Breast Thermogram Analysis Using Classifier Ensembles and Image Symmetry Features

Bartosz Krawczyk and Gerald Schaefer, Member, IEEE

Abstract—Breast cancer is the most commonly diagnosed form of cancer in women. Thermography, which uses cameras with sensitivities in the thermal infrared, has been shown to provide an interesting alternative to the standard modality of mammography for detecting breast cancer as it is able to detect smaller tumors and hence can lead to earlier diagnosis. In this paper, we present an approach to breast thermogram analysis that extracts features describing bilateral symmetries from an image and then utilizes a classifier ensemble for decision making. Importantly, our classification approach addresses the problem of imbalanced class distribution that is common in medical data analysis. We do this by constructing feature subspaces from balanced data subsets and train different classifiers on different subspaces. To combine the individual classifiers, we investigate two different strategies. The first dynamically assigns classifier weights based on an evolutionary algorithm, while the second uses a neural network for classifier fusion. Both approaches are shown to work well and to lead to significantly improved performance compared to canonical classification systems.

Index Terms—Breast cancer diagnosis, imbalanced classification, image features, multiple classifier system (MCS), pattern recognition, thermography.

I. INTRODUCTION

Breast cancer is the most commonly diagnosed form of cancer in women, accounting for about 30% of all cases [1]. Despite earlier less encouraging studies, which were based on low-capability and poorly calibrated equipment, thermal infrared imaging has been shown to be well suited for the task of detecting breast cancer [2], [3]. Thermography uses a camera with sensitivities in the thermal infrared to capture a temperature distribution of the human body or parts thereof. In contrast to other modalities such as mammography, it is a non-contact, passive, and radiation-free technique. The radiance from human skin is a 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 [4]. Asymmetrical temperature distributions, as well as hot or cold spots, are known to be strong indicators of an underlying dysfunction [5].

For breast cancer diagnosis, it has been shown that thermography has advantages over the standard modality of mammography, particularly when the tumor is in its early stages or in dense tissue. Early detection is crucial as it provides significantly higher chances of survival [6], and in this respect, infrared imaging outperforms the standard method of mammography. While mammography can detect tumors only once they exceed a certain size, even small tumors can be identified using thermography due to the high metabolic activity of cancer cells which leads to an increase in the local temperature that can be picked up in the infrared. Reportedly [7], the average tumor size undetected by mammography is 1.66 cm compared to only 1.28 cm by thermography.

In this paper, we present an effective approach for breast cancer detection from thermographic breast images. The basis of our approach is the characterization of image features that describe the bilateral symmetry between the two breast regions since tumors will lead to asymmetries between two sides. The image features that we employ include basic statistical features, histogram features, image moments, and various texture features.

The derived image features are then used in a pattern classification stage, and this is where the main contribution of this paper lies, particularly as it is well known that no single classification algorithm works best for all kinds of problems [8]. The reason for this is that each classifier typically has a different domain of competence. However, if we have a pool of different classifiers, the common incompetence area (i.e., the subset of feature space where all classifiers make incorrect decisions) is typically small as illustrated in Fig. 1.

It follows that we would wish to exploit the strengths of individual classifiers which can be beneficial for the overall
The application of a correlation-based feature selection stage
resulting in hot spots and a change in vascular pattern and hence
of asymmetry between the two sides, we adopted this concept
and use the cross-correlation between the two
ranges of the thermograms. In this paper, we construct normal-
left and the right breast.

II. BREAST THERMOGRAM FEATURE ANALYSIS

While thermograms for breast cancer diagnosis can be taken
based on frontal and/or lateral views, in this paper, we restrict
our attention to frontal view images. As has been shown [10],
affective approach to automatically detect cancer cases is to
study the symmetry between the left and the right breast. In the
case of cancer presence, the tumor will recruit blood vessels,
resulting in hot spots and a change in vascular pattern and hence
in an asymmetry between the temperature distributions of the
two breasts. On the other hand, symmetry typically identifies
healthy subjects.

We follow this approach and (manually) segment the areas
corresponding to the left and the right breast from the ther-
grams. Once segmented, we convert the breast regions to a polar
coordinate representation to simplify the calculation of several
of the features that we employ. A series of statistical features is
then calculated to provide indications of symmetry between the
regions of interest (i.e., the two breast areas) [11].

The simplest feature to describe the temperature distribution
captured in thermograms is to calculate its statistical mean. As
we are interested in symmetry features, we calculate the mean
for both breasts and use the absolute difference between the
two. Similarly, we calculate the standard temperature deviation
and use the absolute difference as a feature. Furthermore, we
employ the absolute differences of the median temperatures and
the 90-percentiles.

Image moments [12] are defined as

\[
m_{pq} = \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} x^p y^q g(x, y)
\]

where \(x\) and \(y\) define the pixel location and \(N\) and \(M\) define the
image size. We utilize moments \(m_{01}\) and \(m_{10}\) which essentially
describe the center of gravity of the breast regions as well as the
distance (both in the \(x\)- and \(y\)-directions) of the center of gravity
from the geometrical center of the breast. For all four features,
we calculate the absolute differences of the values between the
left and the right breast.

Histograms record the frequencies of certain temperature
ranges of the thermograms. In this paper, we construct normal-
ized histograms for both regions of interest (i.e., the left and
the right breast) and use the cross-correlation between the two
histograms as a feature. From the difference histogram (i.e.,
the difference between the two histograms), we compute the
absolute value of its maximum, the number of bins exceeding a
certain threshold (0.01), the number of zero crossings, energy,
and the difference of the positive and negative parts of the
histogram.

Co-occurrence matrices have been widely used in texture
recognition tasks [13] and can be defined as

\[
\gamma_{T_i, T_j}^{(k)} (I) = \text{PR}_{p_1 \leq I_{T_i}, p_2 \leq I_{T_j}} [p_2 \in I_{T_j}, |p_1 - p_2| = k]
\]

with

\[
|p_1 - p_2| = \max \{|x_1 - x_2|, |y_1 - y_2|\}
\]

where \(T_i\) and \(T_j\) denote two temperature values and \((x_k, y_k)\)
denote pixel locations. In other words, given a temperature \(T_i\) in
the thermogram, \(\gamma\) gives the probability that a pixel at distance
\(k\) away is of temperature \(T_j\). In order to arrive at an indication
of asymmetry between the two sides, we adopted this concept
and use the cross-co-occurrence matrix [14] defined as

\[
\gamma_{T_i, T_j}^{(k)} (I^1, I^2) = \text{PR}_{p_1 \leq I_{T_i}^1, p_2 \leq I_{T_j}^2} [p_2 \in I_{T_j}^2, |p_1 - p_2| = k]
\]

i.e., temperature values from one breast \((I^1)\) are related to the
temperatures of the second side \((I^2)\). From this matrix, we can

---

Fig. 2. Schematic of an MCS.

---

performance of the recognition system, leading to what is
known as a multiple classifier system (MCS), combined clas-
sifier, or classifier ensemble [9] and illustrated in Fig. 2.

Consequently, in this paper, we employ an MCS approach to
classifying the features derived from breast thermograms. We
address several crucial issues when designing classifier ensembles,
including how to select a pool of diverse and complementary
individual classifiers for the ensemble and how to design a
fusion method that optimally combines base classifiers.

Importantly, we also address the problem of class imbalance
that is frequently encountered in medical decision making due
to the relatively low number of malignant cases compared
to benign ones. Rather than reverting to cost matrices, we
approach this by constructing feature subspaces from balanced
data subsets. Each created subspace then forms the basis of an
individual classifier.

We present two methods for combining the results of these
classifiers, one that assigns weights derived using an evolution-
ary algorithm and a fuser that employs a neural network. The
results obtained on a large data set of breast thermograms prove
that our approach achieves high sensitivity without sacrificing
specificity while clearly outperforming canonical classifiers.
The application of a correlation-based feature selection stage
is also shown to further improve the results.

The remainder of this paper is organized as follows. Section II
describes the image features that we extract for each
breast thermogram. Section III gives the necessary background
on pattern classification, including the challenges of classify-
ing imbalanced data, while Section IV presents our proposed
approach. Section V gives experimental results, and Section VI
concludes this paper.
extract several features [13], and the ones that we are using are

\begin{align}
\text{Homogeneity } G &= \sum_k \sum_l \frac{\gamma_{k,l}}{1 + |k - l|} \\
\text{Energy } E &= \sum_k \sum_l \gamma_{k,l}^2 \\
\text{Contrast } C &= \sum_k \sum_l |k - l| \gamma_{k,l} \\
\text{Symmetry } S &= 1 - \sum_k \sum_l |\gamma_{k,l} - \gamma_{l,k}|.
\end{align}

We further calculate the first four moments \(m_1\) to \(m_4\) of the matrix

\[ m_p = \sum_k \sum_l (k - l)^p \gamma_{k,l}. \]

The mutual information \(MI\) between two distributions can be calculated from the joint entropy \(H\) of the distributions and is defined as

\[ MI = H_L + H_R + H \]

with

\[ H_L = - \sum_k P_L(k) \log_2 p_L(k) \]
\[ H_R = - \sum_l P_R(l) \log_2 p_R(l) \]
\[ H = \sum_k \sum_l P_{LR}(k, l) \log_2 p_{LR}(k, l) \]
\[ p_{LR}(k, l) = x_{k,l} / \sum_{k,l} x(k, l) \]
\[ p_L(k) = \sum_l p_{LR}(k, l) \]
\[ p_R(k) = \sum_k p_{LR}(k, l) \]

which we use as a further descriptor.

Finally, we calculate the Fourier spectrum and use the difference of absolute values of the image spectra. The features that we adopt are the difference maximum and the distance of this maximum from the center.

In summary, we characterize each breast thermogram using the following set of features: four basic statistical features, four moment features, eight histogram features, eight cross-cop-occurrence features, mutual information, and two Fourier descriptors. We further apply a Laplacian filter to enhance the contrast and calculate another subset of features (the eight cross-cop-occurrence features together with mutual information and the two Fourier descriptors) from the resulting images. In total, we hence compile a set of 38 descriptors per breast thermogram designed to describe the asymmetry between the two sides.

III. PATTERN CLASSIFICATION

Pattern classification aims at assigning an input object to one of a number of predefined categories (or classes). This is done on the basis of training data, consisting of learning examples, i.e., observations of features together with their correct classifications. In the context of this paper, the features are those described in Section II and are concatenated into a feature vector \(x \in X\). A pattern classification algorithm \(\Psi\) thus maps the feature space \(X\) to the set of class labels \(M\)

\[ \Psi : X \rightarrow M. \]

A. Multiple Classifier Systems

When creating MCSs, or classifier ensembles, the main efforts are concentrated on combining the outputs of elementary classifiers, a concept first presented in [15]. MCSs can improve the robustness and performance of the best base classifiers since they are able to exploit the unique strengths of the individual classifiers [9].

Let us assume that we have a pool of \(N\) classifiers \(\Psi^{(1)}, \Psi^{(2)}, \ldots, \Psi^{(N)}\). For a given feature vector \(x\), each of the individual classifiers makes a decision with respect to class \(i \in M = \{1, \ldots, M\}\). The combined classifier \(\Psi\) then makes a decision based on

\[ \bar{\Psi}(\Psi^{(1)}(x), \Psi^{(2)}(x), \ldots, \Psi^{(N)}(x)) = \arg \max_{\ell \in M} \sum_{l=1}^N \delta(j, \Psi^{(l)}(x))w^{(l)}\Psi^{(l)}(x) \]

where

\[ \delta(j, i) = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \]

and \(w^{(l)}\) is the weight assigned to the \(l\)-th classifier.

The weights used in (14) play a key role in establishing the quality of \(\bar{\Psi}\). Much research has been dedicated to weight configurations. If all weights are set to 1, then (14) becomes the well-known majority voting rule. In our presented approach, a dynamic method of weight assignment is used, instead of the conventional static one proposed in [9].

B. Classification of Imbalanced Data Sets

A data set is imbalanced if the classification categories are not (approximately) equally represented [16]. Often, real-world data sets are predominately composed of “normal” examples with only a small percentage of “abnormal” cases as depicted in Fig. 3.

The performance of classification algorithms is usually evaluated using predictive accuracy. However, this is not appropriate when the data are imbalanced as disproportions in the number of objects between the classes may lead to severe deterioration of the classification accuracy for the minority class.
Fig. 3. Example of imbalanced data set.

In medical diagnosis, there are typically far fewer malignant cases than there are benign ones, and consequently, conventional classification approaches often suffer from low sensitivity due to the skewed class distribution. In our application, we have a data set of 146 breast thermograms of which only 29 are malignant cases and the remaining 117 are benign; that is, there are roughly four times as many benign samples than there are malignant ones.

Several approaches have been proposed to deal with imbalanced data sets. One of the most popular is SMOTE [16], which is based on the oversampling of the minority class, and its variants [17], [18]. However, in cases of non-Euclidean or non-metric data, these do not work well, and a ghost point approach was developed in [19] for such data. Neural networks were also adapted for this problem with a new dynamic threshold learning algorithm [20]. In the context of classifier ensembles dedicated to imbalanced classification [21], an approach consisting of boosting support vector machines (SVMs) was presented in [22], while an effective method of combining MCSs with oversampling was presented in [23]. Another approach, using cost-sensitive ensembles, was introduced in [24].

IV. PROPOSED METHOD

In this paper, we present an effective MCS for the diagnosis of breast cancer from thermograms. In particular, we base our classification on the image symmetry features described in Section II while focusing on an alternative to class oversampling and cost-sensitive classification.

Instead of trying to find an optimal single classifier, we combine a set of weaker learners to obtain improved overall accuracy. Of course, to achieve this, the ensemble should be created from classifiers with good individual accuracy and diversity at the same time. There are several approaches to achieve the diversity of a base classifier pool. For example, one could use different partitions of a data set or generate various data sets by data splitting so that different base classifiers are trained on different inputs and hence complementary.

Our approach is based on the idea of object space partitioning where each classifier is trained on a different subspace. Crucially, subspaces are constructed in a way to counter the original imbalance in the data set. In addition, we perform correlation-based feature selection to boost the classification performance. For combining the outputs of the base classifiers, we investigate two different approaches, one that employs an evolutionary algorithm to assign the weights and one that is based on a neural network.

A. Space Partitioning

As mentioned, for our particular case, the data set is imbalanced so that there are approximately four times more malignant thermograms than benign ones. Using classical approaches, the majority class is typically identified well (as it has sufficient training instances to learn from) while classification for the minority class (and hence sensitivity) is often poor.

Our approach to address this problem is based on object space division. We create a number of subspaces, and a pool of classifiers \( \Pi^\Psi = \{\Psi^{(1)}, \Psi^{(2)}, \ldots, \Psi^{(N)}\} \) is then constructed by training a single classifier on each of the subspaces.

We use space partitioning to balance the unfavorable distribution of classes using a random undersampling method. This means that each of the newly created subspaces contains a smaller number of objects, randomly drawn from the main data set, and is created in such a way that the number of objects in each of the subspaces is equal for both classes. The choice of objects from the malignant class can be determined in two ways:

1) that each subspace contains all of the objects from the malignant class;
2) that each subspace contains only some randomly selected samples from the malignant class.

Moreover, we will compare these two methods in our experimental evaluation.

An illustration of the proposed method for an example of eight objects and creation of three subspaces is given in Fig. 4. As can be seen, each of the subspaces has a balanced data distribution using all minority class samples and an equal number of randomly drawn objects from the majority class.

Objects of the minority class are randomly sampled and removed from the training set. Subspaces are then created as long as there are objects in the majority set. In this way, we eliminate the imbalanced class distribution which, in turn, should lead to an improvement of the recognition rate for the minority class.

B. Feature Selection

Feature selection algorithms transform the original feature space \( X \) into a smaller subspace \( X' \) which ideally should consist only of relevant features. Feature selection is acknowledged as a very important yet nontrivial step in pattern recognition and is employed for a variety of reasons, including improving the overall accuracy, speeding up the training process, and reducing the computational complexity of the data model [25].

In our approach, we use a feature selection algorithm to select features independently for each of the created subspaces. Therefore, in each of the subspaces, the feature subsets may vary, leading to an increased overall diversity of the classifier pool and, consequently, to a better ensemble. In particular, we employ the fast correlation-based feature filter (FCBF) [26] due to its typically good performance and low computational time. In FCBF, the relations between features and classes and between pairs of features are considered. The algorithm proceeds at two levels. First, a ranking algorithm based on the symmetric uncertainty coefficient index is used to estimate
class–feature relevance, and a threshold is used to select pre-
dominant features. In the second part, features that are re-
dundant are removed.

C. Classifier Fusion

Classifier fusion is an important aspect of classifier ensem-
bles, and the choice of fusion method, which is responsible
for the collective decision making process, is hence crucial.
Fusion methods can, in general, be divided into two main
groups: voting approaches which make decisions on the basis
of outputs of individual classifiers [27], and methods which
construct discriminant functions on the basis of continuous
outputs (supports) of individual classifiers [28].

Instead of using traditional approaches such as majority
voting or static weight assignment, in this paper, we implement
dynamic approaches to combine the outputs of base classifiers
created on different object subspaces. In particular, we present
dynamic approaches to combine the outputs of base classifiers
of all the ensemble under consideration, which consists of weights as-
adjusted model of chromosomes, fitness function, mutation,
and crossover operators. A chromosome $Ch$ represents a vector of proposed values for the weight function
of a neural network as a fuser. In both cases, we train the fuser
weights using an evolutionary algorithm and one that employs
two different methods, one which derives the values of the
outputs of individual classifiers [28], and methods which
leads to an increase of the overall accuracy [29].

A selection of individuals from the

5) Crossover: The crossover operator generates one off-
spring member from two parents following the two-point
rule.

6) Population update: A selection of individuals from the
population is formed by merging the descendant population
and a set of individuals created by mutation and crossover. The probability of selection $P_{s}$ of a particular
individual is proportional to the value of its fitness. We
employ a tournament selection scheme.

The control parameters of the algorithm are as follows:

1) $N_{c}$: the upper limit of generations;
2) $N_{p}$: the number of individuals in the population;
3) $\beta$: the mutation probability;
4) $\gamma$: the crossover probability;
5) $\Delta_{m}$: the mutation range factor;
6) $V$: the upper limit of algorithm iterations without quality
improvement.

2) Neural Fuser: Our second approach uses a one-layer per-
ceptron as a trained fuser for the classifier ensemble, originally
proposed in [28]. For this approach, all base classifiers must
make decisions based on the values of discriminant functions.
Let $F^{(l)}(i, x)$ denote such a function for the $i$-th class with a
given value of $x$, which is used by the $l$-th base classifier from
the pool $\Pi^{\Psi}$. The compound classifier $\Psi$ makes a decision on the basis of

$$
\Psi(x) = i \Leftrightarrow \hat{F}(i, x) = \max_{k \in M} \hat{F}_{k \in M}(k, x)
$$

(16)

where

$$
\hat{F}(i, x) = \sum_{l=1}^{N} w^{(l)} F^{(l)}(i, x) \quad \text{and} \quad \sum_{l=1}^{N} w^{(l)} = 1.
$$

(17)

Based on a training process, the fuser needs to iden-
ty $W = \{W_{1}, W_{2}, \ldots, W_{N}\}$, which consists of weights as-
signed to each classifier and each of the $M$ classes, $W_{i} =
[w^{(l)}(1), w^{(l)}(2), \ldots, w^{(l)}(M)], T$.

The aim is to find a fuser which assures the lowest misclas-
sification rate of $\Psi$ which is solved using a neural network
with a canonical learning approach. The idea of such a neural
fuser is presented in Fig. 5. One perceptron fuser is constructed
for each of the classes under consideration. The perceptron
may be trained with any standard procedure used in neural

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Fig_4.png}
\caption{Example of class rebalancing by creating subspaces.}
\end{figure}
network learning, and the input weights established during the learning process are then the weights assigned to each of the base classifiers.

V. EXPERIMENTAL RESULTS

As mentioned, we use a data set of 146 thermograms of which 29 cases have been confirmed as malignant whereas the other 117 cases were benign. This is the same data set that was used in earlier work [14], [11] and is significantly larger than those used in other studies (e.g., [10]). For all thermogram pairs, we extract the 38 features described in Section II.

As base classifier, we use SVMs [30] implemented in the kernlab package [31], particularly its ksvm() function [32], which includes the C-SVM classification algorithm as well as a bound constraint version of C classification, C-BSVM. As kernel, we chose a Gaussian RBF, and we performed classifier tuning [31] to obtain optimal parameters for our setup which lead to $\sigma = 0.1$ and $C = 10$.

For our approach, we examined the proposed method with and without feature selection, with weights set according to (14), with weights established by the evolutionary algorithm (see Section IV-C1), and with weights obtained from the neural network fuser (see Section IV-C2), respectively. For the genetic and neural fusers, which require a training step, we utilized the same training set that was used for the classifier training step. For the evolutionary weight optimization, we used the following settings: $N_e = 200$, $N_p = 100$, $\beta = 0.7$, $\gamma = 0.3$, $\Delta_{nn} = 0.2$, and $V = 20$.

In order to put the obtained results into context, we have performed classification using several canonical classifiers, namely, a standard single SVM, bagged SVM [33], boosted SVM [34], and a random forest classifier [35]. Furthermore, we implemented two recent approaches of ensemble algorithms for dealing with imbalanced data, SMOTEBagging [36] and SMOTEBoost [23].

First, for our approach, we wanted to identify the optimal number of malignant class samples in a subspace. For this, all possible (and sensible) combinations were considered, ranging from 10 objects (fewer would lead to the creation of too many subspaces) to 28 objects (i.e., all malignant cases except one for testing). Since our main interest is in achieving a high prediction accuracy for malignant cases, we compare the different set sizes based on this criterion. As the validation procedure for assessing the optimal number of objects from the minority class in the subspace, we used leave-one-out cross-validation. Fig. 6 presents the individual accuracy based on different cardinalities of malignant objects in a subspace.

As we can see from there, the best results were achieved for the maximum number (i.e., 28) of malignant objects in each of the subspaces. Therefore, in the following experiments, we used this setting, and subspaces comprise the maximum possible (i.e., available in a given fold) number of malignant objects and an equal number of benign cases. The number of classifiers in the proposed ensemble method varies between the folds, as it is dependent on the number of objects in a given fold (subspaces corresponding to classifiers are created as long as there are objects from the majority class unassigned to any subspace). For the examined data set, the number of base classifiers in the ensemble pool varied from three to seven.

We used a statistical significance test, namely, a combined $5 \times 2$ CV F test [37], to compare the individual accuracies and judge if their differences are statistically significant with respect to sensitivity. The process of subspace creation and feature selection were repeated independently for each of the folds. The results are given in Table I for all classifiers. For each approach, we report 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).

The results from Table I allow us to draw several interesting conclusions. The proposed feature set provides a good basis for the subsequent classification. As discussed in Section II, we have extracted a wide range of features to provide a diverse input for the classification algorithm. Employing feature selection, as described in Section IV-B, allows to discard redundant or unhelpful features while promoting those that have high discriminative power. Each of the base classifiers was built based on a subset of different features, which assures initial diversity among the pool of predictors and hence leads to improved ensemble quality.

While the presented method did not significantly outperform the canonical approaches when considering the overall accuracy, as mentioned, this is not an appropriate metric for the
investigated problem. More importantly, we can notice clearly improved sensitivity results which confirms that our approach is able to correctly identify the more important (i.e., malignant) cases well. In fact, our best performance is more than three times higher than the best canonical approach, demonstrating the strength of our proposed method. On the other hand, the results for the benign class drop somewhat which is, however, more than compensated by the classification performance on malignant samples. Using the neural fuser combined with feature selection results in a clear increase of all of the accuracies, including those of our earlier approach [38] (which essentially corresponds to the SPLIT method of Table I).

Also interesting is the comparison with SMOTEBoosting and SMOTEBoost. In comparison to these, which are also ensembles dedicated to imbalanced data, our method is shown to provide a small but statistically significant improvement. Our proposed subspace splitting method does not require any form of oversampling, on which the other two ensembles are based, and hence eliminates the need of creating too many artificial objects, which, in turn, may lead to a shift of class distribution or to overfitting.

Our presented approach also clearly outperforms previous approaches in the literature of breast thermogram analysis. In [11], the same image features as in this paper were employed together with a cost-sensitive fuzzy if-then rule-based classifier which was optimized by a genetic algorithm. The best results reported there were a sensitivity of 79.86% with a specificity of 79.49%. Similar results, on the same data set and using the same features, were achieved in [14] based on a neural network classifier coupled with a feature selection stage, while an overall classification accuracy of 79.52% was achieved using an ant colony optimization classifier in [39] (which essentially corresponds to the SPLIT method of Table I).

VI. Conclusion

The early identification of breast cancer leads to a crucial increase in survival rate and makes thermal medical imaging a modality of interest and a competitive alternative to mammography. In this paper, we have presented an effective approach for medical decision making from breast thermograms. Based on a set of asymmetry features, we have developed an MCS that is able to deal appropriately with imbalanced data sets.

In particular, we train the base classifiers of the ensemble on different object subspaces where, in each subspace, a balance between the classes is maintained. The decisions of the individual classifiers are then fused into an ensemble, and we have shown that a neural network fusion approach leads to excellent classification performance on a data set of about 150 thermograms while maintaining high sensitivity. It gives better results compared to a second fusion strategy based on evolutionary computing, while both techniques clearly outperform various canonical classification algorithms as well as previous approaches of breast thermogram analysis from the literature. We have also shown that adding a feature selection stage leads to a further improvement.

REFERENCES


Bartosz Krawczyk received the B.Sc. Engineering degree in computer science and the M.Sc. degree (both with distinctions) from the Wroclaw University of Technology, Wroclaw, Poland, and was recognized as the best M.Sc. graduate by the rector of the university. He is currently working toward the Ph.D. degree in the Department of Systems and Computer Networks, Wroclaw University of Technology.

His research is focused on machine learning, multiple classifier systems, one-class classifiers, and interdisciplinary applications of these methods. He has published more than 45 papers in international journals and conferences. Mr. Krawczyk is a Chairman of the Symposium Advances in Artificial Intelligence and Applications and has served on the program committees of over 20 conferences.

Gerald Schaefer (M’04) received the Ph.D. degree in computer vision from the University of East Anglia, Norwich, U.K.

He worked at the Colour and Imaging Institute, University of Derby, Derby, U.K. (1997–1999), in the School of Information Systems, University of East Anglia, Norwich (2000–2001), in the School of Computing and Informatics, Nottingham Trent University, Nottingham, U.K. (2001–2006), and in the School of Engineering and Applied Science, Aston University, Birmingham, U.K. (2006–2009), before joining the Department of Computer Science, Loughborough University, Loughborough, U.K., where he leads the Vision, Imaging and Autonomous Systems Research Division. His research interests are mainly in the areas of color image analysis, image retrieval, physics-based vision, medical imaging, and computational intelligence. He has published extensively in these areas with a total publication count exceeding 350.

Dr. Schaefer has been invited as keynote or tutorial speaker to numerous conferences, is the organizer of several international workshops and special sessions at conferences, and is the editor of several books, conference proceedings, and special journal issues.