Novartis benchmarking initiative: making sense of AI

Mark Baillie
AMLD EPFL, Lausanne
January 28th, 2020

NOVARTIS | Reimagining Medicine

DataArt launches SkinCareAI app to detect early melanoma signs

SHARE

https://www.medicaldevice-network.com/news/dataart-launches-skincareai-app/

How do we know it works?

The five rights are at the heart of clinical drug development

Right drug

Right patient

Right time

Right dose

Right route

Discovery | Early development | Full development | Post-market
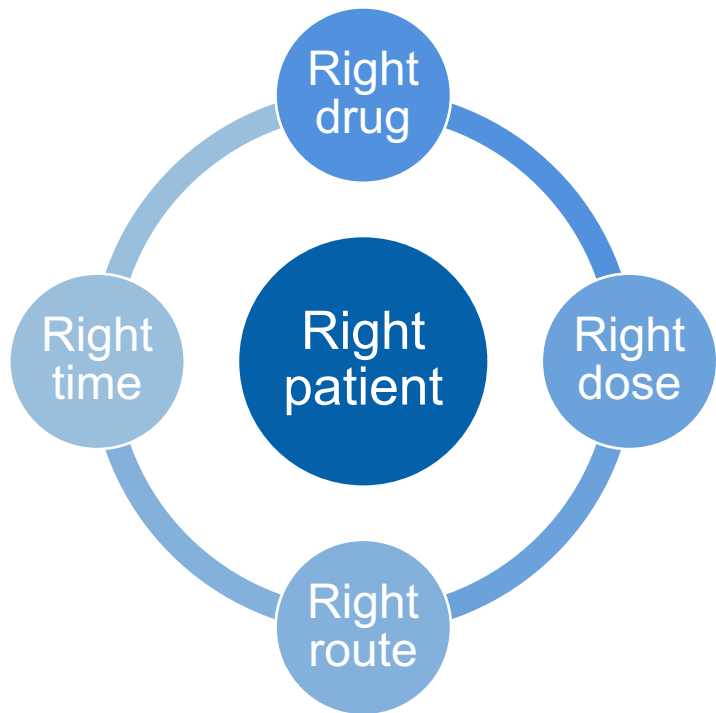
I work here

"Personalized [health care] … means knowing what works, knowing why it works, knowing whom it works for, and applying that knowledge for patients." (Personalized Medicine Coalition)

Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition

Julia K. Winkler, MD; Christine Fink, MD; Ferdinand Toberer, MD; Alexander Enk, MD; Teresa Deinlein, MD; Rainer Hofmann-Wellenhof, MD; Luc Thomas, MD; Aimilios Lallas, MD; Andreas Blum, MD; Wilhelm Stolz, MD; Holger A. Haenssle, MD
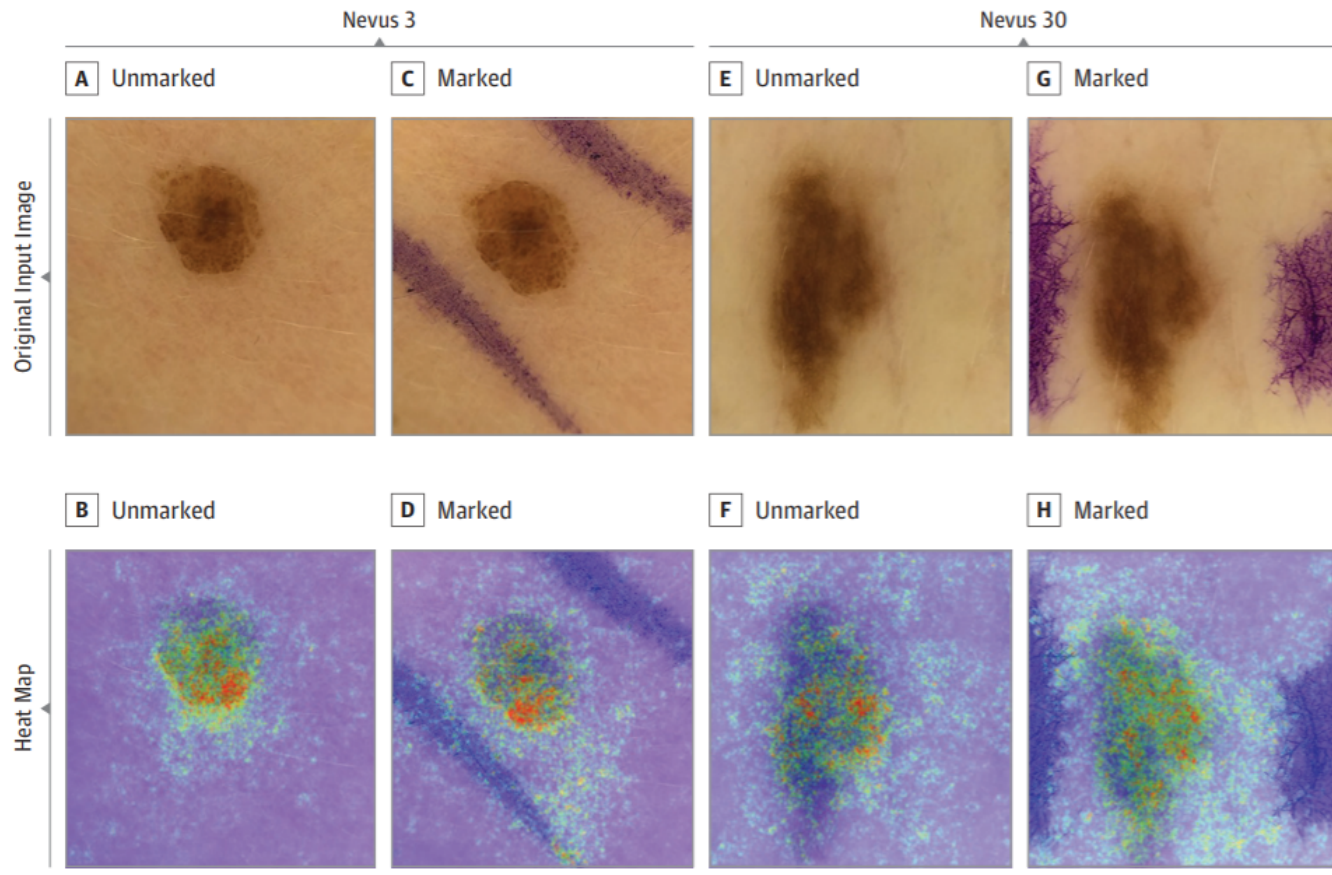
+ Supplemental content

IMPORTANCE Deep learning convolutional neural networks (CNNs) have shown a performance at the level of dermatologists in the diagnosis of melanoma. Accordingly, further exploring the potential limitations of CNN technology before broadly applying it is of special interest.
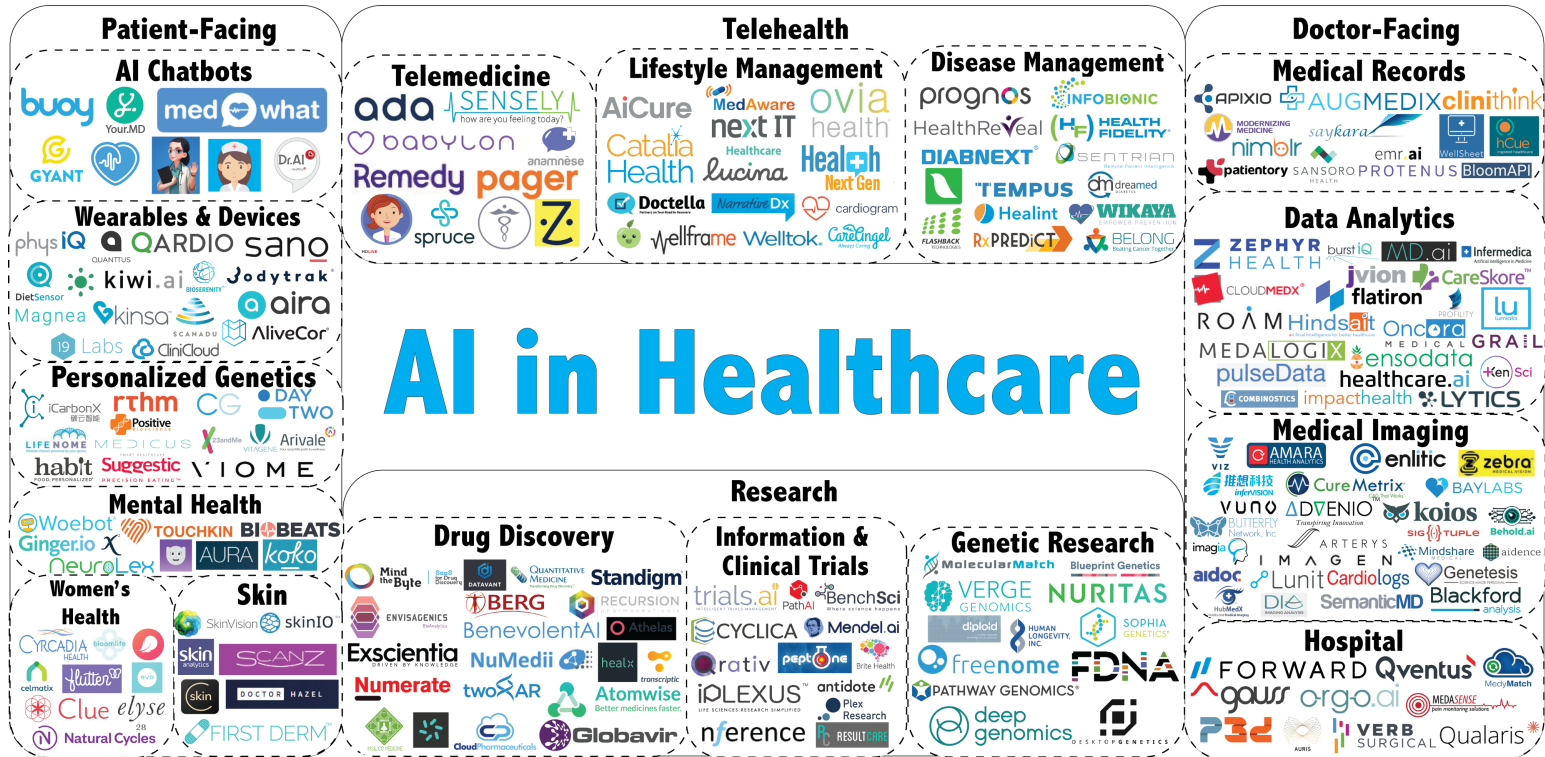
https://jamanetwork.com/journals/jamadermatology/fullarticle/2740808

ɯ NOVARTIS | Reimagining Medicine

Figure 3. Heat Maps of 2 Benign Nevi With Unchanged Melanoma Probability Scores After Addition of In Vivo Skin Markings

Nevus 3

A Unmarked **C** Marked

E Unmarked **G** Marked

Nevus 30

Original Input Image

B Unmarked **D** Marked

F Unmarked **H** Marked

Heat Map

ining Medicine

How do we know it works?

How do we systematically evaluate?

- A standard process for benchmarking «innovtation»:
  - Common task framework
  - Reporting guidelines
- This process aims to:
  - **evaluate** and **compare** «innovtation» on relevant tasks
  - **de-risk** engagement
  - **reduce** internal resources for evaluation

Common task framework

Discussion

50 Years of Data Science

David Donoho ✉

Pages 745-766 | Received 01 Aug 2017, Published online: 19 Dec 2017
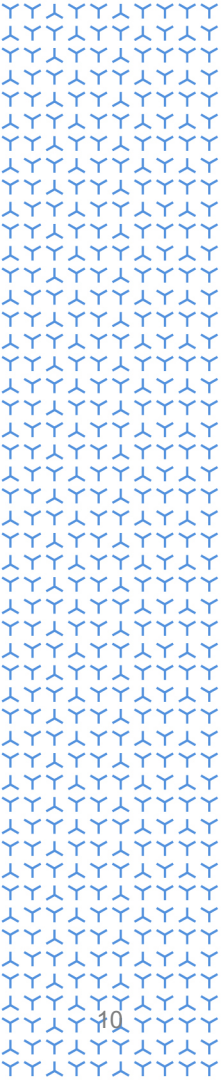
❝ Download citation ↗ https://doi.org/10.1080/10618600.2017.1384734

Common task

Shared data

Standard evaluation

https://www.tandfonline.com/doi/full/10.1080/10618600.2017.1384734

An approximate answer to the right question is worth a great deal more than a precise answer to the wrong question.

- John Tukey

https://projecteuclid.org/download/pdf_1/euclid.aoms/1177704711

Reporting guidelines

https://www.tripod-statement.org/

Why reporting guidelines such as TRIPOD?

- TRIPOD is an evidence-based, minimum set of recommendations for reporting prediction modeling studies in biomedical sciences.

- TRIPOD is part of a wider set of guidelines under the https://www.equator-network.org/ including CONSORT for clinical trials

- TRIPOD includes both prognostic and diagnostic prediction models as well as prediction model development, validation, updating or extending studies (i.e. the core of AI/ML).

- TRIPOD offers a standard way for reporting the results of prediction modeling studies and thus aiding their critical appraisal, interpretation and uptake by potential users.

- TRIPOD and other related reporting guidelines have been adopted by many top tier scientific journals

Putting it all together

| Challenge issuance | Transfer data | Q&A call | Challenge | Report and Evaluation | Debrief |

- We have been evaluating the approach as a proof of concept
  - Issue issuance document with detailed information on challenge
  - Transfer data through secured service on receipt of signed document
  - Set up introductory call
  - Participant submits a short report documenting solution
  - Evaluation primarily based on the <u>TRIPOD</u> guidelines
  - Debrief call

U NOVARTIS | Reimagining Medicine

Progress and learnings so far

- Learnings
- Black boxes
- Next steps

Confessions of a pragmatic statistician

Chris Chatfield

University of Bath, UK

In summary, the pragmatic statistician realizes that the really important actions during a statistical study include

(a) exploring the *context*—obtaining sufficient background information to formulate the problem carefully,
(b) collecting the necessary *data* in a valid way,
(c) carrying out a preliminary examination of the data,
(d) formulating an appropriate *model* and being willing to revise it,
(e) checking the predictive accuracy of the model by using out-of-sample results wherever possible,
(f) taking active steps to avoid trouble and
(g) communicating the results clearly.

Black boxes?

Predictive analytics in health care: how can we know it works? 🔓

Ben Van Calster ✉, Laure Wynants, Dirk Timmerman, Ewout W Steyerberg, Gary S Collins

https://academic.oup.com/jamia/advance-article/doi/10.1093/jamia/ocz130/5542900

Next steps: scaling up

- Develop new challenges

- We have tested this approach, the next step is to scale up across the wider organization (i.e. all development units, countries, etc.)

- Proactively engage scientifically community on methodology research where there is a need to develop new benchmarks which reflect real world

Thank you

Mark Baillie (with Conor Moloney & Janice Branson)
AMLD EPFL, Lausanne
January 28th, 2020

NOVARTIS | Reimagining Medicine