2017 CMBEC40 Conference
Winnipeg MB
May 23–26, 2017

A COMPARISON OF CLASSIFIERS FOR DETECTING TUMOURS USING MICROWAVE SCATTERING IN NUMERICAL BREAST MODELS

Jorge Sacristán[1], Stephen Pistorius, PhD.[1,2]

1: Dept. of Physics & Astronomy, University of Manitoba

2: Dept. of Radiology, University of Manitoba

INTRODUCTION

Current breast cancer screening, using X-ray mammography has various drawbacks. These include the use of ionizing radiation, the need for breast compression, high cost and the difficulty in implementing this technology in rural communities. Some studies have shown a reduction in mortality associated with x-ray mammography breast cancer screening [1, 2]. However, more recent papers have challenged these findings and have argued that the benefits associated with mammography are inconclusive [3, 4].

The reported incidence of breast cancer is higher in the developed nations. However, people in emerging economies have lower survival rates. For instance, the five-year survival rate for breast cancer is less than 50% in Gambia, Uganda, and Algeria, while it is nearly 90% in the United States [5]. In Manitoba, Canada, women from rural areas have a cancer incidence-to-mortality rate of 60%, while their urban counterparts have a rate of 37% [6].

The research presented in this paper is part of an ongoing project that aims to improve the availability of breast cancer screening by providing a portable device that is suited to the needs of low and middle-income countries and rural communities. The feasibility of a portable breast microwave sensing (BMS) system is being evaluated. This portable and inexpensive system, which does not require highly trained operators, is being developed to bring breast cancer screening to remote communities that might not otherwise have access [7].

EXPERIMENTAL SYSTEM

The proposed system classifies the microwave responses from a patient's breast into one of two classes, those containing tumours and those without. The classification algorithm consists of these steps:

1. The patient places her breast between a transmitting antenna and an array of receiving sensors.
2. The transmitting antenna sends a microwave pulse towards the breast while the scattered signals are captured by the sensors and logged into the on-board computer.
3. The servo-motor rotates the chamber that holds the transmitting antenna and sensors.
4. Steps 2 and 3 are repeated until the scattered signals have been acquired at five different rotation angles.
5. An array containing the values of all scattered signals from all different angles is given as an input to a classifier that was previously trained. The classifier indicates if a tumour is present.

To test the feasibility of the proposed experimental system, Richmond's frequency-domain simulation was performed [8], and the results were fed into different classifiers for evaluation. The datasets were generated using an algorithm that simulates the scattering of the electric field given a 2D cross-sectional model of an arbitrarily shaped scatterer [8]. In our case, the scatterer consists of the breast tissues and the system's chamber.

The simulations were performed using random variables for breast diameter, tumour sizes, fibro-glandular densities, and breast misalignment with respect to the chamber's center of rotation.

TISSUES AND PROPERTIES

In the absence of tissues, the sensors would only detect the electrical field produced by the transmitting antenna and scattered by the PVC walls of the chamber. As tissue is introduced into the experiment, the electrical field is perturbed and the values logged by the sensors are affected.

The degree to which electrical field scattering occurs is determined by the relative permittivity of the introduced material. The relative permittivity of the phantom material is a function of frequency, moisture, temperature and pressure [9]. However, if the measurement is carried out at constant values of moisture, temperature, and pressure, the accepted method for modeling a material's relative permittivity is by fitting known values of the Cole-Cole model into a single equation [10], which has the following form [11]:

$$\varepsilon_r = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1 + (j\omega\tau)^{1-\alpha}} + \frac{\sigma_s}{j\omega\varepsilon_0} \tag{1}$$

BREAST MODELS AND DATASETS

The dataset used in this paper is composed of 871 randomly generated breast models. To simulate the chamber's rotational capabilities, each model was rotated in increments of 12° for a total of five breast orientations.

Each breast model was subjected to five frequencies (2.3 GHz, 3.35 GHz, 4.4 GHz, 5.45 GHz, and 6.5 GHz) and the scattered field values were captured at 12 sensor locations, resulting in 60 values per breast orientation.

The dataset was classified using three different modalities:

- Mode 1 assumes every breast orientation is a separate experiment, resulting in 4355 samples of 60 features.
- Mode 2 selects only the first orientation and disregards the rest, resulting in 871 samples of 60 features
- Mode 3 takes all five positions as a single experiment, for a total of 871 samples of 300 features.

SCATTERING SIMULATION

The method used to simulate the scattering phenomena of the electric field with the different materials is an implementation of the algorithm described in Richmond's classic paper [8]. This algorithm uses a single frequency and different permittivity values on a 2D grid. The total field at any point is found by solving a system of algebraic equations. The field intensity at each of the sensor locations is taken from these results. While this is an idealised situation, it is a reasonable approximation to the measurements that are obtained by using solid-state sensors that are similar in size to the spatial resolution of the simulation. This algorithm generates results that compare well with the exact solution [8].

Five frequencies were used in these simulations. Since the permittivity of the materials change as a function of frequency, this algorithm was run once for each frequency. A grid size of 2 mm was chosen as a balance between processing time and spatial resolution.

CLASSIFICATION

SVMs, as well as other machine learning techniques, have been successfully used to classify electromagnetic signals scattered from breast tissue [16]. A total of three classifiers were chosen for this paper: A K-Nearest Neighbour, a Support Vector Machine, and a Neural Network.

The pre-processing algorithm for each classifier was as follows:

- A cross-validation train-test split of 80%-20% was carried out.
- A Principal Component Analysis (PCA) was used to weight the features, according to the apparent correlation between each feature and the true class of the sample. No feature reduction was carried out.
- A Grid-search algorithm found the best-performing values for the different classifier's parameters. A simultaneous 5-

fold cross-validation on the training data was carried out.

- The remaining 20% of the data that represent the test set were used by the classifier to obtain its predictive power.

The scores of a classifier vary with each run, and an average score was obtained by running the algorithms 20 times with each dataset's mode. The machine learning module for Python "sklearn 0.16.1" was used to perform the classification. This software package includes the code necessary to implement and train the different classifiers [17, 18].

The classification metrics used to compare each algorithm were the following:

$$DA = \frac{tp + tn}{tp + tn + fp + fn} \tag{2}$$

$$tpr = \frac{tp}{tp + fn} \tag{3}$$

$$tnr = \frac{tn}{tn + fp} \tag{4}$$

Where *DA* stands for Diagnostic Accuracy, *tpr* is the True positive rate, *tnr* the true negative rate, *tp* the true positive fraction, *tn* the true negative fraction, *fp* the false positive fraction, and *fn* the false negative fraction.

The Receiver Operating Characteristic (ROC) shows a classifier's performance by comparing the sensitivity and the false positive rate. The Area under the Curve (AUC) of the ROC was chosen as a metric to evaluate a classifier's overall performance.

RESULTS

Three machine learning algorithms were used to classify the simulated experiments: K-Nearest Neighbors, Support Vector Machine with a Radial Basis Function Kernel, and a Neural Network.

The following tables show the average performance metrics obtained by each classifier on the three different dataset modes.

Table 1: Performance metrics: KNN

Dataset mode	**Diag. Accuracy**	***Sensitivity***	***Specificity***	***ROC AUC***
1	**74%**	**60%**	**86%**	**83%**
2	**64%**	**41%**	**85%**	**70%**
3	**53%**	**5%**	**98%**	**63%**

Table 2: Performance metrics: SVM

Dataset mode	**Diag. Accuracy**	***Sensitivity***	***Specificity***	***ROC AUC***
1	**88%**	**86%**	**91%**	**95%**
2	**74%**	**73%**	**76%**	**82%**
3	**58%**	**79%**	**40%**	**67%**

Table 3: Performance metrics: Neural Networks

Dataset mode	**Diag. Accuracy**	***Sensitivity***	***Specificity***	***ROC AUC***
1	**90%**	**89%**	**92%**	**78%**
2	**74%**	**71%**	**77%**	**63%**
3	**61%**	**54%**	**67%**	**52%**

An example of a ROC curve obtained from the classifiers is shown in Figure 1.

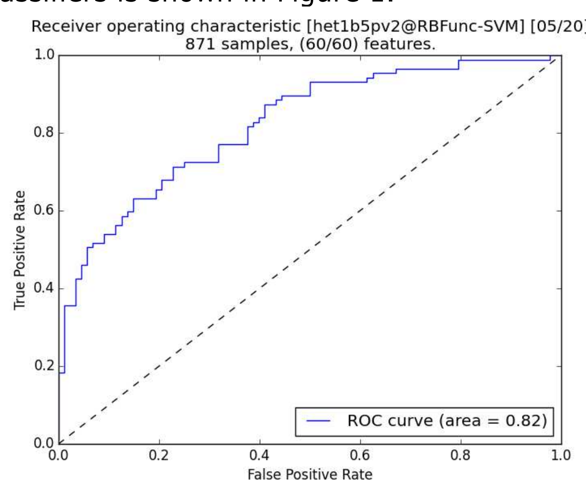

Figure 1: ROC curve of SVM at mode 2

DISCUSSION

In this paper, we have evaluated the feasibility of a breast tumor detection device, which uses machine learning to identify the presence of a lesion from the scattered microwave data. The data used for classification was simulated using 2D models of breasts and the scattering patterns they produced. This information was gathered into a dataset of 871 randomly generated models that was further processed in three possible modes.

Each of the modalities produced a different samples-to-features ratio. This ratio has a direct impact on classification performance. Generally, a higher ratio allows for better classification rates. While Mode 1 has the biggest samples-to-features ratio of all dataset modes, it assumes all five positions of a single breast model are separate phenomena. This cannot occur in practice, as a priori information from a patient will not be available. Mode 2 more realistically, takes only one position into account and produces ROC AUC values similar to x-ray mammography.

The third mode showed poor performance that can be attributed to the low number of samples in comparison to the feature array. More simulations need to be carried out to improve the classification performance in mode 3. Additional pre-processing techniques that make use of data from all five positions are being investigated to improve classification rates even further.

ACKNOWLEDGEMENTS

This research was supported by:

- Grand Challenges Canada
- CancerCare Manitoba
- NSERC: Natural Sciences and Engineering Research Council of Canada

REFERENCES

[1] G. Van Schoor et al. "Increasingly strong reduction in breast cancer mortality due to screening". In: British Journal of Cancer (2011).

[2] J. Fracheboud, S J Otto, J A A M van Dijck, M J M Broeders, A L M Verbeek, and H J de Koning. "Decreased rates of advanced breast cancer due to mammography screening in The Netherlands". In: British Journal of Cancer (2004).

[3] Anthony B Miller, Claus Wall, Cornelia J. Baines, Ping Sun, Teresa To, and Steven A. Narod. "Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial". In: BMJ (2014).

[4] JB Harford et al. "Problem solving for breast health care delivery in low and middle resource countries (LMCs): consensus statement from the Breast Health Global Initiative". In: Elsevier: The Breast (2011).

[5] "Cancer in Africa". In: American Cancer society. (2011).

[6] Department of Epidemiology and CancerCare Manitoba. Cancer Registry. Cancer in Manitoba, 2012 Annual Statistical Report. 2015.

[7] A. Cole. Winnipeg woman, 26, shares story of breast cancer diagnosis. http://www.cbc.ca/news/canada/manitoba/winnipeg-woman-26-shares-story-of-breastcancer-diagnosis-1.3017319. Apr. 2015.

[8] J. Richmond. "Scattering by a Dielectric Cylinder of Arbitrary Cross Section Shape." In: IEEE Transactions on Antennas and Propagation (1965).

[9] V. Vyacheslav. Handbook of Dielectric and Thermal Properties of Materials at Microwave Frequencies. Plastics International, Artech House, 2012.

[10] D.K. Ghodgaonkar and A.B.f Daud. "Calculation of Debye parameters of single Debye relaxation equation for human skin in vivo". In: 4th National Conference of Telecommunication Technology (2003).

[11] M Lazebnik et al. "A large-scale study of the ultrawideband microwave dielectric properties of normal breast tissue obtained from reduction surgeries". In: Physics in Medicine and Biology (2007).

[12] S. Alekseev and MC. Ziskin. "Human Skin Permittivity Determined by MillimeterWave Reflection Measurements". In: Bioelectromagnetics (2007).

[13] Abas Sabouni, Sima Noghanian, and Stephen Pistorius. "A global optimization technique for microwave imaging of the inhomogeneous and dispersive breast". In: Canadian Journal of Electrical and Computer Engineering (2010).

[14] M Lazebnik et al. "A large-scale study of the ultrawideband microwave dielectric properties of normal, benign and malignant breast tissues obtained from cancer surgeries". In: Physics in Medicine and Biology (2007).

[15] J. Sacristán, B. L. Oliveira, and S. Pistorius. "Classification of electromagnetic signals obtained from microwave scattering over healthy and tumorous breast models". In: 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). May 2016, pp. 1–5. DOI: 10.1109/CCECE.2016.7726761.

[16] R. Conceição. "The Development of Ultra Wideband Scanning Techniques for Detection and Classification of Breast Cancer". In: National University of Ireland Galway (2010).

[17] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: Journal of Machine Learning Research 12 (2011), pp. 2825–2830.

[18] Lars Buitinck et al. "API design for machine learning software: experiences from the scikit-learn project". In: ECML PKDD Workshop: Languages for Data Mining and Machine Learning. 2013, pp. 108–122.