

Automated Classification of Pigmented Lesions Towards Melanoma Detection

Elisa Liu, Alex Lu, Winston Wang

1 Introduction

Melanoma is the deadliest form of skin cancer, with increasing incidence over the last few decades¹. However, melanoma has a relatively high 5-year survival rate, especially when detected early. Notably, diagnosis can be achieved from visual inspection alone, aided by the widespread adoption of epiluminescence microscopy (dermoscopy).

Dermoscopy provides a non-invasive method for examining potentially malignant skin lesions, enhancing microscopic features that may otherwise be missed by an un-aided visual examination, and has been shown to improve diagnostic accuracy when performed by an expert clinician². In some cases, inter-clinician disagreement over subtle features may lead to differential diagnosis. Given the complexity of features in skin lesions, implementing an automated dermoscopic assessment algorithm may increase diagnostic accuracy and boost throughput.

Many computer-aided techniques have been studied for automated skin lesion detection and classification. Current publications use complex algorithms such as Wavelet decomposition³, grey-level co-occurrence matrix⁴, or Fourier power spectrum⁵ to extract features that are of non-clinical significance. However, algorithms based on non-medical features are difficult to implement in a clinical setting because practice physicians have difficulty accepting and understanding such methods, and are often of lower diagnostic accuracy⁶. While extensive work has been dedicated towards this subject, there is little discussion on developing a robust feature extraction algorithm that can recapitulate an expert clinician's approach.

2 Method

We propose to use a combination of texture and color descriptors to extract features of clinical significance (feature extraction techniques were adapted from Riaz et al⁷). Our implementation follows a workflow described here:

1. Feature extraction
 - a. Dynamically down-scale images to a convenient size
 - b. Extract texture features
 - c. Extract color features
2. Support vector machine

2.1 Local binary pattern for texture descriptors

One of the first symptoms of melanoma is a change in the texture of a skin lesion⁸ -- popular clinical dermatology heuristics, including the 7-point checklist and ABCDE rule, mark features such as irregular border and/or peripheral globules. Analysis using local binary pattern (LBP) can characterize the local texture at a pixel scale level. LBP compares the relative grey-scale value of a pixel against its neighbors. By assigning a binary value according to relative magnitude of each of 8 pixels in a circle around the center pixel, LBP yields a grey-scale invariant 8-bit code⁹.

Computation of this 8-bit code can also be formulated to confer rotational invariance by incorporating a measure of uniformity, known as uniform LBP (ULBP). Rotation and grey-scale invariance are both key characteristics in extracting textural features from dermoscope images which may be heterogeneous in orientation and lighting.

ULBP provides a quantification of image texture which captures position and relative change in intensity without incorporating irrelevant contexts, preserving identification of spots, corners, and edges.

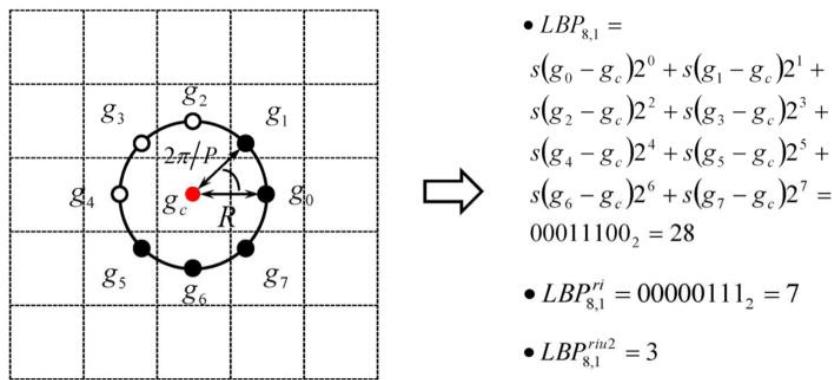


Figure 1. Description of LBP implementation, contrasting default from uniform coding.

2.2 Hue-saturation-value (HSV) color descriptors

Another key clinical presentation of melanoma is skin lesion color. However, typical lesions contain a limited color range not efficiently represented in RGB color space. Further, the presence shadows, hairs, glare, and other undesirable image content can clutter our representation of lesion color. We perform 3-color segmentation on the images and isolate the center segment with the assumption that the segments with the lesions include the center pixel.

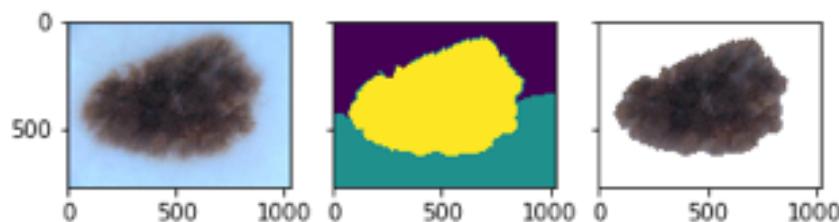


Figure 2. Example segmentation

Color features within the center segmented are then expressed in hue-saturation-value (HSV) color space as an L1-normalized count of the values falling into 16 hue buckets, 4 saturation buckets, and 4 value buckets, resulting in a 256-vector per image.

2.3 Support vector machine for binary classification

Because the purposes of a dermoscopy evaluation are to recommend future clinical actions, i.e. biopsy, a one-versus-rest binary classification of melanoma versus benign is well-suited. Prominent classification methods include logistic regression, k-nearest neighbors, neural networks, and decision trees. These provide a binary classification as well as a probability of

class membership. However, papers reporting on these methods have described variable outcome¹¹.

A support vector machine is another major approach for making a binary diagnostic model. While it loses the probability of class membership, its performances are on par with other machine learning algorithms¹⁰. We chose SVM for its ease of implementation and popularity within literature.

3 Results

We evaluated our automated classification of dermoscopy images on data obtained from the ISIC 2017 challenge¹¹. The training data contains 2000 images, of which 374 are classified as malignant melanoma.

We performed 10-fold cross validation with 2000 images, obtaining an accuracy of 81%. However, our classifier suffers heavily from the loss of texture and color feature encoding, and is heavily impacted by the result of class imbalance. Consider the following visualizations:

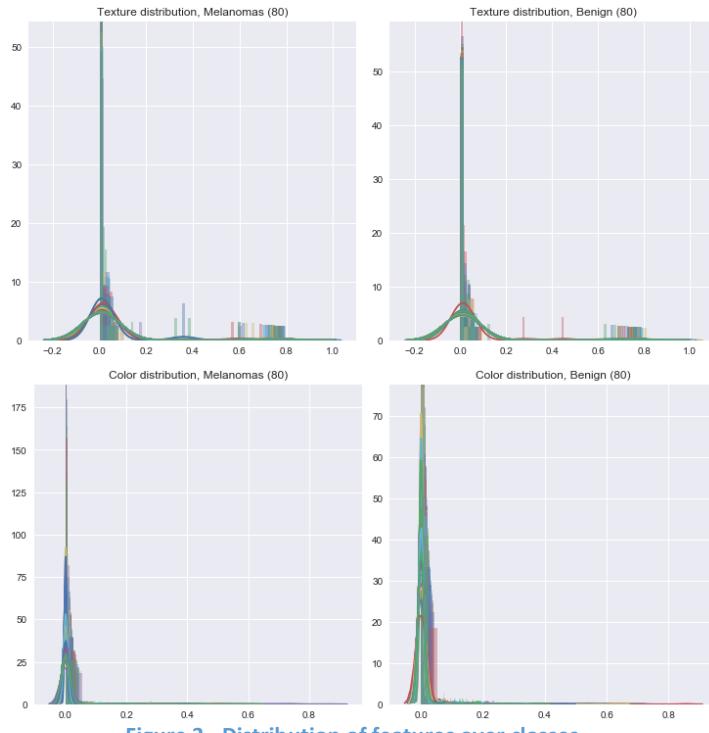


Figure 3. Distribution of features over classes

Here, we can see that the distribution of our L1-normalized counts is very similar between classes, with sparsity manifesting as a strong peak around 0 in all four histograms.

We attempted to account for the effect of class imbalance by weighting the melanoma class differently and altering the error penalty coefficient. Our tuning results can be seen in figure 4.

Overall, our max accuracy was 81.3%.

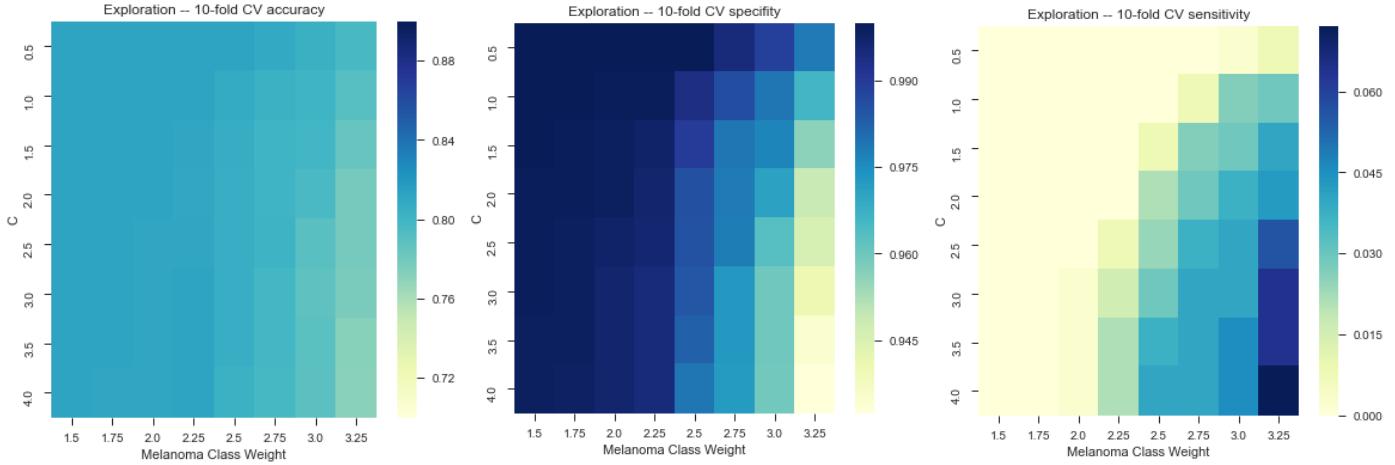


Figure 4. Graph Exploration for Tuning Parameters – 10-fold cross validation

4 Discussion

4.1 Challenges

Dermoscopy images prepared in dermatology clinics are fundamentally noisy and heterogeneous, often featuring glare, bubbles, large shadows, and foreign objects. The heterogeneity of images in the ISIC 2017 Imaging Challenge training set is no exception – with differences in inclusion of foreign objects and shadows, as well as class imbalance, and image size heterogeneity. Altogether, these image characteristics introduce unique challenges that impact extraction of color and texture features, as well as subsequent classification steps.

Originally, we set out to label each pixel with both texture and color information to achieve a position-specific understanding of lesion features. However, differences in image size and the presence of foreign objects or hair introduce significant amounts of noise and computational difficulties. To circumvent image sizing issues, textural and color features are represented as L1-normalized bucketed counts.



Figure 5. Examples of challenging images

To reduce the impact of foreign objects – shadows, hair, rulers, ink markings made in clinic, band-aids, etc. – we employ a segmentation pre-processing that identifies the lesion border, isolates this region of interest, and preserves only color information from within this region. This segmentation, although useful, was not perfect across all of the examples.

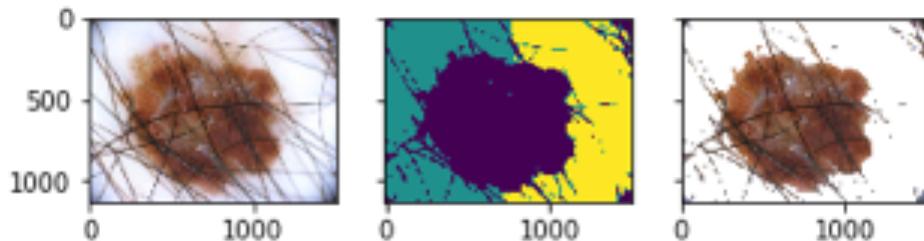


Figure 6. Example of an unsuccessful segmentation

Various approaches to remove pre-process images have been explored. Hair removal-specific methods have also been implemented in practice, including fast marching method¹² and curvilinear detection of hair and gap filling¹³. Others have attempted general purpose filters, such as a Gaussian or median filter to smooth images^{4,14}. While these approaches have been shown to lead to better diagnostic results, they have high computational requirement or introduce additional artifacts into the images.

Ultimately, extraction of features as L1-normalized count vectors is a lossy method that sacrifices significant amounts of data. Coupled with the class imbalance in the dataset, we were unable to achieve significant accuracy beyond the baseline. Our biggest gain in this project was the recapitulation of a computer science maxim – “garbage in, garbage out.” Regardless of how we tune or manipulate the features that we have already extracted, our tradeoff between sensitivity and specificity will always look horrendous.

4.2 Future Directions

Currently, no automated algorithm can replace an expert clinician’s intuition. However, improvements to current approaches may eventually facilitate the diagnosis of melanoma from dermoscope images at accuracies greater than previously achieved. In addition, with the increased availability of commercial dermoscope attachments for “at home diagnosis”, developing automated and distributed algorithms will greatly improve early detection. Therefore, refining our algorithm to include additional features, especially those of clinical significance, may bring us closer to accessible, automated diagnosis.

One area for future development is to include dynamic changes from sequential dermoscopic images. Buhl et al. integrates static and dynamic features, showing significant correlation between melanoma diagnosis and dynamic architectural changes¹⁵. Such algorithms have only recently been explored, but represent a powerful way to incorporate the dynamic criteria of clinical scores into an automated process.

Another area for refinement is to take the strength of local texture patterns into account. Riaz et al. use a novel methodology to use scale adaptive LBPs for feature extraction, attributing its increased performance to its ability to characterize local contrast and reflect the differential attention clinicians pay to high contrast textures⁷.

Finally, another possible avenue to explore to increase diagnostic accuracy is to use convolutional neural networks (CNNs). These networks are often used to classify images, with the advantage of not having to manually extract features. Instead, the network progressively narrows down the image into its most important features, and this set of features is fed into another neural network for classification. CNNs tend to be very hard to train, but have the ability to detect salient features that are not hard-coded.

Team Member Contributions

All members contributed equally in the implementation of the code and the writing of the report.

Citations

¹ Maglogiannis, Ilias, Elias Zafiropoulos, and Christos Kyranoudis, ‘Intelligent Segmentation and Classification of Pigmented Skin Lesions in Dermatological Images’, Lecture Notes in Computer Science (presented at the Hellenic Conference on Artificial Intelligence, Springer, Berlin, Heidelberg, 2006), pp. 214–23 <https://doi.org/10.1007/11752912_23>

² Key Statistics for Melanoma Skin Cancer. <https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html>.

³ J. Sikorski, “Identification of malignant melanoma by wavelet analysis,” in *Proceedings of Student/Faculty Research Day*, Pace University, 2004.

⁴ Maglogiannis, Ilias, Elias Zafiropoulos, and Christos Kyranoudis, ‘Intelligent Segmentation and Classification of Pigmented Skin Lesions in Dermatological Images’.

⁵ Tanaka, T., S. Torii, I. Kabuta, K. Shimizu, M. Tanaka, and H. Oka, ‘Pattern Classification of Nevus with Texture Analysis’, in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2004, I, 1459–62
<<https://doi.org/10.1109/IEMBS.2004.1403450>>

⁶ Masood, Ammara, and Adel Ali Al-Jumaily. *Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms*.

⁷ Riaz, F., et al. “Detecting Melanoma in Dermoscopy Images Using Scale Adaptive Local Binary Patterns.” *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 6758–61, doi:[10.1109/EMBC.2014.6945179](https://doi.org/10.1109/EMBC.2014.6945179).

⁸ First Symptoms of Melanoma. 2016, <https://www.skinvision.com/articles/first-symptoms-of-melanoma-the-three-symptom-types-to-watch-out-for>.

⁹ Kwak, Jin Tae, et al. “Efficient Data Mining for Local Binary Pattern in Texture Image Analysis.” *Expert Systems with Applications*, vol. 42, no. 9, June 2015, pp. 4529–39, doi:[10.1016/j.eswa.2015.01.055](https://doi.org/10.1016/j.eswa.2015.01.055).

¹⁰ N. S.-T. J. Cristianini, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, Cambridge, UK, 2000.

¹¹ Covalic. <https://challenge.kitware.com/#phase/5840f53ccad3a51cc66c8dab>.

¹² M. G. Fleming, C. Steger, J. Zhang et al., “Techniques for a structural analysis of dermatoscopic imagery,” *Computerized Medical Imaging and Graphics*, vol. 22, no. 5, pp. 375–389, 1998.

¹³ Q. Abbas, I. Fondo'n, and M. Rashid, “Unsupervised skin lesions border detection via two-dimensional image analysis,” *Computer Methods and Programs in Biomedicine*, vol. 104, no. 3,

p. -e15, 2011.

¹⁴ B. Erkol, R. H. Moss, R. J. Stanley, W. V. Stoecker, and E. Hvatum, “Automatic lesion boundary detection in dermoscopy images using gradient vector ow snakes,” *Skin Research and Technology*, vol. 11, no. 1, pp. 17–26, 2005.

¹⁵ Buhl, Timo, et al. “Integrating Static and Dynamic Features of Melanoma: The DynaMel Algorithm.” *Journal of the American Academy of Dermatology*, vol. 66, no. 1, Jan. 2012, pp. 27–36, doi:[10.1016/j.jaad.2010.09.731](https://doi.org/10.1016/j.jaad.2010.09.731).