Utilizing convolutional neural networks to detect skin cancer: a review of initial trials

Matthew Ladda (BSc(Pharm))[1]; Trevor Champagne (MD, FRCPC)[1,2]

[1]Faculty of Medicine, University of Toronto, Medical Sciences Building, 1 King's College Circle, Toronto, ON, Canada, M5S 1A8.
[2]Division of Dermatology, Women's College Hospital, 76 Grenville St, Toronto, ON, Canada, M5S 1B2.

Abstract

A convolutional neural network (CNN) is a type of deep, feed-forward artificial neural network able to detect, recognize, and classify visual imagery. Given the visual nature of dermatology, recent research has explored the ability of trained CNNs to detect skin cancers. The ability of CNNs to correctly categorize biopsy-confirmed images of skin lesions has been compared to that of dermatologists in several studies discussed herein. This article will review several studies that evaluated the ability of CNNs to detect skin cancer. Given the diagnostic accuracy that CNNs have demonstrated in studies to date, and the ease of which they could be incorporated into smartphones and digital dermatoscopes, CNNs have the potential to significantly improve the detection of skin cancer and potentially other dermatological diseases. Further research is needed regarding how CNNs could be effectively integrated into clinical practice.

Introduction

Skin cancer is one of the most common cancers worldwide and is the most common cancer affecting white-skinned individuals.[1,2] Skin cancers can be divided into two categories: melanoma and non-melanoma skin cancers. Non-melanoma skin cancers, particularly basal cell carcinomas and squamous cell carcinomas, are very common. Melanoma, while less common, carries a 5-year survival rate of approximately 15% if detected in its latest stage.[3] Early detection of skin cancer can allow for earlier treatment intervention and improved outcomes.[4]

Skin cancers are usually first identified visually. Accordingly, there has been interest in developing a computer-based method of image analysis that could detect skin cancers with a high degree of specificity and sensitivity. However, the appearance of a lesion in a photographic image can vary depending on the circumstances in which the photograph is taken. For example, factors such as lighting, angle, focus, and zoom can introduce variability and challenges for image classification software.[5] Traditional image classification algorithms also relied on common features identified by humans, such as colour or border information.[6,7] For example, characteristic features of melanoma first noted by the human eye, such as border irregularity, would be included in the traditional algorithm as a key factor. As human recognition of different lesion appearances is limited, this meant that the resulting algorithms would also be limited.

Image classification has long been a challenging frontier in computer vision. Recently, deep learning convolutional neural networks (CNNs) have emerged as an effective means to recognize, classify, and detect visual imagery.[8] While CNNs are complex and a detailed description of their design is outside the scope of this review, a basic understanding of CNNs can be gleaned by an understanding of the visual cortex in the brain. Inspired by biology, CNNs take inputs (such as an image of a skin lesion), process them through a series of unidirectional steps, and then arrive at an output (e.g. a diagnosis). CNNs are also capable of learning and improving their accuracy by adjusting the weight of different features to a particular output.[8] Therefore, CNNs can be "trained" with large amounts of high-quality images to fine-tune important identifying features.

Diagnosing a Picture – CNNs versus Dermatologists

In a pivotal study by Esteva et al., the ability of a trained CNN to classify skin lesions was compared to that of 21 board-certified dermatologists.[9] The CNN in this study was trained using a dataset of 129,450 images representing 2,032 individual diseases. In the later stages of algorithm validation, the investigators tested the CNN's ability to distinguish previously unseen biopsy-confirmed keratinocyte carcinomas from benign seborrheic keratoses, and malignant melanoma from benign nevi. The first case represents the ability to diagnose the most common skin cancers, where the second case represents the ability to diagnose the most aggressive skin cancers. In the case of distinguishing malignant melanoma from benign nevi, the CNN was given both standard images and dermoscopic images. For each image, dermatologists were also asked independently to either biopsy or treat the lesion, or reassure the patient. Sensitivity and specificity were measured for both dermatologists and the CNN in regards to detecting keratinocyte carcinoma and melanoma on standard images, and melanoma under dermoscopy. In all cases, the CNN marginally outperformed both the average of the 21 dermatologists and the majority of the dermatologists individually.

Results from the study by Esteva et al. showed that their CNN had an ability similar to board-certified dermatologists at classifying skin cancers based on images. However, lesions are not solely diagnosed or categorized based on their appearance. Dermatologists take contextual factors into account when creating a differential diagnosis, such as the clinical history, lesion size, tactile feel, and three-dimensional appearance. As the clinical context of each lesion was not provided in this study, performance of the CNN compared to an in-person assessment by a dermatologist could not be assessed.

Recently, a study by Haenssle et al. incorporated some clinical context of a lesion into the comparison between a CNN and dermatologists.[10] This study assessed the ability of a CNN to detect dermoscopic melanoma in comparison to an international group of 58 dermatologists, including 30 who had over 5 years of experience with dermoscopy. In this study, a pre-trained CNN was additionally trained with over 100,000 images and corresponding disease labels. Dermatologists for this study were invited from a mailing list of the International Dermoscopy Society, and a response rate of 33.7% was received. The investigators created a set of 100 validated dermoscopic images, none of which had been used to train the CNN. In phase 1, dermatologists were asked whether each lesion was a benign nevus or melanoma, and what their management decision would be (excision, short-term follow-up, or to reassure the patient). The option to select short-term follow up was included as it can help identify lesions that are rapidly changing, and selection of this option was considered to be a true positive in sensitivity calculations and a true negative in specificity calculations. The mean (± standard deviation) diagnostic sensitivity and specificity for the dermatologist group was 86.6% (±9.3%) and 71.3% (±11.2%), respectively. With regards to dermatologist management in phase 1, sensitivity was 98.8% (±2.9%) and specificity was 64.6% (±13.6%). In phase 2, the same group of dermatologists were provided with the same 100 case images 4 weeks later in addition to age, sex, and body site, as well as close-up images. When provided additional clinical information and close up images in phase 2 of the study, diagnostic sensitivity increased to 88.9% (±9.6%) and specificity to 75.7% (±11.7%). Management sensitivity and specificity in phase 2 was 98.6% (±2.8%) and 66.7% (±12.4%), respectively. Using the dermatologists' mean diagnostic sensitivity of 86.6% in phase 1 as a benchmark on the CNN receiver operating characteristics curve, the CNN was found to have a higher specificity of 82.5% ($p<0.01$). Using the dermatologists' phase 2 diagnostic sensitivity of 88.9% as the operating point on the CNN receiver operating characteristics curve, the CNN also had a higher specificity of 82.5% ($p<0.01$).

This study by Haenssle et al. provides additional evidence supporting the effectiveness of trained CNNs in detecting melanoma. By providing some clinical information about a lesion to the dermatologist group, they were better able to mimic a real life patient encounter. That said, only the age, sex and body site was provided to the dermatologists. A comprehensive clinical history was not given for each case, and it is possible that the provision of a patient history could have improved the diagnostic accuracy of the dermatologists. Nonetheless, this study demonstrated that their trained CNN had relatively high sensitivity and specificity for the detection of melanoma.

Diagnosing a Human – CNNs in Clinical Practice

To date, early trials have shown CNNs to be a powerful method of detecting skin cancer.[9,10] After an extensive training period, it appears that CNNs are able to classify images of skin cancers about as accurately as a dermatologist. Though the specific architecture of a CNN can be complicated to construct, once created it can be easily integrated into smartphone applications or a digital dermatoscope. The ways in which this novel research could be translated to clinical practice has been the subject of much discussion.[9-12]

Several concerns must be addressed before CNNs can be implemented in clinical practice. Since it is essential to train CNNs to recognize images, an extensive training method should be developed. The sensitivity and specificity of CNNs trained with different large image sets should also be compared to ensure consistency and reproducibility. Image sets used in CNN training should also include lesion appearances across a spectrum of skin colours, as well as atypical lesion presentations to improve diagnostic ability. Prior to commercial use, approval from jurisdictional regulatory agencies should be sought to ensure the product has been appropriately assessed for effectiveness and safety. Ideally, data from prospective clinical trials assessing the congruency of CNN diagnostic predictions to biopsy results should be obtained.

Additionally, it must be decided how this technology will be used – whether or not it will be largely restricted to use by dermatologists, or if access via a smartphone application will be made available to the general public. In the hands of a dermatologist, a trained CNN can complement a dermatologist's independent assessment of a patient and help confirm diagnoses, similar to how medical imaging currently assists in the diagnosis of many diseases. CNN interpretations could also be used to increase or decrease a dermatologist's index of suspicion for indeterminate lesions. If also made available directly to patients, it is possible that concerning lesions could be detected faster and the patient could be rapidly triaged to a dermatologist. However, there is concern that a false benign diagnosis from a CNN could delay a patient in seeking medical attention. In this event, who would be liable? Furthermore, for a patient to appropriately use a CNN to classify a lesion, the patient must first notice the lesion. An overreliance on a smartphone CNN application could lead to delayed diagnosis of lesions in areas that are often overlooked by patients, such as the back and behind the ears.[13]

Future Directions

Though CNNs offer great promise, additional research is required before they can be implemented in clinical practice. Further research is required to determine the most effective method of training a CNN to diagnose skin cancer, and how this technology can be incorporated most effectively into existing practice.

The utility of CNNs to make alternative dermatologic diagnoses should be explored. Currently, CNNs have also been shown to have accuracy similar to dermatologists at detecting onychomycosis.[14] Given the mechanism of CNNs, almost anything that can be identified visually could theoretically be identified by a CNN. As such, it is likely that CNNs have the potential to identify many other dermatological diseases.

Conclusions

Trained CNNs have demonstrated the ability to detect skin cancer in an image with a sensitivity and specificity similar to that of a dermatologist. Though further research is required to fully comprehend their uses and limitations, it is possible that CNNs will eventually be a powerful diagnostic tool in the armamentarium of dermatologists.

References

1. Cancer [Internet]. World Health Organization. [cited 2018]. Available from: http://www.who.int/news-room/fact-sheets/detail/cancer
2. Lomas A, Leonardi-Bee J, Bath-Hextall F. A systematic review of worldwide incidence of nonmelanoma skin cancer. Br J Dermatol. 2012;166(5):1069–80.
3. Survival statistics for melanoma skin cancer - Canadian Cancer Society [Internet]. www.cancer.ca. [cited 2018]. Available from: http://www.cancer.ca/en/cancer-information/cancer-type/skin-melanoma/prognosis-and-survival/survival-statistics/?region=on
4. Conic RZ, Cabrera CI, Khorana AA, et al. Determination of the impact of melanoma surgical timing on survival using the National Cancer Database. J Am Acad Dermatol. 2018;78(1):40-46.e7.
5. Ramlakhan K, Shang Y. A Mobile Automated Skin Lesion Classification System. In: Proceedings of the 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence [Internet]. Washington, DC, USA: IEEE Computer Society; 2011 [cited 2018 Oct 18]. p. 138–141. (ICTAI '11). Available from: http://dx.doi.org/10.1109/ICTAI.2011.29
6. Ferris LK, Harkes JA, Gilbert B, et al. Computer-aided classification of melanocytic lesions using dermoscopic images. J Am Acad Dermatol. 2015;73(5):769–76.
7. Chang W-Y, Huang A, Yang C-Y, et al. Computer-aided diagnosis of skin lesions using conventional digital photography: a reliability and feasibility study. PloS One. 2013;8(11):e76212.
8. Rawat W, Wang Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Comput. 2017;29(9):2352–449.
9. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–8.
10. Haenssle HA, Fink C, Schneiderbauer R, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol. 2018;29(8):1836–42.
11. Abbott LM, Smith SD. Smartphone apps for skin cancer diagnosis: Implications for patients and practitioners. Australas J Dermatol. 2018;59(3):168–70.
12. Mar VJ, Soyer HP. Artificial intelligence for melanoma diagnosis: how can we deliver on the promise? Ann Oncol. 2018;29(8):1625–8.
13. Kantor J, Kantor DE. Routine dermatologist-performed full-body skin examination and early melanoma detection. Arch Dermatol. 2009;145(8):873–6.
14. Han SS, Park GH, Lim W, et al. Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network. PloS One. 2018;13(1):e0191493.