32nd URSI GASS, Montreal, 19-26 August 2017

Support Vector Machines to Aid Breast Cancer Diagnosis Using a Microwave Radar Prototype

Raquel C. Conceição* (1), Daniela M. Godinho (1), Dallan Byrne (2), and Ian Craddock (2)
(1) Instituto de Biofísica e Engenharia Biomédica, Faculdade de Ciências, Universidade de Lisboa, 1749-016-Lisboa, Portugal
(2) Department of Electrical & Electronic Engineering, University of Bristol, UK

Abstract

In this paper we use Support Vector Machines (SVM) to aid breast cancer diagnosis using the microwave imaging radar prototype from the University of Bristol. We extract features from the electromagnetic signals collected with a microwave imaging system and classify them with a SVM classifier. The classifier will give an indication on whether the electromagnetic signals are more likely to belong to a focal point in the breast corresponding to healthy tissue – a "miss" – or to tumour tissue – a "hit". The proposed work has the potential to ultimately aid microwave imaging and help avoid false positives and detect tumours that were imaged as false negatives.

1. Introduction

Previous studies have addressed the use of classifiers to determine whether high scattering regions obtained by radar microwave imaging of the breast correspond to benign or malignant tumours. Several algorithms have been used for this purpose, including Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis, Support Vector Machines (SVM) [1], Spiking Neural Networks [2] and Self-Organising Maps [3]. Other studies have also looked at classifying whether a breast is healthy or has any tumours [4, 5].

This paper extends our previous work in [6] in which we used Linear Discriminant Analysis to assess the backscattered electromagnetic signals recorded with a multistatic radar microwave imaging prototype ahead of using imaging beamformers to reconstruct the energy profile of the breast. Each synthetic focal point of the breast profile will be calculated as the probability of tumour occurrence, hence creating a "probabilistic map" of the breast. We extend our previous work by using a machine learning algorithm – SVM – as the classifier in this study, and by applying a threshold to our "probabilistic map" so that we have a binary map of the breast that includes only healthy and tumour tissue.

The materials used are presented in section 2, the methodology is presented in section 3, the results are presented in section 4 and conclusions in section 5.

2. Materials

2.1 Microwave Imaging System Description

We used a Microwave Imaging prototype system, described elsewhere [7]. This system includes a Rohde-Shwarz ZVT 8 port VNA – operating in the 3-8GHz frequency range – and a conformal array of 60 wide-slot antennas embedded in a low-loss matching ceramic shell. A total of 36 breast phantoms were tested in this system for this study. Each receiving antenna m recorded the electromagnetic signals (E_{nm}) transmitted by each transmitting antenna n. The skin was removed via a differential measurement acquired with signals collected after a 10º rotation of the phantom on the prototype.

2.2 Phantoms

All 36 breast phantoms were developed with layers of Tissue Mimicking Material (TMM), which exhibit similar dielectric properties to the internal breast tissue. The skin was considered as an evenly distributed layer of TMM with a thickness varying between 1 and 3 mm. Each breast phantom contains different tumour phantoms (which can have different sizes, location and number) filled with heterogeneous and/or homogeneous TMM. The breast phantoms were completed with a chest-wall TMM – a liquid which was mantained at the base of the breast phantom.

The dielectric properties at 6 GHz for each TMM layer are approximately the following: $\varepsilon_r = 30$ and $\sigma = 4\ s/m$ for the skin; $\varepsilon_r = 50$ and $\sigma = 7\ s/m$ for the tumour tissue; $\varepsilon_r = 30$ and $\sigma = 3.9\ s/m$ for the heterogeneous breast tissue; and $\varepsilon_r = 9.3$ and $\sigma = 0.22\ s/m$ for the matching liquid.

3. Methodology

The model considered in this study contains a total of 1558 Synthetic Focal Points (SFPs), extracted from a total of 1,925,455 measurements of "hits" and "misses" on 36 different phantoms. 779 of these SFPs are "hits" and the remaining SFPs are "misses", as represented in the synthetic breast model in Fig. 1.

Twenty-four features were extracted from the measured data of each focal point and a SVM classifier was used to distinguish “hits” and “misses”.

The dataset was then divided into training and testing sets and using cross-validation. A k-fold cross validation with k=10 was implemented and the calculated metrics correspond to the sum of the classification results for each fold. The classification performance was evaluated by metrics such as accuracy, sensitivity, specificity, positive predictive value and negative predictive value.

SVM classifier maps the input vectors (i.e. the 24 features for each focal point) into high-dimensional feature spaces according to the chosen kernel, creating a hyperplane that separates the data into classes. The kernel considered in this study was the Radial Basis Function (RBF) which has a scaling factor γ which needs to be optimised for each dataset. There is also another important parameter of SVM, usually represented with C, which is a penalty parameter of the error term. Both γ and C need to be optimised to guarantee good classification results. We performed a parameter grid-search [8] to optimise the parameters C and γ for the present database. We compared the performance metrics with several combinations of C and γ, and $C = 2$ and $\gamma = 4$ were found to yield the best classification results.

For each focal point, the SVM classifier returns a classification value which we normalised from 0 to 1. We have both binary and regression results. For binary results, we apply a threshold of 0.5, and so any classification value below 0.5 becomes “0” (which represents “miss” or a healthy point), or else it becomes “1” (which represents “hit” or a tumour point). The regression results have all values ranging between 0 and 1.

4. Results

The reconstructed profile of the synthetic breast model considered was the same used in [6]. The results of SVM classifier are shown in Fig. 2 (binary results) and Fig. 3 (regression results).

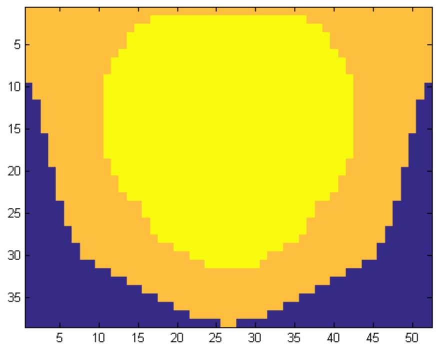

Figure 1. An example of a “hit” and a “miss” synthetic focal point in the synthetic breast model. Orange and yellow pixels represent true “miss” and “hit” SFPs, respectively.

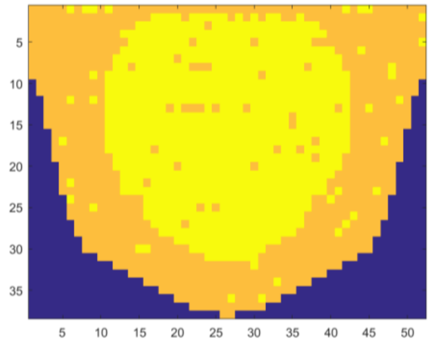

Figure 2. Reconstructed profile of the synthetic breast model using the SVM classifier in the binary mode. Orange and yellow pixels represent the classified “miss” and “hit” SFPs, respectively.

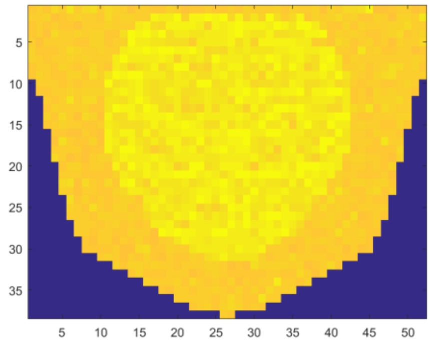

Figure 3. Reconstructed profile of the synthetic breast model using the SVM classifier in the regression mode. Orange and yellow pixels represent the classified “miss” and “hit” SFPs, respectively.

5. Conclusions

In general, the approach with SVM classifier presented in this paper outperforms the regression classification results with LDA, presented in [6], and will be considered in future related work.

6. Acknowledgements

This paper is supported by Fundação para a Ciência e a Tecnologia – FCT / MEC (PIDDAC) – under the Strategic Programme UID/BIO/00645/2013. This paper has been developed in the framework of COST Action TD1301, MiMed.

7. References

1. R. C. Conceição *et al*, "Numerical Modelling for Ultra Wideband Radar Breast Cancer Detection and Classification", *PIER B*, **34**, 2011, pp. 145-171, doi: 10.2528/PIERB11072705

2. M. O’Halloran *et al*, "Spiking Neural Networks for Breast Cancer Classification in a Dielectrically Heterogeneous Breast", *PIER*, **113**, 2011, pp. 413-428, doi: 10.2528/PIER10122203

3. M. Jones *et al*, "Self-Organising Maps for Breast Cancer Detection Using Ultra-Wideband Radar", *ICONS-IARIA*, Seville, Spain, 2013, pp. 46-51
4. D. Byrne *et al*, "Breast Cancer Detection Based on Differential Ultrawideband Microwave Radar", *PIER M*, **20**, 2011, pp. 231-242, doi: 10.2528/PIERM11080810
5. A. Santorelli *et al*, "Investigation of Classifiers for Tumor Detection with an Experimental Time-Domain Breast Screening System", *PIER*, **144**, 2014, pp. 45-57, doi: 10.2528/PIER13110709
6. R. C. Conceição *et al*, "Initial study for the investigation of breast tumour response with classification algorithms using a microwave radar prototype", *EuCAP*, Davos, Switzerland, 2016, pp., doi: 10.1109/EuCAP.2016.7481464
7. T. Henriksson *et al*, "Clinical Trials of a Multistatic UWB Radar for Breast Imaging", *LAPC*, Loughborough, UK, 2011, pp. 1-4, doi: 10.1109/LAPC.2011.6114004
8. C. Hsu *et al*, *A Practical Guide to Support Vector Classification*, *in:* www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf