REGULAR RESEARCH PAPER

An ensemble classification approach for melanoma diagnosis

Gerald Schaefer · **Bartosz Krawczyk** ·
M. Emre Celebi · **Hitoshi Iyatomi**

Abstract Malignant melanoma is the deadliest form of
skin cancer, and has, among cancer types, one of the most
rapidly increasing incidence rates in the world. Early diag-
nosis is crucial, since if detected early, its cure is simple.
In this paper, we present an effective approach to melanoma
identification from dermoscopic images of skin lesions based
on ensemble classification. First, we perform automatic bor-
der detection to segment the lesion from the background
skin. Based on the extracted border, we extract a series of
colour, texture and shape features. The derived features are
then employed in a pattern classification stage for which we
employ a novel, dedicated ensemble learning approach to
address the class imbalance in the training data and to yield
improved classification performance. Our classifier com-
mittee trains individual classifiers on balanced subspaces,
removes redundant predictors based on a diversity measure
and combines the remaining classifiers using a neural net-
work fuser. Experimental results on a large dataset of der-
moscopic skin lesion images show our approach to work
well, to provide both high sensitivity and specificity, and our
presented classifier ensemble to lead to statistically better
recognition performance compared to other dedicated clas-
sification algorithms.

1 Introduction

Malignant melanoma is the deadliest form of skin can-
cer, and has, among cancer types, one of the most rapidly
increasing incidence rates in the world. For example, in the
United States 76,690 cases and 9,480 deaths are predicted
for 2013 alone [24]. Early diagnosis of melanoma is partic-
ularly important since melanoma can be cured with a simple
excision if detected early.

Dermoscopy has become one of the most important tools
in diagnosing melanoma and other pigmented skin lesions. It
is a non-invasive skin imaging technique that involves optical
magnification, along with optics to minimise surface reflec-
tion, making subsurface structures more easily visible when
compared to conventional clinical images [2]. This in turn
reduces screening errors and provides greater differentia-
tion between difficult lesions such as pigmented Spitz nevi
and small, clinically equivocal lesions [25]. However, der-
moscopy may also lower the diagnostic accuracy in the hands
of inexperienced dermatologists [3]. Thus, in order to min-
imise diagnostic errors resulting from the difficulty and sub-
jectivity of visual interpretation, computerised image analy-
sis techniques are highly sought after [11].

Computer-aided approaches to diagnose melanoma typi-
cally proceed in three main stages: border detection, feature
extraction and classification [6]. In this paper, we also follow
this strategy, but pay particular attention to the final stage,

G. Schaefer (✉)
Department of Computer Science, Loughborough University,
Loughborough, UK
e-mail: gerald.schaefer@ieee.org

B. Krawczyk
Department of Systems and Computer Networks, Wroclaw
University of Technology, Wroclaw, Poland

M. E. Celebi
Department of Computer Science, Louisiana State University,
Shreveport, USA

H. Iyatomi
Department of Applied Informatics, Hosei University, Tokyo, Japan

i.e. the pattern classification task. We first perform automatic border detection using a variant of the JSEG image segmentation algorithm. Based on the obtained lesion definition, we then extract a set of shape features, while colour and texture features are calculated based on a division of the image into clinically significant regions using a Euclidean distance transform.

The obtained features are then employed in a pattern classification stage. For this, typically a (single) classifier is trained and applied. In this paper however, in order to provide improved and more robust performance, we use a multiple classifier system rather than relying on a single predictor. Our proposed novel ensemble approach to classification is carefully constructed for this purpose. In particular, we address the present class imbalance which stems from the fact that far fewer malignant samples are available for training compared to benign cases. We do this by training individual classifiers on subspaces that are balanced, and combine their decisions using a neural network fuser. Redundant classifiers are removed, based on a fuzzy diversity measure, to arrive at a less complex classification system that at the same time delivers improved recognition performance. The selected classifiers are then combined using a trained fuser implemented using neural networks. Experimental results on a large dataset of 564 skin lesion images show our approach to work well, giving a sensitivity of 93.76 % coupled with a specificity of 93.84 %, and also confirm our approach to statistically outperform several state-of-the-art ensemble classifiers dedicated to imbalanced classification.

The remainder of the paper is organised as follows. Section 2 explains how we segment the lesion from the background and defines the features that we extract. Section 3 then presents in detail our ensemble approach for skin lesion classification and hence melanoma identification. Section 4 gives experimental results, while Sect. 5 concludes the paper.

2 Segmentation and feature extraction

Automated border detection is typically the first step in the automated analysis of dermoscopy images [7] and is crucial for two main reasons. First, the border structure provides important information for accurate diagnosis, since clinical features, such as asymmetry and border irregularity, can be calculated directly from the border. Second, the extraction of other important clinical features such as atypical pigment networks, globules, and blue-white areas, critically depends on the accuracy of border detection.

In our approach, we perform automated border detection using the technique from [5] which in turn is based on the JSEG algorithm [10]. Following a pre-processing step to smooth the image and a colour quantisation stage, the image is thresholded to arrive at an approximate outline of

the lesion. This is further refined using region growing on a local homogeneity channel and colour-based region merging. Finally, in a post-processing step, background regions and isolated areas are removed and the remaining regions merged to give the final segmentation.

From the segmented lesion area, we then extract a series of features including shape, colour and texture features, where some of the colour and texture features are calculated based on the definition of three significant image regions—lesion, inner and outer periphery—which are obtained using a Euclidean distance transform. In particular, the following descriptors are extracted (for details see [6]):

- *Shape features:* lesion area; aspect ratio of lesion; two asymmetry features; compactness; maximum lesion diameter; eccentricity; solidity; equivalent diameter; rectangularity and elongation of the object-oriented bounding box.
- *Colour features:* mean and standard deviation of each channel in RGB, rgb, HSV, l1l2l3 and CIEL*u*v* colour spaces; ratios and differences of mean and standard deviation from the different image regions for all colour spaces; two colour asymmetry features each for R, G, and B channels; centroidal distances for each channel of all colour spaces; CIEL*u*v* L_1 and L_2 histogram distances between the different image regions.
- *Texture features:* maximum probability, energy, entropy, dissimilarity, contrast, inverse difference, inverse difference moment, and correlation of the normalised gray-level co-occurrence matrix [13] (averages over the four major orientations); ratios and differences of the same co-occurrences features from the different image regions.

In total, we end up with 11 shape features, 354 colour features and 72 texture features for each image.

3 Melanoma classification

3.1 Pattern classification

A pattern classification algorithm Ψ maps the feature space $\mathcal{X}$ to the set of class labels $\mathcal{M}$. Typically, for melanoma analysis a single classifier is employed for this decision making stage.

In contrast, in this paper we propose the use of a multiple classifier system (MCS), also referred to as ensemble classifiers, for enhanced and more robust classification. MCSs can improve the performance of the best base classifier, since they can exploit the strengths and eliminate the weaknesses of the individual classifiers [19] as illustrated for a toy problem in Fig. 1.

Assume that we have n classifiers $\{\Psi^{(1)}, \Psi^{(2)}, \ldots, \Psi^{(n)}\}$. For a given object $x \in \mathcal{X}$, each individual classifier makes
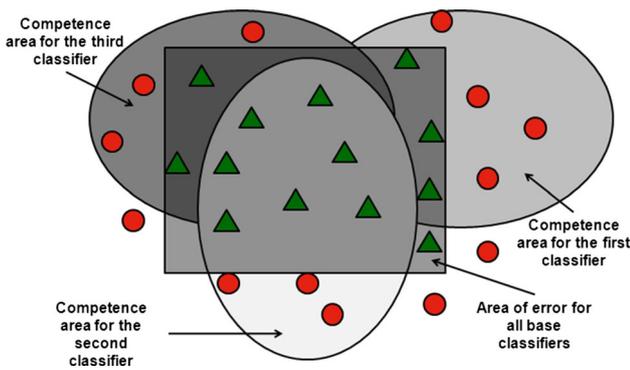
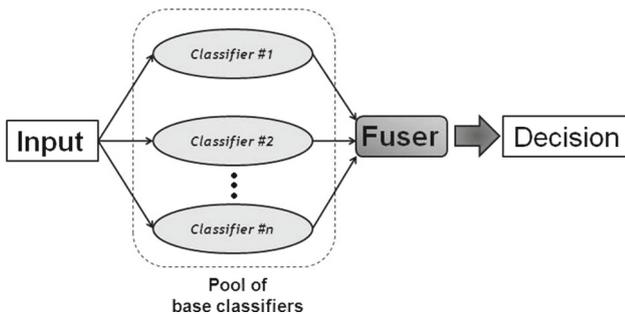**Fig. 1** Decision areas of three different classifiers for a dichotomy toy problem

Fig. 2 Schematic of a multiple classifier system

Fig. 3 Example of bias towards the majority class in linear classification of an imbalanced problem. The established decision boundary (*line*) would give poor prediction for minority class samples

a decision for class $i \in \mathcal{M} = \{1, \ldots, M\}$ based on the values of discriminants. Let $F^{(l)}(i, x)$ denote a function that is assigned to class i for a given value of x, and that is used by the l-th classifier $\Psi^{(l)}$. The combined ensemble classifier Ψ, illustrated in Fig. 2, makes a decision based on [29]

$$\Psi(x) = i \quad \text{if} \quad \hat{F}(i, x) = \max_{k \in M} \hat{F}(k, x), \tag{1}$$

where

$$\hat{F}(i, x) = \sum_{l=1}^{n} w^{(l)} F^{(l)}(i, x) \quad \text{and} \quad \sum_{i=1}^{n} w^{(l)}(i) = 1. \tag{2}$$

The weights can be set dependent on the classifier and class number: weight $w^{(l)}(i)$ is assigned to the l-th classifier and the i-th class.

The novel ensemble approach to classification that we present in this paper is carefully constructed and adapted to the problem at hand in order to lead to improved performance not only compared to single predictors but also in comparison with other MCSs from the literature.

3.2 Imbalanced classification

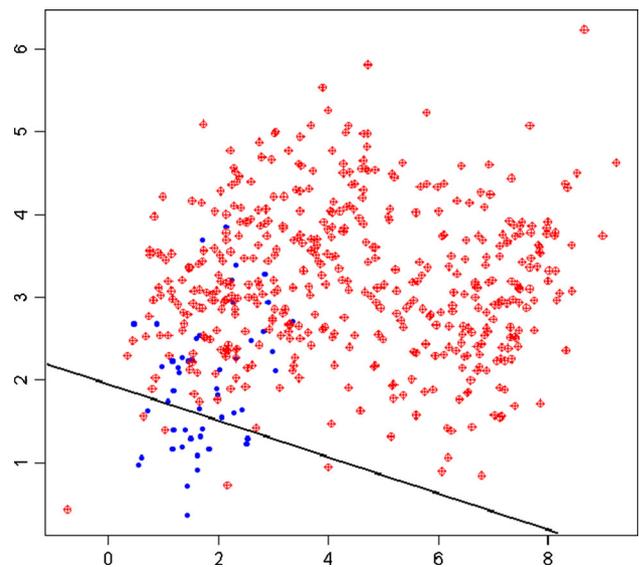In medical decision making problems such as the one we are addressing in this paper, datasets are often predominantly composed of "normal" or benign examples with only a small percentage of "abnormal" or malignant cases. This class imbalance often presents a challenge for classification algorithms [8]. Since the performance of classifiers is typically evaluated and tuned based on overall classification accuracy, this leads to predictors that are biased towards the majority class as illustrated in Fig. 3, while in medical diagnosis it is clearly the minority (malignant) class that is of higher importance.

Techniques that address the problems associated with imbalanced datasets can in general be divided into data level approaches, classifier level approaches and cost-sensitive approaches. Data level approaches work, in a pre-processing stage, directly on the data space, and attempt to re-balance the class distributions, often through oversampling or undersampling. Oversampling methods [8] however may also lead to other problems, such as class distribution shift when running too many iterations (since new artificial objects are being created on the basis of previously introduced samples). Classifier level approaches try to adapt existing algorithms to the problem of imbalanced datasets and bias them towards favouring the minority class. One possibility is to perform one-class classification, which can learn the concepts of the minority class by treating majority class objects as outliers [15]. Cost-sensitive approaches can use both data modifications (by adding a specified cost to the misclassification) and modifications of the learning algorithms (to adapt them to the possibility of misclassification) [23]. Here, a higher misclassification cost is assigned for minority class objects and classification performed so as to reduce the overall learning cost.

Ensemble classifiers addressing class imbalance typically combine an MCS algorithm with one of the above techniques. Examples of a combination of oversampling and classifier ensembles are SMOTEBagging [28] and SMOTEBoost [9] which introduce new objects into each of the bags/boosting iterations separately. In contrast, in Underbagging [21] base classifiers are trained on downsampled data sets to overcome any existing class imbalance. IIvotes [4] is an approach which fuses a rule-based ensemble with a SPIDER pre-processing scheme so as to be more robust with respect to atypical data distributions in minority classes and to automatically find an optimal number of bags. A fusion of MCSs and one-class classifiers constructed with respect to maintaining their diversity has also been shown to be effective for imbalanced classification [17]. Cost-sensitive MCSs are mostly based on adjusting the object weights in a boosting scheme [26], although ensembles based on cost-sensitive decision trees have also been exploited [18]. EasyEnsemble [20] uses bagging as the main concept; since for each of the bags AdaBoost is employed as the base model, it can be viewed as an ensemble of ensembles.

3.3 Ensemble classification of skin lesion features

In our ensemble approach, we address class imbalance by training classifiers on balanced object subspaces [16]. We thus construct an MCS that is dedicated to imbalanced classification, and proceed in four main steps:

1. Creation of a number of subspaces consisting of minority class and under-sampled majority class objects.
2. Construction of a pool of classifiers by training a single classifier on each of the subspaces. Optionally, a feature selection algorithm can be employed in this stage which is applied independently for each of the subspaces.
3. Diversity-based pruning of a pool of classifiers to select complementary models for the committee.
4. Trained fusion of the outputs of the elementary classifiers.

In the following we explain these steps in detail.

3.3.1 Space partitioning

Using classical approaches, the majority class is typically identified well, while classification for the minority class is often poor. In our approach, we address this problem by object space division where we first construct a number of balanced subspaces and then train one base classifier on each subspace to create a pool of predictors $\Pi^\Psi = \{\Psi^{(1)}, \Psi^{(2)}, \ldots, \Psi^{(n)}\}$.

We use space partitioning to balance the unfavourable class distribution using a random undersampling method. Each of the newly created subspaces contains a smaller number of objects, randomly drawn from the dataset, so that the number of objects in each of the subspaces is equal for both classes. 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.

3.3.2 Classifier construction

Each of the generated subspaces forms the basis of one of the classifiers in the committee. While in principle any classification algorithm can be used as base classifier, in this paper we utilise support vector machines (SVMs) [27]. Each SVM uses a polynomial kernel and we perform classifier tuning [14] to obtain optimal parameters for the degree of the polynomial (in the range $[1; 6]$) and the cost parameter (in the range $[0; 10]$).

In addition, we perform feature selection for each of the subspaces. Therefore, in each of the subspaces the derived feature subsets may vary, leading to an increased overall diversity of the pool of classifiers. We employ the fast correlation-based feature filter (FCBF) [30], which considers the relations between features-classes and between pairs of features. The algorithm first uses a ranking algorithm based on the symmetric uncertainty coefficient index to estimate class-feature relevance, and to identify a threshold coefficient for selecting predominant features. In the second part, features that are redundant to the predominant features are removed.

3.3.3 Classifier selection

Different base classifiers will have different areas of competence and hence may provide different contributions to the committee. Therefore, careful classifier selection should be conducted in order to choose the most valuable individual models. This can be performed based on the ensemble's diversity as a decision criterion, so as to choose predictors that are as different as possible from each other. This is motivated by the fact, that adding similar classifiers to the committee does not improve its quality but only increases its complexity. On the other hand, diverse models might be mutually supplementary and hence allow to exploit different areas of competence.

In our approach, we do this based on a non-pairwise (global) measure of diversity, in particular a fuzzy Shannon diversity measure [12]. Assuming that there are n classifiers in the pool, of which s classifiers correctly classify a given training object x_j, we can define a fuzzy membership function $\mu_{x_j} = \frac{s}{n}$ for a given object, with $0 \leq \mu_{x_j} \leq 1$. The obtained membership function is given to a Shannon function to measure its fuzziness and thus acts as a diversity measure

$$DIV_S(\Pi^\Psi) = \left\{(x_j, \mu_{x_j}) | x_j \in X\right\} \rightarrow \sum_j \left\{-\mu_{x_j} \ln(\mu_{x_j})\right.$$
$$\left. -(1 - \mu_{x_j}) \ln(1 - \mu_{x_j})\right\}. \tag{3}$$

The measure gives a value in the interval [0, 1], where 0 corresponds to identical classifiers and 1 to the highest possible diversity respectively.

Diversity-based pruning is achieved via an exhaustive search over all possible combinations of committee members to identify the ensemble with maximal diversity.

3.3.4 Classifier fusion

Classifier fusion is an important aspect of classifier ensembles, and the choice of fusion method, which is responsible for the collective decision making process, is hence crucial. In our approach, we make decisions based on discriminant functions as indicated in Eq. (2) with the fusion method determining the weights $w^{(l)}(i)$ in there.

The trained fuser we employ is a neural fuser implemented as a one-layer perceptron [29] as illustrated in Fig. 4. The values of support functions given by each of the base classifiers serve as input, while the output is the weighted support for each of the classes. One perceptron fuser is constructed for each of the classes, and may be trained with any standard procedure used in neural network learning (we use the Quickprop algorithm). The input weights established during the learning process are then the weights assigned to each of the base classifiers.

4 Experimental results

In our experiments, we use a dataset of 564 skin lesion images which was originally introduced in [6]. The samples stem from three university hospitals (University of Graz, University of Naples and University of Florence) [2] and from the Sydney Melanoma Unit [22]. All images are true-colour images, with a typical resolution of 768×512 pixels, from which the 437 features from Sect. 2 are extracted. Of the 564 cases, 88 are melanoma while the remaining 476 are benign, justifying our dedicated approach to address class imbalance.

In order to put the obtained results into context, we have also performed classification using a single SVM (of the same type as used in our ensemble) and an SVM combined with SMOTE [8] to address class imbalance.

For our proposed ensemble, we evaluate four different configurations to be able to judge the significance of the various components of our algorithm:

- the ensemble without feature selection, without pruning and using majority voting instead of the proposed neural fuser (noFS,noPR,MV);
- the ensemble with feature selection, without pruning and using majority voting (FS,noPR,MV);
- the ensemble with feature selection, with pruning and using majority voting (FS,PR,MV);
- the proposed ensemble, i.e. with feature selection, pruning and neural network based classifier fusion.

In addition, we implemented several classifier ensembles that are dedicated to imbalanced classification, namely SMOTEBagging [28], SMOTEBoost [9], IIvotes [4],
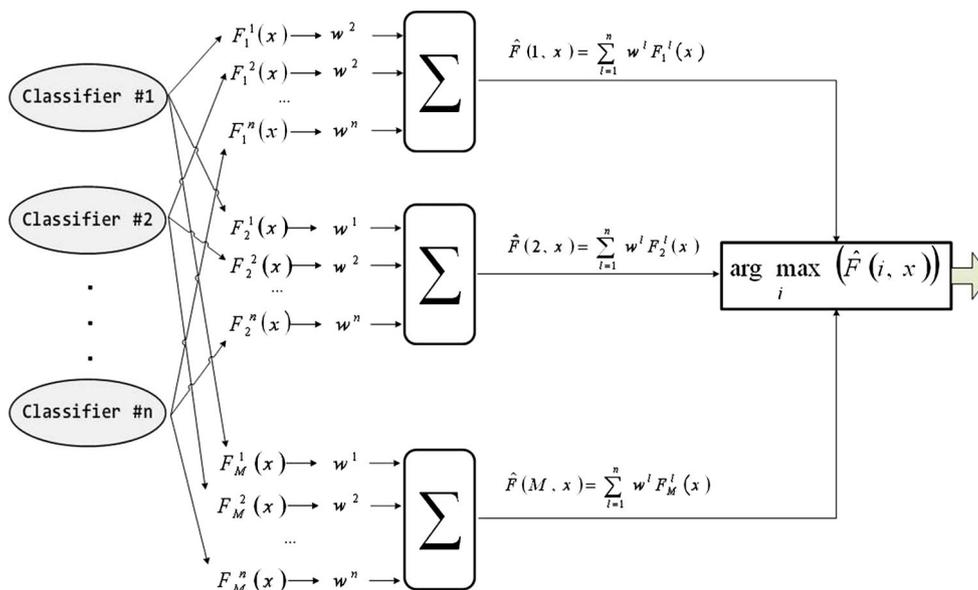

Fig. 4 Classifier fuser implemented as a one-layer perceptron

EasyEnsemble [20] and Underbagging [21]. For all ensembles, we use the same kind of base classifiers, i.e. support vector machines with polynomial kernels.

A combined 5×2 cross-validation F test [1], with a significance level of 0.05 and repeated ten times, was carried out to assess statistical significance of the obtained results in terms of sensitivity and overall accuracy. Since clearly there is a trade-off between sensitivity and specificity (and hence accuracy), to give an overall measure, a classifier is assumed as statistically significantly better compared to another one if one of the following is true:

- its sensitivity is statistically significantly better (as evaluated by a 5×2 CV F test) and its overall accuracy is not statistically significantly worse (again, as evaluated by a 5×2 CV F test);
- its overall accuracy is statistically significantly better (as evaluated by a 5×2 CV F test) and its sensitivity is not statistically significantly worse (again, as evaluated by a 5×2 CV F test).

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.

Looking at the results, we can first of all notice that a canonical approach to classification such as a standard SVM, while leading to a reasonable overall classification accuracy, is not appropriate for the problem at hand, as the rather low sensitivity of only about 25 % demonstrates. Addressing the

Table 1 Classification results for all tested algorithms

	Sensitivity	Specificity	Accuracy
SVM	25.84	92.63	82.09
SVM+SMOTE	91.34	92.25	92.10
SMOTEBagging	92.54	93.06	92.98
SMOTEBoost	91.85	92.89	92.73
IIVotes	93.05	93.56	93.48
EasyEnsemble	91.85	92.89	92.73
UnderBagging	93.39	93.12	93.16
Ensemble (noFS, noPR, MV)	87.24	86.36	86.50
Ensemble (FS, noPR, MV)	89.30	90.45	90.26
Ensemble (FS, PR, MV)	90.86	91.72	91.58
Proposed ensemble	93.76	93.84	93.83

Table 2 Results of statistical significance tests

	SVM	SVM+SMOTE	SMOTEBagging	SMOTEBoost	IIVotes	EasyEnsemble	UnderBagging	Ensemble (noFS,noPR,MV)	Ensemble (FS,noPR,MV)	Ensemble (FS,PR,MV)	Proposed ensemble
SVM		−	−	−	−	−	−	−	−	−	−
SVM+SMOTE	+		−	−	−	−	−	+	+	+	−
SMOTEBagging	+	+		+	−	+	−	+	+	+	−
SMOTEBoost	+	+	−		−		−	+	+	+	−
IIVotes	+	+	+	+		+		+	+	+	−
EasyEnsemble	+	+	−		−		−	+	+	+	−
UnderBagging	+	+	+	+		+		+	+	+	−
Ensemble (noFS,noPR,MV)	+	−	−	−	−	−	−		−	−	−
Ensemble (FS,noPR,MV)	+	−	−	−	−	−	−	+		−	−
Ensemble (FS,PR,MV)	+	−	−	−	−	−	−	+	+		−
Proposed ensemble	+	+	+	+	+	+	+	+	+	+	

A + signifies that the algorithm listed in this row statistically outperforms the algorithm listed in this column, a − indicates a statistically inferior performance

class imbalance through oversampling immediately leads to significantly improved performance—combining an SVM with SMOTE gives a sensitivity of about 91 % coupled with a specificity of about 92 %.

Further improved performance is achieved through application of ensemble techniques. All implemented approaches from the literature, i.e. SMOTEBagging, SMOTEBoost, IIvotes, EasyEnsemble, and Underbagging, achieve both better sensitivity and better specificity although the difference is not always statistically significant. Of these approaches, Underbagging performs best. This confirms that ensemble classifiers indeed lead to better classification.

For our proposed ensemble, we first inspect the influence of the various components of our approach. The ensemble without feature selection or pruning and based on classifier fusion using majority voting gives a sensitivity of 87.24 % with a specificity of 86.35 % and thus, while performing better than a single classifier, clearly lacks behind the other MCS approaches.

Classification performance improves through application of the integrated feature selection step. Feature selection thus not only allows for less complex and hence more efficient base classifiers (of the 437 features on average 74 are retained as input for the classifiers) but also supports improved recognition yielding a sensitivity of 89.30 % and a specificity of 90.45 %. This is due to identifying significant features but also to the increased diversity of the ensemble since feature selection is performed separately for each classifier; while the average diversity of the ensemble without feature selection was 0.526, afterwards it increased to 0.659.

A further gain is obtained by employing a pruning stage to remove redundant base classifiers. While the initial ensemble comprises 11–13 classifiers (depending on the fold of CV), the pruned ensembled consists of 6–7 predictors. At the same time, sensitivity improves to 90.86 % and specificity to 91.72 %.

Clearly, the best classification results are achieved by the full classifier that relies on an advanced trained fusion strategy implemented as a neural network. The achieved sensitivity of 93.76 % and specificity of 93.84 % are both the highest among all methods, while Table 2 shows that our method also statistically outperforms all other approaches. This confirms that each step of our carefully crafted ensemble is essential for delivering good classification performance.

Overall, our experiments clearly demonstrate that a carefully constructed ensemble classifier provides a powerful method for classifying skin lesion attributes, and that our proposed approach, by effectively addressing the inherent class imbalance and appropriate combination of feature selection, classifier selection and classifier fusion, outperforms other state-of-the-art classifier ensembles for this task.

5 Conclusions

In this paper, we have presented an effective method for the automated identification of melanoma from dermoscopic skin lesion images. We first segment the area of the lesion using an approach based on thresholding, region growing and region merging. Based on the delineated lesion area, we then extract a set of shape features, while colour and texture features are derived based on the definition of three clinically important image areas obtained through a Euclidean distance transform. Finally, the extracted features are analysed in a pattern classification stage. For this, we employ a carefully crafted ensemble classifier that addresses the encountered class imbalance by training individual classifiers on balanced subspaces. Non-contributing classifiers are removed based on a fuzzy diversity measure, while the remaining predictors are combined using a trained perceptron fuser. Based on a dataset of 564 skin lesion images, our approach is shown to work very well, giving a sensitivity of 93.76 % coupled with a specificity of 93.84 %, while we further demonstrate that it gives statistically better classification performance compared to several state-of-the-art ensemble classifiers dedicated to imbalanced classification.

References

1. Alpaydin E (1999) Combined 5 × 2 CV F test for comparing supervised classification learning algorithms. Neural Comput 11(8):1885–1892

2. Argenziano G, Soyer HP, De Giorgi V (2002) Dermoscopy: a tutorial. EDRA Medical Publishing & New Media, Milan

3. Binder M, Schwarz M, Winkler A, Steiner A, Kaider A, Wolff K, Pehamberger H (1995) Epiluminescence microscopy. A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. Arch Dermatol 131(3):286–291

4. Blaszczynski J, Deckert M, Stefanowski J, Wilk S (2010) Integrating selective pre-processing of imbalanced data with Ivotes ensemble. In: 7th International conference rough sets and current trends in computing, pp 148–157

5. Celebi ME, Aslandogan YA, Stoecker WV, Iyatomi H, Oka H, Chen X (2007) Unsupervised border detection in dermoscopy images. Skin Res Technol 13(4):454–462

6. Celebi ME, Kingravi H, Uddin B, Iyatomi H, Aslandogan A, Stoecker WV, Moss RH (2007) A methodological approach to the classification of dermoscopy images. Comput Med Imaging Graph 31(6):362–373

7. Celebi ME, Iyatomi H, Schaefer G, Stoecker WV (2009) Lesion border detection in dermoscopy images. Comput Med Imaging Graph 33(2):148–153

8. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357

9. Chawla NV, Lazarevic A, Hall LO, Bowyer KW (2003) SMOTEBoost: improving prediction of the minority class in boosting. In: 7th European conference on principles and practice of knowledge discovery in database, pp 107–119

10. Deng Y, Manjunath BS (2001) Unsupervised segmentation of color-texture regions in images and video. IEEE Trans Pattern Anal Mach Intell 23(8):800–810

11. Fleming MG, Steger C, Zhang J, Gao J, Cognetta AB, Pollak I, Dyer CR (1998) Techniques for a structural analysis of dermatoscopic imagery. Comput Med Imaging Graph 22(5):375–389

12. Golestani A, Azimi J, Analoui M, Kangavari M (2007) A new efficient fuzzy diversity measure in classifier fusion. In: IADIS international conference of applied computing, pp 722–726

13. Haralick RM (1979) Statistical and structural approaches to texture. Proc IEEE 67(5):786–804

14. Karatzoglou A, Smola A, Hornik K, Zeileis A (2004) Kernlab, an S4 package for kernel methods. R J Stat Softw 11(9):1–20

15. Koch MW, Moya MM, Hostetler LD, Fogler RJ (1995) Cueing, feature discovery, and one-class learning for synthetic aperture radar automatic target recognition. Neural Netw 8(7–8):1081–1102

16. Krawczyk B, Schaefer G (2012) Effective multiple classier systems for breast thermogram analysis. In: 21st International conference on pattern recognition, pp 3345–3348

17. Krawczyk B, Schaefer G, Wozniak M (2013) Combining one-class classifiers for imbalanced classification of breast thermogram features. In: 4th International workshop on computational intelligence in medical imaging, 2013. Held as part of IEEE symposium series on computational intelligence

18. Krawczyk B, Schaefer G, Wozniak M (2013) A cost-sensitive ensemble classifier for breast cancer classification. In: IEEE 8th international symposium on applied computational intelligence and informatics, pp 427–430

19. Kuncheva LI (2004) Combining pattern classifiers: methods and algorithms. Wiley-Interscience, New Jersey

20. Liu X, Wu J, Zhou Z (2009) Exploratory undersampling for class-imbalance learning. IEEE Trans Syst Man Cybern Part B Cybern 39(2):539–550

21. Liu Y, Chawla NV, Harper MP, Shriberg E, Stolcke A (2006) A study in machine learning from imbalanced data for sentence boundary detection in speech. Comput Speech Lang 20:468–494

22. Menzies SW, Crotty KA, Ingwar C, McCarth WH (2003) An atlas of surface microscopy of pigmented skin lesions: dermoscopy, 2nd edn. McGraw-Hill, Sydney

23. Nakashima T, Yokota Y, Ishibuchi H, Schaefer G, Drastich A, Zavisek M (2007) Constructing cost-sensitive fuzzy rule-based classification systems for pattern classification problems. J Adv Comput Intell Intell Inf 11(6):546–553

24. Siegel R, Naishadham D, Jemal A (2013) Cancer statistics, 2013. CA: Cancer J Clin 63(1):11–30

25. Steiner K, Binder M, Schemper M, Wolff K, Pehamberger H (1993) Statistical evaluation of epiluminescence dermoscopy criteria for melanocytic pigmented lesions. J Am Acad Dermatol 29(4):581–588

26. Sun Y, Kamel MS, Wong AKC, Wang Y (2007) Cost-sensitive boosting for classification of imbalanced data. Pattern Recognit 40(12):3358–3378

27. Vapnik VN (1998) Statistical learning theory. Wiley, New York

28. Wang S, Yao X (2009) Diversity analysis on imbalanced data sets by using ensemble models. In: IEEE symposium on computational intelligence and data mining, pp 324–331

29. Wozniak M, Zmyslony M (2010) Designing combining classifier with trained fuser—analytical and experimental evaluation. Neural Netw World 20(7):925–934

30. Yu L, Liu H (2004) Efficient feature selection via analysis of relevance and redundancy. J Mach Learn Res 5:1205–1224