Breast Cancer Identification Based on Thermal Analysis and a Clustering and Selection Classification Ensemble

Bartosz Krawczyk[1], Gerald Schaefer[2], and Shao Ying Zhu[3]

[1] Department of Systems and Computer Networks,
Wroclaw University of Technology, Wroclaw, Poland
bartosz.krawczyk@pwr.wroc.pl

[2] Department of Computer Science,
Loughborough University, Loughborough, U.K.
gerald.schaefer@ieee.org

[3] School of Computing and Mathematics,
University of Derby, Derby, U.K.
s.y.zhu@derby.ac.uk

Abstract. Breast cancer is the most common form of cancer in women. Early diagnosis is necessary for effective treatment and therefore of crucial importance. Medical thermography has been demonstrated an effective and inexpensive method for detecting breast cancer, in particular in early stages and in dense tissue. In this paper, we propose a medical decision support system based on analysing bilateral asymmetries in breast thermograms. The underlying data is imbalanced, as the number of benign cases significantly exceeds that of malignant ones, which will lead to problems for conventional pattern recognition algorithms. To address this, we propose an ensemble classifier system which is based on the idea of Clustering and Selection. The feature space, which is derived from a series of image symmetry features, is partitioned in order to decompose the problem into a set of simpler decision areas. We then delegate a locally competent classifier to each of the generated clusters. The set of predictors is composed of both standard models as well as models dedicated to imbalanced classification, so that we are able to employ a specialised classifier to clusters that show high class imbalance, while maintaining a high specificity for other clusters. We demonstrate that our method provides excellent classification performance and that it statistically outperforms several state-of-the-art ensembles dedicated to imbalanced problems.

Keywords: breast cancer diagnosis, medical thermography, pattern recognition, multiple classifier system, imbalanced classification, clustering and selection.

1 Introduction

Medical thermography uses cameras with sensitivities in the thermal infrared to capture the temperature distribution of the human body or parts thereof.

K. Imamura et al. (Eds.): BHI 2013, LNAI 8211, pp. 256–265, 2013.

In contrast to other modalities such as mammography, it is a non-invasive, non-contact, passive and radiation-free technique, as well as relatively inexpensive. The radiance from human skin is an exponential function of the surface temperature, which in turn is influenced by the level of blood perfusion in the skin. Thermal imaging is hence well suited to pick up changes in blood perfusion which might occur due to inflammation, angiogenesis or other causes [1].

Thermography has also been shown to be well suited for the task of detecting breast cancer [2,3]. Here, thermal imaging has advantages in particular when the tumor is in its early stages or in dense tissue. Early detection is crucial as it provides significantly higher chances of survival [4] and in this respect infrared imaging can outperform the standard method of mammography. While mammography can detect tumors only once they exceed a certain size, even small tumors can be identified using thermal infrared imaging due to the high metabolic activity of cancer cells which leads to an increase in local temperature that can be picked up in the infrared [5].

In this paper, we propose a medical decision support system based on analysing bilateral asymmetries in breast thermograms. Our approach is based on extracting image symmetry features from the thermograms and employing them in a pattern recognition stage for which we use a multiple classifier system. Multiple classifier systems (MCSs), or ensemble classifiers, utilise more than one predictor for decision making [6], and thus provide several advantages:

- The process of forming an ensemble does not differ significantly from the canonical pattern recognition steps [7], while the design of a classifier ensemble aims to create a set of complementary/diverse classifiers and to employ an appropriate fusion method to merge their decisions.
- MCSs may return an improved performance in comparison with a standard single classifier approach. This is due to their ability to exploit the unique strengths of each of the individual classifiers in the pool. Additionally, an MCS protects against selection of the worst classifier in the committee [8].
- Ensembles may be more robust and less prone to overfitting, because they utilise mutually complementary models with different strengths.

At the same time, there are a number of issues that have to be considered when designing an MCS, namely:

- How to select a pool of diverse and mutually complementary individual classifiers.
- How to design interconnections between classifiers in the ensemble, i.e. how to determine the ensemble topology.
- How to conduct the fusion step to control the degree of influence of each classifier on the final decision.

In this work, we particularly focus on the first problem. Our classifier selection assumes a local specialisation of individual classifiers. Following this, a single classifier that achieves the best results is chosen from a pool for each demarcated partition of the feature space. Its answer is treated as the system answer