DermFollow

A System For Better Diagnosis and Treatment of Skin Cancer

Matt May
Georgia Institute of Technology
maym@gatech.edu

Thanh Dang
Georgia Institute of Technology
thanh.dang@gatech.edu

Stefano Fenu
Georgia Institute of Technology
sfenu3@gatech.edu

Apurv Verma
Georgia Institute of Technology
apurvverma@gatech.edu

Matt Cimino
Georgia Institute of Technology
ciminomc@gatech.edu

1. INTRODUCTION

The current standard-of-care for diagnosing skin cancers (such as melanoma) is in-person care in a dermatologist's office. A dermatologist performs a physical exam to review the patient for probable cancerous lesions. However, follow-up with the patient when he or she is outside of the office is minimal, resulting in unmonitored growth of potentially cancerous lesions over time.

In this paper we describe DermFollow, an application we have built to solve this problem. Patients use their computer or smartphone to upload pictures of their lesions over time. The system uses the knowledge learned from several thousand images via deep neural networks to compute a risk score for the uploaded image and present the most high-risk images, and therefore high-risk patients, to the clinician.

For each uploaded image, we also present the most similar images from the training set to the provider, to explain why (or why not) the network considers an image as a high-risk case. The clinician can use this added information and schedule an appointment with these patients on a priority basis.

2. PROBLEM DEFINITION

Our goal was to build a web application that promotes more effective diagnosis and treatment of skin cancer through machine learning-based analysis of patients' skin lesions over time.

Our system enables this through these key innovations: (1) risk analysis (classification) of suspected skin cancer lesions using deep residual and convolutional neural networks (CNNs), (2) dynamic patient risk scoring based on demographics, medical history, and patient image analysis, and (3) an interactive body map feature that encourages physician-patient communication. We hope to improve patient outcomes, raise patient satisfaction, and improve the efficiency of medical practice.

3. SURVEY

The current state of the art for automated skin cancer analysis generally involves feature detection [14], [8], [16]. In dermatology, there is an ABCDE algorithm (asymmetry, border, color, diameter, evolution over time) [23], [15], [20] that is used by doctors to assess potential skin cancer lesions. Existing approaches often attempt to imitate this algorithm via detection of the same features [24]. Some of these models have yielded poor results on the more diverse images encountered in practice, limiting adoption.

Initially, we also explored the most straightforward method and used simple ABCDE feature extraction (Figure 1) to analyze images of skin lesions [1]. After observing the relatively poor performance on images taken in a variety of sub-optimal lighting and focus conditions, we explored a neural network-based approach to the problem.

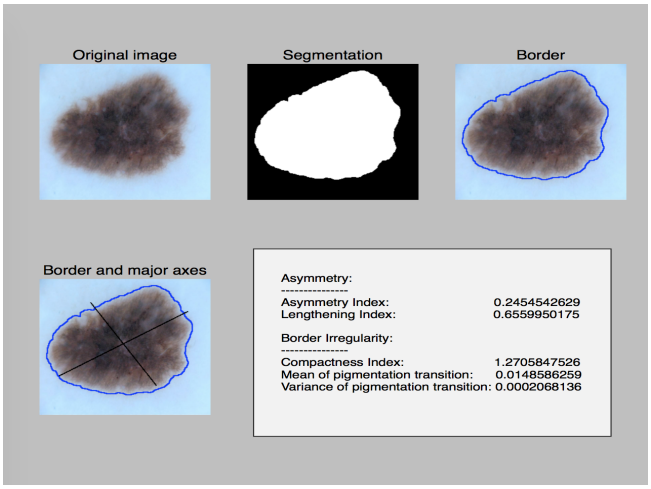

Figure 1: Sample output from feature extraction module

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

Neural network-based approaches have also been used for skin cancer classification [18], [6], [3], [4]. Kreutz et al. [12] use neural networks and feature extraction to classify skin lesions. Sheha et al. [17] use a multilayer perceptron to classify melanoma, attaining 92% accuracy. Esteva et al. [7] use an ensemble of CNNs to attain 90% binary classification (malignant/benign). While these results are impressive, many of these models were trained on images that lack histological (microscopic) verification, the gold standard for determining malignancy.

4. PROPOSED METHOD

4.1 Intuition

Low physician and patient engagement, limited integration into clinical workflows, poor algorithm performance on actual patient images, and expensive hardware requirements have plagued existing approaches.

To increase provider and patient engagement, DermFollow presents an interactive user interface, leveraging JavaScript d3 [2] to allow the physician to review a *body map* of the patient that displays the patient's images superimposed over the relevant point on the body. DermFollow also allows physicians to instantly send short messages to the patient after he or she uploads an image, giving the patient instant feedback.

DermFollow integrates seamlessly with the largely electronic clinical workflow, as it is Web-based and operates on the desktop and mobile devices.

To improve algorithm performance, we use an ensemble of deep CNNs and residual neural networks, which have been shown to be very effective at image classification [13], [22], [19]. Lastly, we use an image dataset with 100% histological verification.

4.2 Description

4.2.1 User Interface

The application has two interfaces, namely the patient interface and provider interface.

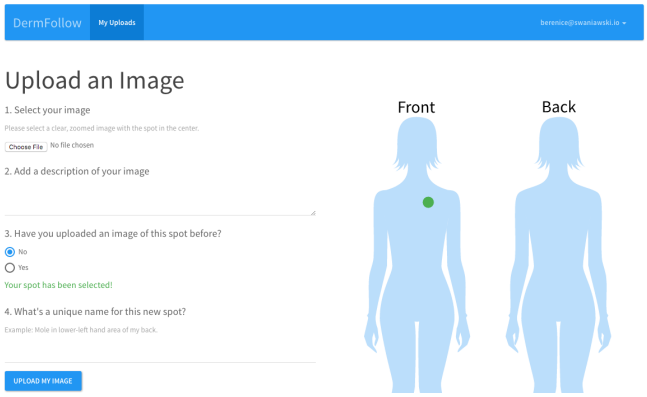

Figure 2: Patient Form

Figure 2 shows the patient interface for uploading an image of a skin lesion. Images are grouped by *spots*, which are unique skin lesions. A user can add a new image of a previously identified spot by choosing the existing spot, or interactively select a point on a 2D body map that identifies the new lesion. Once uploaded, the provider can see all of the patient's lesions on the body map, allowing easy interaction with the data (Figure 3).

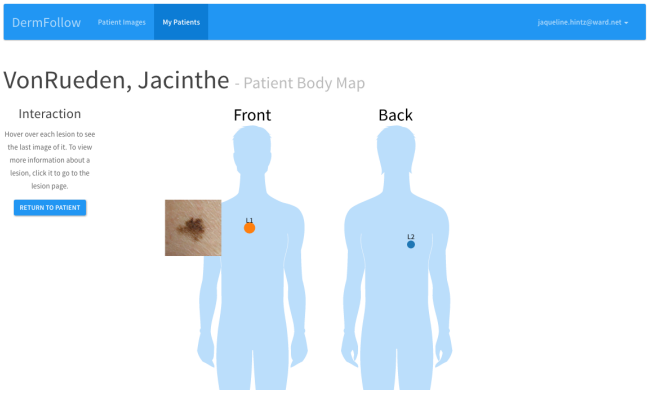

Figure 3: Provider View of Patient Body Map

Figure 4: Provider View of Patient

Figure 4 displays the interface as seen by a dermatologist for a patient. The physician can easily see all the patient's images and their computed risk scores, as well as a global risk score for the patient.

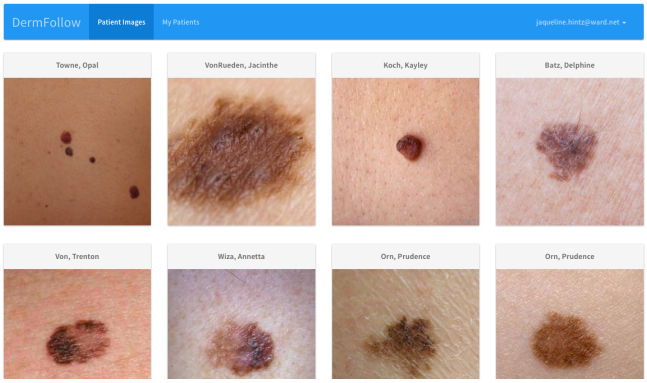

Figure 5: Provider View of Patient Uploads

We also provide an interface (Figure 5) for the dermatologist to view all his or her patients' recent uploads. If

an upload is flagged as high-risk (malignant) by our CNN model, we display an asterisk next to the patient name above the image. We received substantial interest in this feature from the dermatologist community.

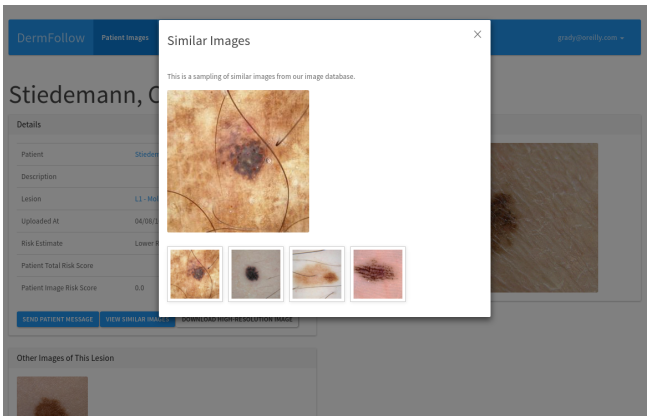

Figure 6: Images similar to the current image

4.2.2 *Neural Network Architecture*

For classification, we use an ensemble of neural networks, including the VGG-16 [19], Inception [21] (Figure 7) and ResNet [9] (Figure 8) architectures.

The VGG-16 is a type of convolutional neural network with 16 convolutional layers and three fully-connected layers. It uses a very small 3×3 kernel for convolutions. The Inception model is a 27-layered network. It uses a dropout of 0.7 and convolution kernels of size 5×5. The ResNet is a deep CNN with shortcut connections which performs identity mappings.

The VGG-16 and Inception models were fine-tuned for the task by lobotomizing the output layers of the network, adding an appropriately sized linear output layer and softmax layer, and retraining. The ResNet was trained de novo on the entirety of the publicly available International Skin Imaging Collaboration (ISIC) dermatological image dataset [10] over the course of two days. The 3,387 high-resolution images from this dataset were augmented 50-fold by a series of scale transformations, blurs, and rotations, that were precomputed for resource efficiency. The resulting 169,350 images were then used to train and cross-validate the models.

The final probability score of an uploaded image is the weighted average of scores from individual models, where the weight assigned to each model is equal to its testing accuracy. We obtained a binary classification (benign/malignant) accuracy of approximately $89.1 \pm 1.8\%$ from 10-fold cross-validation on our ensemble model. Figures 9, 10, and 11 present the training and testing accuracies for the three models over 200 epochs.

4.2.3 *Risk Score*

For each patient, a risk score is computed which is based on a cutaneous lifetime melanoma risk algorithm from [5]:

$$risk\_score = \alpha \times image\_score + (1 - \alpha) \times info\_score \quad (1)$$

The *info_score* is computed from parameters such as skin type, hair color, and medical history. This contributes a

Figure 7: Inception Architecture

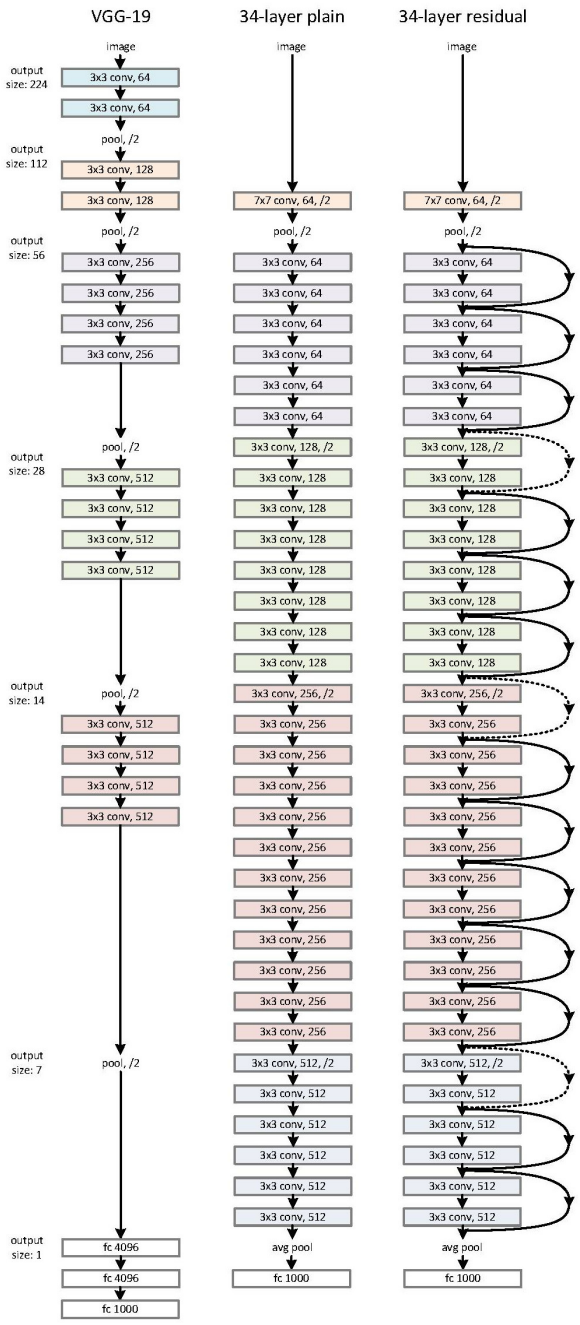

Figure 8: Residual Network Architecture

relatively small proportion of the overall risk score. The *image_score* is a score computed by the neural network model. A risk score close to one indicates a high risk for developing skin cancer, whereas a score close to zero indicates relatively small risk.

Each new patient upload updates the *image_score* for the patient as a weighted average of the current *image_score*, and the probability score (determined from the neural network) of the new image.

It is computed as follows:

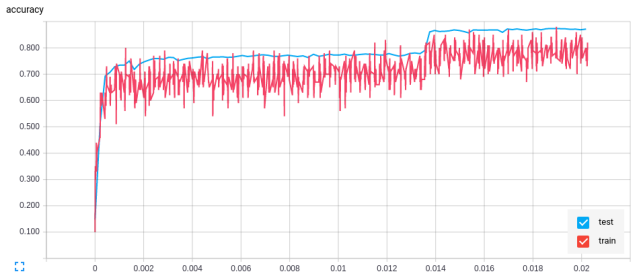

Figure 9: Train/Test Accuracy (Inception)

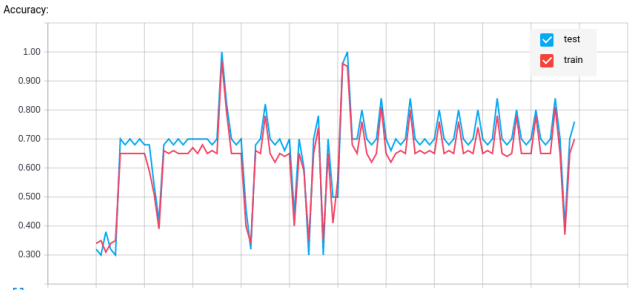

Figure 10: Train/Test Accuracy (VGG-16 Model)

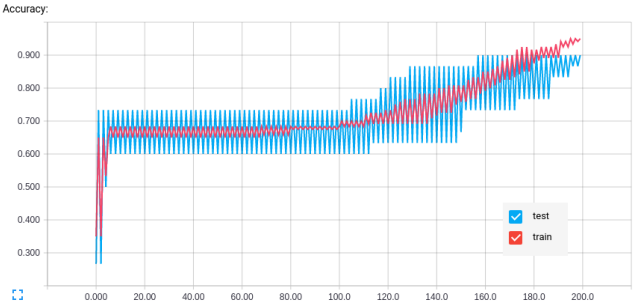

Figure 11: Train/Test Accuracy (ResNet)

$$\begin{cases} (1-3\beta) \times score_{curr} + 3\beta \times score_{new\_image} & score > 0.5 \\ (1-\beta) \times score_{curr} + \beta \times score_{new\_image} & score \leq 0.5 \end{cases}$$

The first update equation assigns a higher weight to the new image and decays the cumulative score rapidly. This ensures that if the new image is a high-risk image, the cumulative score is reflective of that change. Here, α and β are constants. We used a value of 0.8 and 0.25 for α and β respectively.

5. EXPERIMENT AND EVALUATION

5.1 Method

The study adheres to all guidelines for human subjects research. Patients were asked to fill out a pre-experiment questionnaire to obtain basic details such as age, gender, and medical history, as well as to consent to the use of their anonymized images for the study. This was done before providing them an account for DermFollow, to have a baseline

understanding of their skin cancer risk and satisfaction with their dermatologist.

Following this, patients were briefed regarding the application and its features and were given a patient account. Patients uploaded multiple lesion images over the span of a week, after which they were asked to complete a post-experiment questionnaire regarding usability, integration into their daily activities, and value provided from using the application.

Clinician study participants were also given a one-week window to test the application, during which they reviewed patient image analyses and interacted with the visualizations.

5.2 Metrics

Post-experiment patient and provider surveys each contained Likert-scale questions, in which the patient or provider responded to a statement on a scale of one to five, with one being strongly disagree and five being strongly agree.

We computed the mean and standard deviation for the various questions. We additionally asked three questions allowing long-form responses in which the patient and provider were asked to give their comments on the value provided from using the application.

Survey questions were grouped into three major categories:

1. Usability/application performance
2. Workflow/integration into daily life
3. Value provided by the application

6. RESULTS

Table [1] and [2] show the responses collected from the patients before and after the experiment respectively. While 19 patients filled out with pre-experiment questionnaire, only 12 patients uploaded images to DermFollow and completed the post-experiment survey. Table [3] describes the responses collected from the single physician participant.

Pre-study patient responses indicated that patients felt generally neutral when it came to feeling empowered with regard to their health. Patient surveys also showed that patients felt relatively neutral with regard to satisfaction with their dermatologist.

In post-study surveys, patients indicated that the application was generally easy to use, reliable, intuitive, and enjoyable to use. Patients indicated that the body map feature was useful, and that DermFollow would increase communication with the patient's physician if it were used in clinical practice.

A few representative free-form responses collected from patients are as follows:

"I will still see the doctor but this enhances existing communication. It may get some patients to the doctor earlier when time is a critical factor."

"Clearly there is value in being able to have a dermatologist do a quick assessment of a mole/spot on the skin without having to schedule an appointment and visiting the office."

"This application would save time going to physician. Appointments are hard to get with a dermatologist. People don't want to take the time to go for check-ups. This application would be useful for this simple fact."

Table 1: Patient Pre-Study Surveys (N=19)

Category	Statement	Mean (SD)
Cancer awareness	Awareness of skin cancer risk	3.16 (1.64)
Cancer awareness	Estimated skin cancer risk for self	2.58 (1.26)
Health empowerment	Sense of empowerment in health	3.05 (1.27)
Health empowerment	Satisfaction with dermatologist	3.05 (1.31)

Table 2: Patient Post-Study Surveys (N=12)

Category	Statement	Mean (SD)
Usability	Easy to use	4.6 (0.67)
Usability	Performed reliably	4.6 (0.67)
Usability	Intuitive to use	4.3 (0.98)
Usability	Enjoyable to use	4.4 (0.9)
Workflow	Did not interfere with my daily routine	4.5 (0.8)
Workflow	Fit into my daily routine	4.7 (0.49)
Workflow	The body map was useful	4.42 (0.79)
Value	Would enhance communication with my doctor	4.58 (0.9)
Value	Would increase the quality of follow-up	4.3 (0.89)
Value	Would increase the quality of care	4.3 (0.89)
Value	Raised my level of health awareness	4 (1.48)

The provider indicated that the application was easy to use, performed reliably, and was intuitive and enjoyable to use. The provider was neutral on whether the application would easily integrate into the existing clinical workflow. The provider did feel that DermFollow would provide more data for clinical decision-making, and would enhance patient-provider communication.

It should be noted that patient and provider sample sizes were very small (N=12, and N=1, respectively); thus, only limited conclusions can be drawn as a result of the data collected.

Table 3: Provider Post-Study Survey (N=1)

Category	Statement	Mean (SD)
Usability	Easy to use	5
Usability	Performed reliably	4
Usability	Intuitive to use	4
Usability	Enjoyable to use	4
Workflow	Did not interfere with clinical workflow	4
Workflow	Enhanced my clinical workflow	4
Workflow	Would easily integrate in clinical workflow	3
Value	Would enhance patient-provider communication	4
Value	Provided more data for decisionmaking	4
Value	Accuracy of application's risk assessment	3

7. DISTRIBUTION

All team members contributed equally to this project. Stefano developed the neural network model. Matt Cimino conducted user studies. Matt May built the web application, helped with user studies, and developed the visualizations.

Thanh developed the risk score and feature extraction code. Apurv focused on report deliverables and data collection.

8. CONCLUSIONS AND DISCUSSION

DermFollow helps dermatologists make more informed diagnostic and treatment decisions by performing risk analysis on images of skin lesions. It also performs patient risk scoring, and presents an interactive, user-friendly interface for both the physician and patient.

Outcomes for patients with skin cancer can be significantly improved by early diagnosis [11]. Applications like DermFollow can serve as important tools that promote better patient care and follow-up, by allowing providers to monitor their patients more closely over time.

Further studies will be needed to more quantitatively assess the effects of DermFollow in a clinical setting. Additionally, further research is needed to determine the most effective algorithmic approach to perform risk analysis of skin lesions. As the images encountered in practice are extremely diverse, a robust ensemble approach will likely be most fruitful.

9. APPENDIX

Many thanks to our wonderful physician-advisers, Justin Ko, MD, MBA of Stanford Dermatology and Ben Stoff, MD, MS of Emory Healthcare. Their assistance throughout this project was invaluable.

10. REFERENCES

[1] B. Amaliah, C. Fatichah, and M. R. Widyanto. Abcd feature extraction of image dermatoscopic based on morphology analysis for melanoma skin cancer diagnosis. *Jurnal Ilmu Komputer dan Informasi*, 3(2):82–90, 2012.

[2] M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2301–2309, 2011.

[3] R. Bostock, E. Claridge, A. Harget, and P. Hall. Towards a neural network based system for skin cancer diagnosis. In *Artificial Neural Networks, 1993., Third International Conference on*, pages 215–219. IET, 1993.

[4] C.-L. Chang and C.-H. Chen. Applying decision tree and neural network to increase quality of dermatologic diagnosis. *Expert Systems with Applications*, 36(2):4035–4041, 2009.

[5] J. R. Davies, Y.-m. Chang, D. T. Bishop, B. K. Armstrong, V. Bataille, W. Bergman, M. Berwick, P. M. Bracci, J. M. Elwood, M. S. Ernstoff, et al. Development and validation of a melanoma risk score based on pooled data from 16 case–control studies. *Cancer Epidemiology Biomarkers & Prevention*, 24(5):817–824, 2015.

[6] F. Ercal, A. Chawla, W. V. Stoecker, H.-C. Lee, and R. H. Moss. Neural network diagnosis of malignant melanoma from color images. *Biomedical Engineering, IEEE Transactions on*, 41(9):837–845, 1994.

[7] A. Esteva, B. Kuprel, and S. Thrun. Deep networks for early stage skin disease and skin cancer classification. Unpublished manuscript, 2015.

[8] H. Ganster, A. Pinz, R. Röhrer, E. Wildling, M. Binder, and H. Kittler. Automated melanoma recognition. *Medical Imaging, IEEE Transactions on*, 20(3):233–239, 2001.

[9] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[10] ISIC. International skin imaging collaboration archive. https://isic-archive.com/, 2016.

[11] A. F. Jerant, J. T. Johnson, C. Sheridan, T. J. Caffrey, et al. Early detection and treatment of skin cancer. *American family physician*, 62(2):357–386, 2000.

[12] M. Kreutz, M. Anschütz, S. Gehlen, T. Grünendick, and K. Hoffmann. *Bildverarbeitung für die Medizin 2001: Algorithmen — Systeme — Anwendungen*, chapter Automated Diagnosis of Skin Cancer Using Digital Image Processing and Mixture-of-Experts, pages 357–361. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[14] H. Lee and Y. P. Chen. Image based computer aided diagnosis system for cancer detection. *Expert Syst. Appl.*, 42(12):5356–5365, 2015.

[15] D. S. Rigel, R. J. Friedman, A. W. Kopf, and D. Polsky. Abcde—an evolving concept in the early detection of melanoma. *Archives of dermatology*, 141(8):1032–1034, 2005.

[16] P. Rubegni, G. Cevenini, M. Burroni, R. Perotti, G. Dell'Eva, P. Sbano, C. Miracco, P. Luzi, P. Tosi, P. Barbini, et al. Automated diagnosis of pigmented skin lesions. *International Journal of Cancer*, 101(6):576–580, 2002.

[17] M. A. Sheha, M. S. Mabrouk, and A. Sharawy. Automatic detection of melanoma skin cancer using texture analysis. *International Journal of Computer Applications*, 42(20):22–26, 2012.

[18] S. Sigurdsson, P. A. Philipsen, L. K. Hansen, J. Larsen, M. Gniadecka, and H. C. Wulf. Detection of skin cancer by classification of raman spectra. *Biomedical Engineering, IEEE Transactions on*, 51(10):1784–1793, 2004.

[19] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[20] S. M. Strayer and P. Reynolds. Diagnosing skin malignancy: assessment of predictive clinical criteria and risk factors.(research findings that are changing clinical practice). *Journal of family practice*, 52(3):210–219, 2003.

[21] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *CVPR 2015*, 2015.

[22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. *arXiv preprint arXiv:1512.00567*, 2015.

[23] L. Thomas, P. Tranchand, F. Berard, T. Secchi, C. Colin, and G. Moulin. Semiological value of abcde

criteria in the diagnosis of cutaneous pigmented tumors. *Dermatology*, 197(1):11–17, 1998.

[24] E. Zagrouba and W. Barhoumi. A prelimary approach for the automated recognition of malignant melanoma. *Image Analysis and Stereology Journal*, 23(2):121–135, 2004.