STAT 542 Group Project: Skin Cancer Classification

Team leader: Xianbin Cheng (xianbin2); Team members: Ziqin Xiong (zxiong8), Jingyu Li (jli173)

December 16, 2019

Contents

Project Description and Summary

Skin cancer, including melanoma and nonmelanoma skin cancer, is one of the most prevalent carcinomas worldwide. Clinal confirmatory test of skin cancer often requires procedure such as a biopsy, which is invasive and time-consuming. Building reliable models that can classify skin cancer based on dermoscopic images may extensively accelerate the diagnosis process and prevent the suffering of a biopsy. This dataset contains 300 images taken with different equipment with different magnification. They have different resolution; some appears to be redder and others look bluer; some have black frames and ruler on it and others don't, etc. Hence before training models, images were preprocessed with cropping, resizing, blurring, and dilation to eliminate irrelevant objects such as black frames, hair, bubble, and ruler on the lens.

For classification model, logistic regression with LASSO, support vector machine and random forest were used. Input of each image is the concatenation of RGB chanel of processed 300*225 image, so the total dimension is 202500. Each model were tuned for best performance. And the accuracies are around 70%.

With the goal of enhancing interpretability of the models, skin lesion was automatically identified by border detection and 22 features were engineered to reflect skin lesion's shape characteristics (perimeter, area, border irregularity, shape compactness) and color characteristics (pixel mean and standard deviation of normalized RGB, HSV, CIE L*ab* color channels). Decision tree and logistic regression with LASSO model were then applied on these new features, which yield an overall accuracy of 60% and 74% respectively. These models have better interpretability and may facilitate rapid screening of skin cancer.

Question 1

Data Processing

The raw data are consisted of 300 color photos of different sizes. As the photos were taken with a dermoscope on skin with a variety of conditions, the goal of photo preprocessing is to eliminate irrelevant objects in the view, such as the black frame, the dark area beyond the ocular lens, the hair, the bubbles, and the ruler on the object lens. To start with, black frames and dark areas were identified by converting photos to grayscale and checking the pixels in each row/column. A row was considered a black frame when over 50% of the pixels

were black (normalized pixel value < 0.1) while a column was considered intersecting a dark area when over 20% of the pixels were black (Celebi *et al.*, 2008). These rows and columns were cropped from the original photo. Next, all the photos were further cropped and resized to 300 x 225 pixels with linear interpolation to reduce computation time. To remove hair and the ruler, image dilation was conducted with a size-3 square structuring element (Do Hyun Chung & Sapiro, 2000). Furthermore, bubbles were blurred by applying a 5 x 5 median filter (Masood & Al-Jumaily, 2013). Eventually, each photo was converted into a vector of 202,500 pixels, with pixels in R, G, B channels concatenating one another sequentially. An example is provided here to demonstrate the preprocessing effect (**Fig.1**, **Fig.2**).
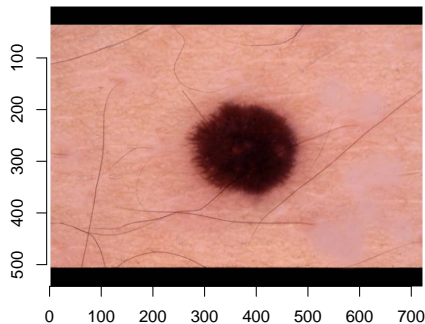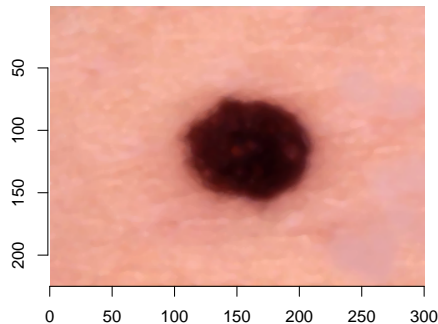
Fig.1 An example image before preprocessing.   **Fig.2 An example image after preprocessing.**

Classification Models Based On Pixels

To train the models, 70% of the data were treated as the training set and 30% of the data were treated as the testing set. As a preprocessing step, the data were scaled by centering and scaling. The seed was set at 123 to ensure reproducibility.

Logisitic Regression with LASSO

We used 10-fold cross validation with logistic regression and set "m" (Malignant) as the positive level. We let the function choose and tune the best tuning parameter, `lambda`. The `lambda.min` which achieves the minimum mean cross-validated error is 0.0352914, and we chose this lambda to validate the model (**Fig.4**).

Applying the model associated with the lambda we chose on the testing data, the accuracy of predicting the tumor to be benign and malignant achieved a moderately high accuracy,0.7, indicating that the logistic regression model was well-performed at identifying tumors. Meanwhile, the sensitivity was 0.6875, which was still moderately high, and it indicated this model was moderately effective in identifying malignant tumors. The specificity, 0.7142857 was similar to the sensitivity, and indicated this model was also moderately effective in identifying benign tumors. The ROC plot also demonstrated the relationship between sensitivity and specificity. We could see from the plot that both of them could achieve a relative high level at the same time, when both were around 0.7 ~ 0.75 (**Fig.5**).

Moving deeper, we found with lasso penalty, there were in total 33 pixels that had non-zero coefficients, indicating they were contributing to the final classification. 3 of them were from Red in RGB channels, 7 of them were from Green, and 23 of them were from Blue (**Table 1**). Just from the number of non-zero coefficients in the model, we assumed the color blue in each pixel would be more significant than others when identifying tumors. However, the number of color channel with nonzero coefficients couldn't tell much information about the relationship between pixel RGB and the tumor, so we need more intuitive features to identify them.

	Red	Green	Blue
# of contributing pixels	7	5	30

Table 1: number of contributing pixels in different channels

Fig.4 Plot of lambda versus coefficients.

Fig.5 ROC plot of the LASSO GLM model.
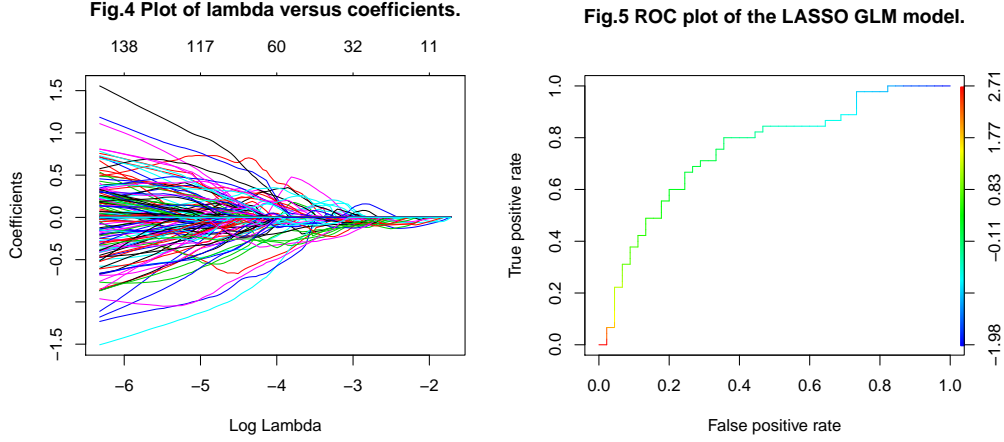
SVM with Radial Basis Function

A SVM model was trained with the radial basis function kernel. The parameter "cost" was tuned over 0.25, 0.5, 1 by 10-fold cross-validation and the model performance was evaluated by ROC. The tuning result suggested that the cost should be set at 1 as the cross-validation ROC was maximized at this point (**Fig.6**).

The final SVM model with cost = 1 could achieve a relatively high test accuracy of 0.7111111. Specifically, this model had a sensitivity of 0.7333333 and a specificity of 0.6888889, which indicates that the model would perform better at predicting a malignant melanoma than predicting a benign melanoma. On one hand, this model would be useful in malignant cancer screening as it has a lower false negative rate. On the other hand, its high false positive rate could impose a noticeable financial burden on patients who may have to go through unnecessary sequential testing.

Fig.6 Parameter tuning for SVM model
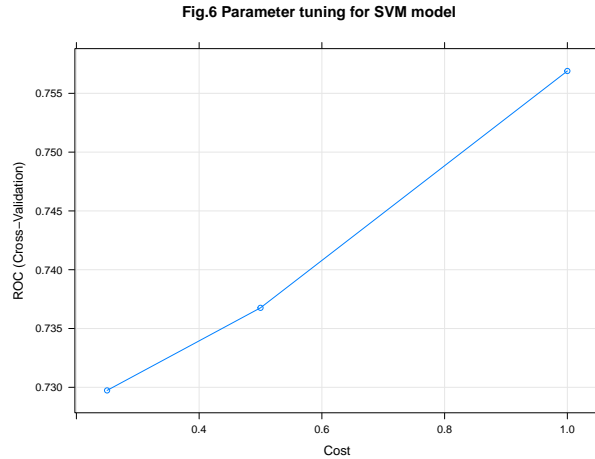
Random Forest

We also fitted random forest models. Here we used Python to fit the model in consideration of efficiency. Since the input here has extremely large dimension (p = 202500), the key parameters to tune are mtry (max

parameter to consider we split each tree) and max depth of each tree (max node $= 2^{\text{max depth}}$). We first tune the max depths when fixing mtry to be 1, i.e., we choose all features and no bootstrap is performed. We can see below that when max tree depth is 4 (max node size 16), the accuracy reach maximum of 0.7. (when training a random forests of 100 trees)
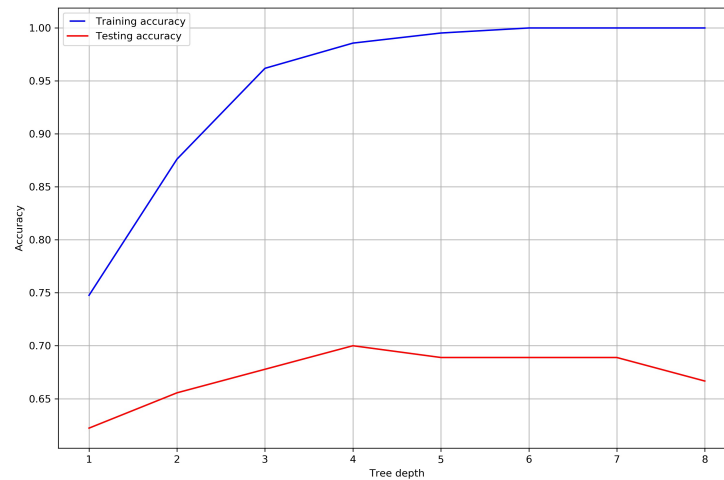
We then tune the mtry parameter: fixing tree depth to be 4 and train model with 20%, 40%, 60%, 80%, 100% features to consider when split the tree. We can see below that when we consider all features, the accuracy is highest. But the difference is small, considering that the randomness is still very high when we have 202500 dimensions and fit only 100 trees.
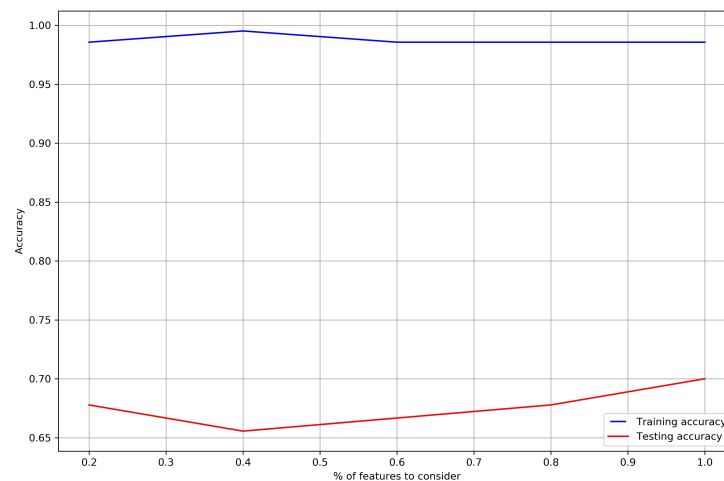
The performance of this final forest is:

F1 Score	Accuracy	Sensitivity	Precision	F1 Score
Final Random Forest Model	0.700	0.702	0.717	0.710

And below is a summary of Question 1.

Table 2: Summary of models in question 1.

Model	Parameters	Test_Accuracy	Sensitivity	Specificity
Logistic Regression with Lasso	lambda = 0.035	0.7000000	0.6875000	0.7142857
SVM with Radial Basis Kernel	sigma = 4.6e-06, C = 1	0.7111111	0.7333333	0.6888889
Random Forest	max.tree.depth = 4, n.trees = 100, mtry = 202500	0.7000000	0.7021277	0.6976744

Question 2

Literature Review

Among all the clinical diagnostic criteria proposed in the literature, the "ABCD" criteria stand out as the most widely-adopted method. The acronym stands for asymmetry, border irregularity, color variegation, and diameter > 6 mm. Alternative criteria like the Glasgow 7-point checklist also mentions relevant characteristics such as varying size, shape, color (blue-white veil structure) and the observation of bleeding and inflammation. (Rigel, Russak, & Friedman, 2010). It has also been reported that features such as vascular growth, thickness of the lesion, evolving lesion size are associated with higher risk of developing skin cancer (Martinez & Otley, 2001). Therefore, the shape and color of skin cancer may be crucial to classification.

Feature Engineering

Based on the literature review, several shape and color features have been engineered with goal of capturing the essence of each skin cancer image.

Before developing features, the skin lesion or the region of interest (ROI) was identified and the border of skin lesion was detected. Specifically, each image was converted to a grayscale image, which was then binarized by an automatically determined threshold. A grayscale pixel was converted to 1 if it exceeded the threshold and 0 if otherwise. This resulted in a black-and-white image where the black region represented the ROI and the white represented the background. An example of border detection is shown in **Fig.7**.

Shape features such as perimeter and area were first engineered. Perimeter (P) of an ROI was defined as the number of pixels on the border and area (A) was defined as the number of pixels covering the entire ROI. On top of these 2 basic shape features, irregularity index ($I = \frac{P^2}{4\pi A}$) and compactness (the ratio between the area of the ROI and the area of a circle with the same perimeter as the ROI, $C = \frac{4\pi A}{P^2}$) were constructed (Celebi *et al.*, 2007; Golston, Stoecker, Moss, & Dhillon, 1992).

Color features such as the mean and standard deviation of pixels in each color channel were constructed. Generally, each image was split by its 3 color channels and the pixels within the ROI were extracted and subjected to calculation of the mean and the standard deviation. To create color features possessing such characteristics as invariance to illumination intensity, perceptual uniformity, and decoupling of chrominance and luminance, we have chosen 3 different color spaces (normalized RGB, HSV, CIE Lab) to calculate the means and standard deviations (Celebi *et al.*, 2007; Green, Martin, Pfitzner, O'Rourke, & Knight, 1994).

In summary, a total of 22 features were engineered for each image including 4 shape features and 18 color features.

Fig.7 Demonstration of border detection.
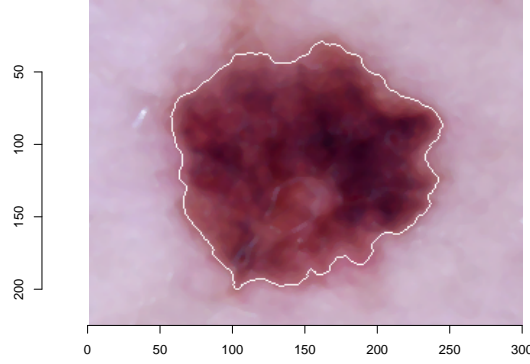
Classification Models Based On New Features

Tree Model

A single tree model was trained with the complexity parameter as the tuning parameter. 5-fold cross-validation was used to tune over such complexity parameter values as 0, 0.001, 0.005, 0.01, and 0.05. The cross validation result shows that the ROC was maximized when `cp = 0.01` and thus this was chosen for the final tree model (**Fig.8**). The final model has an overall accuracy of 0.6 with the sensitivity being 0.6222222 and the specificity being 0.5777778. Evidently, this model has a probability of 62% to correctly classify a malignant mole at the cost of producing many false positives with a probability of 42%.

A major advantage of using a tree model is that the insights on how classification is performed can be revealed by analyzing the decision tree (**Fig.9**). Clearly, perimeter played a major role in classification where a perimeter $>= 662$ pixels might lead to higher probability (69%) of having a malignant skin cancer. This is consistent with the "ABCD" rule in the literature that associates malignancy with a larger diameter. Further down the branch on the right, the mean values of the normalized blue and the green channel as well as the irregularity index were crucial for classification. Generally, a mole is classified as malignant with a probability of at least 71% when its image's mean value of normalized blue channel is < 0.5 (which means darker in blue), its mean value of normalized green channel is $>= 0.6$ (which means brighter in green), and its irregularity index is $>= 5.2$. These decision rules are also consistent with the "ABCD" rule that color variegation, blue-white veil, and irregular border may indicate malignancy (Rigel et al., 2010). With the branch on the left, the standard deviations of hue and value were the driving forces for classification. Evidently, larger standard deviations of hue ($>= 96$) and value ($>= 0.097$) may lead to classifying a model as malignant with a probability of 69%.

In summary, our tree model may be beneficial to doctors who need assistance with rapid screening for malignant skin cancer as it provides easily-understandable decision rules and relatively satisfactory accuracy.
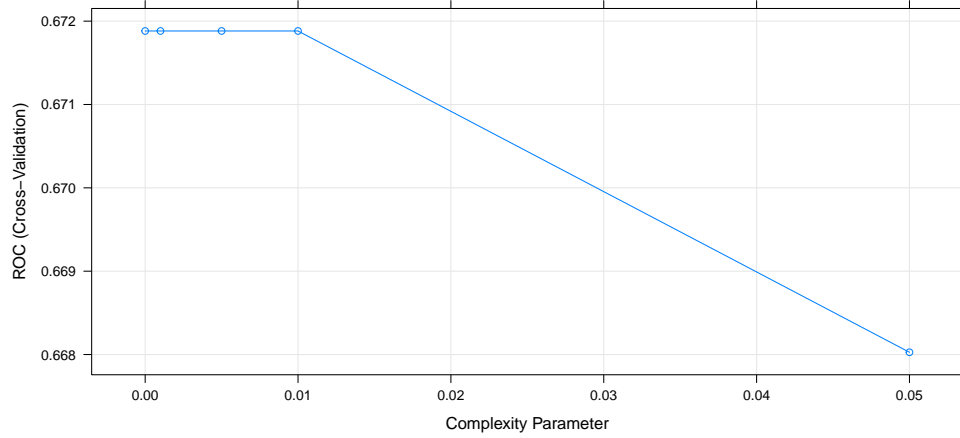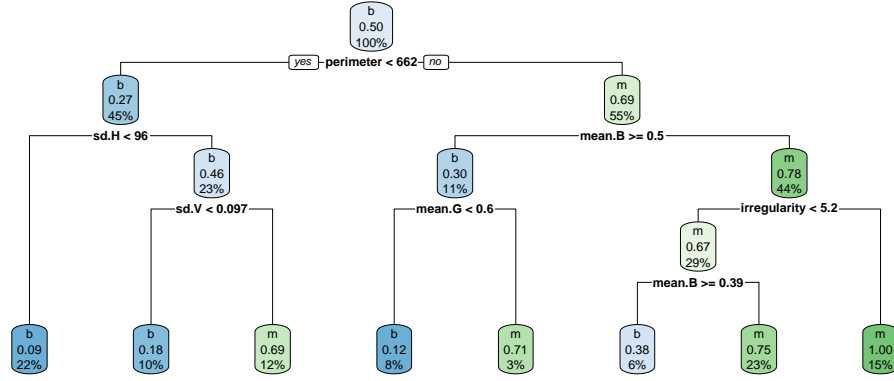
6

Fig.8 Cross–validation of the tree model.

Fig.9 The decision tree.

Logistic Regression with LASSO

We again used 10-fold cross validation with logistic regression and set "m" (Malignant) as the positive level. We again let the function choose and tune the best tuning parameter, lambda. The lambda.min which achieves the minimum cross-validated error is 0.0352914, and we chose this lambda to validate the model (**Fig.10**).

The accuracy of predicting the tumor for the testing data was 0.7444444, a little higher than that in Question 1. Both sensitivity and specificity increased compared to those in Question 1. The ROC plot also showed a more desired curve indicating the sensitivity and specificity. By looking at these statistics, we assumed that the feature engineering process successfully helped increase the accuracy of identifying tumors (**Fig.11**). Lasso helped us to do variable selection. Under the lambda we chose, only 9 out of the 22 features had nonzero coefficients and thus contributed to the final identification. They are "perimeter", "area", "compactness", "mean.B", "sd.B", "mean.H", "sd.H", "sd.S" and "sd.V". Then we took a deeper look at their individual coefficients (**Table 3**). We surprisedly found that the coefficients of mean.B (mean value of blue channel) and sd.S (standard deviation of saturation) had much more lower coefficients than any one else, and both coefficients were negative. This meant, when the mean value of the blue channel or the standard deviation of the saturation value of a mole was high, then the mole was much more likely to be a benign one. The third feature with relatively significant coefficient was compactness (ratio between the area of the mole and the area of the circle with the same parimeter), with coefficient about -3, which indicated that higher compactness of the mole contributed to a higher probability of being benign.

Table 3: Nonzero coefficients of the logistic regression with LASSO model.

	Coefficient
(Intercept)	3.1546678
perimeter	0.0007479
area	0.0000305
compactness	-2.9700950
mean.B	-7.2892043
sd.B	-0.3887360
mean.H	-0.0005017
sd.H	0.0051566
sd.S	-7.0189157
sd.V	0.2424047

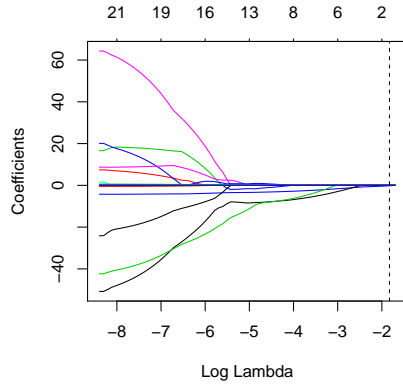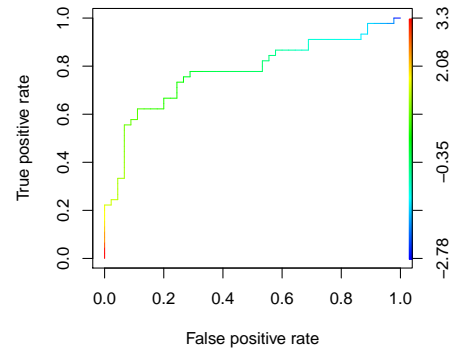**Fig.10 Coefficients VS lambda.**

Fig.11 ROC curve of the logistic regression with LAS

Here is a summary table showing model performance for question 2.

Table 4: Summary of models in question 2.

Model	Parameters	Test_Accuracy	Sensitivity	Specificity
Tree model	Cp = 0.01	0.6000	0.6222	0.5778
Logistic Regression with LASSO	lambda = 0.015	0.7444	0.7292	0.7619

References

Celebi, M. E., Kingravi, H. A., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V., Moss, R. H., ... Menzies, S. W. (2008). Border detection in dermoscopy images using statistical region merging. Skin Research and Technology: Official Journal of International Society for Bioengineering and the Skin (ISBS) [and] International Society for Digital Imaging of Skin (ISDIS) [and] International Society for Skin Imaging (ISSI), 14(3), 347–353. https://doi.org/10.1111/j.1600-0846.2008.00301.x

Celebi, M. E., Kingravi, H. A., Uddin, B., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V., & Moss, R. H. (2007). A methodological approach to the classification of dermoscopy images. Computerized Medical Imaging and Graphics, 31(6), 362–373. https://doi.org/10.1016/j.compmedimag.2007.01.003

Do Hyun Chung, & Sapiro, G. (2000). Segmenting skin lesions with partial-differential-equations-based image processing algorithms. IEEE Transactions on Medical Imaging, 19(7), 763–767. https://doi.org/10.1109/42.

875204

Golston, J. E., Stoecker, W. V., Moss, R. H., & Dhillon, I. P. S. (1992). Automatic detection of irregular borders in melanoma and other skin tumors. Computerized Medical Imaging and Graphics, 16(3), 199–203. https://doi.org/10.1016/0895-6111(92)90074-J

Green, A., Martin, N., Pfitzner, J., O'Rourke, M., & Knight, N. (1994). Computer image analysis in the diagnosis of melanoma. Journal of the American Academy of Dermatology, 31(6), 958–964. https://doi.org/10.1016/S0190-9622(94)70264-0

Martinez, J.-C., & Otley, C. C. (2001). The Management of Melanoma and Nonmelanoma Skin Cancer: A Review for the Primary Care Physician. Mayo Clinic Proceedings, 76(12), 1253–1265. https://doi.org/10.4065/76.12.1253

Masood, A., & Al-Jumaily, A. A. (2013). Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms. International Journal of Biomedical Imaging, 2013. https://doi.org/10.1155/2013/323268

Rigel, D. S., Russak, J., & Friedman, R. (2010). The evolution of melanoma diagnosis: 25 years beyond the ABCDs. CA: A Cancer Journal for Clinicians, 60(5), 301–316. https://doi.org/10.3322/caac.20074