Sources of uncertainty:

Medical face recognition in rare disease detection

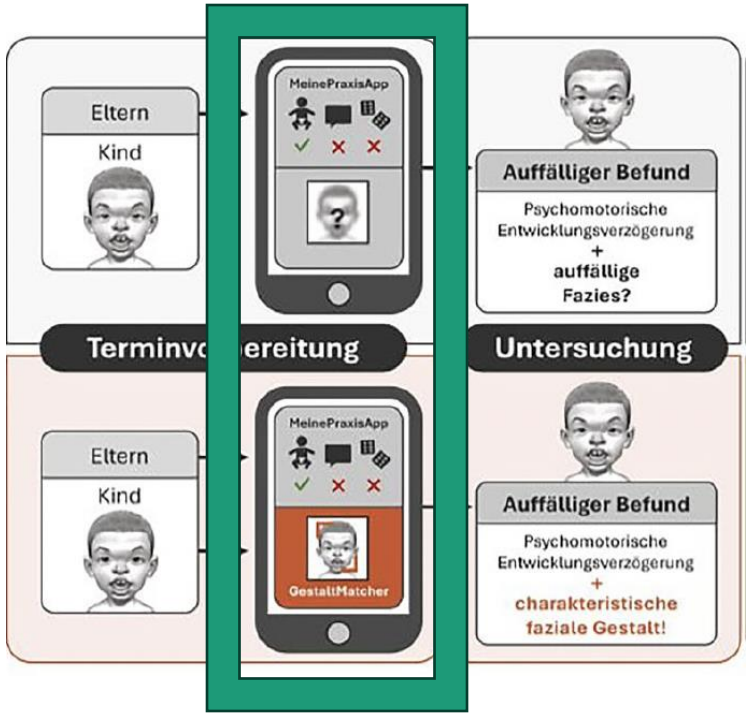
Paul Trauttmansdorff
Technical University of Munich / Ethical Data Initiative

paul.trauttmansdorff@tum.de

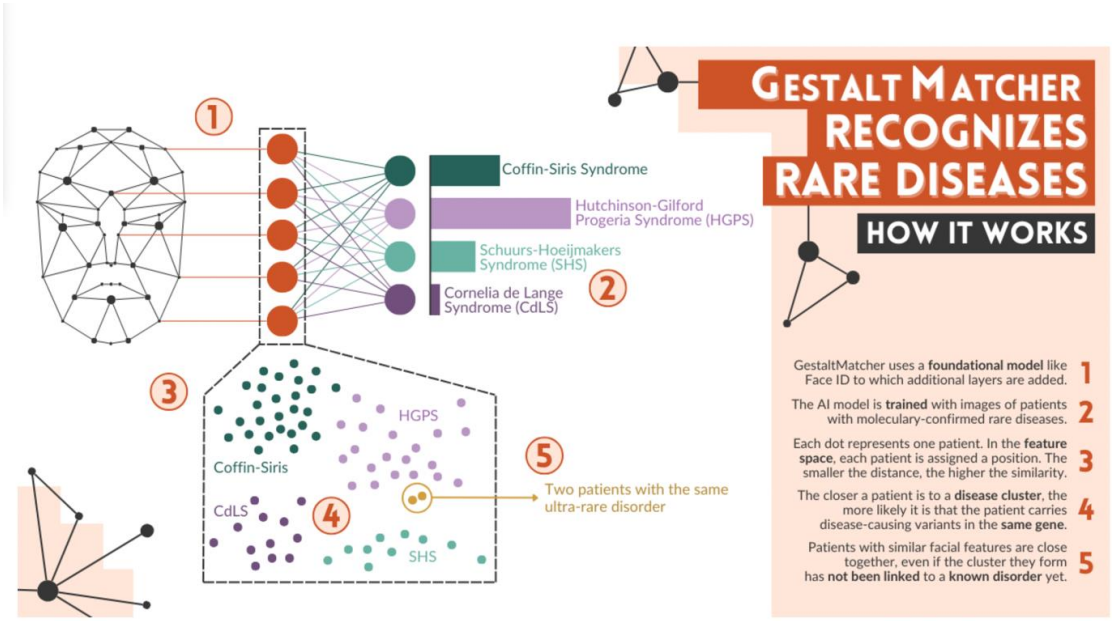
ISHPSSB Porto, July 2025

Automated Face Recognition for Rare Disease Diagnostics

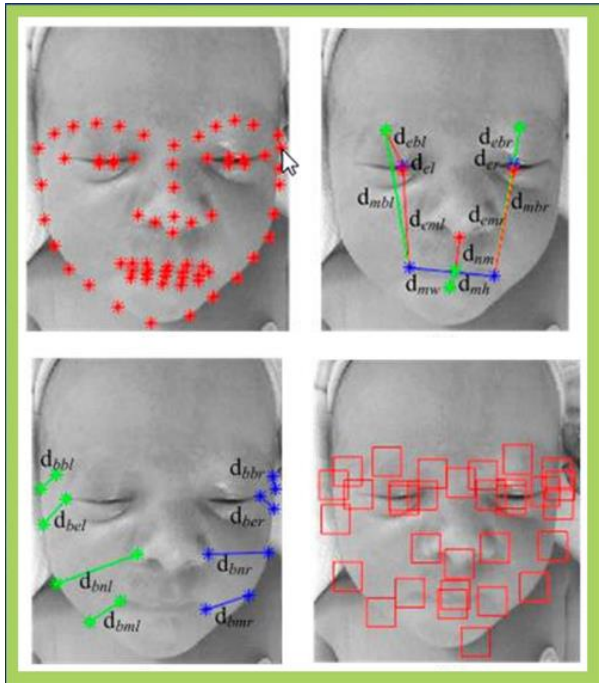
- “Smart phenotyping”: Apps match facial phenotypes from photos to those associated with known disorders / rare diseases.
- Automated assessment of facial features should become essential in the patient’s examination



Patient **Doctor**



Core promises of medical face recognition in rare disease detection



Taken from online exhibition
NIH National Library of Medicine

Core promises

- **Claims about speed, better accuracy, and better efficiency** in recognizing genetic disease through algorithmic facial analysis
- **Easy access** to a diagnosis

- Instead of the intuition of experts/dysmorphologists, it creates **algorithmic evidence**
- Subjective assessment becomes a **standardized facial examination**

Core promises of medical face recognition in rare disease detection

“What previously depended on the intuition and experience of a few experts – the assessment of dysmorphic facial features – is becoming an evidence-based method thanks to artificial intelligence (AI).”

Krawitz/Ruder/Niehues, 2025. own transl.

“[B]efore this kind of technology, the doctor usually only assessed the face or the symptom … [with] their eyes.”

Interview 2024, HS

Epistemic types: how faces become diagnostic knowledge

Clinician/Dysmorphologist's "thick description" and its uncertainties

Acknowledging different approaches

… both my predecessor and I independently annotated thick eyebrows, synophrys, long palpebral fissures, and downslanted palpebral fissures. We also both noted anteverted nares and a short nose. However, I also annotated highly arched eyebrows, prominent eyelashes, underdeveloped nasal alae, and downturned corners of the mouth. These were not noted by the previous grader—**perhaps because they took a more conservative approach**?

Nuanced considerations

That might also explain ptosis: **I considered it plausible**, but perhaps the other annotator would have **preferred to verify it by asking the patient to look up**.

Contextual/situated information?

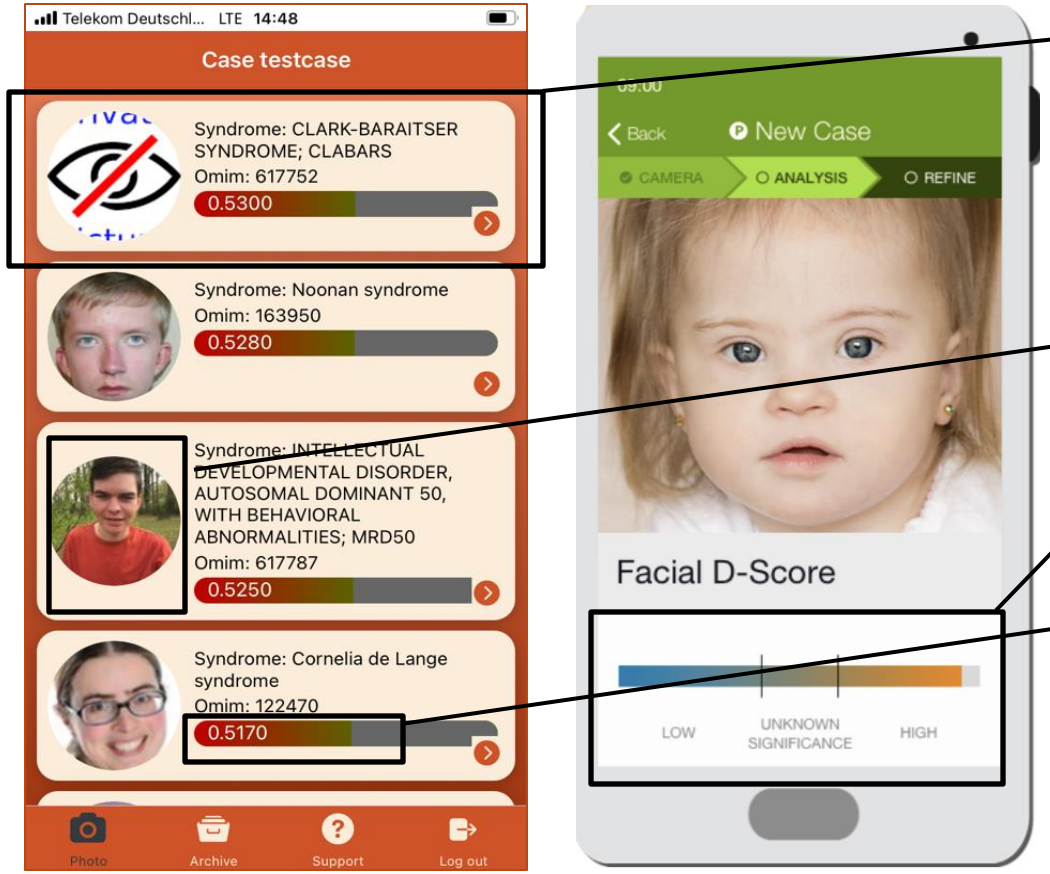
… I believe many features fall into a **gray zone**—there's often **no clear right or wrong**. And sometimes, we just click the wrong thing. (Like me, with "short philtrum"—**I probably meant "long philtrum**," though maybe it just looked that way because of the short nose?)

Human error?

Ambiguity

Epistemic types: how faces become diagnostic knowledge

Algorithmic assessment



Probabilistic results

> "the algorithm ... in the broad sense [… tries] to make the probability [of a disease] only more real"
> *Interview, CM, own transl.*

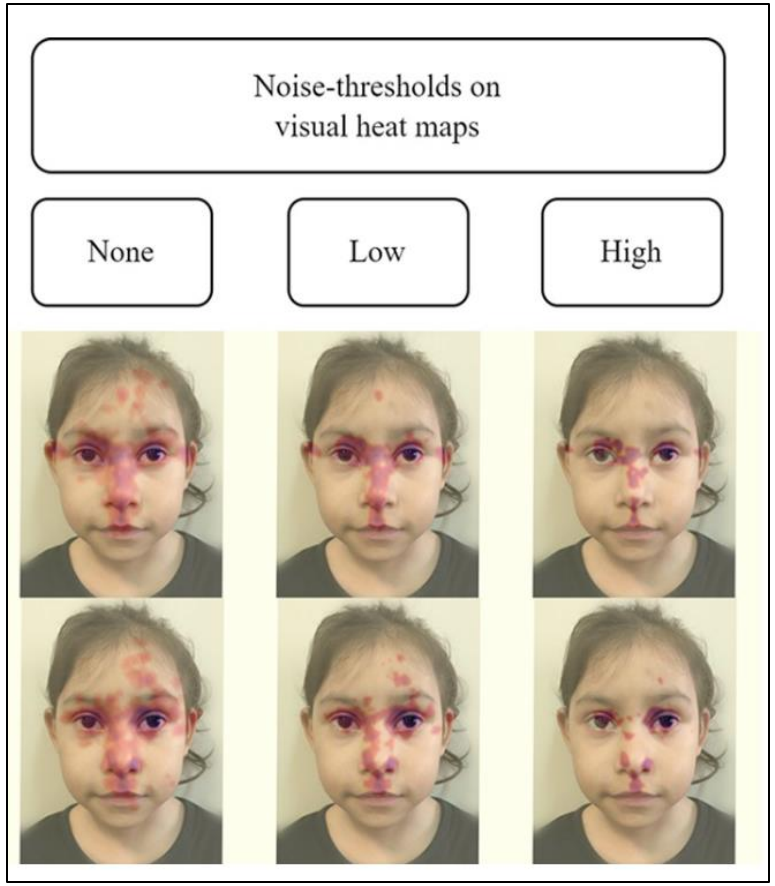
Similarities and likelihoods

> "To provide a kind of dysmorphia score, so to speak."
> *Interview, TP, own transl.*

Encoded binaries: healthy (0) and abnormal (1)

> "It processes it [the face] … to distinguish between healthy and abnormal. And thereby gives an initial indication..."
> *Interview, TP, own transl.*

I. Uncertainties around standardizing the algorithmic method

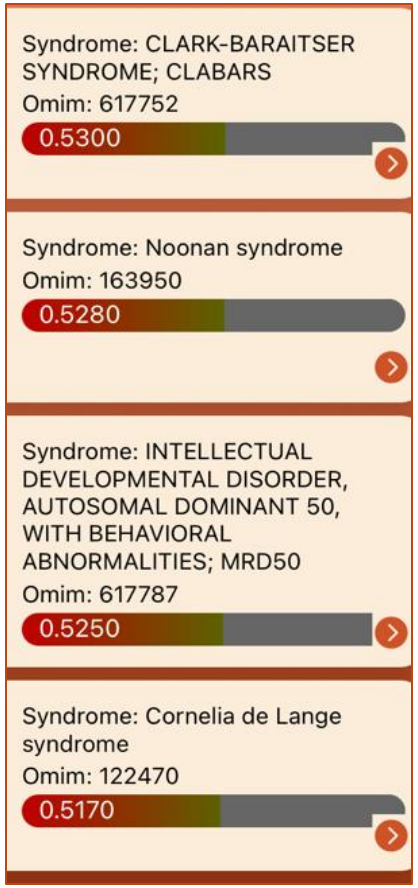


Taken from Duong et al. 2024

- Standardization of frontal images: ignoring other relevant facial traits, e.g. those visible in profiles
- Ignoring contextual nuances: context-dependent facial cues can no longer be reconstructed
- Encoding binary assumptions of abnormal and normal/healthy: ignoring “natural variety”

(Uncertainties articulated in interviews)

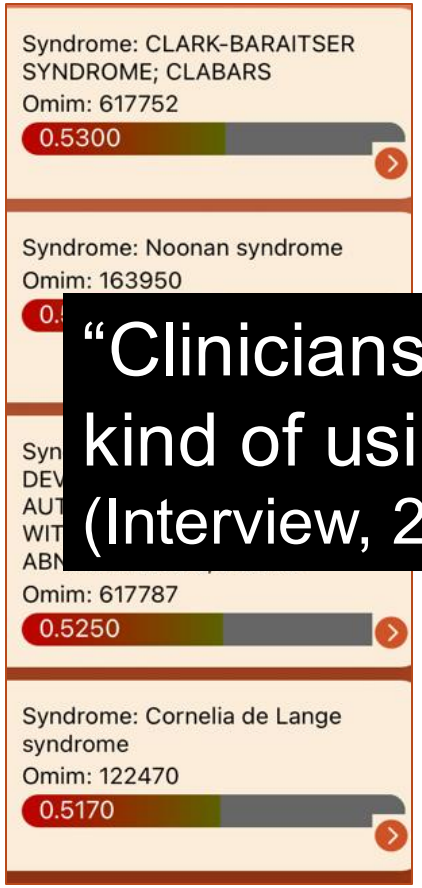
II. Uncertainties around interpreting algorithmic evidence



Cropped photo of Gestaltmatcher, taken by PT

- Non-representativeness of the underlying dataset
  - Incompleteness: "different syndromes are particularly well or poorly recognized" (Interview, 2025)
  - Biases: "There's a huge skewing of the data in terms of where the data is from to train these things […]" (Interview, 2025)
- Making results actionable: how do you interpret and act on probabilistic evidence
  - Interpretation: "the most likely conditions are condition A, condition B, and condition C, that might be true, but it doesn't mean that other conditions aren't the answer right?" (Interview, 2025)
  - Regional differences in clinical/diagnostic practice?

II. Uncertainties around interpreting algorithmic evidence



Cropped photo of Gestaltmatcher, taken by PT

- Non-representativeness of the underlying dataset
  - Incompleteness: "different syndromes are particularly well or poorly recognized" (Interview, 2025)
  - Biases: "There's a huge skewing of the data in terms of where the data is from to train these things […]" (Interview,

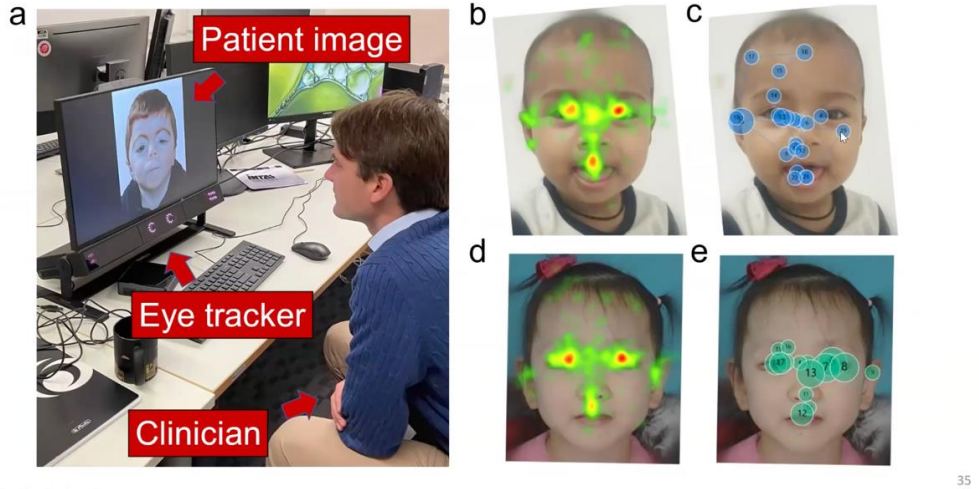
"Clinicians don't really understand enough, … they're kind of using them blindly." (Interview, 2025 BS)

act on probabilistic evidence

  - Interpretation: "the most likely conditions are condition A, condition B, and condition C, that might be true, but it doesn't mean that other conditions aren't the answer right?" (Interview, 2025)
  - Regional differences in clinical/diagnostic practice?

III. Uncertainties around explaining algorithmic results

Machine explainability and human explainability



Taken from Interview 2024, HS

"[T]hese AI algorithms are … a bit of a black box. There is no way to ask the AI: Why do you think this is particularly important now?" (Interview, Hs)

- No way to help explain to clinicians/professionals what syndromic facial features matter in the algorithmic diagnosis
- No way to explain to patients or patient representatives the validity of the results

"So, you cannot just tell the patient like: oh, this AI machine tells you that you have the disease, and we don't know why" (Interview, DD)

Some concluding reflections

- Medical Face Recognition generates new types of facial knowledge
  - Standardized/automated forms of diagnostic knowledge production
  - Epistemically grounded in facial data
  - Explainability of a diagnosis (or lack thereof)

- Leads to novel diagnostic uncertainties rather than clear answers

- However, rare disease diagnosis is already fraught with **uncertainty** and **delay** (diagnostic odysseys for patients)

- Problems lie within **how** we deal with uncertainty
  → **Risk of AI solutionism**: uncertainty is not a bug, it is a feature

  → **Risk of making uncertainty invisible**
    → Means treating algorithmic results as irrefutable “truths”
    → Revives (reminds us of) physiognomic beliefs, under the guise of scientific objectivity

Thank you

Paul Trauttmansdorff

paul.trauttmansdorff@tum.de