Automated Diagnosis of Skin Cancer
Using Digital Image Processing and Mixture-of-Experts

Martin Kreutz[1,2], Maik Anschütz[1,2], Stefan Gehlen[1,2], Thorsten Grünendick[1,2], Klaus Hoffmann[3]

1) Zentrum für Neuroinformatik GmbH, Bochum, Germany
2) ZN Vision Technologies AG, Bochum, Germany
3) Dermatologische Klinik der Ruhr-Universität Bochum im St. Josef Hospital
email: kreutz@zn-ag.de, http://www.zn-ag.de

Abstract. The incidence of malignant melanoma, the most lethal form of skin cancers, has risen rapidly during the last decades. Fortunately, if detected early, even malignant melanoma can be treated successfully. Thus, in recent years, methods for automated detection and diagnosis of skin cancer, particulary malignant melanoma, have elicited much interest. In this paper we present an artificial neural network approach for the classification of skin lesions. Sophisticated image processing, feature extraction, pattern recognition and methods from the field of statistics and artificial neural networks are combined in order to achieve a fast and reliable diagnosis. With this approach, for reasonably balanced training and test sets, we are able to obtain above 90% correct classification of malignant and benign skin lesions coming from the DANAOS data collection.

1 Introduction

1.1 Background

Skin cancer has reached the highest rate of increase among all types of cancer. Fortunately, even the deadliest form, malignant melanoma, may be treated successfully. The key is early detection and the key to early detection is regular screening. There exists a clear demand to improve both, the quantity and the quality of skin cancer screening. This requirement, which will increase in the future, cannot be met to the desired extent by current methods alone. On one hand, cutaneous melanoma is unique among cancers in the sense that it is readily accessible and highly contrasted with the surrounding skin. On the other hand, however, diagnosis of malignant melanoma is a difficult task since other skin lesions may have similar physical characteristics. In many cases, dermatologists perform a biopsy to ascertain whether a lesion is malignant or benign. Since this procedure involves some expense as well as morbidity, particulary in patients with multiple atypical moles, alternative early detection techniques are being sought for rapid, convenient skin cancer screening. A first-level screening may distinguish between benign and suspicious skin lesions. Patients having only benign lesions may be given a clean bill of health, whereas patients with a suspicious lesion are referred to a specialized dermatologist or oncologist.

H. Handels et al. (eds.), *Bildverarbeitung für die Medizin 2001*

1.2 Objective

In this paper we present a framework for a classification system that can, first, continuously be trained with new data and, therefore, is able to learn life-long, and, second, allows a statistical interpretation as well as a validation of its performance. Providing a sufficient amount of training data the system's classification promises to deliver a robust diagnosis that reaches or even exceeds the reliability of an experienced dermatologist.

2 Methods

The clinical diagnosis is often based on the mnemonic ABCD rule (asymmetry, border, color, diameter) [1]. Several approaches aim at standardization and reproducability of the diagnosis by transferring the criteria of the ABCD rule into automatically computed quantities [2, 3, 4]. Since melanoma and certain forms of benign skin tumors differ slightly in their physical characteristics a combination of features is required for reliable diagnostic decisions. When examining skin lesions, dermatologists base their decisions on experience, as well as on complex inferences and extensive knowledge. Such experience cannot be condensed into a small set of rules or symbolic knowledge bases [5]. In contrast to this neural networks are capable of experience-based learning which makes them well suited for this task. In order to train a network well, however, examples from a broad population with different skin types is required. Such a data collection was created during the DANAOS study [6] and is used throughout all experiments.

2.1 Feature extraction

After an image of a skin lesion has been taken (and may be cleaned up by some filtering), its outer border has to be determined. This process is referred to as boundary detection or segmentation. Several methods are discussed in the literature [7, 8, 9, 10]. We employ a hybrid method that combines a statistical clustering of the color space and a hierarchical region-growing algorithm. Once the lesion has been successfully segmented, the feature extraction process can be concentrated directly on the lesion and a margin area around it. The *asymmetry percentage* of a tumor is estimated by first finding the principal axes of inertia of the tumor shape in the image [2] and, second, by computing the nonoverlapping areas after an imaginary "folding" operation along these axes divided by the total area of the tumor. The *border irregularity* is expressed by the irregularity index [5] as well as by fractal features of the boundary [11]. One of the most predictive features in identification of malignant melanoma is *variegated color* [5]. Dermatologists define variegated coloring as the swirling together of tan, brown, red and black. The variegation in color is expressed by the variance in red, green and blue color components. *Texture* also represents an important feature and has been considered in literature [12]. Since the position and orientation of a skin lesion as well as the magnification varies the used features have to be invariant against these variations. Some proposed features are either invariant against translation [13], or rotation [14]. We employ features based on specialized Gabor wavelets which are scale, translation and rotation invariant.

2.2 Mixture-of-Experts architectures

In recent years, artificial neural networks have elicited much interest in the field of cancer diagnosis [5, 3, 15, 16]. Altough fully-connected networks are capable in principle of representing complex nonlinear functions, the time required to train a complex network does not always scale well with problem size and the solution obtained does not always reveal the structure in the problem. Moreover, it is difficult to express prior knowledge in the language of fully-connected networks. Achieving better scaling behavior, better interpretability of solutions and better ways of incoporating prior knowledge may require a more modular approach in which the learning problem is decomposed into sub-problems. A general strategy follows the principle of divide-and-conquer where the problem is treated as one of combining multiple models, each of which is defined over a local region of the input space. An application of the divide-and-conquer principle is the *mixture of experts* architecture [17] which involves a set of function approximators (*expert networks*) that are combined by a classifier (*gating network*). These networks are trained simultaneously so as to split the input space into regions where particular experts can specialize, see Fig. 1. The problem of training mixtures of experts can be treated as a maximum likelihood estimation problem. A general technique for this task is the EM that often yields simple and elegant algorithms [18]. For mixtures of experts the EM decouples the estimation process in a manner that fits well with the modular structure of this architecture. It has been empirically shown that the EM for mixture of experts architectures yields significantly faster convergence than gradient ascent [19]. In order to increase the specialization of the experts, optimization criteria have to be employed that introduce competition between the experts. Several criteria have been discussed [17]. We use a criterion that weights the adaptation rate of each experts with its relative performance.

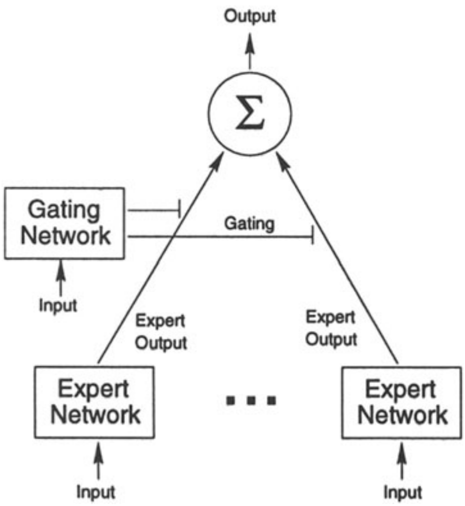

Fig. 1. The *mixture of experts* architecture. The total output is the weighted sum of the expert network outputs, where the weights are the gating network outputs.

3 Results

For the classification of malignant melanoma two different percentages are of interest: sensitivity, the rate of malignant melanoma recognized as malignant, and specifity, the rate of benign lesions recognized as benign. For a first level screening a high sensitivity is needed in order to not miss any malignant lesion. For a significant reduction of unnecessary extrusions, in addition to that, a high specifity is required. In our study we aimed at the first task. Two statistically independent, disjoint sets are used for the training and test sets to allow unbiased results to be obtained on the test set. The networks were trained with a balanced mixture of inputs from each diagnostic class. The data set comprises 423 different cases of skin lesions. Classification results are shown in Tab. 1.

Table 1. Percentages of right classification of *mixture of experts* (MOE) compared with fully-connected multi-layer perceptrons (MLP). The rows show the responses of the networks and the columns the true class.

MOE	malignant	benign	MLP	malignant	benign
malignant	99.2	0	malignant	98.7	0
benign	0	79.4	benign	0	76.5

The results show that for the task of a binary decision "benign or suspicious" a mixture of experts does not outperform a fully-connected multi-layer perceptron. A statistical interpretation, however, is still possible. For more than two output classes no significant improvement could be observed. However, the tendency to overfit was a little bit lower in this case. Overall, the mixture of experts tend less to overfitting than the multi-layer perceptrons since they are inherently able to estimate a local variance for each experts and, therefore, to adjust the mixture perfectly to the variance of the input space.

4 Discussion

Fast and effective methods to separate malignant melanoma from benign tumors are becoming more and more important due to the fact that the incidence of malignant melanoma has risen dramatically in recent years. In this study, we employed a specialized artificial neural network for the classification and diagnosis of melanoma from digitized images of skin lesions. The *mixture of experts* architecture gives two advantages over traditional nonlinear function approximators such as multi-layer perceptrons: a statistical understanding of the operation of the classifier and provision of information about the performance in the form of likelihood information and local error bars. Incorporating ideas from the fields of statistics and artificial neural networks represents a significant step towards *interpretable* learning systems. Overall, experimental evidence is given for the strength of the chosen approach and the classification results obtained on real-world data were found to be very promising.

References

1. R J Friedman, D S Rigel, and A W Kopf. Early detection of malignant melanoma: The role of physician examination and self-examination of the skin. *Ca-A Cancer J Clinicians*, 35:130–151, 1985.
2. W V Stoecker, W W Li, and R H Moss. Automatic detection of asymmetry in skin tumors. *Computarized Medical Imaging and Graphics*, 16:191–197, 1992.
3. R Husemann, S Tölg, W v Seelen, P Altmeyer, P J Frosch, M Stücker, K Hoffmann, and S El-Gammal. Computerised diagnosis of skin cancer using neural networks. In P Altmeyer, K Hoffmann, and M Stücker, editors, *Congress on Skin Cancer and UV Radiation (1996)*, pages 1052–1063. Springer, 1997.
4. W Stolz, R Schiffner, L Pillet, T Vogt, H Harms, T Schindewolf, M Landthaler, and W Abmayr. Improvement of monitoring of melanocytic skin lesions with the use of a computarized acquisition and surveillance unit with a skin surface microscopic television camera. *J Am Acad Dermatol*, 35, 1996.
5. F Ercal, A Chawla, W V Stoecker, H-C Lee, and R H Moss. Neural network diagnosis of malignant melanoma from color images. *IEEE Trans Biomed Eng*, 41:837–845, 1994.
6. G Pott, R Husemann, L Eckert, T Grünendick, S Lux, and P Altmeyer. Danaos - Automated skin cancer diagnosis with neural networks. In *Proc 7th EADV*, 1998.
7. S E Umbaugh, R H Moss, and W V Stoecker. An automatic color segmentation algorithm with application to identification of skin tumor borders. *Computarized Medical Imaging and Graphics*, 16:227–235, 1992.
8. F Ercal, M Mognati, W V Stoecker, and R H Moss. Detection of skin tumor boundaries in color images. *IEEE Trans Medical Imaging*, 12:624–628, 1993.
9. L Xu, M Jackowski, A Goshtasby, C Yu, D Roseman, S Bines, A Dhawan, and A Huntley. Segmentation of skin cancer images. *Image and Vision Computing*, 17:65–74, 1999.
10. P Schmid. Lesion detection in dermatoscopic images using anisotropic diffusion and morphological flooding. In *Proc Int Conf Image Processing (ICIP'99)*, volume 3, pages 449–453. IEEE Signal Process Soc, 1999.
11. E Claridge, P N Hall, M Keefe, and J P Allen. Shape analysis for classification of malignant melanoma. *J Biomed Eng*, 14:229–234, 1992.
12. W V Stoecker, C S Chiang, and R H Moss. Texture in skin lesions: Comparison of three methods to determine smoothness. *Computerized Medical Imaging and Graphics*, 16:179–190, 1992.
13. J-F Liu and J C-M Lee. An efficient and effective texture classification approach using a new notion in wavelet theory. In *Proc ICPR'96*, pages 820–824, 1996.
14. R Porter and N Canagarajah. Robust rotation-invariant texture classification: Wavelet, Gabor filter and GMRF based schemes. *IEEE Proceedings - Vision, Image and Signal Processing*, 144:188ff, 1997.
15. M Binder, H Kittler, A Seeber, A Steiner, H Pehamberger, and K Wolff. Epiluminescence microscopy-based classification of pigmented skin lesions using computerized image analysis and an ANN. *Melanoma Res*, 8:261–266, 1998.
16. H Handels, T Ross, J Kreusch, H H Wolff, and S J Poppl. Feature selection for optimized skin tumor recognition using GAs. *Artif Intell Med*, 16:283–297, 1999.
17. R A Jacobs, M I Jordan, S J Nowlan, and G E Hinton. Adaptive mixture of local experts. *Neural Comput*, 3:79–87, 1991.
18. A P Dempster, N M Laird, and D B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J Royal Stat Soc B*, 39:1–38, 1977.
19. M I Jordan and L Xu. Convergence results for the EM approach to mixture of experts architectures. *Neural Networks*, 8:1409–1431, 1995.