DermFollow: A System For Better Diagnosis and Treatment of Skin Cancer

Matt Cimino, Thanh Dang, Stefano Fenu,
Matt May, Apurv Verma

March 5, 2016

1 What Are You Trying To Do?

Our goal is to build a web application that promotes more effective diagnosis and treatment of skin cancer. Our system enables this through classification of suspected skin cancer lesions, patient risk scoring, and patient image analysis over time. 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 [14], [9], [18]. In dermatology, there is an ABCDE algorithm [24], [16], [22] 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 [25]. Some of these models have yielded poor results on the more diverse images encountered in practice, limiting adoption.

Neural network-based approaches have also been used for skin cancer classification [20], [6], [4], [5]. Kreutz et al. [12] use neural networks and feature extraction to classify skin lesions. Sheha et al. [19] use a multilayer perceptron to classify melanoma, attaining 92% accuracy. Esteva et al. [7] use an ensemble of convolutional neural networks (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.

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 incentives for uploading images of their skin lesions over time, increasing compliance.

Our system will integrate seamlessly with the largely electronic clinical workflow, through a Web-based interface that operates on the desktop and mobile devices. 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 deep CNNs, which have been shown to be very effective at image classification [13], [23], [21]. Lastly, we use an image dataset with 100% histological verification.

4 Who Cares?

It has been shown that healthcare IT can lead to significant increases in patient satisfaction [17]. Patient satisfaction is one of the primary metrics now considered an indicator of health care quality by hospitals [8], [2].

We believe our system will promote a higher standard of patient care, increasing patient satisfaction. The literature has shown that telemedicine technology is effective [10]. By giving doctors better information to make decisions with, 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, improving 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 structures, 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 hard costs are in training models. 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], and Ruby on Rails for the development of our system. From a human resource perspective, we intend to collectively spend 400 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?

We have recruited dermatologists to test our 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 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 incentives and an optimized model.

Thus far, Stefano has led model exploration, examining SVM, CNNs, and pre-trained models [15]. Thanh has developed the UI and patient risk scoring. Matt May has collected datasets (we use the International Skin Imaging Collaboration (ISIC) dataset [11], with $\sim$3400 high-resolution images), secured dermatologist collaborators, and developed the web application. Apurv has done report/presentation generation. Matt Cimino has done user study design.

Going forward, Stefano is focusing on model development and training (80 hours). Thanh is focused on risk score generation and UI design (80 hours). Matt May is building the web application (80 hours). Apurv is focused on data management and report development (80 hours). Matt Cimino is conducting user studies (80 hours). All authors have contributed equally and will contribute equally in the future.

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] 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.

[5] 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.

[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] 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.

[9] 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.

[10] 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.

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

[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] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

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

[22] 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.

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

[24] 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.

[25] 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.