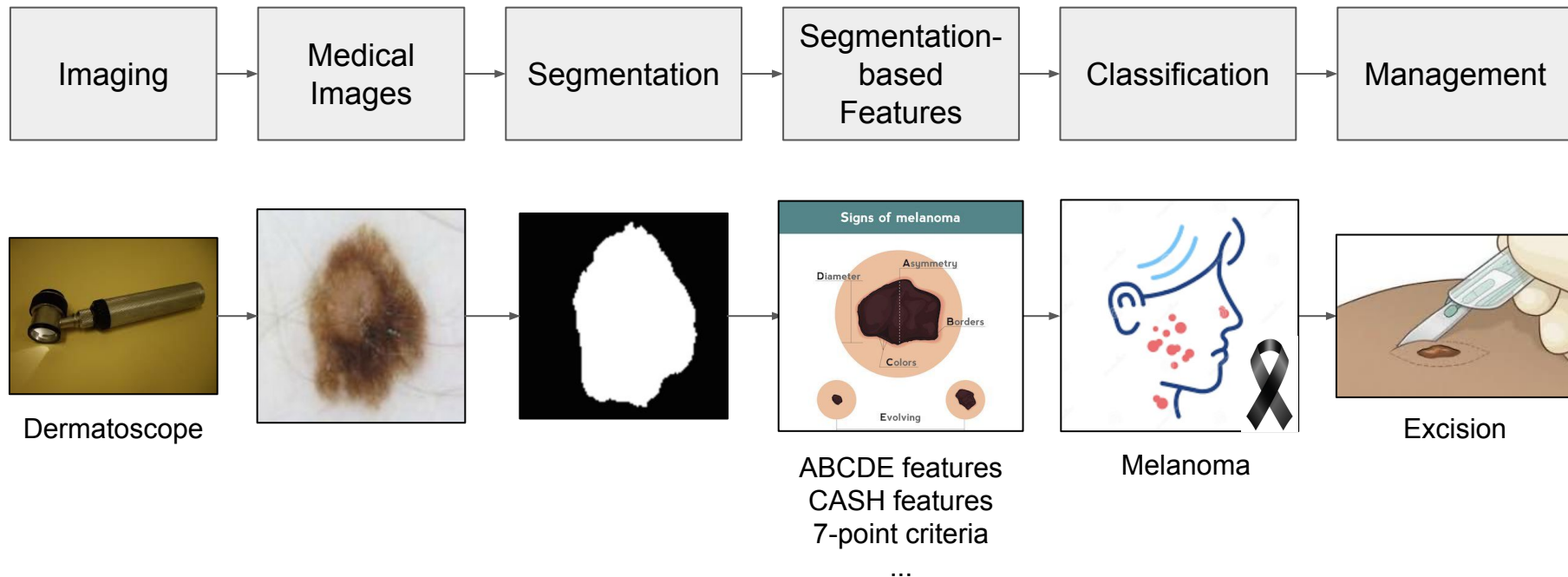# Direct Prediction of Clinical Management Bypassing AI-based Diagnosis: Application to Skin Lesions

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

kabhishe@sfu.ca

December 3, 2020

ARTIFICIAL INTELLIGENCE | DATA MINING | MACHINE LEARNING
COMPUTATIONAL BIOLOGY | GRAPHICS | NETWORKS & SYSTEMS
ARTIFICIAL INTELLIGENCE | DATA MINING | MACHINE LEARNING
CS RESEARCH DAY 2020 SYSTEMS
LEARNING
COMPUTATIONAL BIOLOGY | GRAPHICS | NETWORKS & SYSTEMS
ARTIFICIAL INTELLIGENCE | DATA MINING | MACHINE LEARNING

Medical Image Analysis Pipeline

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

Dermatoscope

ABCDE features
CASH features
7-point criteria
...

Melanoma

Excision

2

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

nature

nature > letters > article

Published: 25 January 2017

Dermatologist-level classification of skin cancer with deep neural networks

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

Original Research

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

Titus J. Brinker [a, b] ✉, Achim Hekler [a], Alexander H. Enk [b], Joachim Klode [c], Axel Hauschild [d], Carola Berking [e], Bastian Schilling [f], Sebastian Haferkamp [g], Dirk Schadendorf [c], Tim Holland-Letz [h], Jochen S. Utikal [i, j, 1], Christof von Kalle [a, 1]
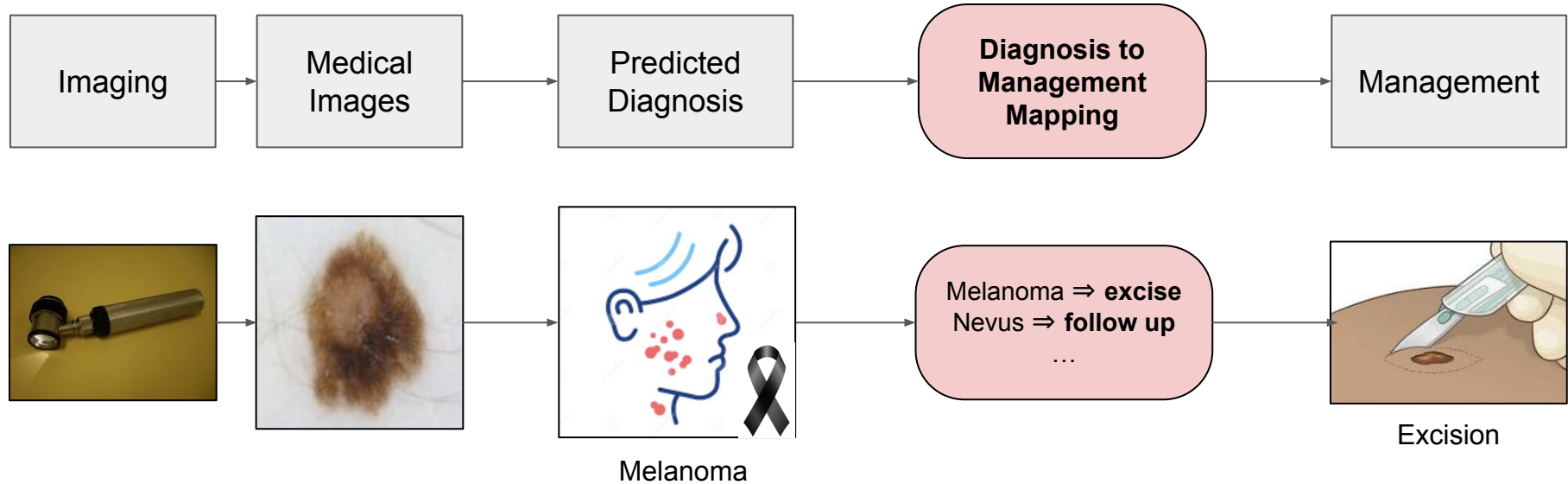

General Dermatology

Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis
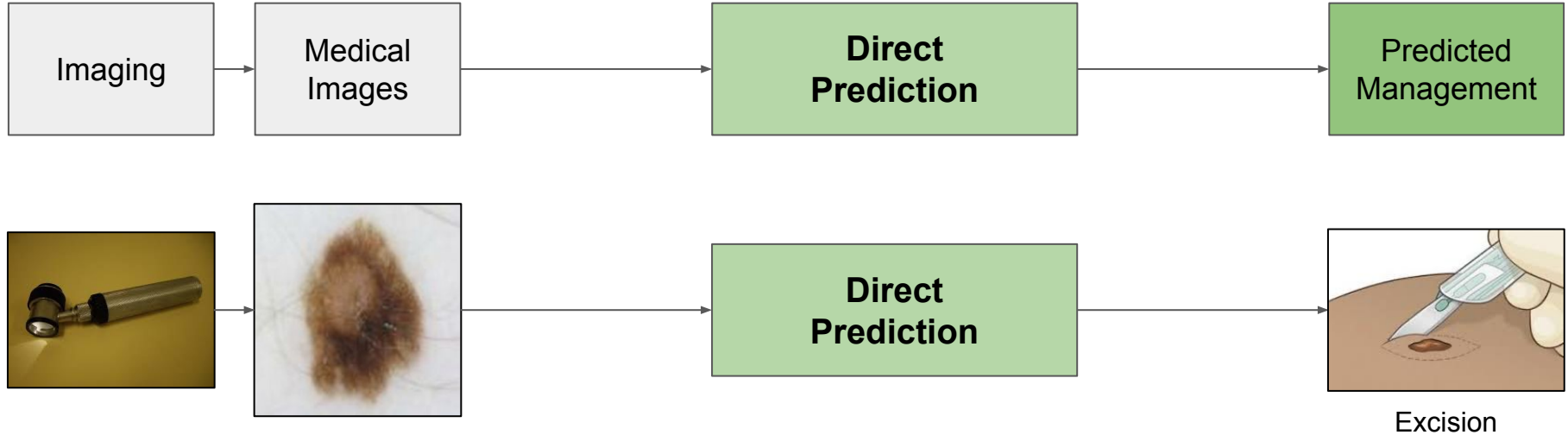
Y. Fujisawa ✉, Y. Otomo, Y. Ogata, Y. Nakamura, R. Fujita, Y. Ishitsuka, R. Watanabe, N. Okiyama, K. Ohara, M. Fujimoto

---

Original Research

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

Roman C. Maron [a, 1], Michael Weichenthal [b, 1], Jochen S. Utikal [c, d], Achim Hekler [a], Carola Berking [e], Axel Hauschild [b], Alexander H. Enk [f], Sebastian Haferkamp [g], Joachim Klode [h], Dirk Schadendorf [h], Philipp Jansen [h], Tim Holland-Letz [i], Bastian Schilling [j], Christof von Kalle [a], Stefan Fröhling [a], Maria R. Gaiser [c, d], Daniela Hartmann [e], Anja Gesierich [j] ... Alexander Thiem

Inferring Management Decisions from Diagnosis Predictions

Melanoma

Melanoma ⇒ **excise**
Nevus ⇒ **follow up**
…

Excision

Directly Predicting Management Decisions

```
┌──────────┐   ┌──────────┐                    ┌──────────────┐                ┌──────────────┐
│ Imaging  │──▶│ Medical  │───────────────────▶│    Direct    │───────────────▶│  Predicted   │
│          │   │  Images  │                    │  Prediction  │                │ Management   │
└──────────┘   └──────────┘                    └──────────────┘                └──────────────┘
```


                                                ┌──────────────┐
                                                │   Direct   │
                                                │ Prediction   │
                                                └──────────────┘

Excision
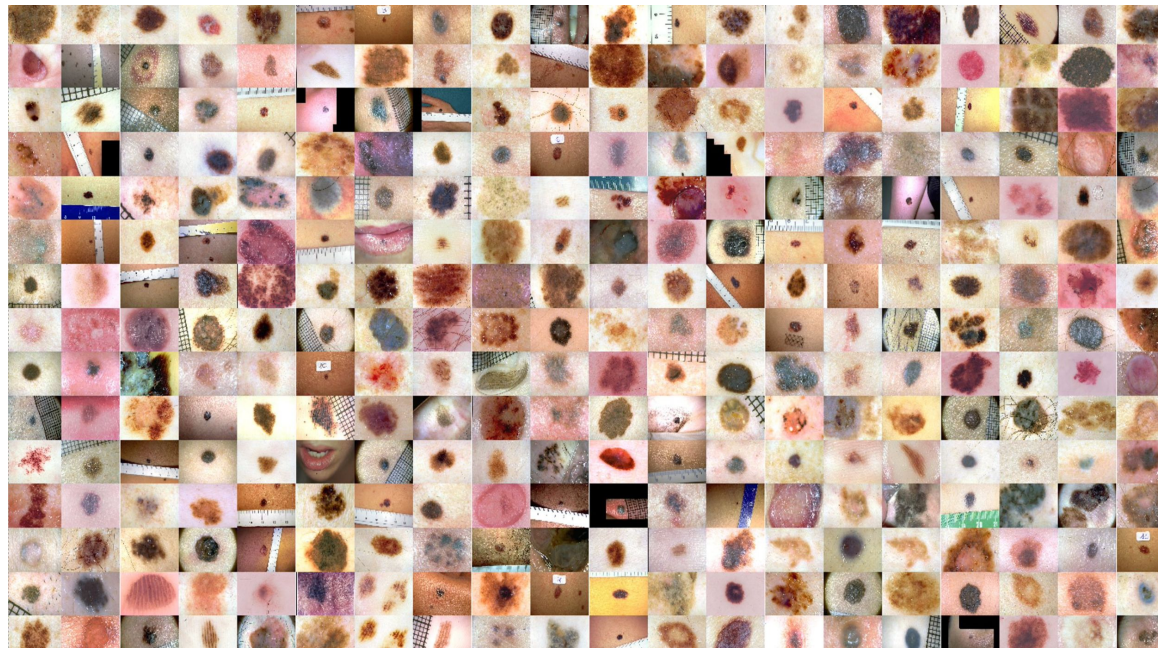
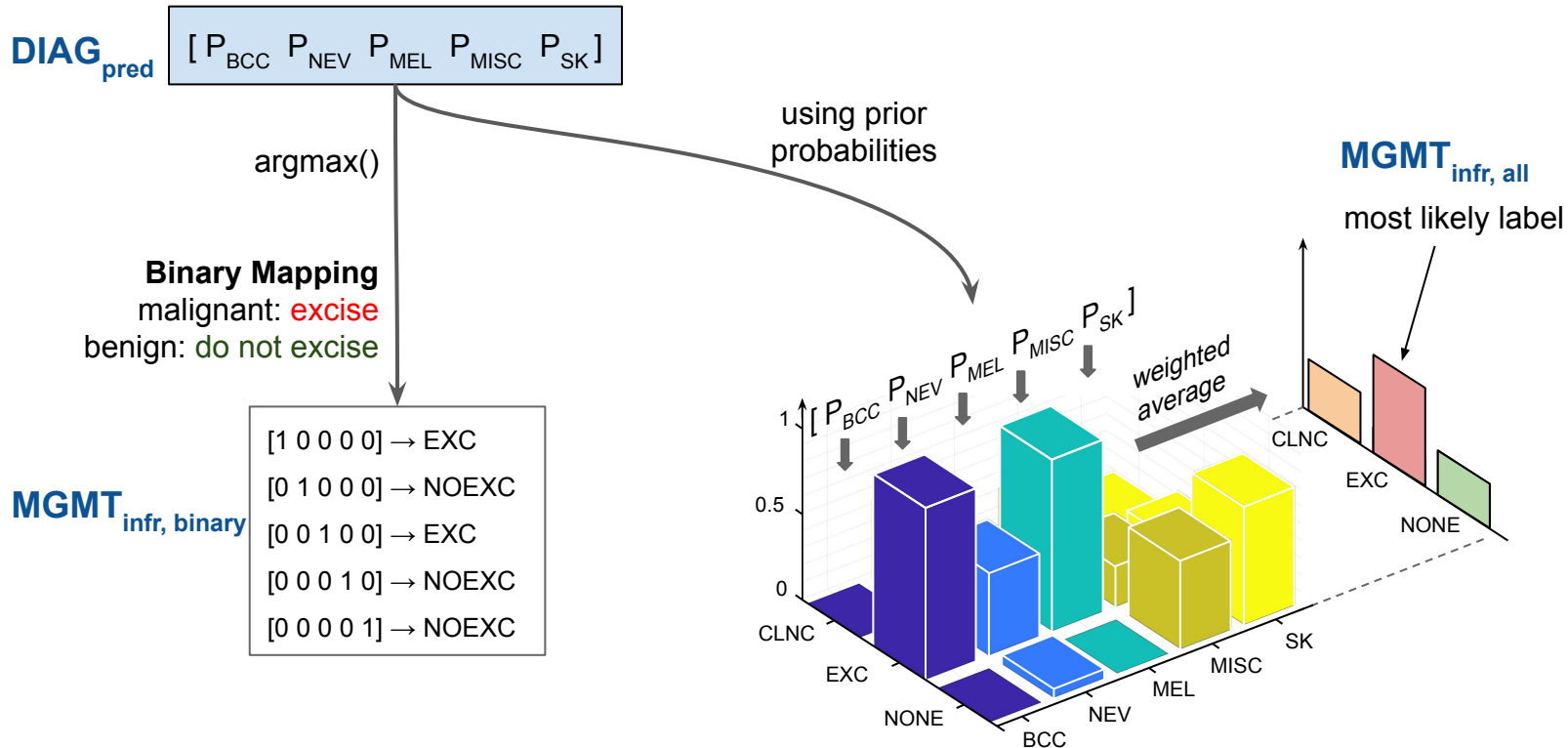# Interactive Atlas of Dermoscopy Dataset

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

$\text{DIAG}_{\text{pred}}$ $[\ P_{BCC}\ \ P_{NEV}\ \ P_{MEL}\ \ P_{MISC}\ \ P_{SK}\ ]$

argmax()

using prior probabilities

$\text{MGMT}_{\text{infr, all}}$

most likely label

Binary Mapping
malignant: excise
benign: do not excise

$\text{MGMT}_{\text{infr, binary}}$

[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

$[\ P_{BCC}\ \ P_{NEV}\ \ P_{MEL}\ \ P_{MISC}\ \ P_{SK}\ ]$

weighted average

1

0.5

0

CLNC

EXC

NONE

BCC

NEV

MEL

MISC

SK

CLNC

EXC

NONE

8

Comparing predicted versus inferred management

Predicted management decisions **are closer to the ground truth** than inferred decisions.
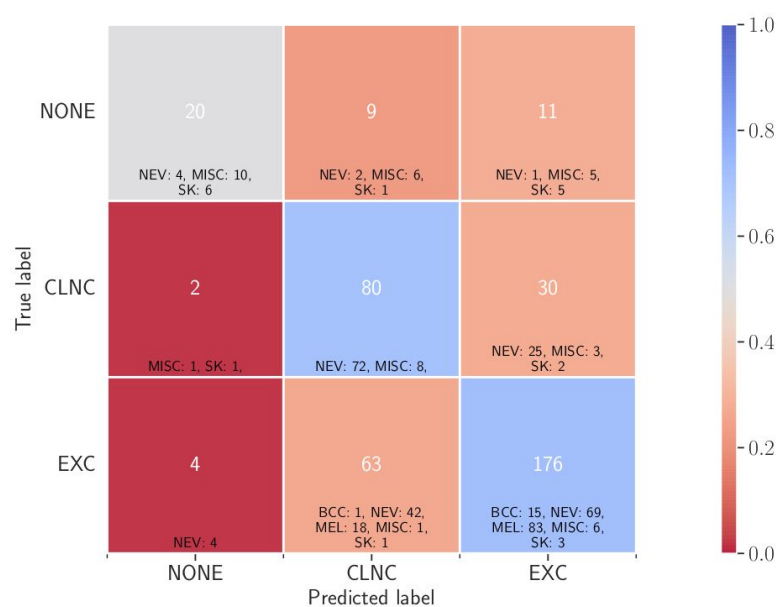(statistically significant at p < 0.001)
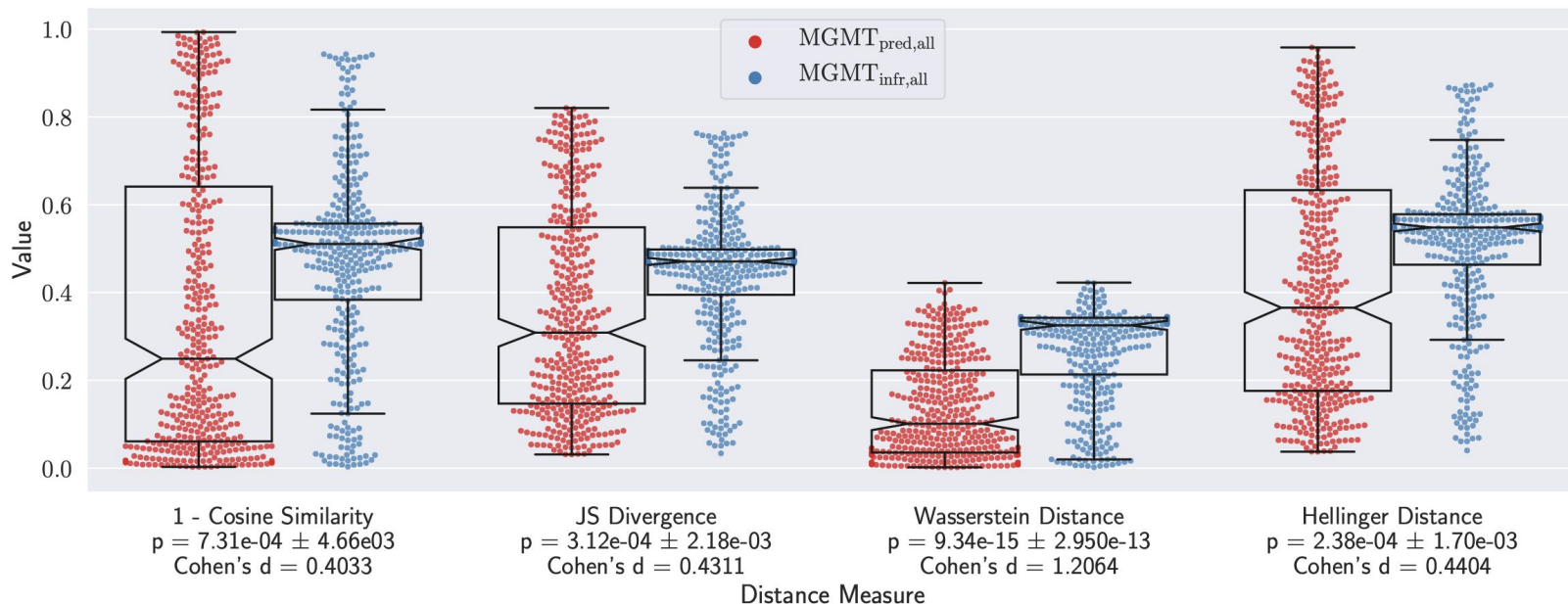
Comparing predicted versus inferred management

Predicted management decisions **are more accurate** than inferred decisions.
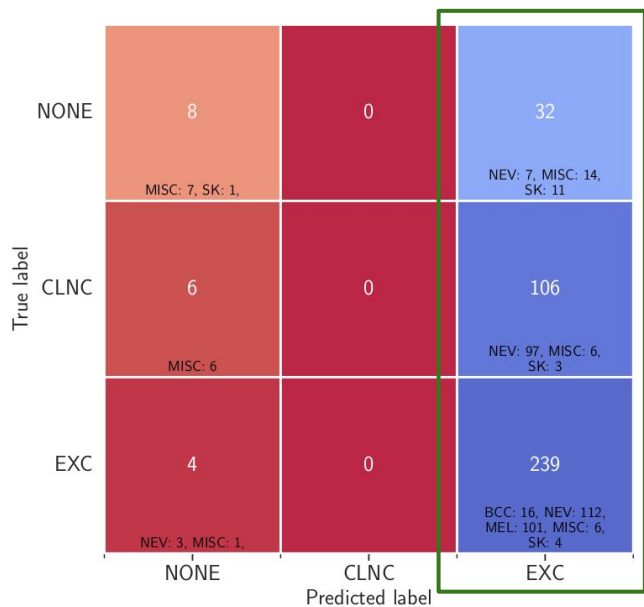
Inferred management

Predicted management

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.

Inferred management

Predicted management
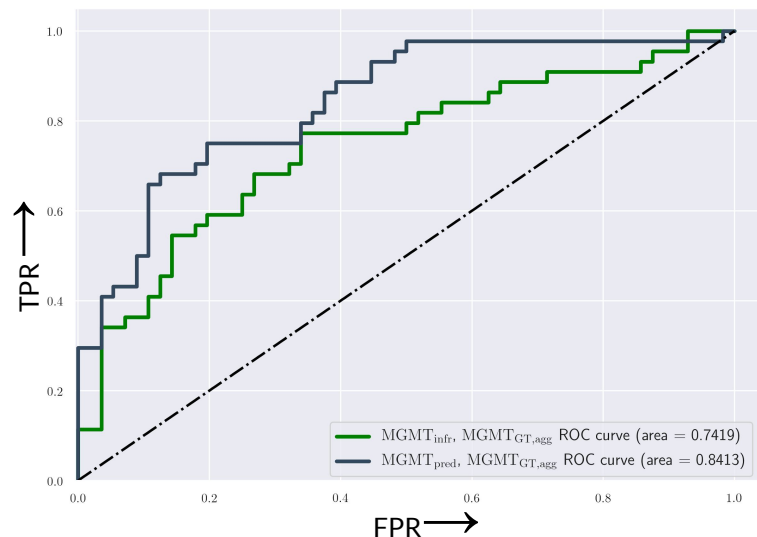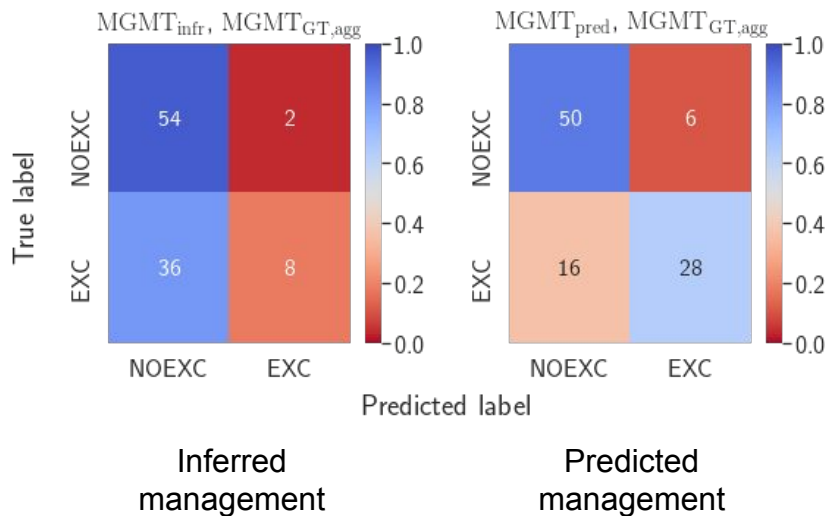
Evaluating Model Agreement with Dermatologists

Two agreement measures

Cohen's kappa (model, derm) **>** Cohen's kappa ($derm_i$, $derm_j$)

0.5424 0.4124

Fleiss' kappa (model, derm) **≈** Fleiss' kappa ($derm_i$, $derm_j$)

0.4080 0.4086

Our model's predictions agree with dermatologists <u>at least as much as</u> 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.