Informatics in Medicine Unlocked 2 (2016) 70–77

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Computational assessment of breast tumour differentiation using multimodal data

Jean Rossario Raj [a,*], Syed Mohammed Khalilur Rahman [a,b], Sneh Anand [a,b]

[a] Centre for Bio-Medical Engineering, Indian Institute of Technology, Delhi, India
[b] Bio-Medical Engineering Unit, All India Institute of Medical Sciences, Delhi, India

ARTICLE INFO

Article history:
Received 18 February 2016
Received in revised form
17 April 2016
Accepted 19 April 2016
Available online 24 April 2016

Keywords:
Breast tumour diagnosis
Symptomatic examination
Mammography
Ultrasound imaging
Computer modelling

ABSTRACT

Early detection of breast cancer requires accurate prediction and reliable diagnostic modalities. This allows physicians to distinguish malignant tumours before proceeding with a painful surgical biopsy. The attributes of three non-invasive primary diagnosing modalities, namely symptomatic examination, ultrasound imaging, and mammographic results, were used for the study. A dataset was created using ten selected features from each modality, after iterations during the training phase. The number of satisfying features was used for the creation of a model, which was further categorised as benign or malignant class. The model was evaluated in the testing phase by comparing biopsy results for benign or malignant classification. The statistical analysis proved it as an efficient approach for non-invasive decision-making. The developed model was tested using supervised learning algorithms with three classifiers for 210 cases by comparing the results with the gold standard biopsy results. The sensitivities for the three classifiers were 80%, 73% and 76.5%, while specificities were 96%, 94.4% and 95%, respectively. This method of breast tumour differentiation using the features of the non-invasive modalities can be widely used in telemedicine applications, helping to reduce confirmatory biopsies.

1. Introduction

International Agency for Research on Cancer states that global cancer burden has increased, with new incidences and cases of breast cancer, the second leading cause of mortality for women in the world [1]. Prior diagnosis leads to reduced mortality rate. Hence, improved diagnostic efficiency is needed for distinguishing the benign tumours (non-cancerous) from malignant (cancerous) without proceeding with a painful surgical biopsy. For women with cancer that has not gone to the metastasis stage, the long-term survival rate has increased, with majority of them surviving many years after diagnosis and treatment [2].

1.1. Diagnostic modalities

During the diagnosis of lesions, women were initially subjected to symptomatic examinations. In suspected cases, they need further imaging using mammography. In certain cases, they were subjected to ultrasonography as well. These modalities assist in determining the possibility of malignancy. The current breast imaging modalities involve screening in the upper inner and outer quadrants, lower inner and outer quadrants, central, upper half, lower half, lateral half and Axilla [3]. Due to limited health literacy, especially in developing countries, many women do not undergo a mammogram until a lump is detected symptomatically, and thus the first mammogram becomes diagnostic [4]. The inconclusive and suspicious cases are followed up with using ultrasound after the mammogram [5]. In one study [6], it was noted that patients with palpable masses used breast ultrasound, while those with bulging masses and/or deformed breast outlines used mammography as the first-line imaging examination.

Occult or controversial findings between various modalities result in vague lesions assessments, leading to core or open breast biopsies. Breast cancer diagnosis in young individuals is a challenge because their breast tissue is often denser than the breast tissue of older women [7]. In addition, there is difficulty in correlating breast cancer features, while there are also diagnostic dilemmas between the mammography and ultrasound imaging techniques. In mammograms, fat is radiolucent and appears dark, but fibro-glandular tissues are radio-dense and appears white [8]. Moreover, studies [9] have shown that a tumour becomes detectable on a mammogram only after 40 cell doublings from the growth of the first tumour cell in breast cancer. Thus, a more efficient correlation mechanism is required for better prediction of malignant tumours.

* Correspondence to: Room No: 299, 2nd Block, Centre for Bio-Medical Engineering, Indian Institute of Technology, New Delhi 110016, India.
E-mail address: bmz108120@cbme.iitd.ac.in (J.R. Raj).

http://dx.doi.org/10.1016/j.imu.2016.04.001

1.2. Supervised machine learning approach

This approach uses training data to produce a function or model for predicting new outcomes. An optimal algorithm can evaluate unseen instances correctly based on the attributes or features, which are the variables of the machine learning process. Here, the accuracy and sensitivity is improved with satisfactory training data, reliable feature selection, proper classification algorithms, and high dimensional input data characteristics. In this study, a supervised learning approach is successfully examined for the prediction of malignant breast cancer using the datasets of non-invasive screening features for symptomatic examination, ultrasound imaging, and mammography.

1.3. Databases for feature extraction

The digital database for mammography screening, the national mammography database, and the mammographic image analysis society database are some of the common databases for analysing the mammographic images [10]. The attributes for machine learning are selected from the standard databases of non-invasive diagnostic modalities, such as breast cancer symptom examinations from the national health services in the United Kingdom, the screening standards for ultrasonography from the American college of radiology, and mammography diagnosing features from the national mammography database. The mammographic imaging features database provide basic data for the initial diagnosis of breast cancer [11]. In another study [12], configurations of the feature vectors in evaluation for mammogram images were extracted from segmented regions on the cranio-caudal and/or medio-lateral oblique views in the study. The automatic extraction of features from the original database reduces input dimensionality and improves the discriminatory performance of the classifier [13].

1.4. Feature selection and classification

The features extracted are given to a classifier for automatic classification into benign and malignant breast conditions [14]. In one study [15], three classes of study for normal, benign, and malignant cancerous cases were not used, as normal cases were excluded. The feature selection was appropriate to handle such exclusion. In the study using a support vector machine (SVM), the best parameters of kernel functions for pattern analysis and categorisation were carried out with the grid search method [16]. Furthermore, the feature optimisation was performed for categorisation parameters using linear and nonlinear kernels [17]. The optimisation of an anisotropic kernel was performed by eliminating features of the low relevance classifier, as per previous research [18].

The feature selection for a sparse representation-based classification algorithm from mammographic grey scale information was analysed and compared in another study [19]. Moreover, the CAD system was found to be an efficient method to eliminate the operator dependency and improve the diagnostic accuracy in breast tumour differentiation and classification [20].

1.5. Machine learning model creation and evaluation

A fitted model was created by combining all existing classified data and is used for prediction and diagnosis of breast cancer. The nodes were the variables in a probabilistic graphical model [21]. The developed parameter model was applied to a validation dataset consisting of patients with tumours and those having healthy non-malignant cases were used as controls [22]. During the testing phase, the supervised learning categorises the input parameters into classification algorithms and is applied to response values as true or false from few known results. The fine tuning of the model's predictive value is carried out by predicting one of the classes with maximum precision [23]. The fractions of representation from each of the classes are used to achieve the boundary shift and asymmetric regularisation to extend to the multi-class scenario [24]. The statistical model was evaluated using different classification models for its specificity, sensitivity, accuracy, F-score, Youden's index and the ROC curve [25]. Using the attributes selected for the model, the areas under the ROC curves are obtained [26]. The model is considered statistically significant if the two-tailed P value is less than 0.05 [27].

1.6. Objective of the study

The objective of the study was to use supervised learning and mathematical modelling approaches with improved sensitivity for the differentiation of malignant and benign tumours. For developing the model, the different malignant and benign features, which are identifiable using symptomatic, mammographic and ultrasound modalities, were used. The developed model was further refined with multiple testing phases. The biopsy results taken for the suspicious malignant cases were used as the gold standard for comparison. The identification of malignant cases can be well-suited using one class because the decision is unaffected by the outliers and the form of the data fit more precisely [28]. In the study, three classifiers were used to understand which fits well for the selected feature set. The outcomes of the three classifiers can be combined to form a single ensemble model to get more accuracy [29]. The novelty of the work is the use of combined feature set of symptomatic, mammography and ultrasound modalities for prediction using machine learning methods.

2. Materials and methods

2.1. Data set and attributes

The training phase included 180 cases. Of these, biopsy showed that 132 cases were benign and 48 were malignant. The evaluation of the classifiers as benign or malignant was done with the 'present' or 'not present' feature selection. The features are selected from the databases described in section 1.3. The features used for the symptomatic examinations are given in Table 1. The invasive procedures were excluded from the feature selection.

The features considered for the mammographic screening are given in Table 2 for classification into benign or malignant.

Similarly, in the case of ultrasound imaging, the features taken into consideration are given in Table 3.

A few of the features on the ultrasound images are given in Fig. 1.

Table 1
Symptomatic features used for the study.

Symptomatic features
Immovable lump or thickening
Change in size, shape or contour
Colour change of the nipple
Puckering or dimpling of the breast skin
Continuous pain or tenderness in breast or armpit
Nipple inversion or change in appearance
Swelling or darkening in the breast
Itchy scaly sore or rash on the nipple
Blood-stained or clear fluid discharge from the nipple
Noticeable flattening or indentation on the breast