Feature Dimension Reduction for Efficient Classification of Dermoscopic Images with Feature Fusion

Rik Das, Anish Anurag, Govind Kumar Jha, and Mahua Banerjee

Abstract Dermoscopic images carry rich information to identify the malignancy in patients at initial stage. Research initiatives in the domain of content-based image classification can be instrumental in identifying fatal diseases like skin cancer by exploring the dermoscopic image database. This paper has carried out feature dimension reduction for representation of significant content-based image descriptors to the classifiers. The approach has resulted in designing an early fusion based classification model with reduced computational overhead to enhance accuracy of malignancy detection at its inception.

Keywords Skin cancer · Melanoma · Computer aided diagnosis · Principal component analysis · Feature fusion

1 Introduction

Recent advancements in medical imaging have kindled the scope for computer-aided diagnosis (CAD) of life-threatening ailments like cancer. Researchers have made commendable progress in the domain of content-based image classification in identifying benign and malignant categories of the terminal disease [1]. Melanoma is widely known as one of the fatal forms of skin cancer which has claimed innumerable lives. However, timely recognition of the disease has resulted in cure for 99% of the cases within an interval of 5 years of survival. The prime reason of delayed

R. Das (✉) · M. Banerjee
Xavier Institute of Social Service, Ranchi, India
e-mail: rikdas78@gmail.com

M. Banerjee
e-mail: mahuabanerjee2015@gmail.com

A. Anurag · G. K. Jha
Vinoba Bhave University, Hazaribag, India
e-mail: anish.jnu08@gmail.com

G. K. Jha
e-mail: jhagovi@gmail.com

M. Chakraborty et al. (eds.), *Trends in Wireless Communication and Information Security*,
Lecture Notes in Electrical Engineering 740,
https://doi.org/10.1007/978-981-33-6393-9_14

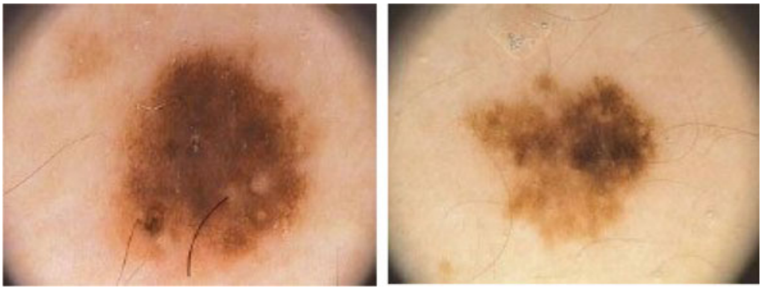

Fig. 1 Illustration of images from PH2 dataset

identification is the extended time for consulting an expert due to their tight schedule of practice [2]. This has stimulated the researchers to design automated system for primary identification of the lethal syndrome by means of content-based image identification. Several attempts are designed to enhance the accuracy of malignancy identification of the disease by designing informative descriptors from the images of the infected portion of the skin. This work has attempted to explore the effect of feature dimension reduction and early feature fusion for identification of malignancy for the melanoma disease. The experiments are conducted with a public dataset named PH2 dataset for which an illustration is shown in Fig. 1 [3].

The results of the experiment have revealed that reduction of feature dimension has resulted in identification of significant feature values influencing classification results. Moreover, reduced feature dimension has lessened the time for classification, which has in turn reduced the runtime of the categorization algorithm.

2 Literature Review

Early skin cancer detection is an emerging field of research in which a comparative study of color constancy, and lesion analysis is carried out [4]. Feature extraction techniques for analysis of dermoscopy images are surveyed to determine the robustness of state-of-the-art techniques [5]. Dermoscopy images are classified using neural networks for melanoma identification [6]. High precision of for melanoma classification is achieved using deep convolutional neural networks [7]. Synergic deep learning techniques are adopted for classification of skin lesion in dermoscopy images [8]. Birthmark mole detection in clinical images are conducted for early melanoma detection [9]. The process has converted the images to monochrome for extraction of significant feature vectors. Feature vectors are extracted from color, shape, and text after performing segmentation of the dermoscopic images [10]. Neural network-based deep ensemble model is evaluated for skin lesion classification in dermoscopic images [11]. Efficient classification of skin lesion is carried out using attention residual learning [12]. Computer-aided diagnosis of skin cancer has outclassed trained dermatologists with the use of deep neural networks [13].

However, implementation of deep neural networks for melanoma detection is mostly infeasible in real time due to its resource hungry nature. The high processing requirements of deep networks makes it challenging to design light weight devices for instant melanoma detection.

In this work, the authors have addressed this issue and have attempted to carry out melanoma detection with lightweight handcrafted features by means of dimension reduction. The results are promising and have revealed high precision for melanoma detection.

3 Classification Techniques

Feature vector extraction is the foremost step and a precursor for the task of content-based image classification. Dermoscopic images in PH2 dataset are preprocessed prior to feature extraction. The dataset comprises of 200 images on the whole spread across 80 images each for common nevi and atypical nevi and the rest 40 images are of melanoma. The images are segmented using ground truth mask available with the dataset as shown in Fig. 2.

Images are rotated in the range of -180° to 180° for generating 12 diverse forms which for each image as in Fig. 3.

This has augmented the dataset and the total number of images becomes 2400. Each image is further resized to 256 * 256 dimension.

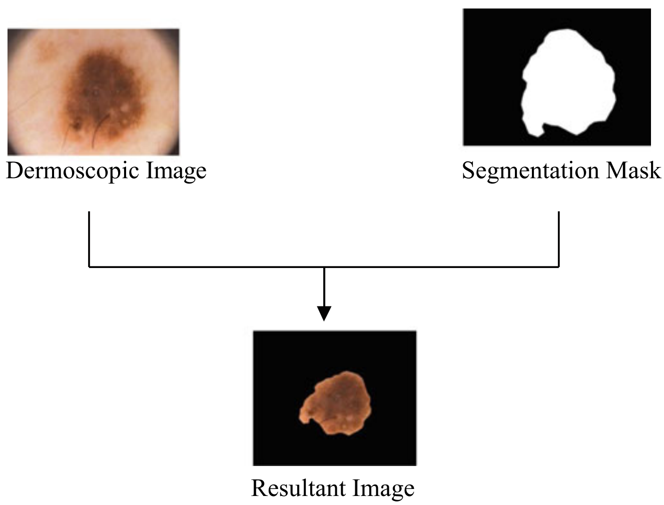

Fig. 2 Image segmentation

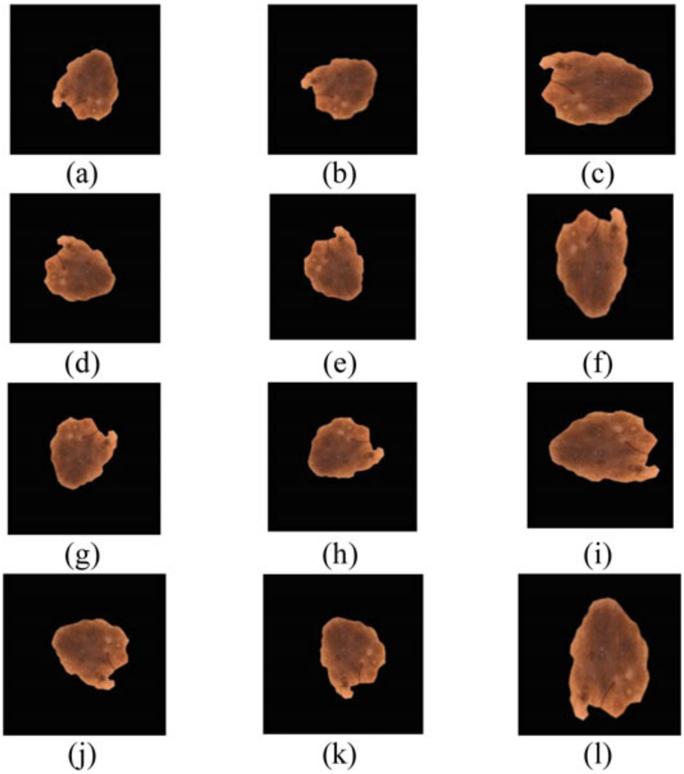

Fig. 3 Image varieties created with angular rotation

Two different feature extraction techniques, namely, histogram of oriented gradients (HOG) and color histogram (CH) are used to extract feature vectors from the dataset.

Henceforth, the feature vectors are reduced in dimension by applying principal component analysis (PCA) [14].

Finally, the two different feature vectors, namely HOG and CH are fused horizontally after their dimensions are reduced.

Classifiers of two different varieties, namely support vector machine (SVM) and logistic model tree (LMT), are applied to evaluate the classification accuracy of the original feature vectors, their reduced dimension varieties, and the fused feature vectors.

4 Classification Techniques

Classification is carried out with tenfold cross-validation for each of the feature extraction techniques. Primarily, the feature vectors with default dimension are evaluated for classification accuracy. Further, PCA is applied for dimension reduction of the feature vectors, and the classification accuracy with reduced dimension of the feature vectors is calculated. Finally, the features with reduced dimension are fused horizontally and are tested for classification accuracies. Time taken to build the classification model in each of the cases is recorded for comparison.

The comparison of accuracies of different feature extraction techniques are shown in Fig. 4.

An integrated comparative result of accuracies, dimensions, and time taken to build classification models is given in Table 1.

The results in Table 1 have revealed highest accuracy for the feature fusion technique in case of both the classifiers. Although, the classification results with single feature (CH) is almost equivalent to feature fusion in case of LMT, but a significant difference is noticed for SVM. This is because the fused features have captured and represented both the color properties and gradient orientation of the images simultaneously to the classifier. Hence, the classifier is able to identify the images with more clarity and consistency across diverse environments compared to any of the

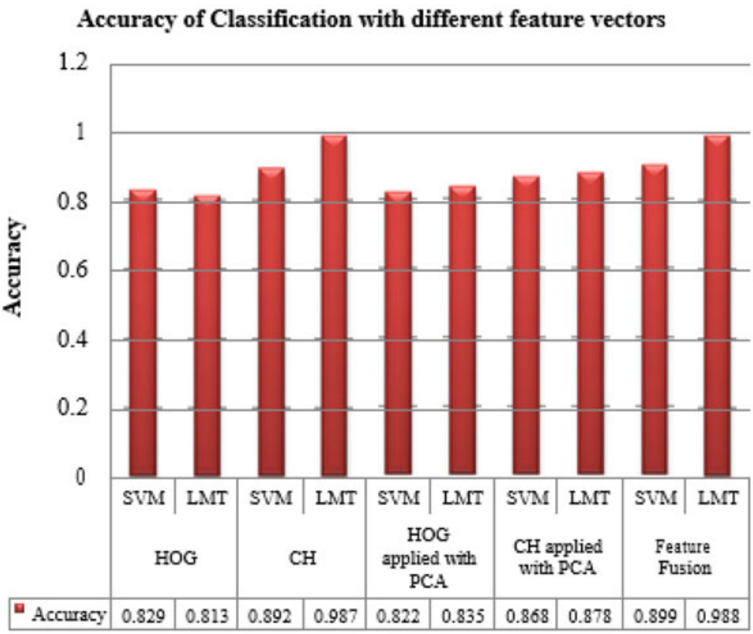

Fig. 4 Comparison of accuracies for different features

Table 1 Comparison of accuracy, dimension of feature vectors, and time taken for building classification model

Features	Classifiers	Accuracy	Dimension	Time (s)
HOG	SVM	0.829	576	13.95
	LMT	0.813		71.64
CH	SVM	0.892	64	0.23
	LMT	0.987		10.23
HOG applied with PCA	SVM	0.822	250	3.36
	LMT	0.835		20.39
CH applied with PCA	SVM	0.868	50	0.21
	LMT	0.878		4.76
Feature fusion	SVM	0.899	300	3.82
	LMT	0.988		47.73

individual feature descriptors which have represented either the color properties or the gradient orientation at a given instance.

Additionally, it is observed that reduction of dimension of the feature vectors with application of PCA has retained significant feature components by eliminating the ones which do not have much influence on classification decision. The time taken for building classification model by the reduced feature vectors is considerably low compared to their original counterparts. Fusion-based features have consumed lesser time in building classification model compared to the original HOG features, but have surpassed the time taken by CH.

Finally, feature dimension in fused feature is much lesser than original HOG features but the accuracy of classification is higher, which indicates toward the efficiency of fusion-based approach.

5 Conclusion

The paper has proposed a design to implement computer-aided diagnosis for classification of melanoma with dermoscopic images. Conventional classification methodologies consider single feature vector as representation of the image data to be classified. However, all the feature values do not have significant weight to influence classification decision. This work has reduced the dimension of features by applying PCA and has reduced the classification time without compromising accuracy. A feature fusion-based approach with reduced features is also shown for enhancing classification accuracy with less computational overhead.

References

1. Hou L, Samaras D, Kurc TM, Gao Y, Davis JE, Saltz JH (2016) Patch-based convolutional neural network for whole slide tissue image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2424–2433
2. Matsunaga K, Hamada A, Minagawa A, Koga H (2017) Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble. arXiv preprint arXiv:1703.03108
3. Fernandes SL, Chakraborty B, Gurupur VP, Prabhu G (2016) Early skin cancer detection using computer aided diagnosis techniques. J Integr Design Process Sci 20(1):33–43
4. Barata C, Celebi ME, Marques JS (2018) A survey of feature extraction in dermoscopy image analysis of skin cancer. IEEE J Biomed Health Inform 23(3):1096–1109
5. Xie F, Fan H, Li Y, Jiang Z, Meng R, Bovik A (2016) Melanoma classification on dermoscopy images using a neural network ensemble model. IEEE Trans Med Imaging 36(3):849–858
6. Demyanov S, Chakravorty R, Abedini M, Halpern A, Garnavi R (2016) Classification of dermoscopy patterns using deep convolutional neural networks. In: 2016 IEEE 13th international symposium on biomedical imaging (ISBI). IEEE, pp 364–368
7. Zhang J, Xie Y, Wu Q, Xia Y (2018) Skin lesion classification in dermoscopy images using synergic deep learning. In: International conference on medical image computing and computer-assisted intervention. Springer, Cham
8. Balakrishnan J, David D (2019) Melanoma classification and birthmark mole detection on clinical images. In: 2019 international conference on vision towards emerging trends in communication and networking (ViTECoN). IEEE, pp 1–5
9. Suganya R (2016)An automated computer aided diagnosis of skin lesions detection and classification for dermoscopy images. In: 2016 international conference on recent trends in information technology (ICRTIT). IEEE, pp 1–5
10. Xie Y, Zhang J, Xia Y (2018) A multi-level deep ensemble model for skin lesion classification in dermoscopy images. arXiv preprint arXiv:1807.08488.
11. Zhang J, Xie Y, Xia Y, Shen C (2019) Attention residual learning for skin lesion classification. IEEE Trans Med Imaging 38(9):2092–2103
12. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, Utikal JS (2019) Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. Eur J Cancer 113:47–54
13. Satheesha TY, Satyanarayana D, Prasad MG, Dhruve KD (2017) Melanoma is Skin Deep: A 3D reconstruction technique for computerized dermoscopic skin lesion classification. IEEE J Transl Eng Health Med 5:1–17
14. Zhang T, Yang B (2016) Big data dimension reduction using PCA. In: 2016 IEEE international conference on smart cloud (SmartCloud). IEEE, pp 152–157