DermFollow: A System For Better Patient Follow-Up In Dermatology

Thanh Dang, Stefano Fenu, Matthew May, Apurv Verma

March 4, 2016

1   What Are You Trying To Do?

Our goal is to build a Web-based system that allows dermatologists to more effectively follow up with patients that have, or are at risk for, skin cancer. Our system 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 patient outcomes, raise patient satisfaction, and improve the efficiency of 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 [14] algorithm 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 [15]. However, the limitation of many of these models has been low-quality results on the more diverse images actually encountered in practice, limiting adoption.

Relatively recently, neural networks have been applied to skin cancer classification. In [8], the authors use a combination of neural networks and feature extraction to classify skin lesions. Sheha et al. [13] use a multilayer perceptron to classify melanoma, attaining 92% accuracy. Esteva et al. [4] use an ensemble of convolutional neural networks (CNNs) to attain 90% binary classification (malignant/benign). While these results are promising, many of these models were trained on images that lack histological (microscopic) verification, the gold standard for determining malignancy.

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

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 provide patients rewards for uploading images of their lesions over time, providing an incentive for compliance.

Our system will integrate easily with existing electronic clinical workflows. It will perform analysis on uploaded patient images and then present the results to the physician in an uncluttered way, allowing immediate patient follow-up.

To improve algorithm performance, we will use state-of-the-art CNNs, which have been shown to be very effective at image classification [9]. By allowing patients to upload images from their smartphone, no additional hardware will be required. Lastly, we use datasets for training and testing with 100% histological verification.

4 Who Cares?

It has been shown that health care IT can lead to significant increases in patient satisfaction [12]. Patient satisfaction is one of the primary metrics now considered an indicator of health care quality by hospitals [5], [2].

By allowing patients to receive a higher standard of care through technology, we believe patient satisfaction will grow. The literature has shown that telemedicine technology is effective [6]. By providing physicians with meaningful data and analysis, we believe patient outcomes and satisfaction will increase.

5 What Difference Will It Make?

To determine whether our approach would have utility in practice, we spoke with several dermatologists. One common thread was that patient follow-up is limited. Patients see their dermatologist infrequently. This creates large time periods where skin lesions continue to change, but there is no supervision.

Our system will enable consistent follow-up, potentially allowing cancerous lesions to be caught earlier. By analyzing lesions over time, we will

generate risk profiles for each patient, bettering clinical decisionmaking. By allowing the physician to schedule an appointment with a patient after a questionable lesion is uploaded, our system will promote better patient follow-up.

6 What Are the Risks and Payoffs?

6.1 Risks

Dermatologists may be hesistant to adopt our technology because in certain reimbursement schemes, physicians are incentivized only to see the patient in the office. If our system leads to a decrease in patient visits, this could slow adoption. Thus, we should focus on hospitals with positive incentives. Secondly, the "saturation" of technology in medicine [3] could mean physicians are unwilling to adopt additional technology.

6.2 Payoffs

The scalability of our system means that it could be deployed to a wide range of settings, having a large impact on the quality of care. Ultimately, better patient follow-up could mean that patient outcomes are improved (i.e., lives saved), practices are managed better, and patients are more satisfied.

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

Initially, our primary costs are in training neural networks. For our initial model development, we estimate approximately 150 hours of Amazon instance utilization (∼$100 total). Our website will run on a t2.micro instance on Amazon, which is free under a promotion.

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, user interface and web application development, user studies, and report generation, which can be estimated at ∼$40,000 in opportunity cost with each hour of our time valued at $100/hour.

8 How Will Progress Be Measured?

To measure our progress, we have recruited dermatologists to test our demonstration system and provide clinical feedback. Moreover, we intend to perform user studies/focus groups in which patients use our system and offer feedback. We will administer Likert-scale surveys and analyze results.

We aim to achieve $\geq 90\%$ in binary skin cancer classification on our test set. Another checkpoint is the version one application in which patients can upload images and their physician can view them and interact with the data. Lastly, in final form there will be patient rewards and an optimized model.

Thus far, Stefano has led model exploration, examining SVM, CNNs, and pre-trained models [11]. Thanh has developed the UI and patient risk scoring. Matt has gathered/cleaned datasets, secured dermatologist collaborators, and developed the web application. Apurv has done report and presentation generation.

Future tasks are split as follows. Stefano is focusing on model development and training (100 hours). Thanh is focused on risk score generation, UI design, and user studies (100 hours). Matt is building the web application (100 hours). Apurv is focused on data management and report 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.