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 for all objects in that partition. This methodology was first described in [9]. While some further proposals based on this idea assume a local specialisation of particular classifiers and only search for locally optimal solutions [10,11], other methods divide the feature space and select/train a classifier for each generated partition [12,13].

In our approach, we propose a modification of the Clustering and Selection ensemble [12] that is dedicated to addressing class imbalance. We partition the feature space into several clusters, and then delegate the most competent classifier from the pool to each of the clusters. We utilise a fixed pool of classifiers, consisting of both standard models and ones dedicated specifically for imbalanced problems, to cope with any class imbalance by assigning a specialised classifier to clusters with uneven distributions, while preserving the good specificity provided by standard classifiers to other clusters. Our approach, tested on a large dataset of breast thermograms, is shown to return excellent classification results and to statistically outperform various classifier ensembles dedicated to imbalanced problems.

2 Breast Thermogram Features

As has been shown, an effective approach to detect breast cancer based on thermograms is to study the symmetry between the left and right breast regions [14]. In the case of cancer presence, the tumor will recruit blood vessels resulting in hot spots and a change in vascular pattern, and hence an asymmetry between the temperature distributions of the two breasts. On the other hand, symmetry typically identifies healthy subjects.

We follow this approach and extract image features that describe bilateral differences between the areas of the left and right breasts extracted from frontal view thermograms. In particular, we employ the image features that were derived in [15], namely:

- Basic statistical features: mean, standard deviation, median, 90-percentile;
- Moment features: centre of gravity, distance between moment centre and geometrical centre;
- Histogram features: cross-correlation between histograms; maximum, number of non-empty bins, number of zero-crossings, energy and difference of positive and negative parts of difference histogram;
- Cross co-occurrence matrix [16] features: homogeneity, energy, contrast, symmetry and the first 4 moments of the matrix;
- Mutual information between the two temperature distributions;
- Fourier spectrum features: the difference maximum and distance of this maximum from the centre.

Each breast thermogram is thus described by 4 basic statistical features, 4 moment features, 8 histogram features, 8 cross co-occurrence features, mutual information and 2 Fourier descriptors. We further apply a Laplacian filter to enhance the contrast and calculate another subset of features (the 8 cross co-occurrence features together with mutual information and the 2 Fourier descriptors) from the resulting images, and consequently end up with a total of 38 features which describe the asymmetry between the two sides and which form the basis for the following pattern classification stage.

3 Imbalanced Classification

Many medical datasets are inherently imbalanced which leads to challenges for pattern recognition algorithms. A dataset is imbalanced if the classification categories are not (approximately) equally represented [17]. Conventionally, predictive accuracy is used to evaluate the performance of classifiers. However, this simple and intuitive measure is not appropriate when dealing with imbalanced data, as it will typically lead to a bias towards the majority class. Consequently, a classifier can display a poor recognition rate for the minority class (and hence, in medical context, a poor sensitivity), while at the same time achieving a high overall accuracy.

The disproportion in terms of number of samples from different classes in the training set is not the sole source of learning difficulties [18]. It has been shown, that if the number of minority samples is sufficient, the uneven class distribution itself does not cause a significant drop in recognition rate [19]. However, the uneven class distribution is usually accompanied by other difficulties such as class overlap, small sample size or small disjuncts in the minority class structure.

Various approaches have been suggested to address class imbalance. In the context of MCSs, which are based on the principle of combining the decisions of several base classifiers [6], they typically combine an MCS with one of the techniques dedicated to dealing with imbalanced data. SMOTEBagging [20] and SMOTEBoosting [21] are the most popular examples of a combination of oversampling and classifier ensembles, and are based on introducing new objects into each of the bags/boosting iterations separately using SMOTE [17]. IIvotes [22] fuses a rule-based ensemble, while EasyEnsemble [23] is a hierarchical MCS that utilises bagging as the primary learning scheme, but uses AdaBoost for each of the bags.

4 A Clustering and Selection Ensemble for Breast Thermogram Analysis

The method that we propose in this paper is based on the Clustering and Selection (CS) algorithm which consists of three main steps [12]:

1. Selecting individual classifiers for a pool.
2. Establishing clustering algorithm parameters and partitioning the learning set according to a given algorithm.
3. Selecting the best individual classifier for each cluster according to a given competence criterion.

A clustering algorithm is applied to partition the feature space by separating subsets of elements from the learning set based on their mutual similarity [24]. There is no restriction on allowing cluster borders to cross the borders separating areas with objects from particular classes. This is a desired effect to separate areas in which these classifiers achieve high classification performance.

In our approach, we propose to modify the CS algorithm in order to make it applicable to imbalanced classification problems. Standard classifiers typically have a bias towards the majority class, while predictors designed specifically to handle imbalanced classes often sacrifice specificity to improve sensitivity. The idea behind our proposed method lies in the intuition that the imbalanced classification task, caused by uneven object distribution, is not present in all parts of the decision space. Thus, by having a pool of classifiers comprising both standard models and predictors dedicated to imbalanced imbalanced, we can boost the minority class recognition rate in highly imbalanced clusters, while maintaining a satisfactory specificity by assigning canonical classifiers to parts of the decision space that are not imbalanced.

Assume that we have K base classifiers which are to be used for building an ensemble system

$$\Pi^{\Psi} = \{\Psi_1, \Psi_2, \ldots, \Psi_g, \ldots, \Psi_K\}, \tag{1}$$

of which g classifiers are, by design, biased towards the recognition of the minority class, while the remaining $K - g$ classifiers are designated to maintain the overall accuracy. One of the key issues for forming Π^{Ψ} is to maintain a good level of diversity, which can be ensured by using different classifier models. Classifiers are trained on all objects from the training set.

Clustering and Selection is based on the idea that exploiting local competencies of classifiers should lead to improved classification accuracy. For that purpose, the feature space is divided into a set of H competence areas (clusters)

$$\mathcal{X} = \bigcup_{h=1}^{H} \hat{X}_h, \tag{2}$$

where $\hat{X}_h$ is the h-th cluster and

$$\forall k, l \in \{1, \ldots, H\} \quad \text{and} \quad k \neq l \quad \hat{X}_k \cap \hat{X}_l = \emptyset. \tag{3}$$

In our approach, $\hat{X}_h$ is represented by the centroid [24]

$$C_h = [C_h^{(1)}, C_h^{(2)}, \ldots, C_h^{(d)}]^T \in \mathcal{X}, \tag{4}$$

and centroids are gathered in a set

$$C = \{C_1, C_2, \ldots, C_H\}. \tag{5}$$

An object x is assigned to the competence area whose centroid is closest to the object, i.e.

$$A(x, C) = \arg\min_{h=1}^{H} D(x, C_h), \tag{6}$$

where D is a distance measure (Euclidean distance in our approach).

The number of generated clusters plays an essential role for the performance of the system. At the same time, it is difficult to define strict rules on how to choose it since the decision is data-dependent. Consequently, it should be selected for each problem separately on the basis of experimental research or a priori knowledge.

For the classifier selection step we propose the approach that is detailed in Algorithm 1.

Algorithm 1. Classifier selection algorithm for partitioned feature space

Input:
$\mathbf{C} \rightarrow$ set of competence areas
$P_m \rightarrow$ set of classifiers dedicated to minority class recognition
$P_a \rightarrow$ set of classifiers dedicated to achieving high (overall) accuracy

Output:
$\mathbf{Q} \rightarrow$ pairs of competence areas and classifiers assigned to them

for all competence areas **do**
  detect the level of imbalance in a given area
  if no minority objects **then**
    measure the accuracy of classifiers from P_a over the objects in the cluster
    assign to this cluster a classifier with the highest competence
  else
    measure the sensitivity of classifiers from P_m over the objects in the cluster
    assign to this cluster a classifier with the highest competence
  end if
end for

return Q

5 Experimental Results

For our experiments, we use a dataset of 146 thermograms of which 29 cases have been confirmed as malignant, whereas the other 117 cases were benign [15]. For all thermograms, the 38 features from Section 2 are extracted.

Our employed CS ensemble consists of two classifiers dedicated to imbalanced data and two standard classifiers. For the former we use a cost-sensitive decision tree [25] and a C4.5 decision tree built with the SMOTE algorithm [17], while for the latter we utilise a standard C4.5 classifier and a support vector machine (SVM) with RBF kernel and parameter optimisation. For comparison, we implemented several state-of-the-art ensemble methods for imbalanced classification, namely SMOTEBoost [21], IIvotes [22] and EasyEnsemble [23], all with C4.5 decision trees as base classifiers. Additionally, we evaluate the individual performances of the classifiers in the CS pool.

Table 1. Classification results for all tested algorithms

	sensitivity	specificity	accuracy
C4.5	7.85	81.50	66.77
SVM	8.34	86.32	71.23
CSTree	70.16	80.05	78.07
SMOTETree	77.84	78.15	78.08
SMOTEBoost	79.03	91.00	88.62
IIvotes	79.56	**91.89**	89.44
EasyEnsemble	80.02	90.17	88.22
C&S Ensemble	**82.55**	**91.89**	**90.02**

A combined 5x2 CV F test [26], was carried out to assess the statistical significance of the obtained results. A classifier is assumed as statistically significantly better compared to another if one of the following is true:

- its sensitivity is statistically significantly better and its overall accuracy is not statistically significantly worse;
- its overall accuracy is statistically significantly better and its sensitivity is not statistically significantly worse.

The results of our experimental comparison are given in Table 1, which lists sensitivity (i.e. the probability that a case identified as malignant is indeed malignant), specificity (i.e. the probability that a case identified as benign is indeed benign) and overall classification accuracy (i.e. the percentage of correctly classified patterns) for each approach. In addition, we provide the results of the statistical significance test in Table 2.

From the results, we can see that our proposed modification of the Clustering and Selection algorithm returns a highly satisfactory performance, balancing excellent sensitivity with high specificity. Furthermore, our algorithm is shown

Table 2. Results of statistical significance test. A + signifies that the algorithm listed in this row statistically outperforms the algorithm listed in this column, a − indicates a statistically inferior performance.

	C4.5	SVM	CSTree	SMOTETree	SMOTEBoost	IIvotes	EasyEnsemble	C&S Ensemble
C4.5		−	−	−	−	−	−	−
SVM	+		−	−	−	−	−	−
CSTree	+	+		−	−	−	−	−
SMOTETree	+	+	+		−	−	−	−
SMOTEBoost	+	+	+	+			−	−
IIvotes	+	+	+	+				−
EasyEnsemble	+	+	+	+	+			−
C&S Ensemble	+	+	+	+	+	+	+	

to statistically outperform three state-of the-art ensembles that are dedicated to imbalanced classification. In all three cases, our approach provides a better sensitivity and a better overall classification accuracy. In fact, as we can see from Table 1, our method gives both the highest sensitivity and the highest specificity (tied with IIvotes) which is quite remarkable.

It is interesting to look at this performance in the light of the classification accuracies achieved by the individual base classifiers. Both C4.5 decision trees and SVMs give a good specificity but this is coupled with an unacceptably low sensitivity. CSTree and SMOTETree are able to boost sensitivity, but at the cost of specificity. On their own, none of the base algorithms returned results that would be considered impressive. However, when combining them into our CS Ensemble and because of the proposed classifier selection step, both sensitivity and specificity are boosted significantly, leading to an excellent overall performance. Also, as specificity/sensitivity increases when using more than one type of dedicated classifier, this signifies that the predictors have different areas of competence, and therefore that combining them allows us to create a diverse and mutually complementary ensemble.

6 Conclusions

In this paper, we have proposed an effective approach to analysing breast thermograms in the context of cancer diagnosis. Our approach extracts a set of image features quantifying asymmetries between the two breast areas in the thermogram, and utilises them in a pattern recognition stage. For classification, we employ an ensemble classifier that is rooted in the Clustering and Selection approach but is dedicated to addressing class imbalance. We do this by training different types of classifiers on different clusters so as to provide both high sensitivity and high specificity. That this leads to a highly successful classifier ensemble is demonstrated by our experimental results which show our approach not only to provide excellent classification performance but also to statistically outperform several state-of-the-art ensembles dedicated to address class imbalance.

References

1. Jones, B.F.: A reappraisal of infrared thermal image analysis for medicine. IEEE Trans. Medical Imaging 17(6), 1019–1027 (1998)
2. Anbar, N., Milescu, L., Naumov, A., Brown, C., Button, T., Carly, C., AlDulaimi, K.: Detection of cancerous breasts by dynamic area telethermometry. IEEE Engineering in Medicine and Biology Magazine 20(5), 80–91 (2001)
3. Head, J.F., Wang, F., Lipari, C.A., Elliott, R.L.: The important role of infrared imaging in breast cancer. IEEE Engineering in Medicine and Biology Magazine 19, 52–57 (2000)
4. Gautherie, M.: Thermobiological assessment of benign and maligant breast diseases. Am. J. Obstet. Gynecol. 147(8), 861–869 (1983)

5. Keyserlingk, J.R., Ahlgren, P.D., Yu, E., Belliveau, N., Yassa, M.: Functional infrared imaging of the breast. IEEE Engineering in Medicine and Biology Magazine 19(3), 30–41 (2000)
6. Kuncheva, L.I.: Combining pattern classifiers: Methods and algorithms. Wiley Interscience, New Jersey (2004)
7. Giacinto, G., Roli, F., Fumera, G.: Design of effective multiple classifier systems by clustering of classifiers. In: 15th Int. Conference on Pattern Recognition, vol. 2, pp. 160–163 (2000)
8. Marcialis, G.L., Roli, F.: Fusion of face recognition algorithms for video-based surveillance systems. In: Foresti, G., Regazzoni, C., Varshney, P. (eds.) Multisensor Surveillance Systems: The Fusion Perspective, pp. 235–250 (2003)
9. Rastrigin, L., Erenstein, R.H.: Method of Collective Recognition. Energoizdat, Moscow (1981)
10. Goebel, K., Yan, W.: Choosing classifiers for decision fusion. In: 7th Int. Conference on Information Fusion, pp. 563–568 (2004)
11. Ruta, D., Gabrys, B.: Classifier selection for majority voting. Information Fusion 6(1), 63–81 (2005)
12. Kuncheva, L.: Clustering-and-selection model for classifier combination. In: 4th Int. Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies, vol. 1, pp. 185–188 (2000)
13. Jackowski, K., Wozniak, M.: Algorithm of designing compound recognition system on the basis of combining classifiers with simultaneous splitting feature space into competence areas. Pattern Analysis and Applications 12(4), 415–425 (2009)
14. Qi, H., Snyder, W.E., Head, J.F., Elliott, R.L.: Detecting breast cancer from infrared images by asymmetry analysis. In: 22nd IEEE Int. Conference on Engineering in Medicine and Biology (2000)
15. Schaefer, G., Zavisek, M., Nakashima, T.: Thermography based breast cancer analysis using statistical features and fuzzy classification. Pattern Recognition 42(6), 1133–1137 (2009)
16. Zavisek, M., Drastich, A.: Thermogram classification in breast cancer detection. In: 3rd European Medical and Biological Engineering Conference, pp. 1727–1983 (2005)
17. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research 16, 321–357 (2002)
18. Sun, Y., Wong, A.K.C., Kamel, M.S.: Classification of imbalanced data: A review. International Journal of Pattern Recognition and Artificial Intelligence 23(4), 687–719 (2009)
19. Chen, X., Wasikowski, M.: Fast: A roc-based feature selection metric for small samples and imbalanced data classification problems. In: ACM SIGKDD Int. Conference on Knowledge Discovery and Data Mining, pp. 124–132 (2008)
20. Wang, S., Yao, X.: Diversity analysis on imbalanced data sets by using ensemble models. In: IEEE Symposium on Computational Intelligence and Data Mining, pp. 324–331 (2009)
21. Chawla, N.V., Lazarevic, A., Hall, L.O., Bowyer, K.W.: SMOTEBoost: improving prediction of the minority class in boosting. In: Lavrač, N., Gamberger, D., Todorovski, L., Blockeel, H. (eds.) PKDD 2003. LNCS (LNAI), vol. 2838, pp. 107–119. Springer, Heidelberg (2003)

22. Błaszczyński, J., Deckert, M., Stefanowski, J., Wilk, S.: Integrating selective pre-processing of imbalanced data with Ivotes ensemble. In: Szczuka, M., Kryszkiewicz, M., Ramanna, S., Jensen, R., Hu, Q. (eds.) RSCTC 2010. LNCS, vol. 6086, pp. 148–157. Springer, Heidelberg (2010)
23. Liu, X., Wu, J., Zhou, Z.: Exploratory undersampling for class-imbalance learning. IEEE Trans. Systems, Man and Cybernetics - Part B: Cybernetics 39(2), 539–550 (2009)
24. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM Comput. Surv. 31, 264–323 (1999)
25. Ling, C.X., Yang, Q., Wang, J., Zhang, S.: Decision trees with minimal costs. In: 21st Int. Conference on Machine Learning, pp. 544–551 (2004)
26. Alpaydin, E.: Combined 5 x 2 CV F test for comparing supervised classification learning algorithms. Neural Computation 11(8), 1885–1892 (1999)