2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Quantitative Comparison of Color Asymmetry Features for Automatic Melanoma Detection

Ruchir Srivastava[1] and Ee Ping Ong[1] and Beng-Hai Lee[1] and Lucinda Siyun Tan[2] and Hong Liang Tey[2,3,4]

***Abstract*— Asymmetry assessment is an important step towards melanoma detection. This paper compares some of the color asymmetry features proposed in the literature which have been used to automatically detect melanoma from color images. A total of nine features were evaluated based on their accuracy in predicting lesion asymmetry on a dataset of 277 images. In addition, the accuracies of these features in differentiating melanoma from benign lesions were compared. Results show that simple features based on the brightness difference between the two halves of the lesion performed the best in predicting asymmetry and subsequently melanoma.**

***Clinical relevance*— The proposed work will assist researchers in choosing better performing color asymmetry features thereby improving the accuracy of automatic melanoma detection. The resulting system will reduce the workload of clinicians by screening out obviously benign cases and referring only the suspicious cases to them.**

I. INTRODUCTION

Melanoma is the most serious form of skin cancer. The risk of developing melanoma during ones lifetime has been increasing worldwide, especially in whites. In the US, melanoma raw incidence rates climbed in 2016 from 22.6 to 23.6 (0.9% CAGR) per 100,000 population [1]. However, if diagnosed early, the survival rate can be more than 98% as compared to only 22.5% if the melanoma progresses to stage four [2].

At an early stage, melanoma can be distinguished from a common mole (also called nevus) based on its visual appearance. However, with the increasing prevalence of melanoma, visual examination can be burdensome for the clinicians, especially when many of the suspected lesions are nevus. This has led to research looking into automatic computer-based image analysis. Such works use features extracted from a color image of the skin lesion to predict whether the lesion is a melanoma or nevus. These features are mostly inspired by the clinical practices. For example, for visual assessment, one of the most important factors is asymmetry of the suspected lesion. Asymmetry is incorporated into some of the standard clinical 'rules' for melanoma assessment such as ABCD (asymmetry, border, color, differential structures) and CASH (color, architecture, symmetry, homogeneity). Consequently, many of the works on automated melanoma detection use features capturing the asymmetry of the lesion.

srivastavar@i2r.a-star.edu.sg
[1]Institute for Infocomm Research, Singapore-138632
[2]National Skin Center, Singapore-308205
[3]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
[4]Yong Loo Lin School of Medicine, National University of Singapore, Singapore

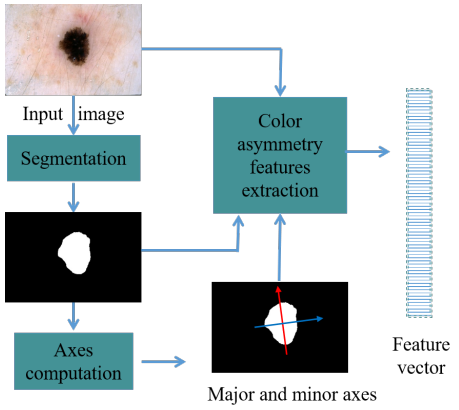


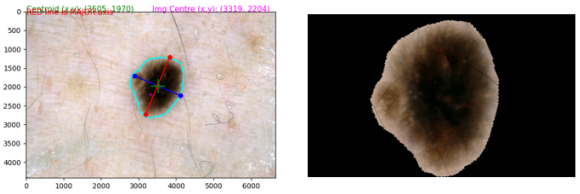
Fig. 1. A flow chart of the method to extract color asymmetry features.

Asymmetry of the lesion is defined with respect to shape, color and structure [3]. However, existing works have mostly looked into shape asymmetry features. Damian et al. [4] studied both shape and color asymmetry features and found former to be better in detecting melanoma as compared to latter. However, they see the potential in using color asymmetry features together with the shape asymmetry features. Considering the importance of color asymmetry features, in this work, we perform a quantitative evaluation of some of the existing color asymmetry features which have been used for performing automatic melanoma detection.

Contributions: There are works which have compared only the shape asymmetry features [3] [5]. However, to the best of our knowledge, there has not been any comparison of just the color asymmetry features in the literature. Clawson et al. [6] have compared only the color asymmetry features that they proposed but not other color asymmetry features. We present a comparison of as many as nine popular color asymmetry features. This will help researchers in choosing better performing color asymmetry features and improving the performance of a melanoma detection system.

II. METHOD

Given a skin lesion image, the lesion is segmented out and the major and minor axes of the lesion are computed (Figure 1). Thereafter, both color and shape asymmetry features are computed. Just one color asymmetry feature is combined with shape asymmetry features to predict the asymmetry of the lesion. The experiment is repeated for all the color asymmetry features. Since shape asymmetry features are common in these prediction experiments, the difference in accuracy is due to a different color asymmetry



(i) (ii)

Fig. 2. (Best viewed in color) (i) Sample image with major and minor axes marked in red an blue respectively, (ii) Extracted ROI.

feature used. This helps to evaluate the effectiveness of individual color asymmetry features. Computed features are also used to predict whether the lesion is melanoma using the ABCD rule in line with the clinical practice. Details of each step are mentioned as follows.

A. Lesion Segmentation

In a color image of the skin lesion, the lesion may occupy only a part of the full image making its segmentation necessary. To this end, we used the work by Ong et al. [7] which uses a deep-learning based segmentation method. The method was originally developed for skin wound segmentation, so we retrained their model on our dataset, the details of which are mentioned in section III-A.

B. Major and Minor Axes Computation

Once the lesion is segmented, the major and minor axes of the lesion were computed using the eigenvector approach where the lesion region was represented as a point cloud in two dimensions. The first principal component resulting from the eigen-analysis corresponded to the major axis while the axis perpendicular to the major axis was taken to be the minor axis. To facilitate further analysis, the image was rotated to align the major and minor axes with the coordinate axes. Then the lesion region was cropped out as the region of interest (ROI). The ROI was obtained by cropping out equal number of rows (or columns) from all the sides of the image. The amount to crop was maximized while ensuring that the lesion was not cropped out and lesion centroid coincided with the image center. In addition, intensities of the pixels outside the lesion region were put to zero to minimize their influence on feature computation (Figure 2). This ROI was used for computation of color asymmetry features.

C. Color Asymmetry Features Used

In order to compare color asymmetry features, we selected five works from the literature based on their citations, availability of the paper and ease of implementation. Due to space constraints, we have summarized the features in Table I with the associated references. Each of these features compare the two halves of the lesion. Since division into the two halves can be either along the major or minor axes, the features were computed for both the scenarios and for each channel (Dimension = 6). Note that for $SSIM_{RGB}$, the features for each channel were averaged, leading to a dimension of 2. $MeanDiffRGB$ would be similar to A_1A_2Color, so it was not computed.

TABLE I

COLOR ASYMMETRY FEATURES COMPARED IN THIS WORK. DIM. STANDS FOR FEATURE DIMENSION.

Feature	Channels	Description	Dim.
KL [3]	RGB	Kullback-Leibler divergence between histograms of two halves	6
Q [4]	RGB	Chi-square distance between histograms of two halves	6
A_1A_2Color [5]	RGB	Color versions of A_1 and A_2	6
$MeanDiffHSV$ [8]	HSV	Mean of the brightness difference of two halves	6
$MeanDiffLab$ [8]	Lab		6
$StdDiffHSV$ [8]	HSV	Std. dev. of the brightness difference of two halves	6
$StdDiffLab$ [8]	Lab		6
$StdDiffRGB$ [8]	RGB		6
$SSIM_{RGB}$ [9]	RGB	Color version of SSIM	2

D. Shape Asymmetry Features Used

In this work, we intend to assess the effect of each color asymmetry feature, used alone, on asymmetry prediction accuracy. However using just one feature, with a small dimension, for prediction might not be sufficiently accurate. To get a more representative set of asymmetry features, we also compute other shape asymmetry features which together can be denoted as SA. Our assumption is that if, for example, using SA and Q gives a better melanoma detection accuracy than using SA and KL, then Q is a better feature than KL. The shape asymmetry features used in this work are computed using the grayscale images or the binary segmentation masks. The shape asymmetry features computed using the grayscale images are SSIM [9] and MSE [9]. In addition, we have used a normalized version of MSE where the feature value is normalized by $I_{max} - I_{min}$, where I_{max} and I_{min} are the maximum and minimum intensity values in the grayscale image. The features computed using the binary segmentation masks are A_1 and A_2 used by Celebi et al. [5] and Circularity, Asymmetry Index, Eccentricity and all the seven Hu's moments used by Damian et al [4].

A total of 70 features were computed from each image including 50 color asymmetry features and 20 shape asymmetry features.

E. Evaluation

The color asymmetry features extracted above were evaluated on two criteria: 1) How well they represent color asymmetry, and 2) How effective are they in detecting melanoma. Accordingly, there were two tests performed as follows:

Test A: One way to evaluate the features based on criterion 1 is to compare their accuracy in predicting asymmetry of the lesion. For such an evaluation, each of the color asymmetry features, taken one at a time, was combined with shape

asymmetry features, SA and used to predict asymmetry. This ensures that the effect of only one color asymmetry feature is captured at a time.

A support vector machines (SVM) classifier predicted the asymmetry in the form of the 'A' score. 'A' score is one of the four scores used in the ABCD rule. The other scores 'B', 'C', and 'D' were not predicted in this work since we focus on asymmetry prediction. The values of these scores were adopted from the ground truth to avoid inaccuracies in their prediction affecting the melanoma prediction accuracy. These scores are combined to predict melanoma as explained in Test B.

Test B: To evaluate the color asymmetry features based on criterion 2, 'A' score predicted in Test A was used to predict melanoma using the ABCD rule which combines 'A', 'B', 'C', and 'D' scores to produce a total dermoscopy score (TDS) defined as [10]:

$$TDS = A \times 1.3 + B \times 0.1 + C \times 0.5 + D \times 0.5 \quad (1)$$

A TDS value less than 4.75 implies a nevus while a value greater than 5.45 imply melanoma. Any value in between indicates a suspicious lesion. We have used the threshold 5.45 in this work.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. Dataset

Our data was extracted from the ISIC 2018 grand challenge dataset [11], [12]. This dataset contains color images of skin lesions and the ground truth regarding the lesion class has been provided based on histology. For a subset of the images, segmentation masks were available and used to train the segmentation model. For feature evaluation, since we are only interested in melanoma and nevus, we chose 277 images belonging to one of these two classes. The images selected had only one lesion fully contained in the image and with no error in lesion segmentation. This was to ensure that the comparison of features wasn't affected due to segmentation errors. Out of the 277 images, 210 were of nevus while 67 were of melanoma.

'A', 'B', 'C', and 'D' scores were assigned to these images by a trained dermatologist. An asymmetry score of 0 was given if the lesion was symmetrical along both major and minor axes. A score of 1 was given if there was color asymmetry observed along one axis only while a score of 2 indicated color asymmetry along both axes. After the annotation, it was found that the number of images with 'A' score of 0, 1, and 2 were 48, 168, and 61, respectively. Since there was a class imbalance, we used the SMOTE algorithm [13] to generate synthetic samples for the minority classes (scores 0 and 2). This resulted in 504 samples. Note that the synthetic samples were only used for training the classifier while during testing only the original samples were used.

Our preliminary experiments indicated that for the prediction of 'A' score, there was a lot of confusion between classes 0 and 1 and between classes 1 and 2, resulting in a lower accuracy and reliability. As a result, we removed the samples with A score equal to 1 leaving us with 2 classes of asymmetry: 0 (symmetrical) and 2 (asymmetrical) and a total of 336 samples. This 2-class asymmetry prediction has also been used by Clawson et al. [6].

B. Evaluation of the Color Asymmetry Features

Our experiments were coded in python. A 4-fold cross-validation was performed to select the model with the highest accuracy across the 4-fold. The selected model was used to predict all the 277 images excluding the samples generated using SMOTE. Due to the unbalanced nature of the data, just monitoring the accuracy (Acc) might be misleading so sensitivity (Sen) and specificity (Spec) were also used for performance evaluation.

As mentioned in section II-E, two tests were performed to evaluate the effectiveness of the extracted color asymmetry features. Details on implementation and the results of each test are as follows:

1) Test A: Prior to 'A' score prediction, the features were normalized to have zero mean and unit variance. Classification was performed using SVM with a radial basis function (RBF) kernel. SVM was implemented using the Scikit-learn library for python [14] with the default settings. Performance of each set of features is presented in Table II. Results show that the best performing feature is $MeanDiffLab$ (97.3%) while the least accuracy is that of KL. In addition, other features derived from the brightness difference between the two halves of the lesion perform better than other feature categories such as Q and KL which measure the similarity of histograms of the two halves. This indicates that simple features based on mean and standard deviation are accurate enough for asymmetry prediction.

Another useful observation is from the results obtained using just the shape asymmetry features, SA. Since SA were present with each of the color asymmetry features, it indicates that with the exception of $SSIM_{RGB}$, KL, and Q, color asymmetry features improved the asymmetry prediction accuracy compared to using just the shape asymmetry features. Moreover, when all the features are used, the accuracy was the best (98.2%) which indicates the usefulness of color asymmetry features.

2) Test B: This test was conducted based on the predictions made in test A and since the ground truth values of 'B', 'C' , and 'D' scores were used, the melanoma prediction accuracy should correlate with the asymmetry prediction accuracy. However, this is not guaranteed since even if TDS score matches the actual TDS score, melanoma prediction can be inaccurate due to a threshold involved in the prediction. Note that the ground truth of melanoma vs lesion was based on histology and not TDS score.

For our experiments, the predicted label (melanoma or nevus) was compared with the ground truth and the resulting accuracy for each of the features is reported in Table II under Test B. Results show that melanoma prediction accuracies correlated well with the asymmetry prediction accuracies (Pearson's correlation coefficient = 0.85, p = 0.003). In a

TABLE II

PERFORMANCE OF EACH SET OF FEATURES IN PREDICTING THE ASYMMETRY SCORE (TEST A) AND MELANOMA (TEST B). $AllFeat$ REPRESENTS RESULTS USING ALL THE FEATURES.

Features	Test A			Test B		
	Sen	Spec	Acc	Sen	Spec	Acc
A_1A_2Color	0.92	0.94	92.7	0.76	0.98	88.1
KL	0.89	0.88	88.1	0.78	0.95	87.2
$MeanDiffHSV$	0.97	0.94	95.4	0.80	1.00	90.8
$MeanDiffLab$	0.97	0.98	97.3	0.82	1.00	91.7
Q	0.89	0.92	89.9	0.78	0.98	89.0
$StdDiffHSV$	0.90	0.94	91.7	0.78	0.98	89.0
$StdDiffLab$	0.89	1.00	93.6	0.78	1.00	89.9
$StdDiffRGB$	0.92	0.92	91.7	0.78	0.95	87.2
$SSIM_{RGB}$	0.87	0.94	89.9	0.76	0.98	88.1
SA	0.89	0.94	90.8	0.78	0.98	89.0
$AllFeat$	0.98	0.98	98.2	0.82	1.00	91.7

TABLE III

3-CLASS PROBLEM: PERFORMANCE OF EACH SET OF FEATURES IN PREDICTING THE ASYMMETRY SCORE (TEST A) AND MELANOMA (TEST B).

Features	Test A			Test B		
	Sen	Spec	Acc	Sen	Spec	Acc
A_1A_2Color	0.80	0.87	72.2	0.63	0.95	87.0
KL	0.69	0.87	69.7	0.63	0.97	88.5
$MeanDiffHSV$	0.84	0.87	75.8	0.60	0.96	87.4
$MeanDiffLab$	0.85	0.89	75.5	0.58	0.96	86.6
Q	0.79	0.84	72.9	0.61	0.96	87.4
$StdDiffHSV$	0.80	0.86	74.4	0.64	0.95	87.7
$StdDiffLab$	0.79	0.87	74.0	0.66	0.95	88.1
$StdDiffRGB$	0.78	0.84	70.8	0.60	0.93	84.8
$SSIM_{RGB}$	0.78	0.84	69.0	0.60	0.92	84.5
SA	0.80	0.87	72.6	0.60	0.94	85.6
$AllFeat$	0.80	0.96	79.8	0.58	0.98	88.5

real application, melanoma prediction also depends on the prediction of 'B', 'C' , and 'D' scores so the correlation observed may not be similar to what we observed.

We also report results for a 3-class classification (Table III) to see if the feature ranking is consistent with 2-class classification. The feature rankings are similar and the accuracies for Test A for 2-class correlate well with those for 3-class (Pearson's correlation coefficient = 0.87, p=0.005).

3) Limitations and future work: The findings presented in this paper do suffer with some limitations and have opened up questions worth investigating. The first question pertains to the dataset and the ground truth. We have used 109 images for training which after SMOTE increased to 336. Note that SMOTE only upsamples the minority class. A larger dataset can provide more samples of the majority class (nevus) and better insights into the problem we have investigated. Moreover, accuracy on a more realistic 3-class asymmetry prediction needs to improve, especially the sensitivity since an actual melanoma case should not be missed.

The second question is regarding the choice of the classifier. Can the results change if we use a different classifier? Are there features which perform consistently well regardless of the classifier? These questions will pave the way for our future work.

IV. CONCLUSIONS

In this paper, nine color asymmetry features were evaluated based on their performance in predicting the asymmetry score and melanoma. Ground truth of asymmetry score was marked by a trained clinician while the label of melanoma or nevus was provided along with the dataset (ISIC 2018). 277 images from the dataset were used for the experiments. It was found that simple features related to mean and standard deviation of the difference in brightness of the two halves of the lesion performed the best in predicting asymmetry (91.7% - 97.3%). The histogram-based features gave an accuracy less than 90%. Including color asymmetry features did improve asymmetry prediction accuracy compared to using shape asymmetry features alone. Once asymmetry score was predicted, the ABCD criterion was used to predict whether the lesion was melanoma or nevus and the best performing features gave an accuracy of 91.7%. The research also opened up avenues for further research.

REFERENCES

[1] A. M. Glazer, R. R. Winkelmann, A. S. Farberg, and D. S. Rigel, Analysis of trends in US melanoma incidence and mortality, JAMA dermatology, vol. 153, no. 2, pp. 225-226, 2017.

[2] A. R. Ali, J. Li, and S. J. O'Shea, Towards the automatic detection of skin lesion shape asymmetry, color variegation and diameter in dermoscopic images, Plos one, vol. 15, no.6, 2020.

[3] R. Chakravorty, S. Liang, M. Abedini, and R. Garnavi, Dermatologist-like feature extraction from skin lesion for improved asymmetry classification in PH-2 database, in 2016 EMBC.

[4] F. A. Damian, S. Moldovanu, N. Dey, A.S. Ashour, and L. Moraru, Feature Selection of Non-Dermoscopic Skin Lesion Images for Nevus and Melanoma Classification, Computation, vol 8, no. 2, p 41, 2020.

[5] M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker,and R. H. Moss, A methodological approach to the classification of dermoscopy images, Computerized Medical imaging and graphics, vol. 31, no. 6, pp. 362-373, 2007.

[6] K. M. Clawson, P. J. Morrow, B. W. Scotney, D. J. McKenna, and O. M. Dolan, Computerised skin lesion surface analysis for pigment asymmetry quantification, in International Machine Vision and Image Processing Conference (IMVIP 2007), pp. 75-82.

[7] E. P. Ong, C. T. K. Yin, and B. H. Lee, Efficient Deep Learning-based Wound-bed Segmentation For Mobile Applications, in EMBC, 2020.

[8] M. J. M. Vasconcelos, L. Rosado, and M. Ferreira, A new risk assessment methodology for dermoscopic skin lesion images, in 2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings, pp. 570-575, May 2015.

[9] A. Gadiraju, S. Keshwani, and E. Minkin, Mole Investigator: Detecting Cancerous Skin Moles Through Computer Vision, 2015.

[10] F. Nachbar et al., The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions, Journal of the American Academy of Dermatology, vol. 30, no. 4, 551-559, 1994.

[11] N. Codella et al., Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC), https://arxiv.org/abs/1902.03368, 2018.

[12] P. Tschandl, C. Rosendahl, and H. Kittler, The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, Sci. Data 5, 180161 doi:10.1038/sdata.2018.161, 2018.

[13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, Journal of artificial intelligence research, vol. 16, 321-357, 2002.

[14] Pedregosa et al., Scikit-learn: Machine Learning in Python, JMLR vol. 12, pp. 2825-2830, 2011.