Automated Diagnosis of Melanoma

A Technical Report
presented to the faculty of the
School of Engineering and Applied Science
University of Virginia

by

Daniel Le (dbl3jf)

with

Kelly Chen (kjc8rx)

May 7, 2021

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signed: Daniel Le

Approved: Jundong Li **Date:** 05/07/2021

Jundong Li, Department of Computer Science

Automated Diagnosis of Melanoma

Kelly Chen and Daniel Le
University of Virginia
kjc8rx@virginia.edu, dbl3jf@virginia.edu

ABSTRACT

With nearly 100,000 cases in the United States every year, melanomas are not only one of the most common, but also one of the most dangerous types of skin cancer. Despite their prevalence, melanomas typically require specialized tools for a professional diagnosis. There currently exist software that utilizes artificial intelligence and deep learning to predict diagnoses based on images, but these programs are not readily available to general consumers. We propose a similar tool with a complete user interface so that others have access to a potentially life-saving insight. The system will be developed on a cloud platform. We will use TensorFlow to implement a model that is trained on image sets of skin cancers to achieve reliably high rates of accuracy. We are currently aiming to create a tool that can predict malignant melanomas with at least 90% accuracy. Introducing a publicly available tool of this nature may benefit potential patients who lack the resources or are unable to get a professional opinion in a timely manner.

INTRODUCTION

According to the National Institute of Health (NIH), melanoma is a type of skin cancer that is derived from pigmented cells called melanocytes. More than 50% of these tumors are located on the skin, but could also develop on mucosal surfaces or the uveal tract. Although the appearance of melanomas vary, most are characterized by asymmetry, border irregularities, discoloration, a relatively large diameter, and continual changes.[1] Currently, the Centers for Disease Control and Prevention (CDC) identifies melanoma as the third most common type of skin cancer, and the primary cause of death among all skin cancers. From 2012 to 2016, about 77,698 cases of melanoma occurred in the United States annually. The overall incidence rate of melanoma was 21.8 per 100,000, and was the most prevalent in white individuals (Figure 1).[2] Given the high incidence of melanoma cases, a considerable healthcare cost is used to diagnose and treat the disease. National treatment costs ranged from $44.9 million to $932.5 million per year across all age groups. Stage IV melanomas generally induced the highest costs.[3] In 2018, the 5-year survival rate for Stage IV melanoma was 22.5%.[4] Due to the low survival rate and high healthcare costs of late-stage melanoma, it is vital for the cancer to be detected as soon as possible.

Race/Ethnicity	Rate (%)	Count
All Races	21.8	77,698
White	24.9	73,395
White, Hispanic	4.6	1,591
White, non-Hispanic	28.0	71,801
Black	1.0	372
American Indian/Alaska Native	5.6	190
Asian/Pacific Islander	1.3	239
Hispanic	4.6	1,725

Figure 1: Average Annual Number and Rate of Invasive Melanoma Cases by Race/Ethnicity in the United States 2012 - 2016

Clinical diagnosis usually involves a physical exam and a biopsy. However, some patients rarely undergo physical exams, and a malignant melanoma may go unnoticed until the cancer is in its late stage. The biopsy may additionally take time or cause complications. Patients have to wait up to two weeks for results. The biopsy site may become infected, and the patient may experience bleeding, redness, warmth, tenderness, pus, or red streaks.[5] Although rare, the biopsy may also further spread the cancer. The occurrence of tumor seeding, or the spread of cancer cells via a biopsy needle, is under 1% but not impossible.[6] Providing alternatives to the standard diagnosis could be more time efficient and decrease such complications.

BACKGROUND

In addition to the time and cost intensive diagnosis procedures that may be conducted too late due to oversight or refusal of treatment, the physical appearance of melanomas vary significantly. The “ABCDE” visual inspection system is not an absolute guideline, as occurrences differ between individuals, as well as share many characteristics with benign skin imperfections. Without additional measures, the human eye cannot reliably

identify skin cancers even with comparison examples. Machine learning models present an opportunity to overcome this inability by comparing thousands of inputs at once. While a medical professional may only be able to make an educated judgement of a particular skin lesion, deep learning algorithms provide discrete evaluations based on proven examples. Given enough training data, arbitrary metrics such as “discoloration” or “border” can be resolved to more exacting standards.

As it could potentially be applied to literal life-and-death situations, a machine learning implementation for melanomas requires extensive training data to provide an acceptably high level of accuracy and minimal mistakes. Cloud services may prove to be an all-encompassing answer to the question of having enough computing power to run the deep learning networks, while also providing a platform to make the program available to a broad audience.

RELATED WORK

Machine learning has been considered as an alternative method of diagnosis of melanoma. There have been few researchers who have created machine learning models that could identify melanomas with a relatively high accuracy. However, none of these projects has been released to the public. One such project, called Deep Ensemble for Recognition of Melanoma (DERM), detected malignant melanomas with a 92.8% accuracy. DERM was trained and tested with 7,102 images of both confirmed melanomas and harmless skin tags, and showed great potential for diagnostic use. However, developers of DERM planned to have the application released as a support tool in clinical settings only. This software is currently not open to the public and is still a work in progress. The researchers mentioned that diagnostic accuracy is still dependent on the experience of the examiners, and requires specialized equipment. Additionally, DERM has no deployment platform nor UI/UX design.[7] Compared to DERM, the proposed technical project would be user-friendly and open to the public.

There are also self-diagnosis software that do not use machine learning. However, in these applications, the patient must communicate with a medical professional and the results are not immediate. For example, Klara is a German-based application that sends pictures of skin-related conditions to certified dermatologists. It will provide a result after 48 hours, as patients must wait for the physician to examine the image. Klara requires a fee of $39 to access the application.[8]

SYSTEM DESIGN

The general basis of the proposed melanoma diagnosis tool is a binary classification algorithm. It will attempt to sort input images into either positive (malignant melanoma) or negative (benign skin lesion) categories. Rather than relying on any single metric to make a decision, the model should be able to aggregate and weigh multiple outstanding characteristics at once.

The website Kaggle.com hosts a dataset from Alexander Scarlat, MD containing ten thousand images of skin lesions previously confirmed to either be melanoma or not melanoma.[9] The files are presorted into training and test sets, but the volume of data can be extended by manipulating individual images to simulate the same lesion being photographed from different angles.

In addition to the given fileset, the team will apply further preprocessing to better normalize the pictures to fit into the model. Images will be resized to 200x200 pixels to ensure uniformity, and each image will have three corresponding copies in which they are flipped horizontally, vertically, and rotated 90 degrees clockwise. This triples the pool of image files from ten thousand unique images to nearly thirty thousand.

The nature of the image-based dataset suggests an implementation of a convolutional neural network (CNN) rather than methods better suited for tabular data such as random forests. The Resnet-50 architecture will be utilized due to its reputable accuracy and resistance to degradation from converging networks.[10] It employs a skip connection method that in short, adds the original input to the output of each convolution layer so that the residuals more easily reach zero, and each subsequent layer shall perform at least as well as the layer under it. This technique also serves to preserve accuracy at deeper layers by eliminating vanishing gradients. The proposed model will use blocks with three layers in a 1x1, 3x3, and 1x1 configuration on each residual function. Typical Resnet-50 implementations utilize the first and third convolutions for restricting output dimensions, while the second is left for bottlenecking.

In order to determine the most optimal hyperparameters for the implementation of Resnet-50, it is necessary to iterate over various combinations and mark the accuracy of each one. Initially the values would be set arbitrarily (batch_size=16, learning_rate=0.0001, optimizer=’adam’, etc.). A function would be created to individually adjust each hyperparameter, create the model, train it on the data subset, and record the validation accuracy. Through this the highest performing configuration of Resnet for this particular dataset can be found.

Once elected, the optimal model can then be executed on the testing data subset, and its resultant accuracy and precision can be evaluated with a confusion matrix. The system should be further tuned to favor false positives in cases where the decision lacks confidence, for the reasoning that patients should always check with a medical professional rather than risk ignoring a potential melanoma.

The proposed tool will be hosted on Amazon Web Services (AWS) and will use multiple cloud services. The

application will be built with React Native and deployed on AWS Amplify. The front-end will include a user login and signup page. After logging in, a home page will display. A navbar at the bottom of the application will contain links to the home page, a user profile page, a record of tests, contact information for a doctor, and a button for uploading new images (Figure 2). Pictures saved as JPEG or PNG formats can be uploaded to the testing page, and will be analyzed with machine learning to see if the image contains a melanoma. Like the training images, the uploaded pictures will be resized to 200x200 pixels. After the image is analyzed, the page will display either a positive or negative result and its confidence in the decision. For both results, the application will provide instructions for further steps.

To save user credentials, all information will be stored in Amazon DynamoDB through Amplify DataStore. A user ID, email, password, and location will be stored in a Users table. When a new user creates an account, a new entry is added to the Users table with the new information. After logging in, the user's credentials are verified against the information stored in the database. Another table in DynamoDB will contain melanoma images. This table will contain a user ID, image, and diagnosis result. This information can be viewed on the user profile page.

The application's machine learning capabilities will be powered by TensorFlow and supported by Amazon SageMaker. By using the "Predictions" category in the Amplify framework, the application will be able to import a SageMaker endpoint.

PROCEDURE

This tool would primarily be accessed via mobile application, and should be compatible with any smartphone. To access the product, the customer must create an account with an email, password, location, and optionally medical provider information. The email and password would be used to log in. The customer may upload an image of their skin lesion. The application will run the input as testing data on the CNN, and output the model's decision on whether the lesion is expected to be a positive or negative diagnosis. If positive, the application will connect the patient with nearby hospitals or surgical centers that can treat the melanoma. It will display these locations as a map with the Google Maps API. If negative, the product will give further recommendations, such as monitoring the skin lesion for further changes. The application will also display the confidence of its predictions. If this value is below 75%, the tool will recommend that the patient seek a clinical diagnosis in addition to the product's diagnosis.

Compared to other machine learning diagnosis applications, this product will have a UI/UX design that is customer-oriented and user-friendly. This application will additionally be free, so it will be open to everyone who has a smartphone that can access the internet.

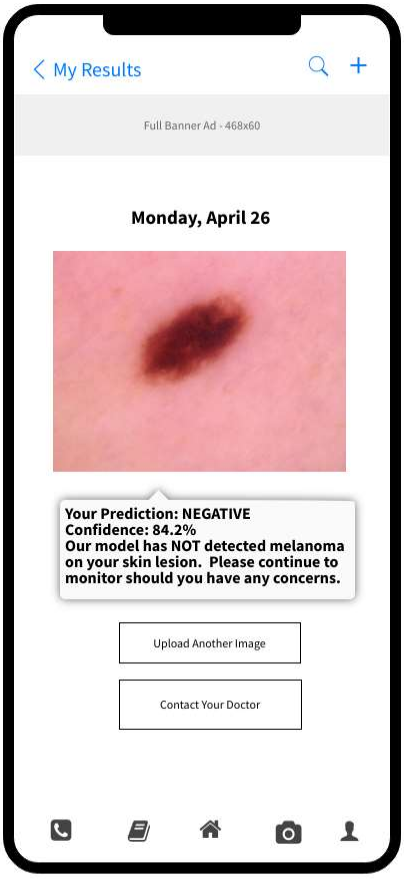

Figure 2: Wireframe for user interface of testing results.

RESULTS

Incorrect results, especially false negatives, may cause harmful consequences for the customer. Thus, the team's goal is to develop an application that can recognize melanomas with at least a 90% accuracy at initial assessment. The product will then be further maintained and trained with more images. The proposed tool would potentially allow customers to easily self-diagnose with internet access, where patients had to experience a lengthy and inconvenient diagnosis process before. While it is undetermined thus far if this tool could replace a complete clinical diagnosis, it will give patients a chance to receive a faster onward referral to a medical practitioner at the very least. As of 2015, the accuracy of shave biopsy on invasive melanoma was 97%.[11] If the application is able to eventually predict with a similar level of accuracy, then it may be considered as a replacement. However, several factors such as the stage of the melanoma and skin complexion would additionally need to be considered.

The team asked several individuals of different demographics and age groups their opinions on this application. A survey was distributed to a sample of 25 people. The survey provided a description of the

application, asked whether the individual would consider using the tool, and an optional section for suggestions and comments. Out of the sample size of 25, 76% would consider using the tool, 8% would not consider the tool, and 16% were unsure. Some suggestions include making the application desktop-compatible, while others expressed their concerns on the accuracy of the application. Three of the survey-takers expressed that they would use the tool, but would not use it as a replacement for a clinical diagnosis.

CONCLUSIONS

The team has proposed a consumer application that evaluates images of skin lesions to predict whether they may be indicative of melanoma. The prediction model utilizes a deep learning network based on the Resnet-50 architecture, and classifies image inputs into either positive or negative cases.

After the primary training of the machine learning model is completed, the service would be hosted on cloud platforms to provide more reliable accessibility to users. Images and results would be stored into a secure database to record and protect user information.

In contrast to currently existing alternatives, the proposed tool aims to be publicly available to consumers, and provides an interface in which users can upload their own images and receive an evaluation, without immediately requiring a medical professional.

FUTURE WORK

Given additional time and computing resources, the tool could likely be further enhanced in terms of prediction performance and usability. Rather than a binary classification, the app may be better suited to a multiclass implementation, one that differentiates between labels such as “benign”, “continue monitoring”, “seek professional opinion”, and “likely malignant”. Having more distinct categories would communicate clearer advice to users, and not downplay the importance of regular health checkups.

On the backend of the application, performance could be enhanced with access to larger datasets as well as more powerful hardware. The team would potentially explore deeper CNNs such as Resnet-152 or DenseNet, and adjust for finer hyperparameters. These models are rather computationally intensive, and thus would take too long to train on local machines, or be too cost prohibitive on larger cloud-hosted instances.

A larger dataset would at least marginally improve the model’s accuracy, but would be comparatively more difficult to acquire than a faster computer. Unless larger collections of images are available online, additional picture files would only be available from medical providers. There may be potential issues relating to patient privacy.

Another improvement to the application’s usability would be to conduct more beta testing to elicit feedback. This would be an ongoing process as users suggest UI enhancements and additional features. The application could additionally be adapted to support a desktop version (Figure 3).

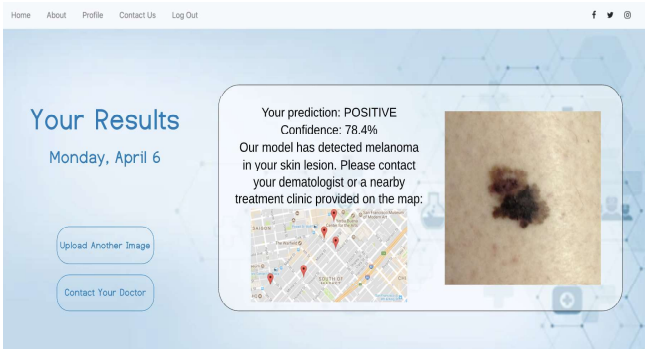

Figure 3: Wireframe for potential desktop-compatible interface.

ACKNOWLEDGMENTS

The team would like to thank Professor Marty Humphrey and Professor Jundong Li for their advice and guidance through the process of writing this report.

REFERENCES

[1] Melanoma Treatment (PDQ®)–Health Professional Version. National Cancer Institute. (n.d.). https://www.cancer.gov/types/skin/hp/melanoma-treatment-pdq#_8_toc.

[2] Centers for Disease Control and Prevention. (2019, June 27). *Melanoma Incidence and Mortality, United States–2012–2016. Centers for Disease Control and Prevention.* https://www.cdc.gov/cancer/uscs/about/data-briefs/no9-melanoma-incidence-mortality-UnitedStates-2012-2016.htm.

[3] Guy, G. P., Ekwueme, D. U., Tangka, F. K., & Richardson, L. C. (2012, November). *Melanoma treatment costs: a systematic review of the literature, 1990-2011.* American journal of preventive medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4495902/.

[4] Stage 4 Melanoma. Melanoma Research Alliance. (n.d.). https://www.curemelanoma.org/about-melanoma/melanoma-staging/stage-4-melanoma/.

[5] WebMD. (n.d.). *Skin Biopsy: Pupose, Procedure, Complications, Recovery.* WebMD. https://www.webmd.com/melanoma-skin-cancer/skin-biopsies.

[6] Staff, ASCO (2021, March 18). *Can a Biopsy Make My Cancer Spread?* Cancer.Net. https://www.cancer.net/blog/2021-03/can-biopsy-make-my-cancer-spread.

[7] Phillips M, Greenhalgh J, Marsden H, Palamaras I. (2019, December 31) *Detection of Malignant Melanoma Using Artificial Intelligence: An Observational Study of Diagnostic Accuracy.* Dermatol Pract Concept. doi:10.5826/dpc.1001a11

[8] Ong, J. (2021, March 19). *The Klara app lets you send photos of your skin problems to certified dermatologists.* TNW | Apps. https://thenextweb.com/news/klara-app-lets-send-photos-skin-problems-certified-dermatologists/amp.

[9] Scarlat, A. (2018, December 2). *melanoma: Augmented dermoscopic pigmented skin lesions from HAM10k.* Kaggle. https://www.kaggle.com/drscarlat/melanoma.

[10] *ResNet (34, 50, 101): Residual CNNs for Image Classification Tasks.* Neurohive. (2019, January 25). https://neurohive.io/en/popular-networks/resnet/.

[11] Zager, J. S., Hochwald, S. N., Marzban, S. S., Francois, R., Law, K. M., Davis, A. H., Messina, J. L., Vincek, V., Mitchell, C., Church, A., Copeland, E. M., Sondak, V. K., & Grobmyer, S. R. (2011). Shave biopsy is a safe and accurate method for the initial evaluation of melanoma. *Journal of the American College of Surgeons*, *212*(4), 454–462. https://doi.org/10.1016/j.jamcollsurg.2010.12.021