

Novel Screening Test for Melanoma Based on Computational Image Analysis

**Amanda Zong
Hunter College High School
71 East 94th street
New York, NY 10128**

Novel Screening Test for Melanoma Based on Computational Image Analysis

Amanda Zong

Hunter College High School, New York City, New York

Mentor– Daniel Gareau (Rockefeller University); Principal Investigator – James G. Krueger (Rockefeller University); Teacher– Gilana Reiss (Hunter College High School)

Melanoma is the deadliest form of skin cancer, resulting in approximately 10,000 deaths annually in the United States. Early diagnosis by dermatologists is essential to decreasing mortality rates and healthcare costs. Currently, patients with skin lesions delay visiting dermatologists due to inability to identify problematic lesions, cost, time, discomfort, and inconvenience. Smartphone applications employing computational tools to detect melanoma from images could dramatically improve early diagnosis. However, the best commercially available computational tool only correctly identifies melanoma lesions 72% of the time (true-positive rate) and benign lesions 36% of the time (true-negative rate), impeding diagnosis for 28% of patients. Consequently, this study developed a novel machine learning system to select red-green-blue features from a dermoscopic dataset of pigmented lesions and implement an ensemble of seven machine learning algorithms to detect melanoma based on images alone. To improve diagnostic accuracy, thirty new melanoma detecting features were developed and added to the existing set of features, surveying for lesion boundary regularity and brightness, image brightness distribution, and lesion color variation. The machine learning model, coupled with twenty-one new statistically significant features ($p < 0.05$), yielded a true-positive rate of 98% and a true-negative rate of 65%, a significant improvement over commercially available tools. The machine learning model reduced the percentage of melanoma patients who risk delayed treatment to 2% (14-fold decrease), significantly improving the early detection of melanoma and increasing patient survival rates. Furthermore, it decreased false positive diagnoses by 29%, reducing unnecessary worry and healthcare costs for many patients.

ACKNOWLEDGEMENTS

I would like to thank the Laboratory of Investigative Dermatology at Rockefeller University for providing me with the space and equipment to conduct my research. Specifically, I thank Dr. Daniel Gareau for mentoring and supervising me throughout the course of this project. Additionally, I would like to thank Dr. Jason Hawkes for teaching me about melanoma detection in clinical practice and Joel Correa da Rosa for teaching me about biostatistics. I thank the Rockefeller Summer Science Research Program for providing me with seminars on scientific methods and ethics. My thanks are also due to Ms. Gilana Reiss and Mr. Frankel, the science research coordinators at my high school, for support and assistance in editing my paper and biography.

STATEMENT

Research involving pre-existing/retrospective data from human subjects in a de-identified/anonymous format was conducted under the supervision of an experienced researcher and followed state and federal regulatory guidance applicable to the humane and ethical conduct of such research. This data was appropriately de-identified before being given to the student researcher and is in compliance with all privacy and HIPAA laws.

TABLE OF CONTENTS

List of Tables.....	1
List of Figures.....	1
Introduction.....	2
Materials and Methods.....	4
Results.....	13
Discussion and Conclusions.....	16
References and Literature Cited.....	18

LIST OF TABLES

TABLE 1	P-values of 30 New Features	11
TABLE 2	Performance of the Machine Learning Model with the Old and New Features	14
TABLE 3	Comparison of Diagnoses	17

LIST OF FIGURES

FIGURE 1	30 Added Features of the Image Processing Algorithm	9
FIGURE 2	Melanoma Classification Model	10
FIGURE 3	Machine Learning Model	10
FIGURE 4	Performance of Machine Learning Model with Old and New Features	13

1. INTRODUCTION

Melanoma is a rare, aggressive form of skin cancer, accounting for almost 75% of all skin cancer related deaths, despite representing only 1% of all skin cancer cases (American Cancer Society, Cancer Facts and Figures 2016). From 1982-2011, the incidence of melanoma doubled in the U.S. and is forecasted to continue to rise (Guy 2015). In 2016, an estimated 75,000 new cases of invasive melanoma and 10,000 melanoma-specific deaths are expected in the U.S. (National Cancer Institute, SEER Stat).

Currently, healthcare practitioners biopsy concerning pigmented lesions for histopathologic evaluation (Rigel 2010; American Cancer Society, Cancer Facts 2016). High mortality rates are associated with a delayed diagnosis (American Cancer Society, Cancer Facts 2016; Hu 2006). The 5-year relative survival rate is 100% for Stage 0 melanoma, yet drops to 60.3% for Stage 3 and 16.2% for Stage 4 (National Cancer Institute, SEER Training). About 17% of melanoma cases are metastatic, diagnosed at either Stage 3 or Stage 4 (National Cancer Institute, SEER Training). Treatment for melanoma diagnosed at later stages often involves complications and severe physical and psychological side effects, since doctors must not only excise the original melanoma site, but also the areas where the melanoma has spread (American Cancer Society, Treatment 2016; National Cancer Institute, Side Effects 2015). In addition, treatment costs are approximately 10 times higher for melanoma cases diagnosed in later stages than for cases diagnosed in earlier stages (Guy 2012). The annual per-patient cost for Stage 4 melanoma is \$34,103–\$152,244 in the United States, while the annual per-patient cost for Stage 1 melanoma is \$2,169 –\$14,499 (Guy 2012). Early diagnosis of melanoma is essential to decreasing mortality rates and healthcare costs.

Early diagnosis is correlated with the frequency of skin examinations (Little 2012; Hu 2006). Although the American Cancer Society recommends a periodic skin-cancer related check-up beginning from age 20, only 9% of people ages 25-44 and 14% of people ages 45-64 visit the dermatologist annually (American Cancer Society 2016; American Cancer Society 2009; National Center for Health Statistics 2009). Many regions in the United States exhibit a severe shortage of dermatologists, with less one dermatologist serving per population of 100,000 (Yoo 2010). A patient must wait up to several weeks for an appointment with their dermatologist from when they first discover a pigmented lesion, delaying diagnosis and increasing stress (Rosenthal 2014; Cancer Council NSW). Moreover, a single skin-cancer

check-up can cost between \$100-200 (Losina 2007). Studies have shown that people with lower socioeconomic status are less likely to receive skin examinations, and that a low socioeconomic status is correlated to a more-advanced stage at diagnosis (Little 2012; Hu 2006). In order to improve early diagnosis rates, it is necessary to develop a melanoma detection tool that people will have equal access to, regardless of socioeconomic or geographic factors. Furthermore, people often report feeling uncomfortable when dermatologists examine their skin, which is a significant factor that causes people to rely on self-examinations in favor of visiting a dermatologist (Geller 2003). In conclusion, current diagnosis methods are expensive, time-consuming, and inconvenient, creating the need for better options.

Automated, self-operated diagnosis tests can increase accessibility, convenience, and efficiency, while decreasing costs. The introduction of home pregnancy tests in the 1950s encouraged women to seek their doctor at an earlier stage in their pregnancies, which increased the early detection rate of ectopic pregnancies three-fold, facilitating treatment (Tone 2012; Lawlor 1993). Self-operated melanoma diagnostic tests would likely encourage patients to seek out dermatologists at earlier stages of melanoma, which would facilitate treatment and increase survival rates. This method of diagnosis would be free, faster, and more convenient than clinical methods. A person who is concerned about a pigmented skin lesion would be able to take and upload a picture of it using their phone, upon which a computer program can extract features from the image of the lesion and classify it as either melanoma or nevi (benign).

The main method used to quantify accuracy of diagnosing melanoma is by plotting a receiver operating characteristic (ROC) curve, which shows the sensitivity and specificity of diagnosis at various thresholds (Hajian-Tilaki 2013; Kumar 2011; Lambert 2016; Zou 2011). For each lesion, the probability of melanoma is computed as a score. A threshold, or cut-off value, is then set so that lesions with scores above the threshold are diagnosed as melanoma and lesions with scores below the threshold are diagnosed as nevi (benign moles) (Hajian-Tilaki 2013). The probability that a diagnostic method correctly identifies melanoma is known as the sensitivity (Zou 2011). The probability that the diagnostic method correctly identifies nevi is known as the specificity (Zou 2011). The AUC, or the area under the ROC curve, measures the overall accuracy of diagnosis. Normally, the AUC is between 0.5 and 1. An AUC of 0.5 means the diagnosis test is no more accurate than random chance, while an AUC of 1 corresponds to perfect diagnosis at all thresholds, with 100% sensitivity and 100% specificity (Zou 2011).

Unfortunately, the existing computer programs do not have high sensitivity or specificity. None of the currently available apps for the detection of melanoma have been validated for diagnostic accuracy by developers (Kassianos 2015). In a recent study, researchers reviewed all melanoma-detection applications currently available for smartphones that implemented image analysis algorithms on uploaded photos of pigmented lesions and predicted the risk of melanoma (Wolf 2013). While the researchers identified these applications, the best of those programs achieved 72% sensitivity at 36% specificity, impeding diagnosis for 28% of patients (Wolf 2013). It is necessary to improve accuracy of diagnosis in order to bring these tools into the market.

In this study, a novel machine learning model was developed that processed features from images of pigmented lesions in order to diagnose melanoma at early stages. Recently, Gareau et al. developed a computational model that examined fifty distinct features of pigmented lesions from photos, such as the degree of asymmetry, pigmented network patterns, and color variation (Gareau). In this study, to improve diagnostic accuracy of that model, thirty new melanoma detecting features were developed and added to the existing set of features, surveying for lesion boundary regularity and brightness, image brightness distribution, and lesion color variation. Seven machine learning algorithms were applied on the data, and the mean probability of melanoma was computed for each lesion. The optimal threshold, or cut-off value, was determined by setting a bar of at least 98% sensitivity. The predictions of the classification model were then compared with the standard histopathologic diagnoses in order to evaluate the sensitivity, specificity, and AUC of the classification model.

2. METHODS

120 dermoscopic images of pigmented skin lesions were obtained for this study. 60 images were of in situ (Stage 0) melanomas and 60 images were of nevi. The image analysis algorithms developed in this study successfully ran on 112 of the 120 images. Data from 8 of the images could not be obtained due to either presence of hair or a lesion border that extended beyond the image border.

Recently, Gareau et al. developed a computational model that examined 50 features of pigmented lesions from images, such as the degree of asymmetry, pigmented network patterns, and color variation (Gareau). 31 of those features were selected in this study to be included in the new model. The discarded 19 features related to the fitting of the variation in brightness at the lesion border with a Gauss error

function. In this study, new sigmoid and linear fitting functions were developed to model variation in brightness at the lesion border.

In total, 30 new melanoma-detection features were developed and added to the existing computational model. 28 of the 30 new features were applied to single color channels. The 2 color identification features were applied to data from the red-green-blue (RGB) channel triad. All of the 30 features mentioned below were conceived of, designed, and implemented in this study after observing distinct characteristics of melanomas and nevi from extensive collections of dermoscopy images (Johr 2004; Bowling 2012). The features are described in detail below:

2.1 Fitting the Lesion Boundary Shape with an Ellipse

The coordinates of the points on the lesion border were obtained using the `bwboundaries` function in Matlab. An ellipse was fitted onto the lesion border using the fitting algorithm, `fminsearch()`, in Matlab. The equation of the ellipse is as follows:

$$Ell(x, y) = \frac{((x - cx) * \cos\theta - (y - cy) * \sin\theta)^2}{a^2} + \frac{((x - cx) * \sin\theta + (y - cy) * \cos\theta)^2}{b^2}$$

where (x,y) are the Cartesian coordinates of a point on the lesion border, (cx,cy) are the Cartesian coordinates of the center of the lesion, a is half of the distance of the major axis, b is half of the distance of the minor axis, and theta is the degree of rotation of the fitted ellipse.

The fitting algorithm found the parameters cx, cy, a, b, and theta while seeking to minimize the error of the fit. A total of 2 features were computed based on the ellipse fits of the lesion border (Figure 1; Table 1).

2.2 Image Pixel Brightness Distribution

The intensity of the all pixels in the image, corresponding to the brightness of the pixels, were scaled to a range of 0 to 1, where 0 is black and 1 is white. The binning algorithm `hist()` was then applied to find 50 evenly spaced bins that represented the full range of pixel intensity values in the image. The locations of the bins with the greatest number of pixels, representing the most common pixel intensities, are denoted by the vector *pks*, while the locations of all of the bins are contained the vector *loc*. A total of 2 features were computed based on the image brightness distribution (Figure 1; Table 1).

2.3 Fitting the Variation in Brightness across the Lesion Boundary

One of the original features included a radial arm that starts from the centroid of the lesion, connects to a point on the lesion border, extends past the lesion border into normal tissue (Postdoctoral Researcher 2016). The radial arm moves through 200 angular analysis positions from 0 to 2π around the lesion. At each angle, the intensities of the pixels along the radial arm are sampled. The code for this radial arm described above was used with consent from Gareau et al. All further methods listed below were designed, coded, and implemented in this study.

2 different types of fits were developed for the pixel intensities sampled at each angle. For both types of fits, the error, $err(x_\theta)$, was defined as follows at each sampling angle:

$$err(x_\theta) = \sum_{\theta=0}^{200} \left| \sqrt{f(x_\theta) - pe(x_\theta)} \right|^2$$

where $f(x_\theta)$ is a fitting function for the intensity of the pixels sampled, $pe(x_\theta)$ is a vector containing the actual intensities of the pixels sampled, and x_θ is a vector containing the numbering of the pixels from 1 to 200.

The first was a sigmoid fit, defined as follows:

$$Sf(x_\theta) = a + \frac{1}{b + e^{cx_\theta}}$$

where a, b, and c are parameters found using the fitting algorithm, `fminsearch()`, which sought to minimize the error of the fit.

In total, 6 features were computed for only the angular analysis positions in which the sigmoid fit was able to converge under 200 iterations.

3 linear functions were also developed to fit the pixel intensities sampled at each angle. At each angle, the beginning linear function, L_{beg} , fit the intensities from the centroid of the lesion to the pixel with the minimum intensity within the lesion. The middle linear function, L_{mid} , fit the intensities from the pixel with the minimum intensity, within the lesion, to the pixel with the maximum intensity, outside the lesion. The end linear function, L_{end} , fit the intensities from the pixel with the maximum intensity to the last pixel outside the lesion. Each linear function resembled the equation:

$$Lf(x_{\theta}) = b + mx_{\theta}$$

where x_{θ} is the numbering of the pixels starting at 1, b is the y-intercept, and m is the slope.

The parameters b and m were found using the fitting algorithm, `fminsearch()`, which sought to minimize the error of the fit. For each of the 3 linear fits, 6 features were computed. Therefore, a total of 18 features were computed based on the linear fits (Figure 1; Table 1).

2.4 Identifying Colors in Lesion

Sample images containing the colors light brown, dark brown, red, blue-gray, black, or white, which are colors most commonly associated with melanoma (Chen 2003; Seidenari 2010), were identified by human eye. For each color, the mean of the RGB intensities of the pixels corresponding to that color was computed, as well as the standard deviation, which were then transformed into CIELAB 1976 $L^*a^*b^*$ coordinates using the Matlab function `rgb2lab()`. CIELAB is regarded as a more perceptually uniform color space, which makes the spatial arrangement of clusters of colors more representative of how similar they are by human perception (Tominaga 1992; Weatherall 1992). In each image, each RGB pixel in the lesion was mapped onto the CIELAB color space. If the distance from the sample pixel to one of the 6 indicative colors was less than the standard deviation threshold for that color, the pixel would be identified as that color. A total of 2 features were computed based on the color identification within the lesion (Figure 1; Table 1).

2.5 Machine Learning Model

First, a 2-tailed t-test was run on values from the 3 color channels for each of the 61 features. For each single-color feature, the p-values from the 3 color channels were compared and the data from the color channel with the most significant p-value was chosen to represent that feature. If the most significant p-value was less than .05, that feature was not included in the machine learning model. The input matrix for the machine learning model contained as many rows as the number of images in the datasets and as many columns as the number of significant features. The machine learning model was composed of 7 algorithms: logistic regression, neural networks, support vector machines (SVM), decision trees, random forest, k-nearest neighbors, and linear discriminant analysis (Figure 3). Logistic regression and neural networks were manually coded for.

The original code written for the machine learning model was published on Github and is linked here: <https://github.com/amqzong/Machine-Learning-Model>

h_θ , the predicted diagnosis of x , is computed as follows for logistic regression:

$$h_\theta = \frac{1}{1 + e^{-\theta x}}$$

where θ is the vector of unknown parameters and x is the input matrix of features. The parameters for logistic regression were found by using the fitting algorithm, `fminsearch()`, in order to minimize the cost function. The cost function, J , for logistic regression was first implemented using the standard equation:

$$J = -\frac{1}{m} \left[\sum_{i=1}^m y_i \log h_\theta(x_i) + (1 - y_i) \log(1 - h_\theta(x_i)) \right] + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2$$

where y is the gold standard diagnosis, m is the number of images, n is the number of features, and λ is the regularization parameter.

Gradient descent was used to minimize the cost function and programmed using the standard equation:

$$\frac{d}{d\theta_j} J(\theta) = (h_\theta(x) - y)x_j + \frac{\lambda}{m} \theta_j$$

The neural network was composed of 1 input layer, 1 hidden layer, and 1 output layer. The input layer contained the same number of neurons as the number of features in the data. The hidden layer contained half as many neurons as the number of features in the input layer. The output layer contained 2 neurons, one for melanoma and one for nevi. Backpropagation was used to compute the gradients for the cost function. The cost function for neural networks was the same as that of logistic regression. For both logistic regression and neural networks, the optimal value for the regularization penalty λ was determined through experimentation from 0.1 to 1 by an increment of 0.05, while seeking to minimize the cost function when applying the parameters found on the test set of images.

The SVM algorithm was implemented using the Matlab function `fitcsvm()` to find the optimal parameters for the input matrix using a linear kernel. The decision tree algorithm was implemented using the Matlab function `fitctree()`. The random forest algorithm was implemented using the Matlab function `TreeBagger()`, with an ensemble of 10 decision trees. For decision tree and random forest, the cost of misclassifying melanomas was set to be twice as great as the cost of misclassifying nevi. This was set to increase sensitivity at the cost of specificity. The linear discriminant analysis algorithm was implemented using the Matlab function `fitcdiscr()`. The k-nearest neighbors algorithm was implemented using the Matlab function `fitcknn()`, which was set to find the five nearest neighbors for classifying each point.

To obtain the melanoma probability for each image, 1000 iterations of the machine learning model were run. In each iteration, 80% of the images were randomly selected as the training set, while the remaining 20% of the images were used as the test set to evaluate the parameters determined from running the algorithms on the training set. For each image, the probability scores from all the iterations were averaged for the final probability of melanoma for each image. The final probabilities were then compared with the standard histopathologic diagnoses for the 112 images. This comparison was used to plot a receiver operating characteristic (ROC) curve. The optimal threshold for classifying melanoma was determined by choosing the highest specificity at 98% sensitivity (Figure 2; Figure 3).

Figure 1: 30 Added Features of Image Processing Algorithm

The features can be separated into four categories, surveying information about lesion boundary regularity, brightness distribution in the image, brightness variation across the lesion boundary, and color identification within the lesion. The diagram also includes the method used to extract features for each category, and the total number of features in each category.

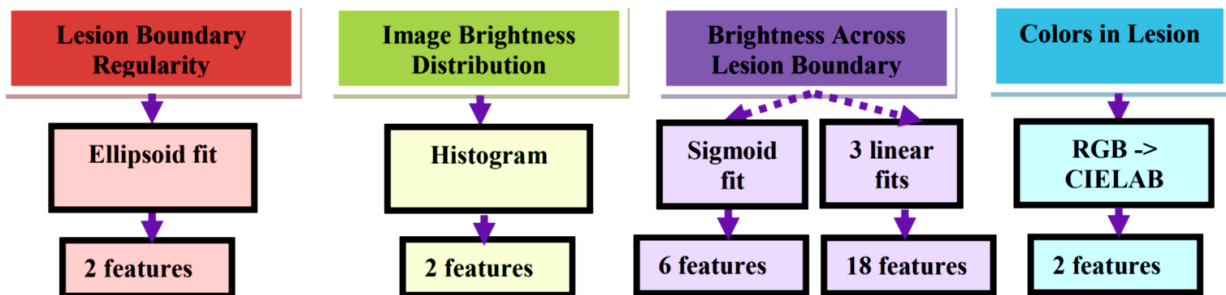


Figure 2: Melanoma Classification Model

Image processing algorithms yielded data for 112 images. A machine learning model then computed the probability that the images were melanoma.

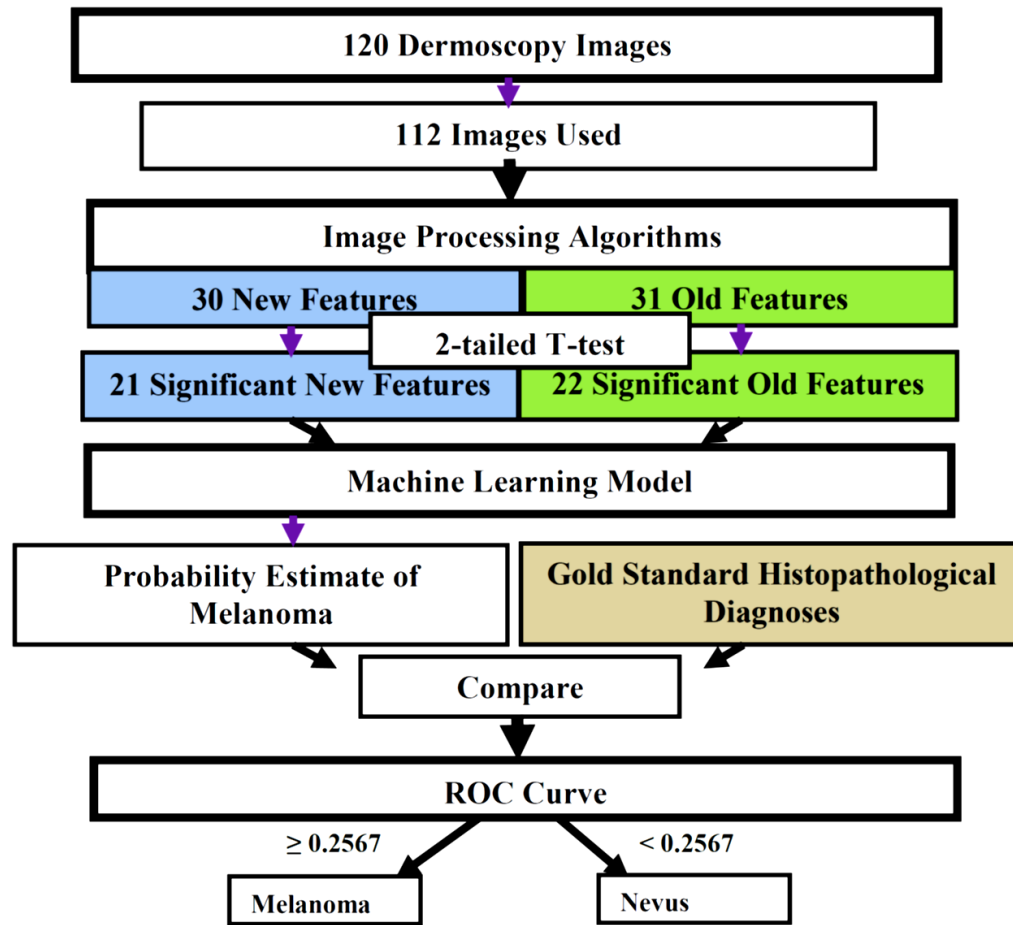


Figure 3: Machine Learning Model

An ensemble of 7 machine learning algorithms was implemented on the images of melanoma and nevi. Each algorithm produced a melanoma probability for each image. The mean of the melanoma probabilities from all 7 algorithms was used for predicting whether the image was of melanoma or nevi.

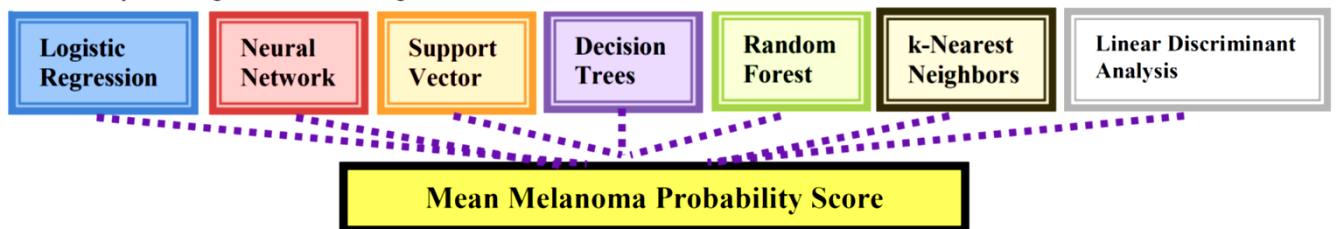


Table 1: P-values of 30 New Features

The 30 added features are listed in order of decreasing significance. The type of feature, equation, parameter obtained, explanation, color channel, and p-value of each feature are provided. The last 4 columns present the mean and standard deviation for melanoma (mel) and for nevi (nev) for each feature.

Type of Feature	Equation	Parameter	Explanation	RG B	p-value	Mel Mean	Mel Std	Nevi mean	Nevi std
Linear Border Fit	$\text{std}_{\text{err,Lmid}} = \sigma(\text{err}_{\text{Lmid}})$	$\text{std}_{\text{err,Lmid}}$	Standard deviation of the error, err, of the middle linear fit, L_{mid} .	3	2.68 E-08	5.05 E+02	1.56 E+02	3.40 E+02	1.33 E+02
Sigmoid Border Fit	$\text{mean}_{\text{err,Sf}} = \frac{\sum_{\theta=0}^{2\pi} \text{err}_{\text{Sf}}(\theta)}{200}$	$\text{mean}_{\text{err,Sf}}$	Mean of the error, err, of the sigmoidal fit, Sf.	3	1.51 E-06	1.26 E+03	4.42 E+02	8.77 E+02	3.44 E+02
Sigmoid Border Fit	$\text{std}_{\text{err,sf}} = \sigma(\text{err}_{\text{sf}})$	$\text{std}_{\text{err,sf}}$	Standard deviation of the error, err, of the sigmoidal fit, Sf.	3	2.65 E-05	4.89 E+02	2.78 E+02	2.98 E+02	1.67 E+02
Linear Border Fit	$\text{mean}_{\text{m,Lend}} = \frac{\sum_{\theta=0}^{2\pi} \text{m}_{\text{Lend}}(\theta)}{200}$	$\text{mean}_{\text{m,Lend}}$	Mean of the parameter m for the end linear fit, L_{end} .	1	2.90 E-05	-1.68 E-01	1.21 E-01	-9.16 E-02	4.74 E-02
Image Brightness Distribution	$\text{maxdif} = \max(pks - \text{median}(\text{loc}))$	maxdif	The maximum of the absolute value of the differences of the locations of the peaks from the median pixel intensity.	1	2.29 E-04	1.79 E-01	7.96 E-02	1.27 E-01	6.11 E-02
Image Brightness Distribution	$\text{sumdif} = \sum(pks - \text{median}(\text{loc}))$	sumdif	The sum of the absolute value of the differences of the locations of the peaks from the median pixel intensity.	1	3.50 E-04	1.91 E-01	1.00 E-01	1.31 E-01	6.88 E-02
Color	--	Cnum	The total number of colors identified in the lesion.	1,2,3	3.71 E-04	3.28 E+00	7.50 E-01	2.84 E+00	5.01 E-01
Linear Border Fit	$\text{mean}_{\text{err,Lmid}} = \frac{\sum_{\theta=0}^{2\pi} \text{err}_{\text{Lmid}}(\theta)}{200}$	$\text{mean}_{\text{err,Lmid}}$	Mean of the error of the fit, err, of the middle linear fit, L_{mid} .	3	6.79 E-04	1.14 E+03	3.75 E+02	9.06 E+02	3.43 E+02
Linear Border Fit	$\text{std}_{\text{err,Lbeg}} = \sigma(\text{err}_{\text{Lbeg}})$	$\text{std}_{\text{err,Lbeg}}$	Standard deviation of the error, err, of the beginning linear fit, L_{beg} .	3	1.12 E-03	2.72 E+02	3.14 E+02	1.13 E+02	1.62 E+02
Linear Border Fit	$\text{mean}_{\text{err,Lbeg}} = \frac{\sum_{\theta=0}^{2\pi} \text{err}_{\text{Lbeg}}(\theta)}{200}$	$\text{mean}_{\text{err,Lbeg}}$	Mean of the error, err, of the beginning linear fit, L_{beg} .	3	2.07 E-03	1.74 E+02	2.26 E+02	7.12 E+01	9.02 E+01
Linear Border Fit	$\text{mean}_{\text{m,Lbeg}} = \frac{\sum_{\theta=0}^{2\pi} \text{m}_{\text{Lbeg}}(\theta)}{200}$	$\text{mean}_{\text{m,Lbeg}}$	Mean of the parameter m for the beginning linear fit, L_{beg} .	3	2.17 E-03	1.63 E+00	1.92 E+00	2.82 E+00	2.06 E+00
Linear Border Fit	$\text{std}_{\text{err,Lend}} = \sigma(\text{err}_{\text{Lend}})$	$\text{std}_{\text{err,Lend}}$	Standard deviation of the error, err, of the end linear fit, L_{end} .	3	2.43 E-03	1.35 E+02	1.60 E+02	6.27 E+01	6.54 E+01
Sigmoid Border Fit	$\text{mean}_{\text{c,Sf}} = \frac{\sum_{\theta=0}^{2\pi} \text{c}_{\text{Sf}}(\theta)}{200}$	$\text{mean}_{\text{c,Sf}}$	Mean of the parameter c for the sigmoidal fit, Sf.	1	2.87 E-03	-3.82 E-02	6.51 E-03	-3.41 E-02	7.82 E-03

Type of Feature	Equation	Parameter	Explanation	RG B	p-value	Mel Mean	Mel Std	Nevi mean	Nevi std
Linear Border Fit	$\text{std}_{m,Lend} = \sigma(m_{Lend})$	$\text{std}_{m,Lend}$	Standard deviation of the parameter m for the end linear fit, L_{end} .	1	4.88 E-03	4.53 E-01	5.45 E-01	2.16 E-01	2.78 E-01
Ellipse Border Fit	$\text{Ell}_{err} = (1 - ec)^2;$	Ell_{err}	Error, err, of ellipsoid fit, Ell	2	5.31 E-03	6.46 E+01	7.81 E+01	3.15 E+01	3.75 E+01
Linear Border Fit	$\text{mean}_{err,Lend} = \frac{\sum_{\theta=0}^{2\pi} \text{err}_{Lend}(\theta)}{200}$	$\text{mean}_{err,Lend}$	Mean of the error err, of the end linear fit, L_{end} .	3	6.69 E-03	7.51 E+01	8.77 E+01	3.95 E+01	3.83 E+01
Linear Border Fit	$\text{mean}_{m,Lmid} = \frac{\sum_{\theta=0}^{2\pi} m_{Lmid}(\theta)}{200}$	$\text{mean}_{m,Lmid}$	Mean of the parameter m for the middle linear fit, L_{mid} .	3	8.09 E-03	8.61 E-01	6.57 E-01	5.96 E-01	3.22 E-01
Linear Border Fit	$\text{mode}_{m,Lbeg} = \text{mode}(m_{Lbeg})$	$\text{mode}_{m,Lbeg}$	Mode of the parameter m for the beginning linear fit, L_{beg} .	3	1.12 E-02	3.23 E+00	9.61 E+00	7.94 E+00	9.69 E+00
Color	--	Cblue	The presence of the color blue-gray in the lesion. 1 indicates presence while 0 indicates absence.	1,2,3	3.08 E-02	1.58 E-01	3.68 E-01	3.64 E-02	1.89 E-01
Sigmoid Border Fit	$\text{mode}_{c,Sf} = \text{mode}(c_{Sf})$	$\text{mode}_{c,Sf}$	Mean of the parameter c for the sigmoidal fit, Sf.	1	3.22 E-02	-4.67 E-02	1.61 E-02	-3.95 E-02	1.91 E-02
Sigmoid Border Fit	$\text{std}_{c,Sf} = \sigma(c_{Sf})$	$\text{std}_{c,Sf}$	Standard deviation of the parameter c for the sigmoidal fit, Sf.	2	4.96 E-02	6.99 E-03	2.92 E-03	5.95 E-03	2.57 E-03
Linear Border Fit	$\text{std}_{m,Lmid} = \sigma(m_{Lmid})$	$\text{std}_{m,Lmid}$	Standard deviation of the parameter m for the middle linear fit, L_{mid} .	2	5.32 E-02	1.09 E+00	2.80 E+00	3.27 E-01	7.53 E-01
Sigmoid Border Fit	$\text{mode}_{err,Sf} = \text{mode}(\text{err}_{Sf})$	$\text{mode}_{err,Sf}$	Mode of the error, err, of the sigmoidal fit, Sf.	1	6.05 E-02	6.82 E+02	4.06 E+02	5.53 E+02	3.05 E+02
Linear Border Fit	$\text{mode}_{m,Lend} = \text{mode}(m_{Lend})$	$\text{mode}_{m,Lend}$	Mode of the parameter m for the end linear fit, L_{end} .	2	8.55 E-02	-2.50 E-01	4.83 E-01	-1.19 E-01	2.90 E-01
Linear Border Fit	$\text{mode}_{err,Lmid} = \text{mode}(\text{err}_{Lmid})$	$\text{mode}_{err,Lmid}$	Mode of the error, err, of the middle linear fit, L_{mid} .	2	1.04 E-01	5.62 E+02	6.87 E+02	3.88 E+02	3.88 E+02
Linear Border Fit	$\text{mode}_{m,Lmid} = \text{mode}(m_{Lmid})$	$\text{mode}_{m,Lmid}$	Mode of the parameter m for the middle linear fit, L_{mid} .	2	1.39 E-01	1.32 E+00	4.68 E+00	3.81 E-01	2.05 E-01
Linear Border Fit	$\text{mode}_{err,Lend} = \text{mode}(\text{err}_{Lend})$	$\text{mode}_{err,Lend}$	Mode of the error, err, of the end linear fit, L_{end} .	3	1.80 E-01	3.65 E+01	1.35 E+02	1.04 E+01	4.76 E+01
Linear Border Fit	$\text{mode}_{err,Lbeg} = \text{mode}(\text{err}_{Lbeg})$	$\text{mode}_{err,Lbeg}$	Mode of the error, err, of the beginning linear fit, L_{beg} .	3	2.95 E-01	3.82 E+00	2.69 E+01	0.00 E+00	0.00 E+00
Ellipse Border Fit	$\text{Ell}_{ecc} = \sqrt{1 - \frac{a^2}{b^2}}$	Ell_{ecc}	Eccentricity, ecc, of the ellipsoid fit, Ell.	1	4.05 E-01	7.13 E-01	1.17 E-01	7.33 E-01	1.39 E-01
Linear Border Fit	$\text{std}_{m,Lbeg} = \sigma(m_{Lbeg})$	$\text{std}_{m,Lbeg}$	Standard deviation of the parameter m for the beginning linear fit, L_{beg} .	1	4.29 E-01	7.85 E+00	4.48 E+00	7.26 E+00	3.31 E+00

3. RESULTS

3.1 Machine Learning Model Achieved 98% Sensitivity at 65% Specificity

From a 2-tailed t-test, 21 of the 30 new features and 22 of the 31 old features were selected as significant, with p-values less than .05. These 43 features were used to distinguish between images of melanomas and nevi. Altogether, the data from these 43 features for the 112 images formed the input matrix for the machine learning model, which achieved 98% sensitivity at 65% specificity, with an AUC of .8801. The threshold for classifying melanoma was determined to be .2567. The overall accuracy was 81% (Table 2). Out of all the individual algorithms, decision trees had the greatest specificity, 54%, at 98% sensitivity, and had an AUC of .8440. Individually, K-nearest neighbors had the greatest AUC. K-nearest neighbors achieved 45% specificity at 98% sensitivity and had an AUC of .8695.

Figure 4: Performance of Machine Learning Model with Old and New Features

The ROC curve was plotted for the mean probability obtained for the old image algorithms and new image algorithms.

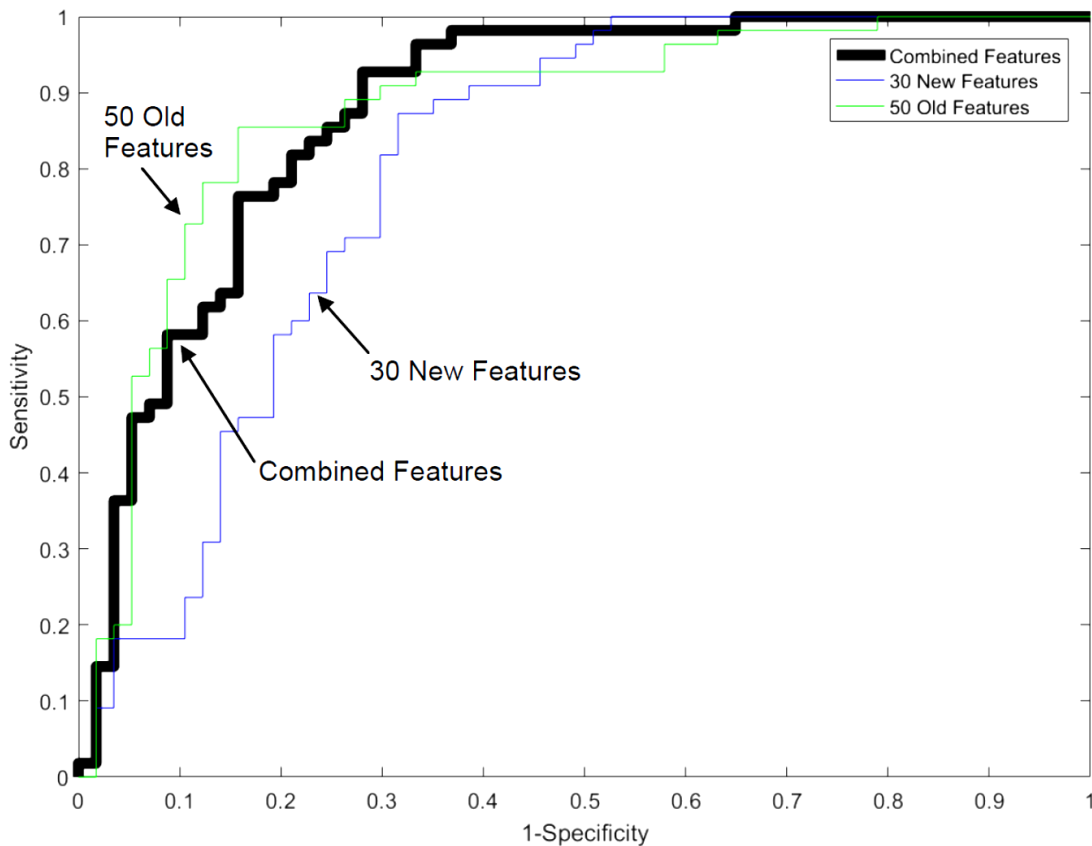


Table 2: Performance of the machine learning model with the old and new features

Specificity at 98% sensitivity, accuracy, threshold, and AUC are computed for the machine learning model, with the combined 30 new and selected 31 old features, then solely with the original set of 50 features, then solely with the new 30 features.

	Original 50 Features	New 30 Features	30 New + 31 Old Features
Specificity at 98% sensitivity	36%	49%	65%
Accuracy at 98% sensitivity	67%	72%	81%
Threshold	0.1867	.2804	.2567
AUC	0.8746	.7984	.8801

3.2 High Regularization Parameter Opposed Overfitting

The regularization penalty λ controls against overfitting the training set, as it penalizes against higher parameter values. The optimal value for the regularization parameter λ was 1 for both logistic regression and neural networks. As λ increased, the cost functions for the classification of the test set for both algorithms decreased, meaning that the algorithms were able to more accurately classify the images in the test set. This regularization is important for optimizing the parameters found by the algorithms such that they may be applicable not only to the images in the training set, but also the images in the test set.

3.3 Combined Features Performed Better than Old Features and New Features Separately

In order to determine whether the added features improved the performance of the classification model, the machine learning model was run separately on solely data extracted from the 30 new features and solely data extracted from the 50 original features. 21 of the 30 new features showed to be significant from a 2-tailed t-test ($p\text{-value} < .05$). With data from these 21 significant new features, the model achieved 98% sensitivity at 49% specificity, with an AUC of .7984 (Figure 4; Table 2). The accuracy at 98% sensitivity was 72%. The machine learning model was also run on data from the original 50 significant features developed by the mentor. 32 features were significant, with $p\text{-values} < .05$. The model achieved 98% sensitivity at 36% specificity, with an AUC of .8746 (Figure 4; Table 2). The accuracy at 98% sensitivity was 67%.

In this study, high sensitivity was prioritized over high specificity, since an undetected melanoma is more harmful than a misdiagnosed nevus. However, high specificity is key to reducing the number of unnecessary biopsies, saving time, cost, worry, and adverse cosmetic effects for patients. The model composed of the 21 new significant features and 22 old significant features achieved almost twice as high specificity as the old model with 32 significant features at 98% specificity (Figure 4). This shows that the additional features conveyed new information about the images and were successful in distinguishing

between melanoma and nevi. The model with solely the new features had a higher specificity at 98%, while the model with solely the old features had a higher AUC (Figure 4). Above 60% specificity, the model with solely the old features has higher sensitivity than the model with solely the new features (Figure 4). The model combining the selected old and new features achieved a higher specificity at 98% sensitivity, AUC, and overall accuracy rate than either of the separate models, performing better than either of the separate models (Figure 4; Table 2).

3.4 Melanomas Exhibited Steeper, More Irregular Transition in Brightness Across the Lesion Border

The most significant new features quantified the variation in brightness across the lesion border (Table 1). The most significant feature was $std_{err,Lmid}$, the standard deviation of the error of the middle linear fit. The mean and standard deviation of the error were much higher, on average, for melanomas than for nevi, for all the fits implemented (Table 1). This suggests that the variation in brightness across the lesion border of melanomas is more irregular.

The middle linear fit is the most revealing of the 3 linear fits, since it fits the variation in the brightness of the pixels starting from the pixel darkest within the pigmented lesion and ending at the brightest pixel outside the lesion. The pixels fit by the beginning fit were inside the lesion, while the pixels fit by the end fit were outside of the lesion. The means of the slopes of the middle and end linear fits, represented by $mean_{err,Lmid}$ and $mean_{m,Lend}$ respectively, were higher for melanomas than for nevi, suggesting that there is a much more prominent difference in brightness between cancerous and normal tissue, corresponding to a steeper transition in brightness across the border for melanomas (Table 1). In contrast, the mean of the slope of the beginning fit, $mean_{err,Lbeg}$, was lower for melanomas (1.63) than for nevi (2.82). This suggests that from within the lesion, melanomas exhibit a smoother transition in brightness, while nevi exhibit steeper transition in brightness.

The parameter, $mean_{c,sf}$, for the sigmoid fit also relates to the steepness of the transition in brightness across the lesion border. A higher value of c in the sigmoid fit corresponds to a steeper transition. While, on average, $mean_{c,sf}$ was higher for melanomas (-0.0382) than for nevi (-0.0341), the difference between these averages was almost negligible (Table 1). More data is needed to confirm this result.

3.5 Melanomas Exhibit a More Irregular Lesion Boundary, and Greater Color and Brightness Variation Within the Lesion

Analysis of the feature Ell_{err} demonstrates that melanomas are more irregular in shape than nevi, using an ellipse as a reference for regularity. Ell_{err} was more than twice as high for melanomas (64.6) than for nevi (31.5). The eccentricity of the ellipsoid fit, Ell_{ecc} , was not a significant feature, with a p-value > .05 (Table 1).

In addition, images of melanomas exhibited a larger range of pixel brightness and a larger variation in color. As shown by the distribution of the features $maxdif$ and $sumdif$, images of melanomas had a higher spread of pixel brightness distribution than nevi, due to the contrast between cancerous tissue and normal tissue (Table 1). Furthermore, of the 6 colors sampled, a greater number of colors were observed for melanomas than for nevi. The color blue-gray was observed, on average, in more images of melanomas (.158) than of nevi (.0364), supporting previous research that it is indicative of melanoma (Chen 2003, Seidenari 2010).

4. DISCUSSION AND CONCLUSIONS

In this project, a novel screening test was developed for melanoma based on an ensemble machine learning approach. The model achieved 98% sensitivity at 65% specificity, with an AUC of .8801. This classification model has the potential to be used as a screening test for melanoma, as it is more successful in identifying melanoma than currently available computational tools. The novelty of this classification model stems from the fact that it is completely automated. It is extremely convenient and efficient for users to upload images of pigmented lesions, and then wait for 10-40 minutes for results. The developed program exceeds the sensitivity and specificity of existing self-operated smart phone applications for melanoma detection (Table 3). In a recent study, researchers reviewed all melanoma-detection applications currently available for smart phones, and found 3 applications that implemented image analysis algorithms on uploaded photos of pigmented lesions and predicted the risk of melanoma (Wolf 2013). While the researchers de-identified these applications, they reported that the best one achieved only 72% sensitivity at 36% specificity. In contrast, the model developed in this study achieved 98% sensitivity at 65% specificity, with an AUC of .8801.

Table 3: Comparison of Diagnoses

Comparisons between sensitivity, specificity, cost, and time for biopsy with histopathological diagnosis (gold standard), dermatologists, primary care physicians, and the model developed in this paper are depicted.

	Biopsy with Histopathological Diagnosis	Dermatologists	Primary Care Physicians	Current Computational Tools	Novel Classification Model
Sensitivity	N/A	57% (Malveyh 2014)	40% (Menzies 2009)	72% (Wolf 2013)	98%
Specificity	N/A	92% (Malveyh 2014)	85% (Menzies 2009)	36% (Wolf 2013)	65%
Cost	\$452 (Losina 2007)	\$100-\$200 (Losina 2007)	\$70 (American Medical Association 2013)	N/A	Free
Time*	A week (Cancer Council NSW 2015)	Weeks (Rosenthal 2014)	Weeks (Rosenthal 2014)	N/A	10-40 minutes

* The total amount of time it takes from when a patient first discovers a pigmented lesion to when he/she receives results for whether the lesion is melanoma. This includes the amount of time patients have to wait when they book appointments with medical professionals.

To validate the results obtained and improve accuracy, the model developed could be tested on other datasets from other regions in the U.S. and note any possible differences in accuracy. In this study, the majority of the images were obtained from one hospital in New York, which limited geographic diversity. Furthermore, the relatively small dataset of images also limited diversity in terms of race, sex, and age. Further datasets should be obtained in order to generalize the model and improve its accuracy.

Another way to improve accuracy would be to improve the specificity of the model. Many nevi, or benign moles, were falsely identified as melanoma. One way to reduce the false positive rate would be to allow people to send in pictures of their non-pigmented skin and develop an algorithm that compares the person's suspicious-looking pigmented lesion with non-pigmented skin, which would serve as a reference. If the person's suspected lesion shares similar attributes with the person's non-pigmented skin, it would likely be benign. If not, it would likely be malignant. Recently, clinical studies have shown that this procedure, known as mole-mapping, improves the diagnostic accuracy of dermatologists by up to 13.5% (Kittler 2001; Gamble 2010). This process could be automated and incorporated into the melanoma classification model developed in this study, in order to evaluate its effect on sensitivity and specificity. Another way to reduce the false positive rate would be to allow people to send in multiple pictures of a single suspicious-looking pigmented lesion over a period of time. An algorithm could be developed to examine the evolution of the lesion over time, which is a significant predictor of melanoma (Abbasi 2004; Rigel 2010).

Unlike other automated diagnosis systems, such as MelaFind™, which requires expensive equipment and can only be operated by dermatologists (March 2015), this program can be transformed into an app and used by anyone with access to a phone. In order to use it, people would just take a picture of the mole on their skin and upload it, whereupon the program would check for features such as border irregularity, variation in lesion border brightness, and color variation. The probability of melanoma would be calculated by evaluating the trained machine learning parameters on the quantified features of the image of the lesion. Using this program, a person can receive results in less than 40 minutes from when they first upload the image. Apps are also free to download and are easy to use. In comparison, biopsy with histologic examination, which is the gold standard for melanoma diagnosis currently, often takes a week to obtain results and costs about \$452 in total (Cancer Council NSW 2015; Losina 2007). By using this program, a person worried about a benign mole can save time and money. This mobile technology would greatly improve melanoma screening surveillance and ultimately save lives.

REFERENCES

- Abbasi, N.R., Shaw, H.M., Rigel, D.S., Friendman, R.J., McCarthy, W.H., Osman, I., Kopf, A.W., & Polsky, D. (2004). Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria. *JAMA*, 292(22), 2771-6. doi:10.1001/jama.292.22.2771
- American Cancer Society. (2009). *Cancer Facts and Figures 2009*. Atlanta, GA: American Cancer Society. Retrieved from <http://www.cancer.org/acs/groups/content/@nho/documents/document/500809webpdf.pdf>
- American Cancer Society. (2016). *Cancer Facts and Figures 2016*. Atlanta, GA: American Cancer Society. Retrieved from <http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-047079.pdf>
- American Cancer Society. (2016). *Treatment of melanoma skin cancer, by stage*. Retrieved from <http://www.cancer.org/cancer/skincancer-melanoma/detailedguide/melanoma-skin-cancer-treating-by-stage>
- Bowling, J. (2012). *Diagnostic Dermoscopy: The Illustrated Guide*. Available from <http://dx.doi.org/10.1002/9781444329834>
- Cancer Council NSW. (2015). *Melanoma diagnosis*. Retrieved from <http://www.cancercouncil.com.au/835/b1000/melanoma-24/melanoma-diagnosis/>
- Chen, J., Stanley, R.J., Moss, R.H., & Van Stoecker, W. (2003). Colour analysis of skin lesion regions for melanoma discrimination in clinical images. *Skin Res Technol.*, 9(2), 94-104. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3196565/>

- Gareau, D., Correa da Rosa, J., Yagerman, S., Carucci, J., Gulati, N., DeFazio, J., Suarez-Farinas, M., Marghoob, A., & Krueger, G. (in press). Novel Imaging Biomarkers for Early Melanoma Detection. *Experimental Dermatology*.
- Geller, A.C., Emmons, K., Brooks, D.R., Zhang, Z., Powers, C., Koh, H.K., Sober, A.J., Miller, D.R., Li, F., Haluska, F., & Gilchrest, B.A. (2003). Skin cancer prevention and detection practices among siblings of patients with melanoma. *J Am Acad Dermatol.*, 49(4), 631-638.
[http://dx.doi.org/10.1067/S0190-9622\(03\)02126-1](http://dx.doi.org/10.1067/S0190-9622(03)02126-1)
- Guy, G.P., Ekwueme, D.U., Tangka, F.K., & Richardson, L.C. (2012). Melanoma Treatment Costs: A Systematic Review of the Literature, 1990–2011. *Am. J. Prev. Med.*, 43(5), 537-545.
<http://dx.doi.org/10.1016/j.amepre.2012.07.031>
- Guy, G.P., Thomas, C.C., Thompson, T., Watson, M., Massetti, G.M., & Richardson, L.C. (2015). Vital Signs: Melanoma Incidence and Mortality Trends and Projections — United States, 1982–2030. *MMWR*, 64(21), 591-596. Retrieved from
<https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6421a6.htm>
- Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian J Intern Med.*, 4(2), 627–635. Retrieved from
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3755824/>
- Hu, S., Soza-Vento, R., Parker, D., & Kirsner, R. (2006). Comparison of Stage at Diagnosis of Melanoma Among Hispanic, Black, and White Patients in Miami-Dade County, Florida. *Arch Dermatol.*, 142(6), 704-708. doi: 10.1001/archderm.142.6.704
- Johr, R., Soyer, H.P., Argenziano, G., Hofmann-Wellenhof, R., & Scalvenzi, M. *Dermoscopy: The Essentials*. UK: Elsevier Limited.
- Kassianos, A., Emery, J., Murchie, P., and Walter, F. (2015). Smartphone applications for melanoma detection by community, patient and generalist clinician users: a review. *Br J Dermatol.*, 172(6), 1507-18. doi: 10.1111/bjd.13665.
- Kittler, H. & Binder, M. (2001). Risks and Benefits of Sequential Imaging of Melanocytic Skin Lesions in Patients With Multiple Atypical Nevi. *JAMA Dermatol.*, 137(12), 1590-1595. doi:10.1001/archderm.137.12.1590
- Kumar, R. & Indrayan, A. (2011). Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatr.*, 48(4), 277-287. doi: 10.1007/s13312-011-0055-4
- Lambert, J. & Chevret, S. (2016). Summary measure of discrimination in survival models based on cumulative/dynamic time-dependent ROC curves. *Stat Methods Med Res.*, 25(5), 2088-2102. doi: 10.1177/0962280213515571
- Lawlor, H.K. & Rubin, B.J. (1993). Early diagnosis of ectopic pregnancy. *West J Med.*, 159(2), 195–199. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1022234/?page=1>
- Little, E.G. & Eide, M.J. (2012). Update on the Current State of Melanoma Incidence. *Dermatol Clin.*, 30(3), 355-361. doi: 10.1016/j.det.2012.04.001

- Losina, E., Walensky, R.P., Geller, A., Beddingfield, F.C., Wolf, L.L., Gilchrest, B.A., & Freedberg, K.A. (2007). Visual Screening for Malignant Melanoma: A Cost-effectiveness Analysis. *Arch Dermatol.*, 143(1), 21-28. doi:10.1001/archderm.143.1.21
- March, J., Hand, M., & Grossman, D. (2015). Practical application of new technologies for melanoma diagnosis Part I. Noninvasive approaches. *J. Am. Acad. Dermatol.*, 72(6), 929-941. <http://dx.doi.org/10.1016/j.jaad.2015.02.1138>
- National Cancer Institute. *SEER Stat Fact Sheets: Melanoma of the Skin*. Bethesda, MD: National Cancer Institute. Retrieved from <http://seer.cancer.gov/statfacts/html/melan.html>
- National Cancer Institute. *SEER Training Modules: Five-Year Survival Rates*. Retrieved from <http://training.seer.cancer.gov/melanoma/intro/survival.html>
- National Cancer Institute. (2015). *Side effects*. Retrieved from <https://www.cancer.gov/about-cancer/treatment/side-effects>
- National Center for Health Statistics. (2009). *National Ambulatory Medical Care Survey Factsheet*. Retrieved from https://www.cdc.gov/nchs/data/ahcd/namcs_factsheet_derm_2009.pdf
- Rigel, D.S., Russak, J., & Friedman, R. (2010). The Evolution of Melanoma Diagnosis: 25 Years Beyond the ABCDs. *A Cancer Journal for Clinicians*, 60(5), 301-316. doi: 10.3322/caac.20074
- Rosenthal, E. (2014). The Health Care Waiting Game. The New York Times. Retrieved from http://www.nytimes.com/2014/07/06/sunday-review/long-waits-for-doctors-appointments-have-become-the-norm.html?_r=0
- Seidenari, S., Ferrari, C., Borsari, S., Benati, E., Ponti, G., Bassoli, S., Giusti, F., Schianchi, S., Pellacani, G. (2010). Reticular grey-blue areas of regression as a dermoscopic marker of melanoma in situ. *Br J Dermatol.*, 163(2), 302-309. doi: 10.1111/j.1365-2133.2010.09821.x
- Tominaga, S. (1992). Color classification of natural color images. *Color Research & Application* 17(4), 230-239. doi: 10.1002/col.5080170405
- Tone, A. (2012). Medicalizing Reproduction: The Pill and Home Pregnancy Tests. *J Sex Res.*, 49(4), 319-327. doi: 10.1080/00224499.2012.688226
- Weatherall, I.L. & Coombs, B.D. (1992). Skin Color Measurements in Terms of CIELAB Color Space Values. *J. Invest. Dermatol.*, 99(4), 468-473. doi: doi:10.1111/1523-1747.ep12616156
- Wolf, J.A., Moreau, J.F., Akilov, O., Patton, T., English, J.C., Ho, J., & Ferris, L.K. (2013). Diagnostic inaccuracy of smartphone applications for melanoma detection. *JAMA Dermatol.*, 149(4), 422-426. doi:10.1001/jamadermatol.2013.2382
- Yoo, J. & Rigel, D. (2010). Trends in Dermatology: Geographic Density of US Dermatologists. *Arch Dermatol.*, 146(7), 779. doi:10.1001/archdermatol.2010.127
- Zou, K. H. (2011). *Statistical Evaluation of Diagnostic Performance: Topics in ROC Analysis* (2nd ed.). New Jersey: John Wiley & Sons, Inc.