = \frac{True\ positive}{True\ positive + False\ positive}$$

$$\text{Recall} = \frac{True\ positive}{True\ positive + False\ negative}$$

$$\text{F1} = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

3 Classification results and analysis

In our study as seen in Table 3, the highest classification accuracy of 92.74% is achieved by using two PCs with the SVM classifier followed by an accuracy of 92.5% with PCA-logistic regression. SVM and logistic regression classifiers show remarkably high accuracy in separating malignant cases from benign when the axes are rotated with PCA. Due to imbalance in data, the ratio between positive and negative support vectors becomes more imbalanced; therefore, samples at the boundary of hyperplane are more likely to be classified as negative and favour predictions of the majority class (benign) on the test samples. Using PCA as a pre-cursor to kNN, we apply kNN to the matrix corresponding to extracted PCs. KNN on its own is slow at classification time. PCA-KNN thus saves computation complexity and time. Naive Bayes classifier is good at predicting negative class tuples, but it is worst amongst all classifiers at predicting positive class tuples. Naive Bayes (NB) is fastest in terms of computation speed. Feature reduction with PCA causes features to become uncorrelated which satisfy the basic independence assumption of naïve Bayes and as a result NB performs much better and has robust results with PCA as a pre-cursor step. Classifiers that are based on comparing the pairwise distances of samples like KNN are hardly affected when the axes are rotated using PCA because the pairwise Euclidean distances remain exactly the same. It is identified that AUC score is significant to consider for correct prediction of breast cancer instead of only considering accuracy. Comparison of our results with past results of literature is shown in Table 4. These results point out that accuracy and specificity of our work is appreciably higher as compared to other works, with a smaller number of misclassifications. Borchartt et al. (2013) results have high sensitivity but very low specificity value is observed due to an unbalanced and small sample set used.

Table 3 Performance evaluation parameters of various classifiers with and without PCA as pre-cursor

Method	*SVM*		*Logistic regression*		*KNN*		*Naive Bayes*	
Parameters	*Without PCA*	*With PCA*	*Without PCA*	*With PCA*	*Without PCA*	*With PCA*	*Without PCA*	*With PCA*
Accuracy %	88.22	92.74	84.35	92.5	82.2	84.11	81.69	88
Sensitivity %	77.77	77.77	44.44	66.67	55.55	66.67	22.22	44.44
Specificity %	89.58	95.83	91.67	97.91	87.5	87.5	87.5	95.83
PPV %	58.33	77.77	50	85.71	45.45	50	33.33	66.66
NPV %	95.55	95.83	89.79	94	91.3	93.33	86.27	90.19
F1 score	0.6667	0.7777	0.4706	0.75	0.5	0.5714	0.2667	0.5333
AUC	0.8368	0.8699	0.6806	0.8229	0.7153	0.7708	0.5486	0.7013
Processing time (sec)	1.5	0.75	0.9	0.48	1.35	0.37	0.36	0.17

Table 4 Comparison between the results obtained in our study and in past literature involving statistical features

Author(s)/ method	*Sensitivity (%)*	*Specificity (%)*	*Accuracy (%)*	*Area under ROC curve (AUC)*	*No of images used*
Our results (PCA-SVM)	77.77	95.83	92.74	0.8699	240 benign, 47 malignant
Localized temperature increase (LTI) (Tang et al., 2008)	93.6	55.7	-	-	70 benign, 47 malignant
Fuzzy classifier and statistical features (Schaefer et al., 2009)	79.86	79.49	79.53	-	29 malignant, 117 benign
Statistical temperature features and SVM classifier (Borchartt et al., 2011)	95.83	25	85.71	0.604	24 unhealthy, 4 healthy
Texture features and SVM classifier (Acharya et al., 2012)	85.71	90.48	88.1	-	50
Gabor coefficients statistical and SVM RBF (Gaber et al., 2013)	-	-	92.06	-	29 healthy, 34 malignant
BEMD and URLBP (Madhavi and Bobby, 2017)	92	73	86	0.82	43 normal, 24 abnormal
Statistical texture features / SVM RBF (Sathish et al., 2017)	87.5	92.5	90	-	40 normal, 40 abnormal
Gabor filter and SVM (Khan and Arora, 2018)	90.52	82.47	84.5	-	35 normal, 35 abnormal

Figure 10 Accuracy and F1 score plots for various machine learning algorithms (see online version for colours)

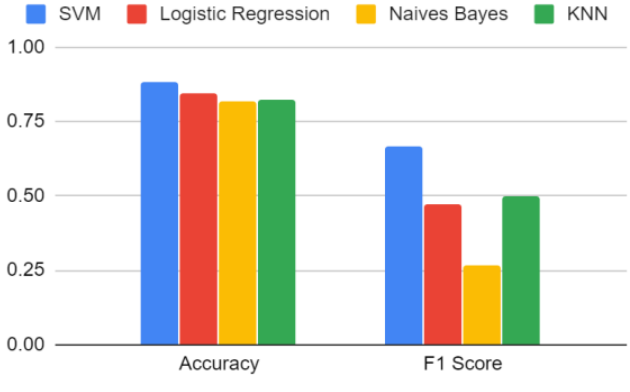


Figure 11 Accuracy and F1 score plots for machine learning algorithms with PCA as pre-cursor (see online version for colours)

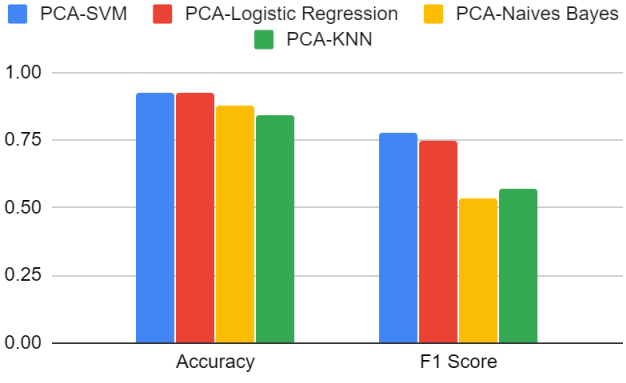
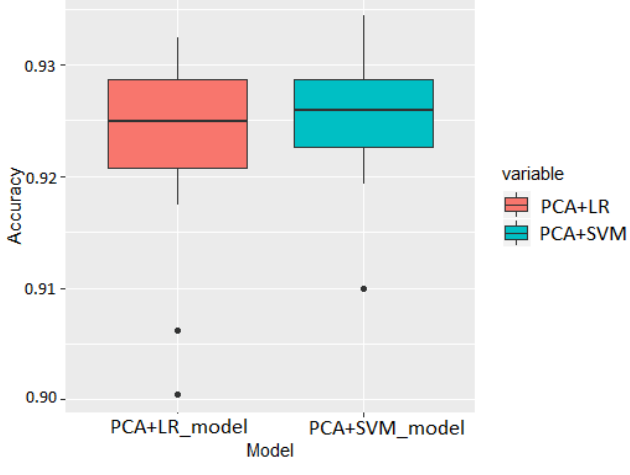


Figure 12 Box plot comparison of accuracies of two best models (see online version for colours)



The accuracy and F1 score of all machine learning algorithms are plotted with and without PCA as a precursor as shown in Figure 10 and Figure 11. The average accuracy of the two best models is about the same (PCA-SVM model has a slightly better average accuracy than PCA-LR), i.e., 92.74% vs. 92.50%. As seen from Figure 12, standard deviation is in favour of the PCA-SVM model, i.e., 3.2% vs. 2.3%.

Higher diagonal values of the confusion matrix imply many correct predictions of both the classes. Both PCA-SVM and PCA-LR misclassify only four samples. However, in case of cancer diagnosis, FP is more acceptable than FN. Thus, PCA-SVM performs better in terms of this aspect than PCA-LR. The confusion matrices for various machine learning algorithms with and without PCA as a pre-cursor are shown in Table 5 and Table 6 respectively.

Table 5 Confusion matrices of various classifiers

Algorithm used	*Predicted class*		
	Actual class	*Malignant*	*Benign*
Naives Bayes	Malignant	2	7
	Benign	4	44
Logistic Regression	Malignant	3	6
	Benign	3	45
KNN	Malignant	5	4
	Benign	6	42
SVM	Malignant	7	2
	Benign	5	43

Table 6 Confusion matrices of various classifiers with PCA as a pre-cursor

Algorithm used	*Predicted class*		
	Actual class	*Malignant*	*Benign*
PCA-Naives Bayes	Malignant	4	5
	Benign	2	46
PCA-logistic regression	Malignant	6	3
	Benign	1	47
PCA-KNN	Malignant	6	3
	Benign	6	42
PCA-SVM	Malignant	7	2
	Benign	2	46

Figure 13 Receiver operating characteristic (ROC) graph for various classifiers with PCA as a pre-cursor (see online version for colours)

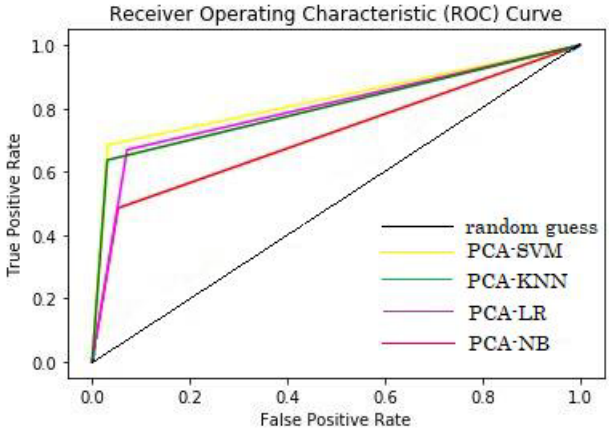


Figure 14 ROC Curve for class 0 (benign) and class 1 (malignant) respectively for DT classifier (see online version for colours)

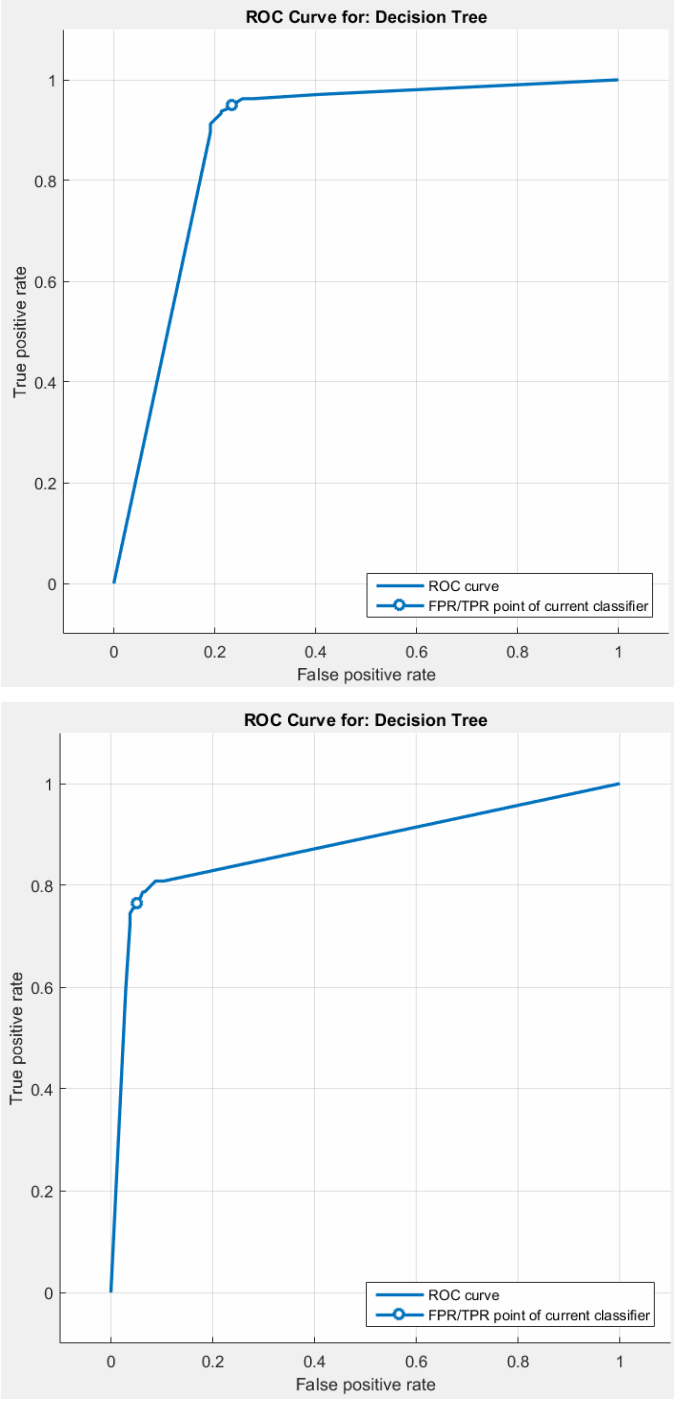


Figure 15 ROC curve for class 0 (benign) and class 1 (malignant) respectively for random forest classifier (see online version for colours)

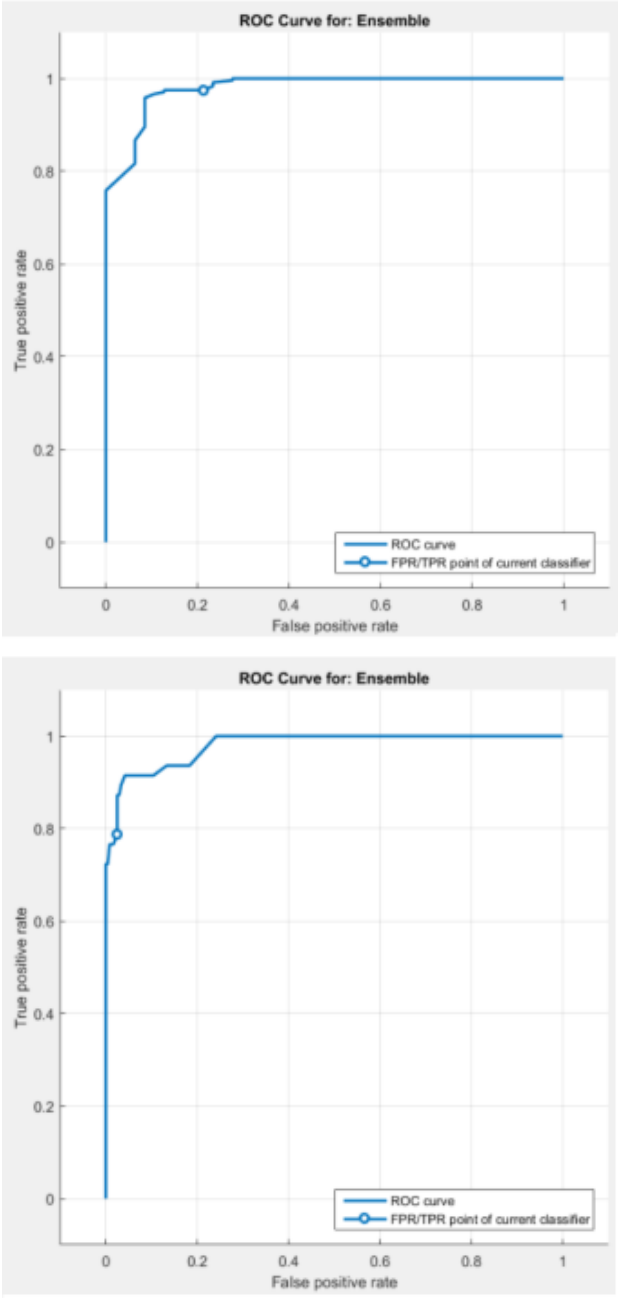


Figure 16 Overall confusion matrix for DT classifier (see online version for colours)

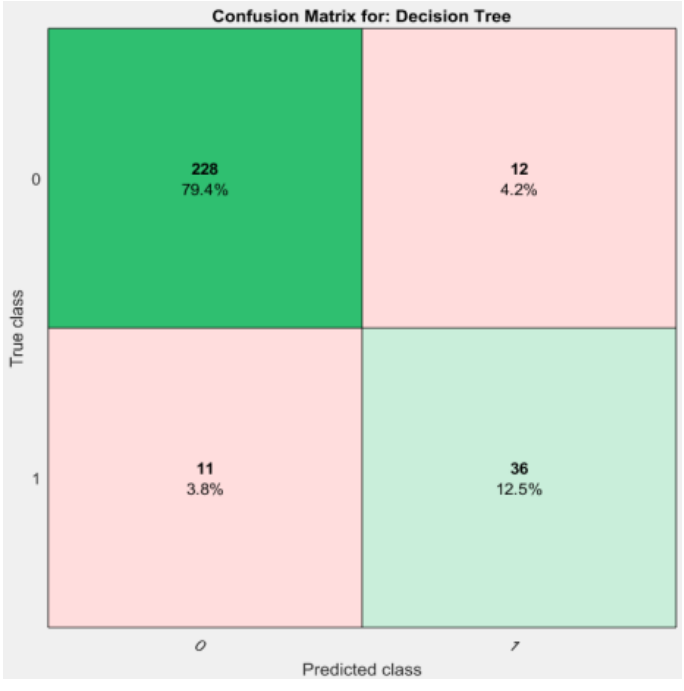
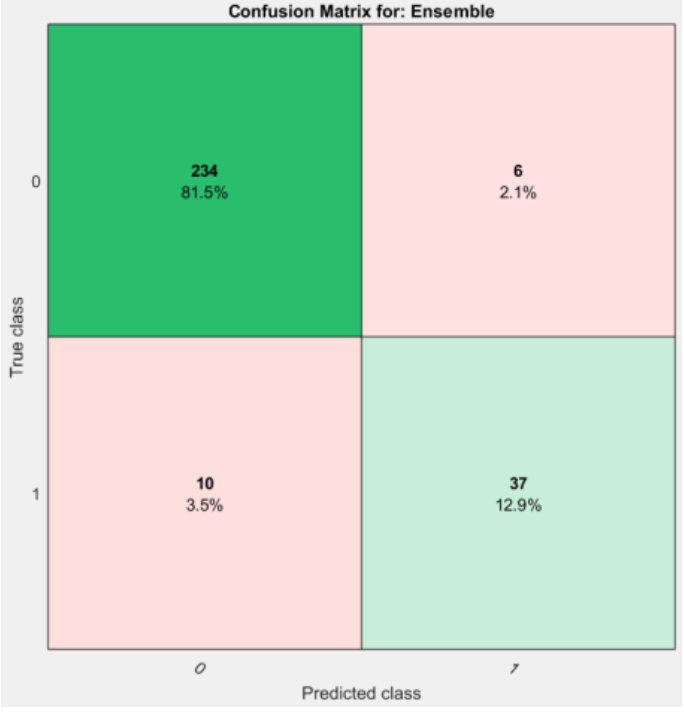


Figure 17 Overall confusion matrix for random forest classifier (see online version for colours)



A study (Mandrekar, 2010) suggests that an AUC of 0.5 reflects almost no discrimination, 0.7 to 0.8 is acceptable, 0.8 to 0.9 is excellent, and more than 0.9 is outstanding for medical diagnosis. ROC analysis shown in Figures 13, 14 and 15 was employed to evaluate the performance of all algorithms. PCA-SVM has the best AUROC value of 0.8699. In tree-based classifiers, random forest classifier has the best accuracy, sensitivity, specificity and AUROC of 94.4%, 97.5%, 78.72% and 0.97961 respectively with five-fold cross validation as seen in Figure 17. For DT classifier, the accuracy, sensitivity, specificity and AUC obtained are 92%, 95%, 76.59% and 0.876729 respectively. Twenty-three samples are misclassified as seen in Figure 16.

4 Conclusions

Breast cancer is an important health problem globally. Thermography is a low-cost potential solution for the early detection of breast cancer and prognosis indication. Despite the presence of mammography as a gold imaging standard in diagnosis of breast cancer, there is a need for promoting additional research in thermography to increase the sensitivity of prognosis in young women with dense breasts. The manual assessment of disease is time-consuming, requiring minute inspection and varies with the perception and the level of expertise of the radiologists. This study presents an approach to classify breast thermograms based on biostatistical features using various machine learning algorithms and increase the reliability of this technique for diagnosis purpose. In this work, SVM classifier with and without PCA as a pre-cursor significantly outperforms other decision-making models viz. Naïve Bayes, kNN and logistic regression on the DMR database. PCA-SVM-based model interprets thermograms as benign or malignant with accuracy of 92.74% for training with 80% data and testing with 20% data which is the most superior as compared to other classifiers. By reducing the dimension of data and time complexity, our approach with PCA helps to improve the results for all classifiers in interpreting thermograms as benign or malignant. Random forest classifier has an average accuracy, sensitivity, specificity and AUROC of 94.4%, 97.5%, 78.72% and 0.97961 respectively with five-fold cross validation which is better than the DT classifier. Results obtained point out at the viability of using simple statistical measures extracted from the breast thermograms as a first approach to aid diagnosis of breast disease. With improved accuracy on large database, thermography can be used in adjunct to ultrasound or mammography in clinical practice. In future work, nonlinear or kernel-based SVM can also be tested on a larger database.

References

Acharya, U.R., Ng, E., Tan, J.H. and Subbhuraam, V.S. (2010) 'Thermography based breast cancer detection using texture features and support vector machine', *Journal of Medical Systems*, Vol. 36, pp.1503–1510 [online] https://doi.org/10.1007/s10916-010-9611-z.

Acharya, U.R., Ng, E.Y.K., Tan, J. and Sree, V. (2012) 'Thermography based breast cancer detection using texture features and support vector machine', *Journal of Medical Systems*, Vol. 36, pp.1503–1510 [online] https://doi.org/10.1007/s10916-010-9611-z.

Borchartt, T., Resmini, R., Conci, A., Martins, A., Silva, A., Diniz, E., Paiva, A. and Lima, R. (2011) 'Thermal feature analysis to aid on breast disease diagnosis', *21st Brazilian Congress of Mechanical Engineering*, 2011.

Borchartt, T.B., Conci, A., Lima, R.C.F., Resmini, R. and Sanchez, A. (2013) 'Breast thermography from an image processing viewpoint: a survey', in *Signal Processing*, Vol. 93, No. 10, pp.2785–2803.

Breast Cancer India (2020) *Pink Indian Statistics* [online] http://www.breastcancerindia.net/statistics/stat_global.html (accessed 21 March 2020).

Breiman, L. (2001) 'Random forests', *Machine Learning Journal Paper*, Vol. 45, pp.5–32.

Bronzino, J.D. (2006) *Medical Devices and Systems*, 3rd ed., pp.25-1–25-17, CRC Press.

Fok, S.C., Ng, E.Y.K. and Tai, K. (2002) 'Early detection and visualization of breast tumor with thermogram and neural network', *J. Mech. Med. Biol.*, Vol. 2, No. 2, pp.185–195.

Francis, S.V., Sasikala, M., Bharathi, G.B. and Jaipurkar, S.D. (2014) 'Breast cancer detection in rotational thermography images using texture features', *Infrared Physics & Technology*, Vol. 67, pp.490–496 [online] https://doi.org/10.1016/j.infrared.2014.08.019.

Gaber, T., Ismail, G., Anter, A., Soliman, M., Ali, M., Semary, N., Hassanien, A.E. and Snasel, V. (2013) *Thermogram Breast Cancer Prediction Approach based on Neutrosophic Sets and Fuzzy C-Means Algorithm*, December [online] https://doi.org/10.5281/zenodo.34913.

Gogoi, U., Majumdar, G., Mrinal, B. and Ghosh, A. (2019) 'Evaluating the efficiency of infrared breast thermography for early breast cancer risk prediction in asymptomatic population', *Infrared Physics & Technology*, Vol. 99, pp.201–211, DOI: 10.1016/j.infrared.2019.01.004.

Gonçalves, C.B., Leles, A.C.Q., Oliveira, L.E., Guimaraes, G., Cunha, J.R. and Fernandes, H. (2019) 'Machine learning and infrared thermography for breast cancer detection', *Proceedings*, Vol. 27, p.45.

Hakim, A. and Awale, R.N. (2020a) 'Thermal imaging – an emerging modality for breast cancer detection: a comprehensive review', *J. Med. Syst.*, Vol. 44, p.136 [online] https://doi.org/10.1007/s10916-020-01581-y.

Hakim, A. and Awale, R.N. (2020b) 'Detection of breast pathology using thermography as a screening tool', *15th Quantitative InfraRed Thermography Conference*, accepted for publication.

Haralick, R.M., Shanmugam, K. and Dinstein, I.H. (1973) 'Textural features for image classification', *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-3, No. 6, pp.610–621.

Head, J.F., Wang, F., Lipari, C.A. and Elliott, R.L. (2000) 'The important role of infrared imaging in breast cancer', *IEEE Engineering in Medicine and Biology Magazine*, Vol. 19, No. 3, pp.52–57 [online] https://doi.org/10.1109/51.844380.

Jones, B.F. (1998) 'A reappraisal of the use of infrared thermal image analysis in medicine', *IEEE Transactions on Medical Imaging*, Vol. 17, No. 6, pp.1019–1027.

Keyserlingk, J., Ahlgren, P., Yu, E. and Belliveau, N. (1998) 'Infrared imaging of the breast: initial reappraisal using high-resolution digital technology in 100 successive cases of stage I and II breast cancer', *The Breast Journal*, Vol. 4, pp.245–251 [online] https://doi.org/10.1046/j.1524-4741.1998.440245.x.

Khan, A.A. and Arora, A. (2018) 'Breast cancer detection through gabor filter based texture features using thermograms images', *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, pp.412–417, DOI: 10.1109/ICSCCC.2018.8703342.

Khandpur, R.S. (1994) *Handbook of Biomedical Instrumentation*, 2nd ed., pp.670–684, Tata McGraw-Hill Education, New Delhi.

Kirubha, S.P.A., Anburajan, M., Venkataraman, B. and Menaka, M. (2018) 'A case study on asymmetrical texture features comparison of breast thermogram and mammogram in normal and breast cancer subject', *Biocatalysis and Agricultural Biotechnology*, Vol. 15, pp.390–401 [online] https://doi.org/10.1016/j.bcab.2018.07.001.

Koprowski, R. (2014) 'Quantitative assessment of the impact of biomedical image acquisition on the results obtained from image analysis and processing', *Biomedical Engineering*, Vol. 13, No. 1, pp.1–21.

Kuruganti, P.T. and Qi, H. (2002) 'Asymmetry analysis in breast cancer detection using thermal infrared images', *Engineering in Medicine and Biology, 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, Proceedings of the Second Joint*, Vol. 2, pp.1155–1156.

Lashkari, A., Pak, F. and Firouzmand, M. (2016) 'Full intelligent cancer classification of thermal breast images to assist physician in clinical diagnostic applications', *Journal of Medical Signals and Sensors*, Vol. 6, No. 1, pp.12–24, DOI: 10.4103/2228-7477.175866.

Lior, R. (2010) 'Ensemble-based classifiers', *Artif. Intell. Rev.*, Vol. 33, pp.1–39.

Lior, R. and Oded, M. (2005) *Decision Trees, Data Mining and Knowledge Discovery Handbook*, pp.165–192, Springer, Berlin.

Madhavi, V. and Bobby, C. (2017) 'Assessment of dynamic infrared images for breast cancer screening using BEMD and URLBP', *International Journal of Pure and Applied Mathematics*, Vol. 114, No. 10, pp.261–269.

Madhu, H., Kakileti, S.T., Venkataramani, K. and Jabbireddy, S. (2016) 'Extraction of medically interpretable features for classification of malignancy in breast thermography', *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp.1062–1065 [online] https://doi.org/10.1109/EMBC.2016.7590886.

Mandal, S.K. (2017) 'Performance analysis of data mining algorithms for breast cancer cell detection using naïve Bayes, logistic regression and decision tree', *Int. J. Eng. Comput. Sci.*, Vol. 6, pp.20388–20391.

Mandrekar, J.N. (2010) 'Receiver operating characteristic curve in diagnostic test assessment', *Journal of Thoracic Oncology*, Vol. 5, No. 9, pp.1315–1316.

Mejia, T., Perez, M., Andaluz, V. and Conci, A. (2015) 'Automatic segmentation and analysis of thermograms using texture descriptors for breast cancer detection', *Computer Aided System Engineering (APCASE) Asia-Pacific Conference*, pp.24–29, DOI: 10.1109/APCASE.2015.12.

Ng, E., Ung, L., Ng, F. and Sim, L.S.G. (2001) 'Statistical analysis of healthy and malignant breast thermography', *Journal of Medical Engineering & Technology*, Vol. 25, pp.253–263, DOI: 10.1080/03091900110086642.

Ng, E.Y.K. and Sudharsan, N.M. (2001) 'Numerical computation as a tool to aid thermographic interpretation', *Journal of Medical Engineering & Technology*, Vol. 25, No. 2, pp.53–60.

NIST (2017) *Image Central Moments Computed using Formulae* [online] https://itl.nist.gov/div898/handbook/eda/section3/eda35b.htm (accessed 24 April 2020).

Nunes, A.P., Silva, A.C. and Paiva, A.C. (2010) 'Detection of masses in mammographic images using geometry, Simpson's Diversity Index and SVM', *International Journal of Signal and Imaging Systems Engineering*, Vol. 3, No. 1, pp.40–51.

Pearl, J. (1998) *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Mateo.

Prasad, S. and Houserkova, D. (2007) 'The role of various modalities in breast imaging', *Biomedical Papers of the Medical Faculty of the University Palacky, Olomouc, Czechoslovakia*, Vol. 151, No. 2, pp.209–218 [online] https://doi.org/10.5507/bp.2007.036.

Qi, H. and Head, J.F. (2001) 'Asymmetry analysis using automatic segmentation and classification for breast cancer detection in thermograms', *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 3, pp.2866–2869.

Qi, H., Snyder, W., Head, J. and Elliott, R. (2000) 'Detecting breast cancer from infrared images by asymmetry analysis', *Proceedings of the 22nd Annual International Conference of the IEEE, Engineering in Medicine and Biology Society*. Vol. 2, pp.1227–1228, DOI: 10.1109/IEMBS.2000.897952.

Rejani, Y.I.A. and Selvi, S.T. (2009) 'Early detection of breast cancer using SVM classifier technique', *Int. J. Comput. Sci. Eng.*, Vol. 1, No. 3, pp.127–130.

Sathish, D., Kamath, S., Prasad, K., Kadavigere, R. and Martis, R. (2017) 'Asymmetry analysis of breast thermograms using automated segmentation and texture features', *Signal, Image and Video Processing*, May, Vol. 11, No. 4, pp.745–752.

Scales, N., Herry, C. and Frize, M. (2004) 'Automated image segmentation for breast analysis using infrared images', *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 3, pp.1737–1740, DOI: 10.1109/IEMBS.2004.1403521.

Schaefer, G., Závišek, M. and Nakashima, T. (2009) 'Thermography based breast cancer analysis using statistical features and fuzzy classification', *Pattern Recognition*, Vol. 42, No. 6, pp.1133–1137.

scikit-learn.org (2019) sklearn.svm.LinearSVC, sklearn [online] http://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html (accessed May 24 2020).

Sh, A., Shahraki, H., Rowhanimanesh, A.R. and Eslami, S. (2016) 'Feature selection using a genetic algorithm for breast cancer diagnosis: an experiment on three different datasets', *Iran J. Basic Med. Sci.*, Vol. 19, No. 5, pp.476–482.

Shahari, S. and Wakankar, A. (2015) 'Color analysis of thermograms for breast cancer detection', *2015 International Conference on Industrial Instrumentation and Control (ICIC)*.

Silva, L.F., Saade, D.C.M., Sequeiros, G.O., Silva, A.C., Paiva, A.C., Bravo, R.S. and Conci, A. (2014) 'A new database for breast research with infrared image', *Journal of Medical Imaging and Health Informatics*, Vol. 4, No. 1, pp.92–100 [online] https://doi.org/10.1166/jmihi.2014.1226.

Tang, X., Ding, H., Yuan, Y-e. and Wang, Q. (2008) 'Morphological measurement of localized temperature increase amplitudes in breast infrared thermograms and its clinical application', *Biomedical Signal Processing and Control*, Vol. 3, No. 4, pp.312–318, DOI: 10.1016/j.bspc.2008.04.001.

US Food and Drug Administration (2011) *Breast Cancer Screening – Thermography Is Not an Alternative to Mammography: FDA Safety Communication*, 2 June [online] https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm257633.htm (accessed 3 March 2020).

Visual Lab (2014) *A Methodology for Breast Disease Computer-Aided Diagnosis using Dynamic Thermography* [online] http://visual.ic.uff.br/dmi (accessed 15 April 2020).

Wakankar, A.T. and Suresh, G.R. (2016) 'Automatic diagnosis of breast cancer using thermographic color analysis and SVM classifier', *Advances in Intelligent Systems and Computing Intelligent Systems Technologies and Applications*, pp.21–32.

Zhang, H. (2004) 'The optimality of naive Bayes', in *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004)*, AAAI Press, Miami Beach, FL, pp.562–567.