DIRECT AI-BASED PREDICTION OF CLINICAL MANAGEMENT BYPASSING DIAGNOSIS: APPLICATION TO SKIN LESIONS

Kumar Abhishek, Jeremy Kawahara, Ghassan Hamarneh

Medical Image Analysis Lab, School of Computing Science, Simon Fraser University, Canada

{kabhishe, jkawahar, hamarneh}@sfu.ca

Existing machine learning approaches that suggest management decisions for skin lesions rely on predicting the underlying skin condition to infer a decision without considering the variability of decisions that may exist within a single condition. We present the first deep learning-based work to predict clinical management decisions directly from images without explicitly predicting the diagnosis. We use clinical and dermoscopic images of skin lesions and patient metadata from the Interactive Atlas of Dermoscopy dataset (1,011 cases; 20 disease labels; 3 management decisions) and demonstrate that predicting management labels directly is more accurate than predicting the diagnosis and then inferring the management decision (13.73±3.93% and 6.59±2.86% improvement in overall accuracy and area under receiver operating characteristic curve [AUROC] respectively). Based on the retrospective analysis, directly predicting management decisions also considerably reduces the over-excision rate as compared to management decisions inferred from diagnosis predictions (24.56% fewer cases wrongly predicted to be excised). Furthermore, we show that training a model to also simultaneously predict the seven-point criteria and the skin lesion diagnosis yields an even higher accuracy (4.68±1.89% and 2.24±2.04% improvement in overall accuracy and AUROC respectively) of management predictions. Finally, we demonstrate our model's generalizability by evaluating on the public MClass-D dataset (100 cases) and show that our model agrees with the clinical management recommendations of 157 dermatologists as much as they agree amongst each other. We believe that such a system has the potential to suggest management decisions to clinicians (as a second opinion) or directly to patients in under-served communities.

Category: 2. Early experiments with well defined objectives/hypotheses
Learning Objective
Can a deep learning algorithm using convolutional neural networks accurately predict the clinical management of skin lesions without explicitly predicting the diagnosis?
Although deep learning-based computer-aided diagnosis of skin lesions from images is approaching the performance levels of dermatologists, very little research has been done towards the direct prediction of clinical management decisions of skin lesions.

Directly predicting management decisions may be more accurate than inferring management using the predicted diagnoses for skin cancer.

Predicted management decisions show a high degree of agreement with the dermatologists.

In addition to serving remote communities, direct prediction of management decisions from images can potentially help dermatologists ascertain if a lesion should be biopsied.
Medical Image Analysis Pipeline

- Imaging
- Medical Images
- Segmentation
- Segmentation-based Features
- Classification
- Management

Dermatoscope → Medical Images → Segmentation → Segmentation-based Features → Classification → Management

ABCDE features
CASH features
7-point criteria

Signs of melanoma

Melanoma

Excision
Why Predict Management Decisions?

- **Management is the ultimate task** and diagnosis can be considered a latent task, so better to dedicate model capacity for predicting management.

- Multiple diagnosis classes may be managed similarly.

- **Images may not have enough visual information** for accurate diagnosis, and dermatologists often request a biopsy for confirmation.
Deep Learning-based Diagnosis of Skin Cancer

**Dermatologist-level classification of skin cancer with deep neural networks**

Andre Esteve, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun

**Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task**


**Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks**

Roman C. Maron, Michael Weichert, Jochen S. Utikal, Achim Hekler, Carola Berking, Axel Hauschild, Alexander H. Erik, Sebastian Haefekamp, Joachim Klode, Dirk Schadendorf, Philipp Jansen, Tim Holland-Letz, Bastian Schilling, Christoph von Kalle, Stefan Fröhling, Maria R. Gaier, Daniela Hartmann, Anja Gasser, Alexander Thiem
Inferring Management Decisions from Diagnosis Predictions

1. Imaging
2. Medical Images
3. Predicted Diagnosis
4. Diagnosis to Management Mapping
5. Management

- Melanoma ⇒ excise
- Nevus ⇒ follow up

Excision
Directly Predicting Management Decisions

Imaging → Medical Images → Direct Prediction → Predicted Management

Direct Prediction → Excision
Interactive Atlas of Dermoscopy Dataset

5 diagnosis classes:

• basal cell carcinoma (BCC)
• nevus (NEV)
• melanoma (MEL)
• seborrheic keratosis (SK)
• others (MISC)

3 management decisions:

• ‘clinical follow up’ (CLNC)
• ‘excision’ (EXC)
• ‘no further examination’ (NONE)
Inferring Management from Predicted Diagnosis

**Binary Mapping**
- malignant: excise
- benign: do not excise

\[
\begin{bmatrix}
P_{\text{BCC}} & P_{\text{NEV}} & P_{\text{MEL}} & P_{\text{MISC}} & P_{\text{SK}}
\end{bmatrix}
\]

argmax()

- [1 0 0 0 0] → EXC
- [0 1 0 0 0] → NOEXC
- [0 0 1 0 0] → EXC
- [0 0 0 1 0] → NOEXC
- [0 0 0 0 1] → NOEXC

**MGMT\textsubscript{infr, binary}**

- weighted average using prior probabilities
- \text{argmax()}

**MGMT\textsubscript{infr, all}**
- most likely label

9
Predicted management decisions are more accurate than inferred decisions. (statistically significant at $p < 0.001$)
Evaluating Model Generalization

Melanoma Classification Benchmark Dataset

100 images

2 diagnosis classes:
- melanoma (MEL)
- nevus (NEV)

2 management decisions:
- ‘biopsy/further treatment’ (EXC)
- ‘reassure the patient’ (NOEXC)

For each image, clinical management recommendations from 157 German dermatologists (43.9% board certified)

Use aggregated decision as ground truth.
Comparing Predicted versus Inferred Management

Predicted management decisions are more accurate than inferred decisions.
Evaluating Model Agreement with Dermatologists

Two agreement measures

- Cohen’s kappa (model, derm) > Cohen’s kappa (dermᵢ, dermⱼ)
  - 0.5424 > 0.4124

- Fleiss’ kappa (model, derm) ≈ Fleiss’ kappa (dermᵢ, dermⱼ)
  - 0.4080 ≈ 0.4086

Our model’s predictions agree with dermatologists at least as much as they agree amongst each other.
Conclusion

- Directly predicting management decisions may be more accurate than inferring management for skin cancer.
- Predicted management decisions show a high degree of agreement with the dermatologists.
Thank you.
Directly predicting the clinical management decisions of skin lesions is more accurate than relying on a predicted diagnosis to infer the management decision.

The management prediction models are able to generalize well to unseen datasets from different sources.

These direct management predictions show a high degree of agreement with those made by 157 dermatologists.

Funding

Direct AI-based Prediction of Clinical Management Bypassing Diagnosis: Application to Skin Lesions

Kumar Abhishek, Jeremy Kawahara, Ghassan Hamarneh
Medical Image Analysis Lab, Simon Fraser University
kabhishe@sfu.ca
Directly predicting the clinical management decisions of skin lesions is more accurate than relying on a predicted diagnosis to infer the management decision.

The management prediction models are able to generalize well to unseen datasets from different sources.

These direct management predictions show a high degree of agreement with those made by 157 dermatologists.
Direct AI-based Prediction of Clinical Management Bypassing Diagnosis: Application to Skin Lesions

Kumar Abhishek, Jeremy Kawahara, Ghassan Hamarneh
Medical Image Analysis Lab

kabhishe@sfu.ca

March 6, 2021