Risk Analysis of Skin Cancer using Deep Learning

Thanh Dang, Stefano Fenu, Matthew May, Apurv Verma

March 3, 2016

1 What Are You Trying To Do?

Our goal is to build a website that allows dermatologists to more effectively follow up with patients with skin cancer, or patients at risk for skin cancer. Our website will do this by providing classification of suspected skin cancer lesions, patient risk scoring, and patient image analysis over time. Ultimately, we hope to improve outcomes, promote happier patients, and improve medical practice.

2 How Is It Done Today?

The current state-of-the-art for automated skin cancer analysis generally involves feature detection [10]. In dermatology, there is a widely used ABCDE (Asymmetry, Border, Color, Diameter, Evolving) [14] algorithm that is used by doctors to assess potential skin cancer lesions. Existing approaches often attempt to imitate this algorithm by "teaching" models how to detect the same features [15]. However, the limitation of many of these models has been verification on a relatively small, highly curated test set of images, leading to low-quality results on broader image sets and limited adoption in clinical practice.

Relatively recently, neural networks have been applied to skin cancer classification. In [8], the authors use a combination of neural networks, image processing, and feature extraction techniques to classify skin lesions. Sheha et al. [13] use a multilayer perceptron to classify melanoma, attaining 92% accuracy on their test set. Esteva et al. [4] use an ensemble of convolutional neural networks (CNNs) to attain 90% binary classification (malignant/benign).

3 What's New In Your Approach? Why Will It Succeed?

The primary limitation of current automated approaches has been limited applicability to actual clinical practice. Low patient engagement, limited integration into clinical workflows, poor algorithm performance on actual patient images, and expensive hardware requirements have plagued existing approaches.

To increase patient engagement, our system will present an interface in which patients receive rewards for uploading images of their skin lesions over time, providing an incentive for patient compliance with their treatment plan.

Due to the recent adoption of electronic medical records (EMR), our Web-based system will easily integrate with existing clinical workflows, which are increasingly electronic. Our system will perform analysis on uploaded patient images and then present the results to the physician in a meaningful, uncluttered way, allowing immediate follow-up with the patient if necessary.

To improve algorithm performance, we will use state-of-the-art convolutional neural networks, which have been shown to be highly effective [9] at difficult image classification problems. Furthermore, by allowing patients to upload images of their skin lesions from their smartphone or computer, there will be no additional hardware purchase required of the patient or physician.

4 Who Cares?

In the literature, it has been shown that the implementation of health care IT can lead to significant increases in patient satisfaction [12]. It has also been shown that the experience of the patient in receiving care is related to overall patient satisfaction [2]. Patient satisfaction is one of the primary metrics now considered an indicator of overall health care quality by physicians and hospital administrators [5].

By allowing patients to receive a higher standard of care through the use of unobtrusive technology, we believe patient satisfaction will grow. Furthermore, it has been shown that telemedicine technology, which this could be considered a form of, is effective and increases patient access to care [6]. It also allows for more efficient use of resources, as physicians can interact with their patients in an asynchronous fashion.

In our research, we have often heard that physicians are frustrated by non-compliant patients. By creating a communication tool that connects the physician with his patients, and provides meaningful data that supports patient follow-up, we believe patient outcomes, as well as patient satisfaction will be improved.

5 What Difference Will It Make?

To determine whether our proposed approach would have utility in practice, we spoke with dermatologists at Stanford Medical Center and Emory University. One common thread that we heard was that patient follow-up is a process that has significant gaps. Patients see their dermatologist at six-month, or in some cases, multi-year intervals, and have little to no contact with their physician in between. This creates large time periods where skin lesions continue to change, but there is little or no supervision of their progress.

Our system will enable follow-up on a consistent basis when the patient is not physically in the office, potentially allowing cancerous lesions to be caught at a much earlier time. Furthermore, by analyzing skin lesions over time, we will be able to generate risk profiles for each patient, increasing the quality of clinical decisionmaking. And by allowing the physician to instantly "ping" the patient to schedule an appointment after a questionable lesion is uploaded, our system will promote more frequent patient follow-up when necessary.

6 What Are the Risks and Payoffs?

6.1 Risks

As with any technology, there are certain risks that should be considered. First, dermatologists may be hesistant to adopt our technology because in certain reimbursement schemes, physicians are compensated primarily when patients come for in-person office visits. If our system leads to a decrease in patient visits because of more effective patient management, this could slow adoption. To mitigate this, we should focus on health care systems with positive incentives for managing patients more effectively.

Secondly, the recent widespread adoption of electronic medical records is a relatively costly and complex procedure that can be frustrating for physicians [3]. Thus, some medical practices may be saturated with technology and unwilling to adopt additional technologies. Lastly, on a more technical note, the relatively small size of our image datasets used for training and testing could promote overfitting, leading to suboptimal results when applied to patient imaging in practice.

6.2 Payoffs

The scalability of our system, and the relatively low cost that it could be offered to medical practices at, means that it could feasibly be deployed to a wide range of clinical settings, having a large impact on the quality of medical care. Ultimately, better patient follow-up could mean that patient outcomes are improved (i.e., lives saved), practices are managed more effectively, and patients are more satisfied with the care they've received, making them more likely to return.

7 How Much Will It Cost? How Long Will It Take?

Our approach leverages open-source technology and also a novel technique known as *transfer learning* [11], in which a pre-trained network is used at the initial basis for feature learning. Thus, we are able to produce a minimally viable product at relatively low cost. More specifically, our primary costs are in training neural networks, which are done at a cost of $\sim$\$0.65/hour on Amazon EC2 dedicated GPU instances. For our initial model development and training, we estimate approximately 150 hours of instance utilization ($\sim$\$100 total). Our website will run on a t2.micro instance on Amazon EC2, which is free under a promotion. If the promotion were to expire, we estimate we could operate the server initially at $\sim$\$0.013/hour, or about \$10/month.

We rely on freely available technologies such as Python, NumPy, TensorFlow [1], Caffe [7], and Ruby on Rails for the development of our system. From a human resource perspective, we intend to collectively spend 400 total hours on the project, split amongst model development, neural network training, user interface development, and user studies, which can be estimated at $\sim$\$40,000 in opportunity cost with each of our time valued at an average of \$100/hour.

8 How Will Progress Be Measured?

To measure our progress, we have recruited dermatologists from two leading institutions, to test our demonstration system and provide feedback from a clinical perspective. Moreover, we intend to perform simple user studies/focus groups in which we recruit patients to use our system and offer feedback on their experience and the likelihood they would use it in reality.

From a more technical perspective, we aim to achieve at least 90% in binary skin cancer classification on our test image set. Another checkpoint is the launch of our application as a minimal product in which patients can upload images and their physician can view them as a stream and interact with the data. Lastly, in final form we intend to provide additional user interface enhancements such as patient rewards for uploads and development of the model using an ensemble approach.

Thus far, Stefano has led model exploration and training, examining a wide range of approaches including simple SVM classifiers (baseline), deep CNNs, and pre-trained models. Thanh has developed the user interface and researched patient risk scoring. Matt has prepared our training and test datasets, secured dermatologist collaborators, and developed the initial web application backend. Apurv has done report and presentation generation, and surveyed the relevant literature for prior art.

Going forward, the tasks are split as follows among the team members. Stefano is focusing on model selection, development, and training (100 hours). Thanh is focused on risk score generation for patients, UI design, and user studies (100 hours). Matt is building the web application and handling integration with the model API (100 hours), as well as interfacing with our dermatologist collaborators. Lastly, Apurv is focused on data pre-processing, component integration, and report and presentation development (100 hours). All authors have contributed and will contributed equally in the future to this work.

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] O. A. Bjertnaes, I. S. Sjetne, and H. H. Iversen. Overall patient satisfaction with hospitals: effects of patient-reported experiences and fulfilment of expectations. *BMJ quality & safety*, 21(1):39–46, 2012.

[3] A. Boonstra, A. Versluis, and J. F. Vos. Implementing electronic health records in hospitals: a systematic literature review. *BMC health services research*, 14(1):1, 2014.

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

[5] J. J. Fenton, A. F. Jerant, K. D. Bertakis, and P. Franks. The cost of satisfaction: a national study of patient satisfaction, health care utilization, expenditures, and mortality. *Archives of internal medicine*, 172(5):405–411, 2012.

[6] D. M. Hilty, D. C. Ferrer, M. B. Parish, B. Johnston, E. J. Callahan, and P. M. Yellowlees. The effectiveness of telemental health: a 2013 review. *Telemedicine and e-Health*, 19(6):444–454, 2013.

[7] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[8] 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.

[9] 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.

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

[11] A. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 806–813, 2014.

[12] M. Roham, A. R. Gabrielyan, and N. P. Archer. Predicting the impact of hospital health information technology adoption on patient satisfaction. *Artificial intelligence in medicine*, 56(2):123–135, 2012.

[13] M. A. Sheha, M. S. Mabrouk, and A. Sharawy. Automatic detection of melanoma skin cancer using texture analysis.

[14] 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.

[15] 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.