

GestaltMatcher: rare disease matching using facial phenotypic descriptors

Peter Krawitz

Institute for Genomic Statistics und Bioinformatics (IGSB)

University Bonn, Germany

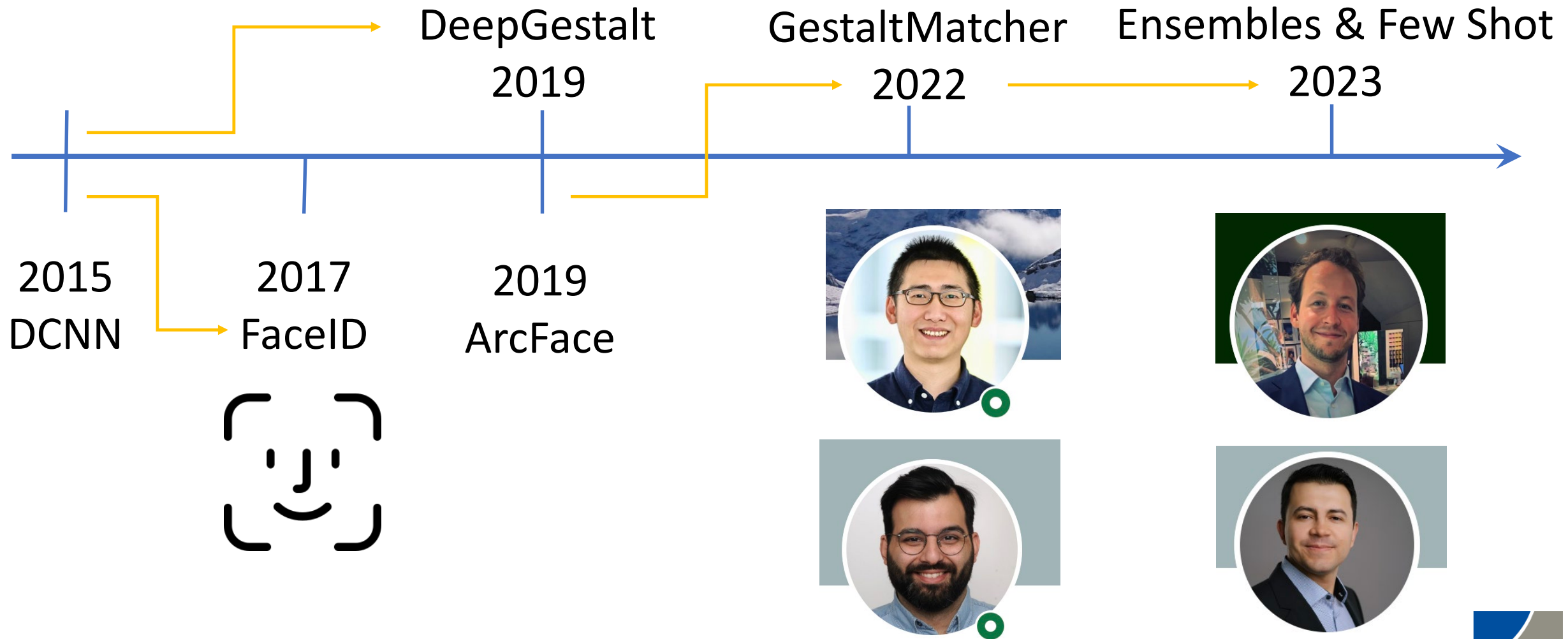
Ljubijana 2. June 2023

What I would like to talk about today

- How can AI support physicians in diagnosing Kleeftstra S?
 - Phenotypic descriptors
- What do computer scientists need to improve AI?
 - FAIR data
- How could patients and families contribute?
 - Share more data



Computer Vision and next generation phenotyping (NGP) in the last decade



Facial analysis by DeepGestalt (Face2Gene)



✓
1st

Kleefstra Syndrome 1; KLEFS1

Differential
 Clinically Diagnosed
 Molecularly Diagnosed

GESTALT FEATURE

2nd

Turner Syndrome

Differential
 Clinically Diagnosed
 Molecularly Diagnosed

GESTALT FEATURE

3rd

Beckwith-Wiedemann Syndrome; BWS

Differential
 Clinically Diagnosed
 Molecularly Diagnosed

GESTALT FEATURE



✓
1st

Hyperphosphatasia ...Tardation Syndrome

Differential
 Clinically Diagnosed
 Molecularly Diagnosed

HPMRS

GESTALT FEATURE

2nd

CHARGE Syndrome

Differential
 Clinically Diagnosed
 Molecularly Diagnosed

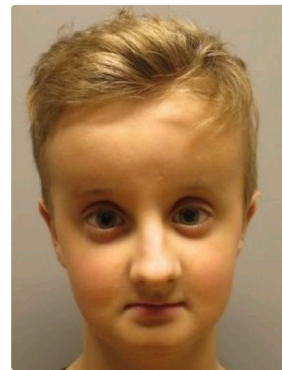
GESTALT FEATURE

3rd

Robinow Syndrome

Differential
 Clinically Diagnosed
 Molecularly Diagnosed

GESTALT FEATURE



1st

Cataract, Congenit...lte Type, Included

Differential
 Clinically Diagnosed
 Molecularly Diagnosed

Not supported yet

GESTALT FEATURE

2nd

Wolf-Hirschhorn Syndrome; WHS

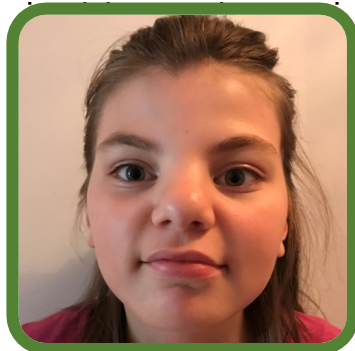
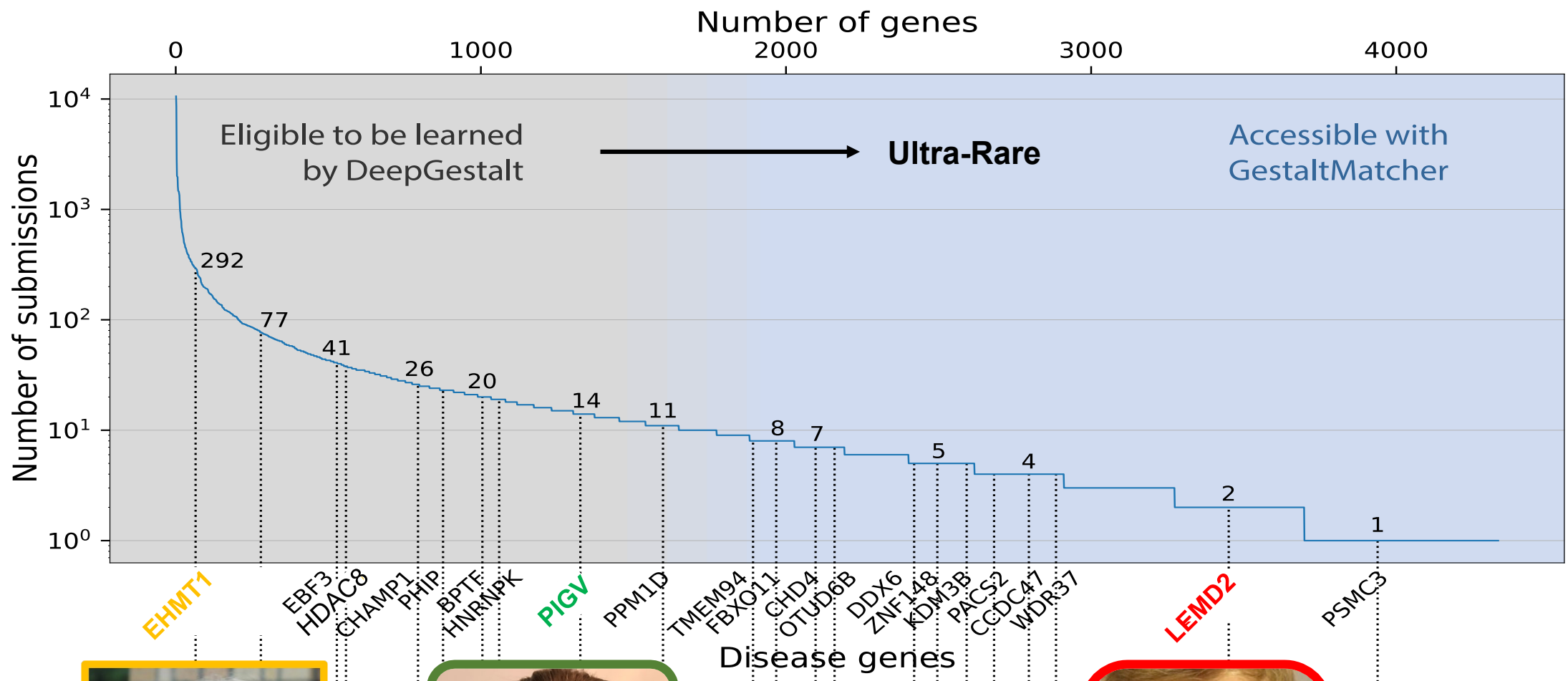
Differential
 Clinically Diagnosed
 Molecularly Diagnosed

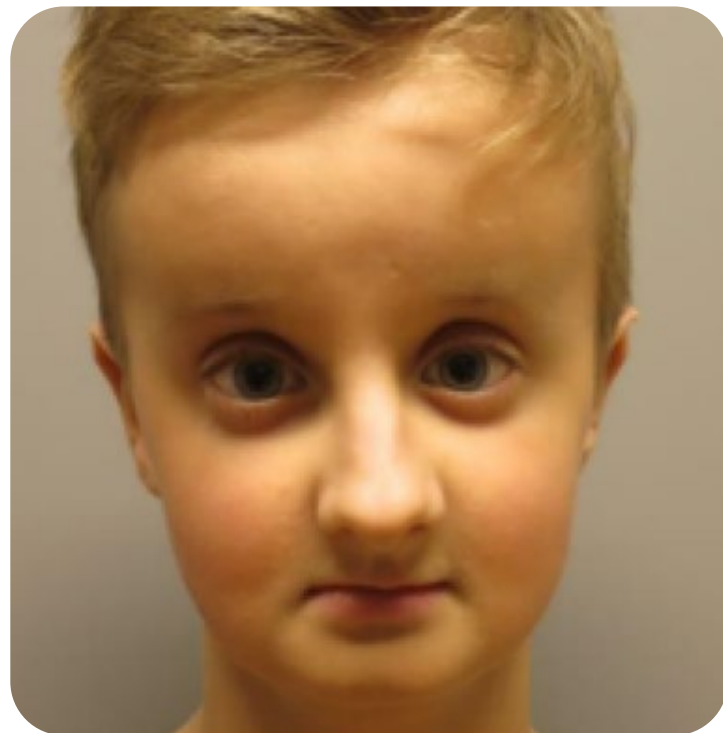
GESTALT FEATURE

Floating-Harbor Syndrome; FLHS

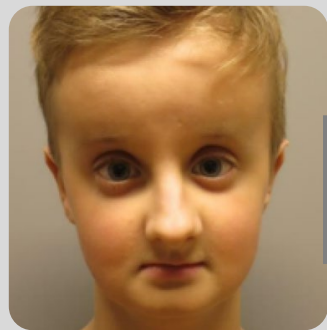
Differential
 Clinically Diagnosed
 Molecularly Diagnosed

GESTALT FEATURE

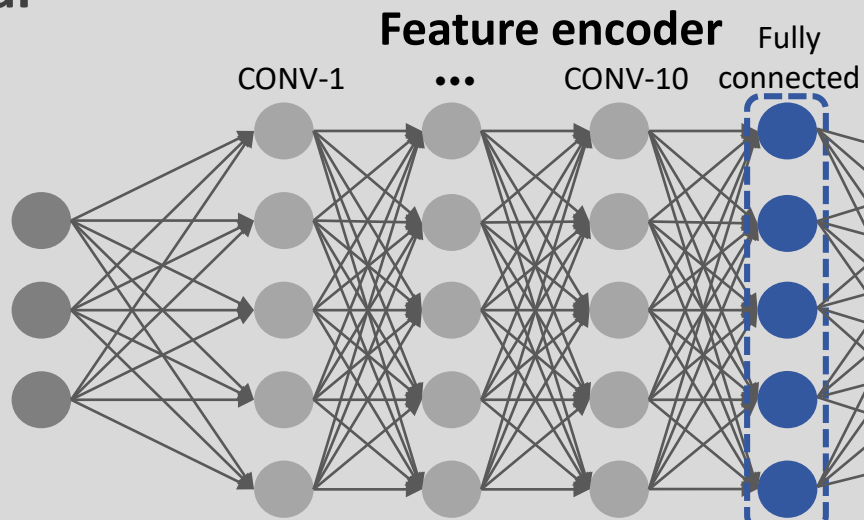




Deep Convolutional Neural Network



Portrait photo of new patient



Classifier



320-dimensional facial phenotypic descriptor (FPD)

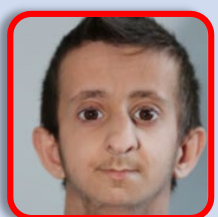


Kleefstra

HGPS

HPMRS

SHS



Patients with the same ultra-rare disease

Clinical Face Phenotype Space

Factors that affect the performance of an AI: Algorithms

Alexander Hustinx



Improving Deep Facial Phenotyping for Ultra-rare Disorder Verification Using Model Ensembles

Alexander Hustinx¹, Fabio Hellmann², Ömer Sümer², Behnam Javanmardi¹,
Elisabeth André², Peter Krawitz¹, Tzung-Chien Hsieh^{1*}

¹ Institute for Genomic Statistics and Bioinformatics, University Hospital Bonn, University of Bonn

² Chair for Human-Centered Artificial Intelligence, University of Augsburg

{ahustinx, b-jav, pkrawitz, thsieh}@uni-bonn.de

{fabio.hellmann, oemer.suemer, andre}@informatik.uni-augsburg.de

Abstract

Rare genetic disorders affect more than 6% of the global population. Reaching a diagnosis is challenging because rare disorders are very diverse. Many disorders have recognizable facial features that are hints for clinicians to diagnose patients. Previous work, such as GestaltMatcher, utilized representation vectors produced by a DCNN similar to AlexNet to match patients in high-dimensional feature space to support “unseen” ultra-rare disorders. However, the architecture and dataset used for transfer learning in GestaltMatcher have become outdated. Moreover, a way to train the model for generating better representation vectors for unseen ultra-rare disorders has not yet been studied. Because of the overall scarcity of patients with ultra-rare disorders, it is infeasible to directly train a model on them. Therefore, we first analyzed the influence of replacing GestaltMatcher DCNN with a state-of-the-art face recognition approach, iResNet with ArcFace. Additionally, we experimented with different face recognition datasets for transfer learning. Furthermore, we proposed test-time augmentation, and model ensembles that

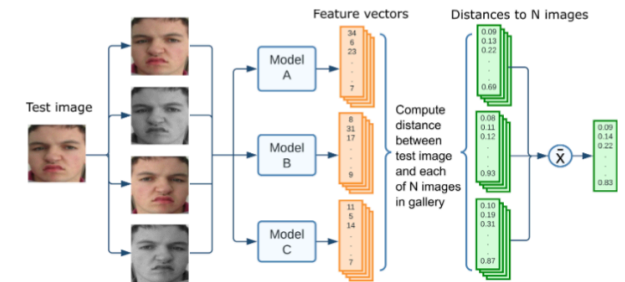


Figure 1. Model ensemble of our approach. We first performed test time augmentation to augment the test image into four images (color and horizontal flip). The four augmented images were further encoded by three different models into 12 representation vectors. We then compared the cosine distance of the 12 representation vectors to the 12 representation vectors from each of the N images in the gallery. It resulted in 12 distance vectors, and each vector contains N cosine distances. In the end, we averaged over 12 distance vectors (\bar{X}) to obtain the final distance vector, which further ranked the N images in the gallery. The gallery image with a smaller distance is more similar to the test image.

Factors that affect the performance of an AI: Algorithms

2023

We first analyzed the influence of replacing GestaltMatcher DCNN with a state-of-the-art face recognition approach, iResNet with ArcFace. Furthermore, we propose model ensembles that mix general face verification models and models specific for verifying disorders to improve the disorder verification accuracy of unseen ultra-rare disorders. For our ensemble, we considered both models that are fine-tuned for disorders and models built for face verification. The face verification models produced strong general features that can be leveraged to verify unseen disorders, while the fine-tuned models were fitted towards features of seen disorders they have been trained on.

Factors that affect the performance of an AI: Algorithms

Model	Dataset	Loss	GMDB-Frequent		GMDB-Rare	
			Top-1	Top-5	Top-1	Top-5
GM-Hsieh2022	CASIA*	CE	15.96%	33.83%	19.26%	36.28%
r50-D/O†	Glint360K*	CE	44.33%	65.76%	29.06%	46.35%
r50-D/O†+ TTA	Glint360K*	CE	47.73%	67.67%	30.29%	46.38%
r100-D/O	Glint360K*	CE	48.37%	71.78%	28.02%	44.32%
r100-D/O + TTA	Glint360K*	CE	51.16%	69.58%	27.92%	46.26%
r100	Glint360K	ArcFace	30.25%	54.81%	33.25%	50.22%
r100 + TTA	Glint360K	ArcFace	35.25%	56.52%	33.47%	51.61%
Model ensemble	n/a	n/a	52.06%	70.70%	34.93%	52.78%
Model ensemble + TTA	n/a	n/a	52.99%	71.01%	35.98%	53.93%

GestaltMatcher in Research



Hannah Klinkhammer · 1.

PhD Student Biostatistics bei IMBIE & IGSB

Lumping and Splitting



Cell Genomics



Volume 2, Issue 5, 11 May 2022, 100131

Article

Lumping versus splitting: How to approach defining a disease to enable accurate genomic curation

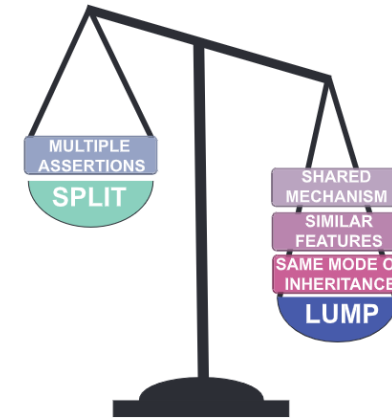
Courtney Thaxton^{1,8}  , Jennifer Goldstein¹, Marina DiStefano², Kathleen Wallace¹, P. Dane Witmer³, Melissa A. Haendel⁴, Ada Hamosh⁵, Heidi L. Rehm^{6,7}, Jonathan S. Berg¹

Defining a disease entity requires careful assessment and weighing of evidence

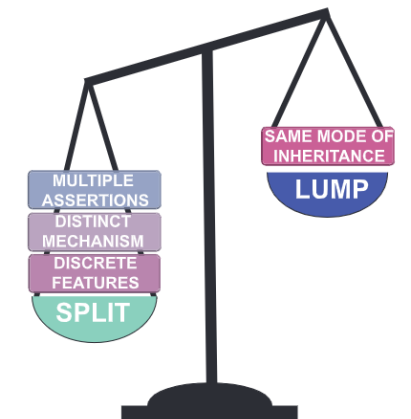
Does evidence favor distinct conditions, or support a single disorder with a phenotypic spectrum?

Gene X has been reported to have causal relationships with 3 conditions.

Gene Y has been reported to have causal relationships with 2 conditions.



Evidence favors a single disease spectrum



Evidence favors 2 distinct entities

Lumper and splitter

Lumper: associate a phenotype to another disorder or a phenotypic series (PS).

We look similar.
We should be in a PS!



Disorder A



Disorder B



Lumper

Splitter: separate a phenotype into two disorders that are linked to different genetic causes.

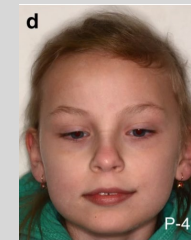
We look different.
We could be different disorders!



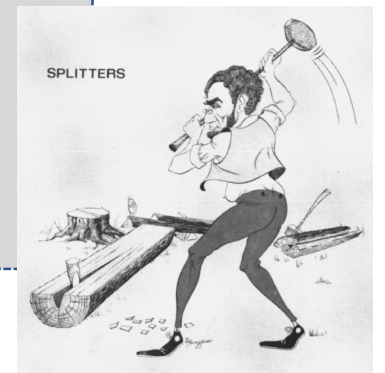
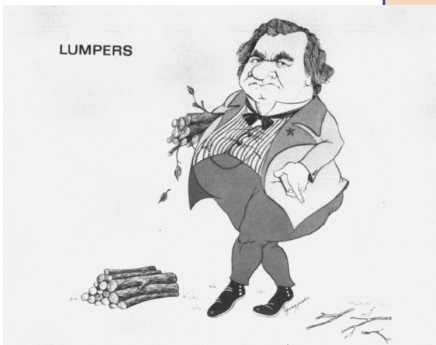
Disorder C



Disorder C

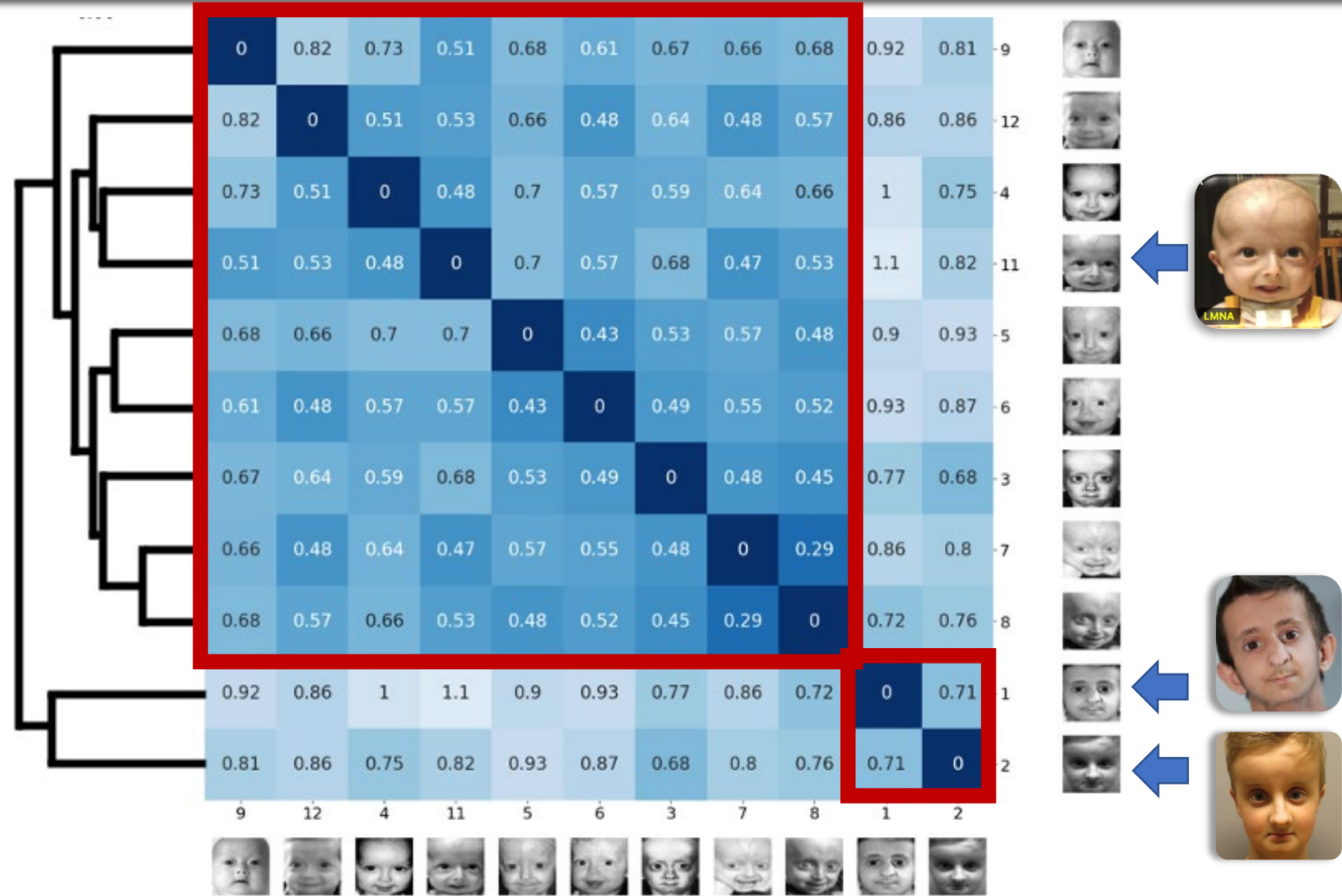


Splitter



McKusick, V. A. 1969. "On Lumpers and Splitters, or the Nosology of Genetic Disease." *Perspectives in Biology and Medicine* 12 (2): 298–312.

Match patients

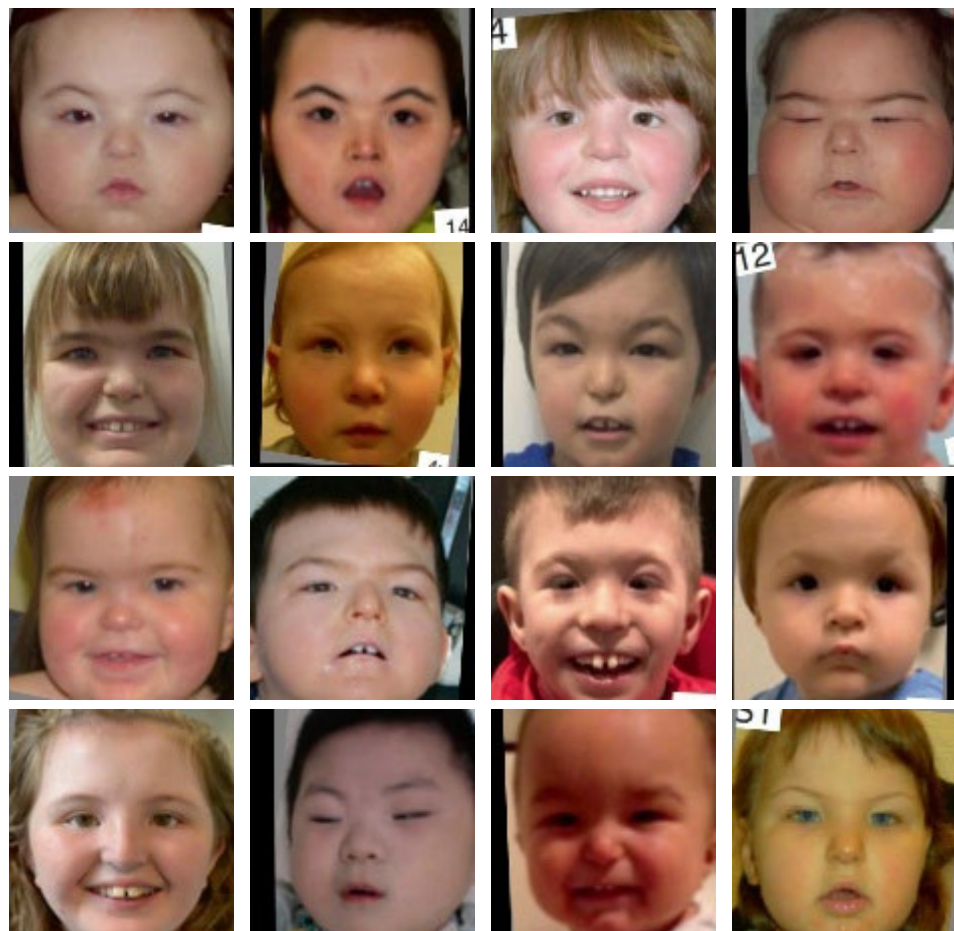


SMARCA2 subgroups with two different episignatures



SMARCA2 subgroups with two different episingatures

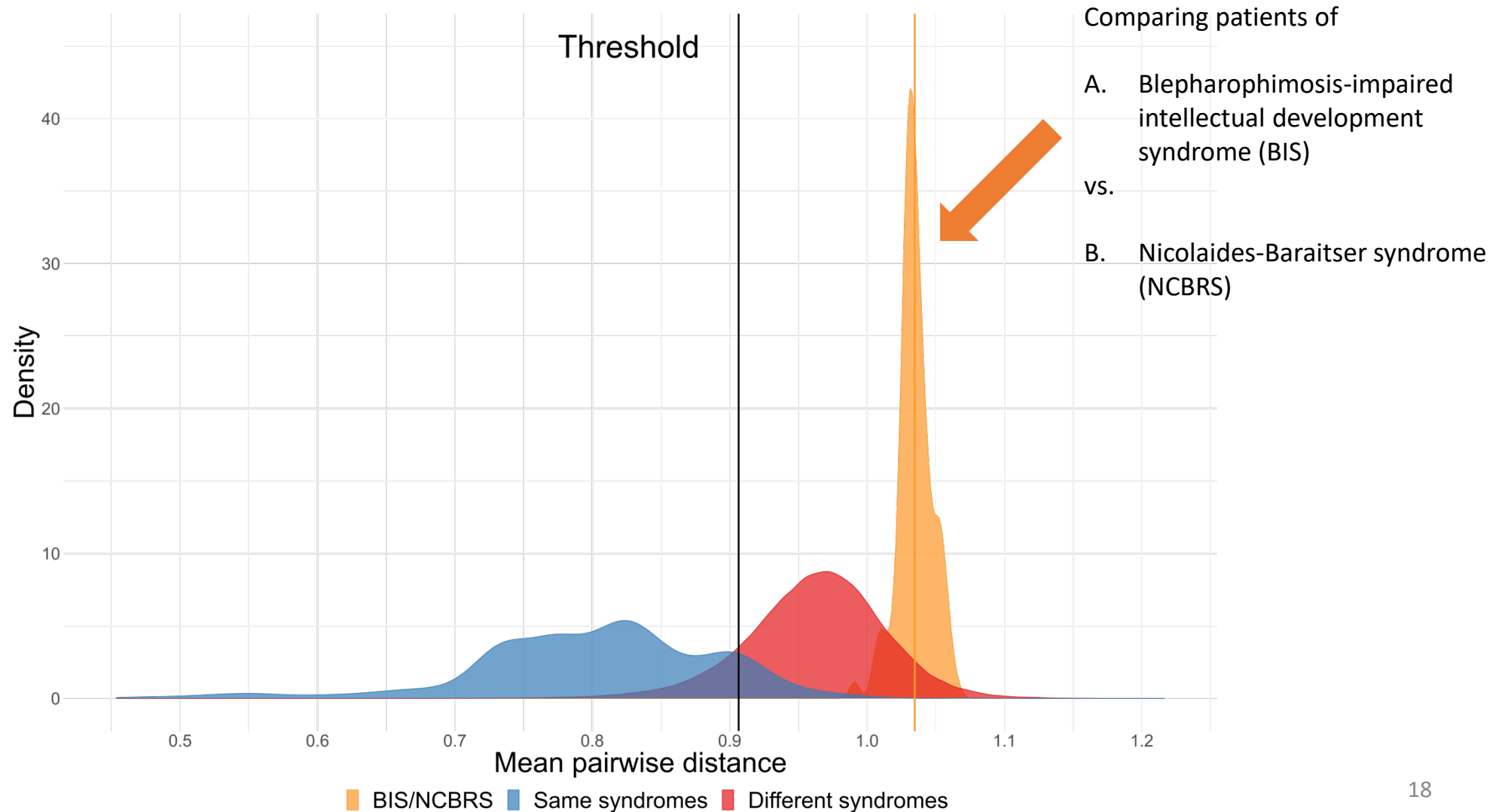
**Blepharophimosis-impaired
intellectual development syndrome**



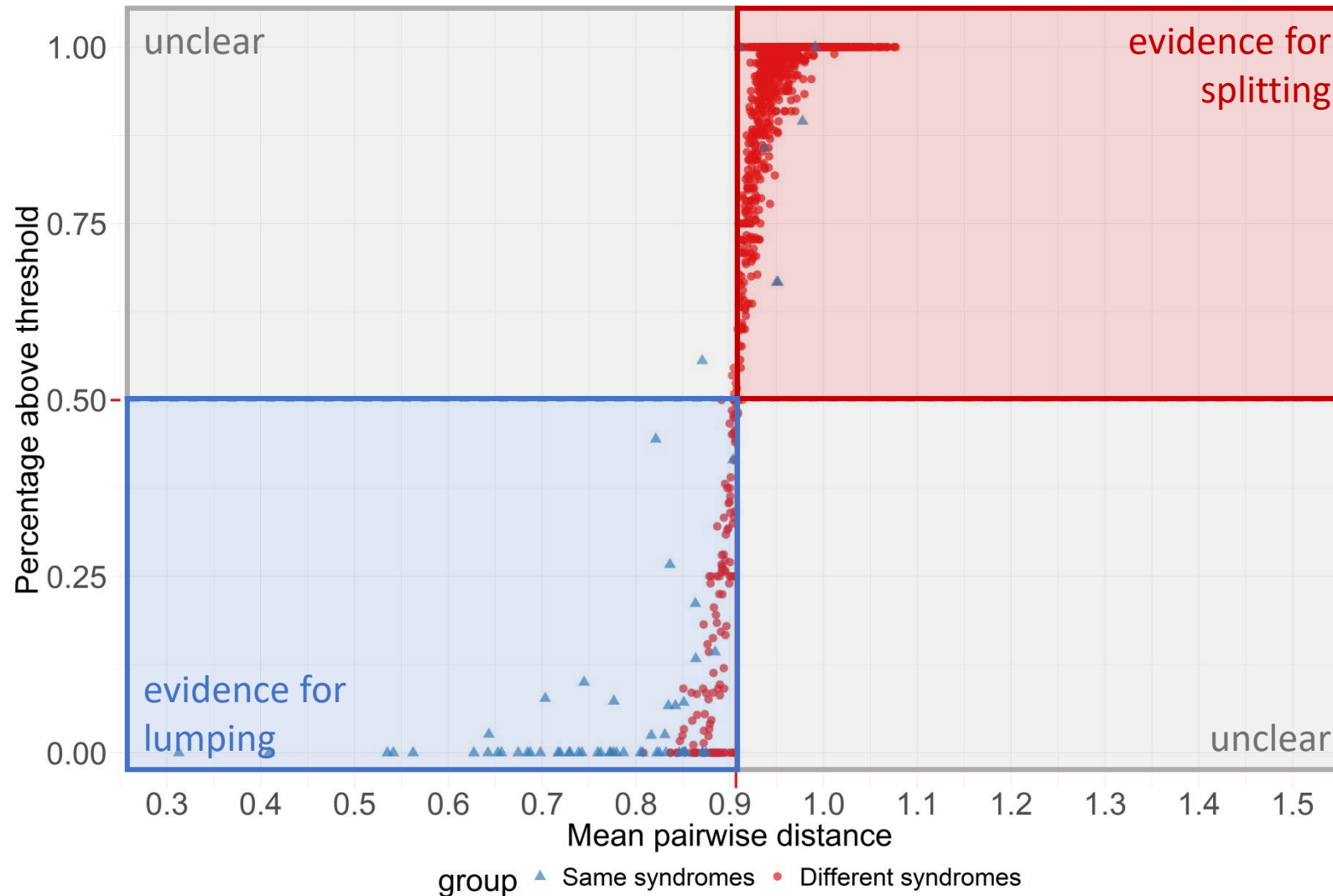
Nicolaides-Baraitser syndrome



SMARCA2 subgroups with two different episignatures



Phenotypic evidence for lumping and splitting



Test set:

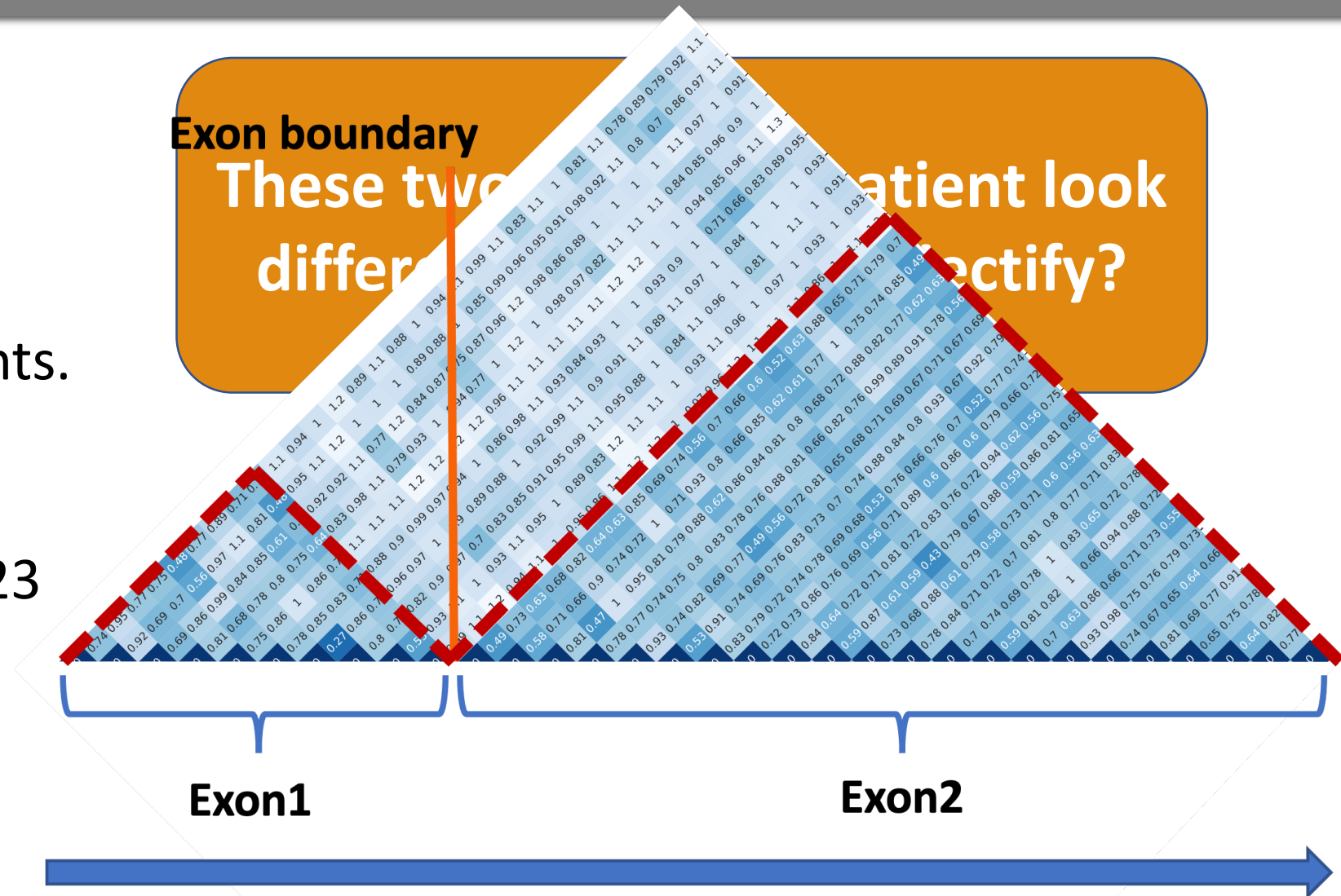
- 61 comparisons of cohorts from the same syndrome
- 88.5 % correctly lumped

Test set:

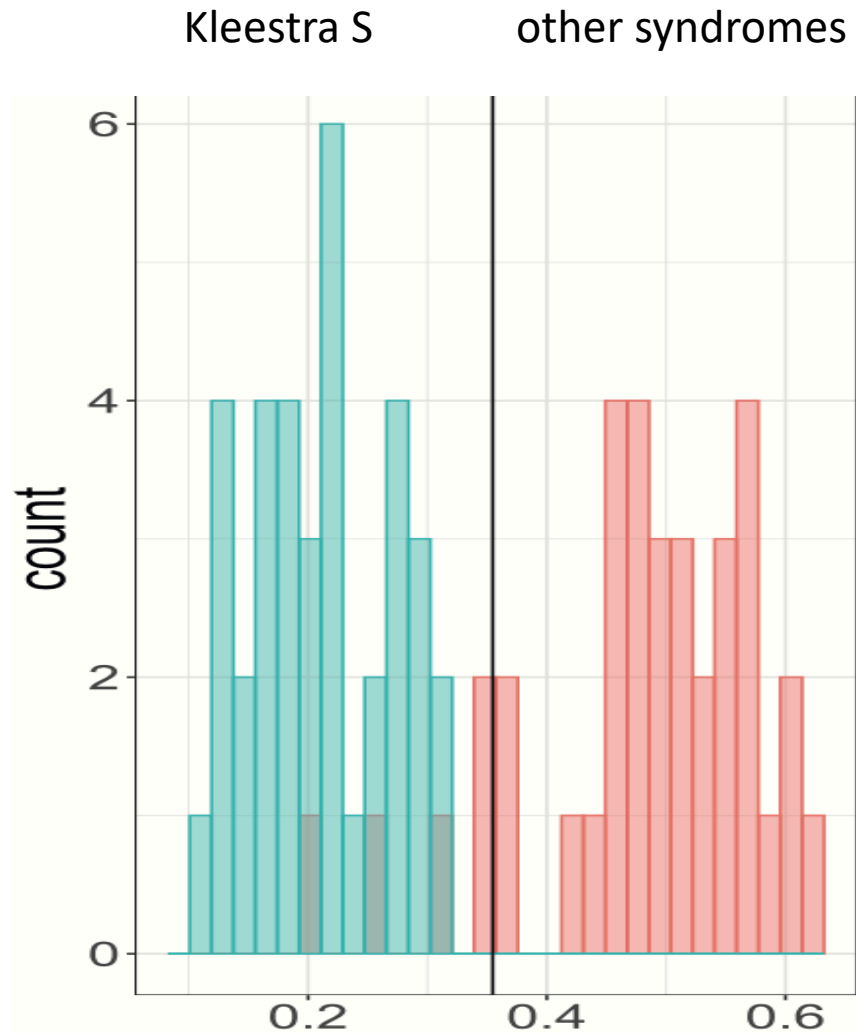
- 1.830 comparisons of cohorts from different syndromes
- 92.3 % correctly splitted

GMDB can do lumpner and splitter analysis

Cohort of *Gene-X*:
33 photos from 32 patients.
(Exon1) **ten** photos from
nine patients.
(Exon2) **23** photos from 23
patients.



Gestalt scores for variant classification



	Kleestra S	other syndromes
Gestalt score $\geq c$	32	1*
Gestalt score $< c$	4	36

$$sensitivity = \frac{true\ positive}{true\ positive + false\ negative} = \frac{32}{32 + 4} \approx 0.89$$

$$specificity = \frac{true\ negative}{true\ negative + false\ positive} = \frac{1}{1 + 36} \approx 0.97$$

$$LR_{CdL}^+ = \frac{sensitivity}{1 - specificity} \approx 32.89$$

Gestalt scores for variant classification

Genetics
inMedicine

ORIGINAL RESEARCH ARTICLE

© American College of Medical Genetics and Genomics

Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework

Sean V. Tavtigian, PhD¹, Marc S. Greenblatt, MD, PhD², Steven M. Harrison, PhD³,
Robert L. Nussbaum, MD⁴, Snehit A. Prabhu, PhD⁵, Kenneth M. Boucher, PhD⁶ and
Leslie G. Biesecker, MD⁷;

on behalf of the ClinGen Sequence Variant Interpretation Working Group (ClinGen SVI)

With a positive likelihood ratio of >30 a „matching phenotype“ is strong evidence for pathogenicity:

Thus, in Kleefstra „PP4“ could be upgraded



GestaltMatcher Database (GMDB)

a FAIR database for medical imaging data of rare disorders

Hellen Lesmann, MD

Institute for Genomic Statistics and Bioinformatics

Institute of Human Genetics

University Bonn



@h_lesmann

FAIR Principles

F - Findable

A - Accessible

I - Interoperable

R - Reusable



GestaltMatcher Database

[Home](#) [Downloads](#) [About](#)

[Sign up](#) [Login](#)

GestaltMatcher
Database

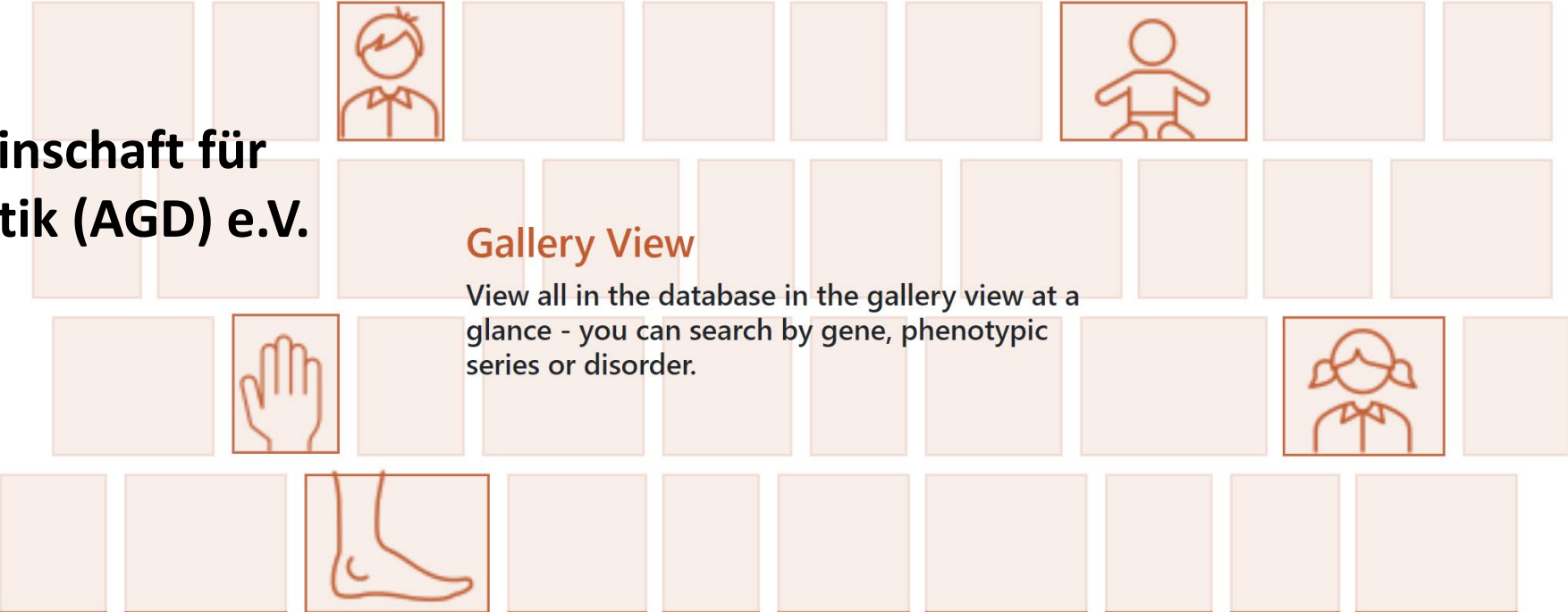

7430


2841


727


719

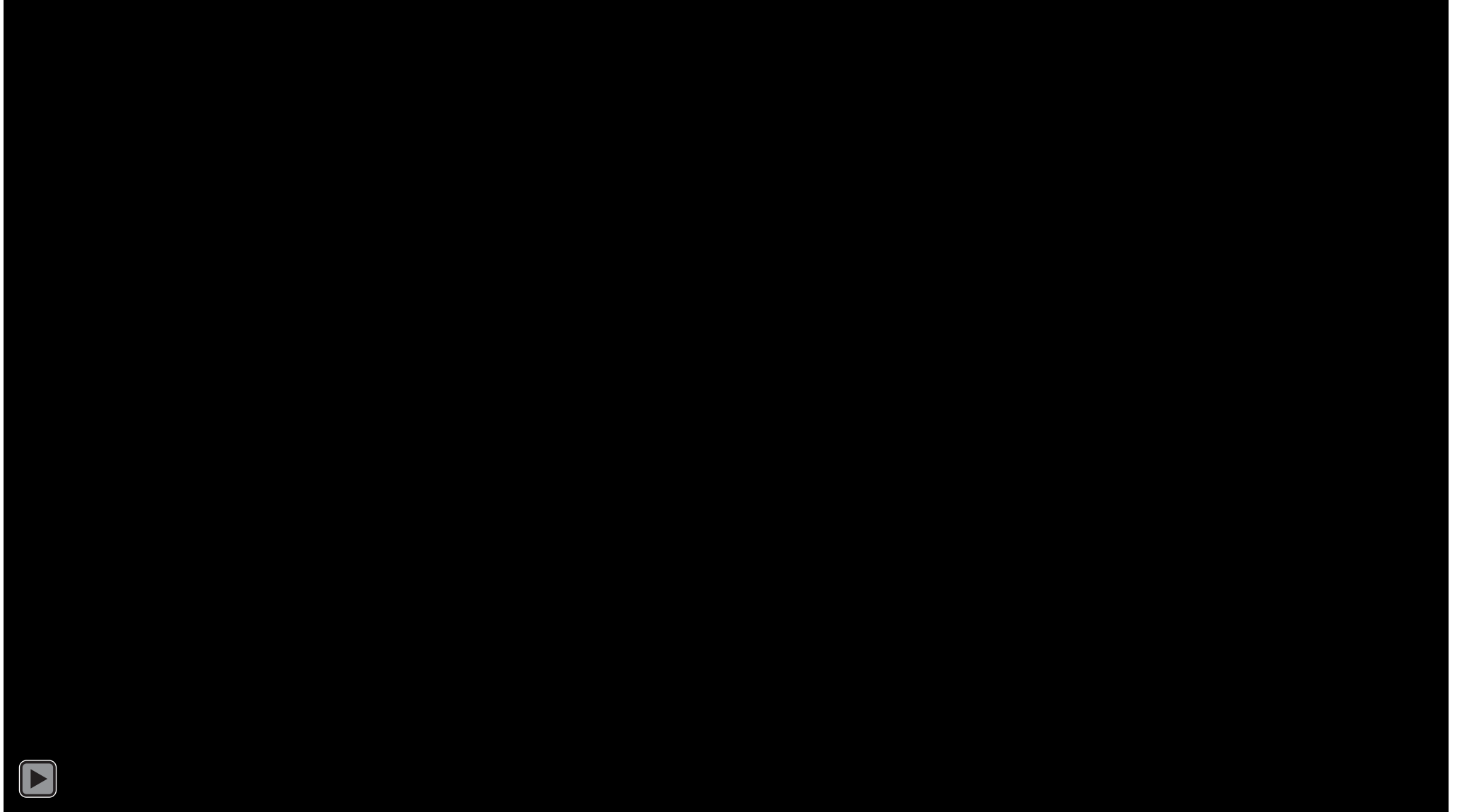
**Governed by
Arbeitsgemeinschaft für
Gen-Diagnostik (AGD) e.V.**



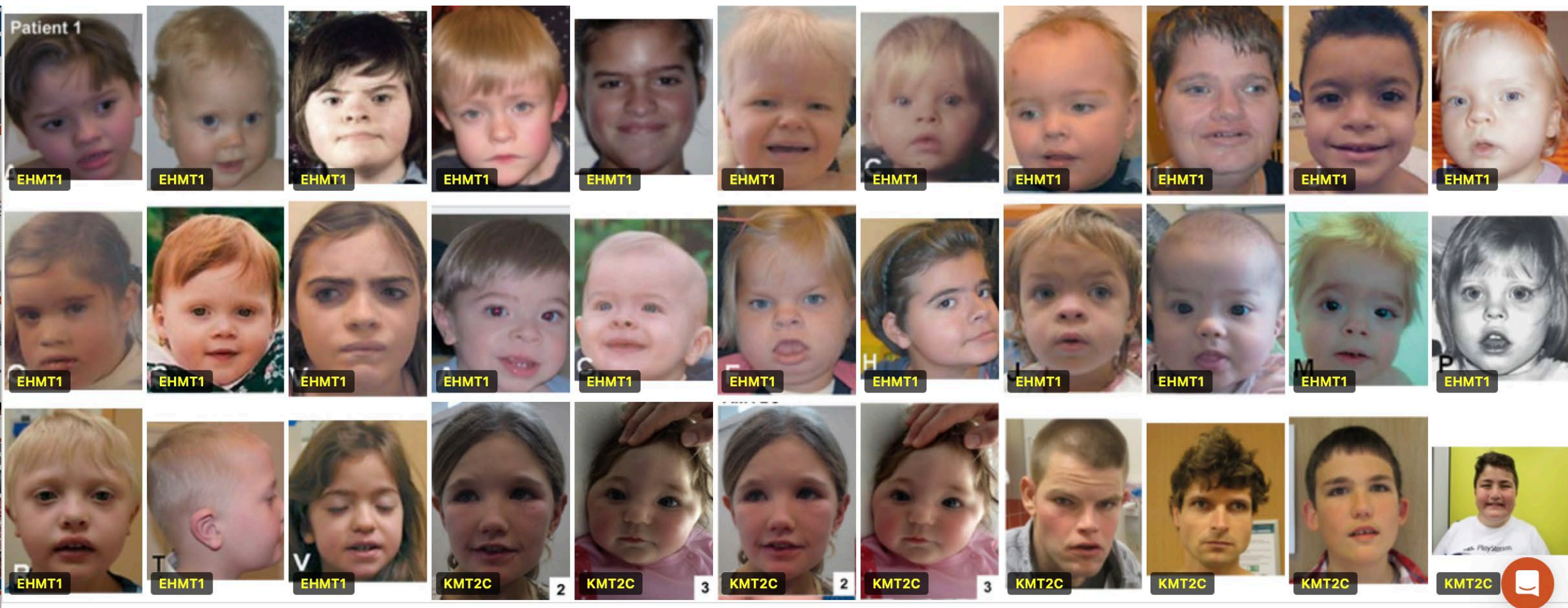
Gallery View

View all in the database in the gallery view at a glance - you can search by gene, phenotypic series or disorder.

Gallery



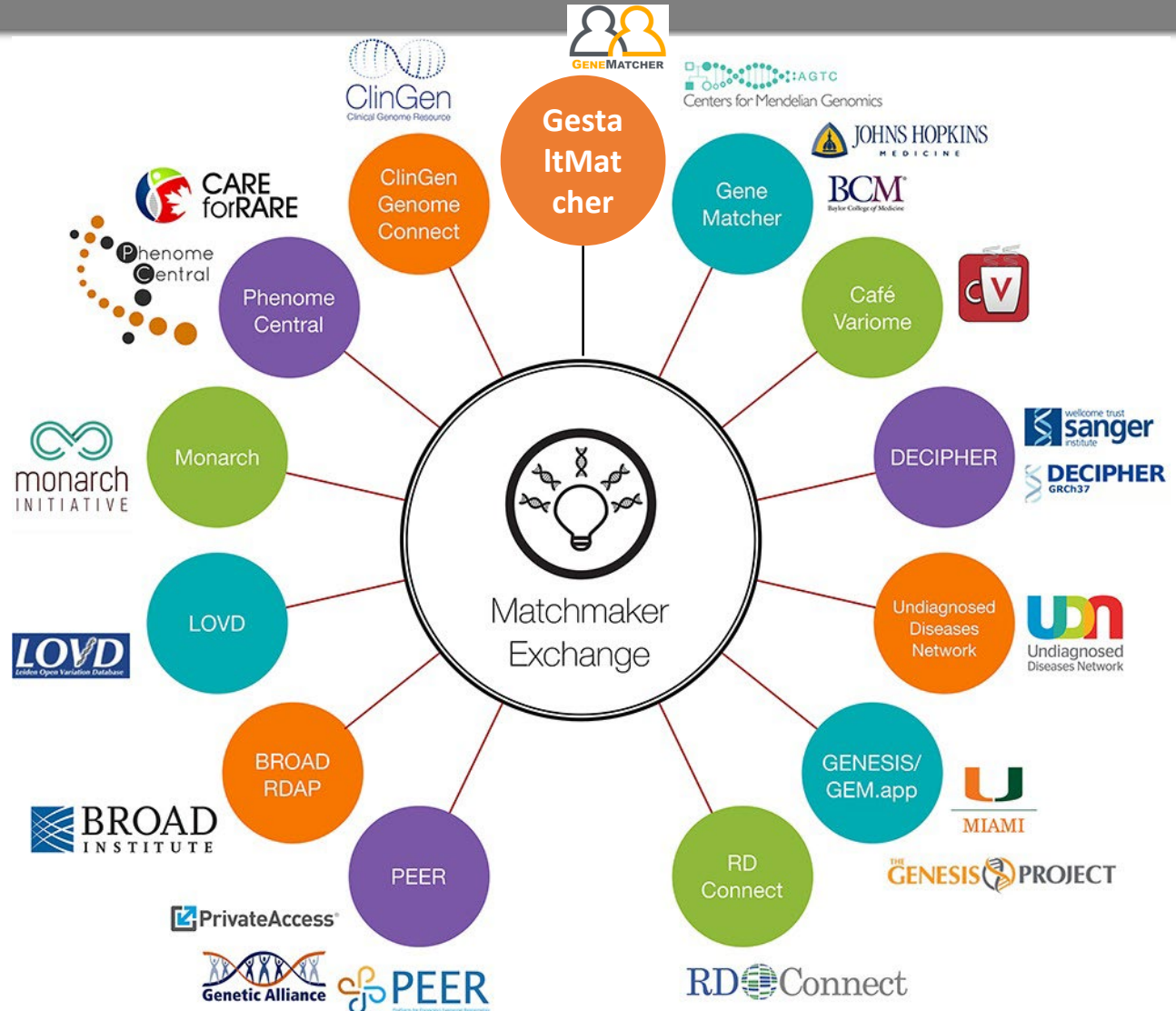
GMDB is a novel publication medium for dynamic case reports



GMDB an image repository and a node in MME

medRxiv

THE PREPRINT SERVER FOR HEALTH SCIENCES



World wide effort from the research community

Join FAIR GMDB and
become a coauthor!



Contact us:
info@gestaltmatcher.org





NICHT FREIGEgeben

HUMANGENETISCHE BEURTEILUNG

München, 03.05.2023

Name, Vorname	██████████	Geb.-Datum	██████████
Geschlecht	männlich	Patienten-ID	██████████
Auftrags-Nr.	██████████	Auftragsbeginn	14.04.2023
Externe Nr.	██████████	Freigabedatum	17.04.2023
Material-Nr.	██████████	Probenart	EDTA-Blut
Probeneingang	14.04.2023	Probenentnahme	keine Angabe
Anforderung/en	1) DMD - Deletion/Duplikation 2) DMD Kompletzsequenzierung		
Indikation	Muskuläre Hypotonie, positives Gowers zeichen, hypertrophe Waden, Creatinkinase wiederholt deutlich erhöht, Verdacht auf Muskeldystrophie		
HPO-Terms	Hypotonia, Gowers sign, Elevated circulating creatine kinase concentration		
Analyse	DMD – Deletion/Duplikation ID 932.00 DMD		

KLINISCHE BEURTEILUNG: AUFFÄLLIG

Diagnose einer DMD-assoziierten Muskeldystrophie.

ERGEBNIS UND INTERPRETATION

Im Rahmen der durchgeführten Analyse konnte eine **Deletion des Exons 45 (c.(6438+1_6439-1)_(6614+1_6615-1)del) hemizygot** im DMD-Gen nachgewiesen.

- Bei der nachgewiesenen Deletion handelt es sich um eine out-of frame Deletion, die den Leserahmen unterbricht. Eine out-of frame Deletion ist im männlichen Geschlecht in der Regel mit der Verlaufsform einer Duchenne'schen Muskeldystrophie verbunden. In den fachspezifischen Datenbanken (www.lovd.nl) ist diese Deletion sowohl mit einer Muskeldystrophie vom Typ Duchenne als auch vom Typ Becker-Kiener beschrieben.

Damit ist der genetische Befund vereinbar mit der Diagnose einer Dystrophin-abhängigen Muskeldystrophie.

Die Erkrankung folgt einem X-gebundenen Erbgang. Wir empfehlen eine genetische Beratung der Mutter sowie die Untersuchung bezüglich einer Anlageträgerschaft. Im Falle einer Anlageträgerschaft der Mutter wird die pathogene Variante mit einer 50%igen Wahrscheinlichkeit an Nachkommen weitergegeben, wobei weibliche Nachkommen Anlageträgerinnen sind und männliche Nachkommen an einer Muskeldystrophie erkranken. Bei fehlendem Nachweis der Anlageträgerschaft bei der Mutter besteht dennoch die Möglichkeit eines Keimzellmosaiks, was

Ärztinnen / Ärzte
Prof. Dr. med. Dipl.-Chem. Elke Holinski-Feder
Prof. Dr. med. Angela Abicht¹
Dr. med. Teresa Neuhaus¹
Dr. med. Iha Ahrens^{2,3}
Dr. med. Stefanie Balg⁴
Dr. med. Kerstin Becker^{1,2}
Dr. med. Anne Behnecke^{1,2}
Dr. med. Carolina Courage^{1,2}
Prof. Dr. med. Isabel Diebold^{1,2,4}
Dr. med. Christian Gebhardt⁵
Dr. med. Antonia Kirchhoff^{6,7}
Dr. med. Felicitas Maier^{1,8}
Dr. med. Yvonne Müller-Koch^{1,9}
Caroline von Plettenberg^{1,3}
Dr. med. Silja Rübner^{1,2,3}
Zelia Schmederer^{1,3}
Dr. med. Miriam Stampfer^{1,2}
Dr. med. Verena Steinke-Lange^{1,2}
Dr. med. Maximilian Witzel^{2,4,10}



Standorte

München
MGZ – Medizinisch Genetisches Zentrum
Bayerstr. 3-5 | Fingerg Schlossstr. 6
80335 München
Telefon: 089 / 30 90 886 - 0

Filiale Albstadt
Zentrum für Kinder und Jugendliche im Schloss
Vincenz-von-Paul-Str. 14 | 83603 Albstadt
Telefon: 089 / 30 90 886 - 0

Filiale Deggenhof
Klinikum Deggenhof
Perlesberger Str. 41 | 94469 Deggenhof
Telefon: 089 / 30 90 886 - 0

Filiale Donauwörth
Onkologisches Zentrum Donauwörth
Naudinger Allee 10 | 86409 Donauwörth
Telefon: 089 / 30 90 886 - 0

Filiale Eggenfelden
Rittal im Klinikum Eggenfelden
Simonsöder Allee 20 | 84307 Eggenfelden
Telefon: 08721 / 983 7224

Filiale Freising
Gynäkologische Gemeinschaftspraxis Freising
Marienplatz 4 | 85354 Freising
Telefon: 089 / 30 90 886 - 0

Filiale Garmisch-Partenkirchen
Klinikum Garmisch-Partenkirchen
Auenstr. 6 | 82467 Garmisch-Partenkirchen
Telefon: 089 / 30 90 886 - 0

Filiale Landshut
Sozialpädiatrisches Zentrum Landshut
Gillperstr. 9 | 84036 Landshut
Telefon: 089 / 30 90 886 - 0

Filiale Lindenberg
MVZ Aligja Andersen
Dr.-Otto-Geßler-Platz 1 | 88161 Lindenberg
Telefon: 089 / 30 90 886 - 0

Filiale Rosenheim
Klinikum Rosenheim
Paternkoflerstr. 10 | 83027 Rosenheim
Telefon: 089 / 30 90 886 - 0

Filiale Straubing
Klinikum St. Elisabeth
Aylburger Str. 19 | 94315 Straubing
Telefon: 089 / 30 90 886 - 0

Filiale Traunstein
Klinikum Traunstein
Carnegigglstr. 21 | 83278 Traunstein
Telefon: 089 / 30 90 886 - 0
info@mgz-muenchen.de
www.mgz-muenchen.de



Die Akkreditierung gilt für die in der 17. Kundenzurufnummer D.M. 13242/06 festgelegten Verfahren.

¹ Fachverband Humangenetik
² F. Anabring
³ Fachverband für Seltene Mutationen im NSE-Netz
⁴ Fachverband für Genetik und Jugendberufshilfe
⁵ im MGZ-Netz praktizierend
⁶ im MGZ-Netz praktizierend
⁷ im MGZ-Netz praktizierend
⁸ im MGZ-Netz praktizierend
⁹ im MGZ-Netz praktizierend
¹⁰ im MGZ-Netz praktizierend

Humangenetische Beurteilung

Name, Vorname	██████████	Auftrags-Nr.	██████████
Geb.-Datum	██████████	Analyse	██████████

im Rahmen einer genetischen Beratung erläutert werden sollte. Der Patient gibt die pathogene Variante an alle seine Töchter weiter (obligate Anlageträgerinnen). Söhne des Patienten sind nicht Träger der pathogenen Variante. Weiteren weiblichen Familienangehörigen kann eine Testung auf eine Anlageträgerschaft im Rahmen einer genetischen Beratung empfohlen werden.

Die nachgewiesene Deletion des Exons 45 ist theoretisch einer aktuell noch in Entwicklung befindlichen molekularen Therapie (Exon-Skipping) zugänglich. Einer Therapie mit Translarna (Ataluren) ist die nachgewiesene genetische Veränderung aufgrund des molekularen Mechanismus nicht zugänglich.

Aufgrund des auffälligen Befundes wurde die ebenfalls angeforderte Untersuchung(en) nicht veranlasst.

Wir empfehlen eine Befundbesprechung im Rahmen eines genetischen Beratungsgesprächs

Validation



elektronisch validiert durch ██████████

Das MGZ ist nach der internationalen Norm DIN EN ISO 15189 "Medizinische Laboratorien - besondere Anforderungen an die Qualität und Kompetenz" akkreditiert. Humangenetische Untersuchungen haben eine sehr hohe aber nicht 100%ige Sensitivität und unterliegen wie alle Laboruntersuchungen einem Fehlerisiko aufgrund präanalytischer, analytischer, postanalytischer und bioinformatischer Prozesse. Dieser Befund basiert auf dem zum Zeitpunkt der Befunderstellung aktuellen Wissensstand. Die Interpretation bei Familienuntersuchungen gilt nur unter der Voraussetzung, dass die angegebenen Verwandtschaftsverhältnisse korrekt sind. Wir empfehlen eine Befundbesprechung im Rahmen eines genetischen Beratungsgesprächs, insbesondere im Falle eines auffälligen Ergebnisses. Bei einer prädiktiven oder vorgeburtlichen genetischen Untersuchung ist die betreffende Person bzw. die Schwangere gemäß Gendiagnostikgesetz (GenDG) vor der genetischen Untersuchung und nach Vorliegen des Untersuchungsergebnisses genetisch zu beraten. Abrechnung nach EBM: 11370x1

PATIENTENINFORMATION

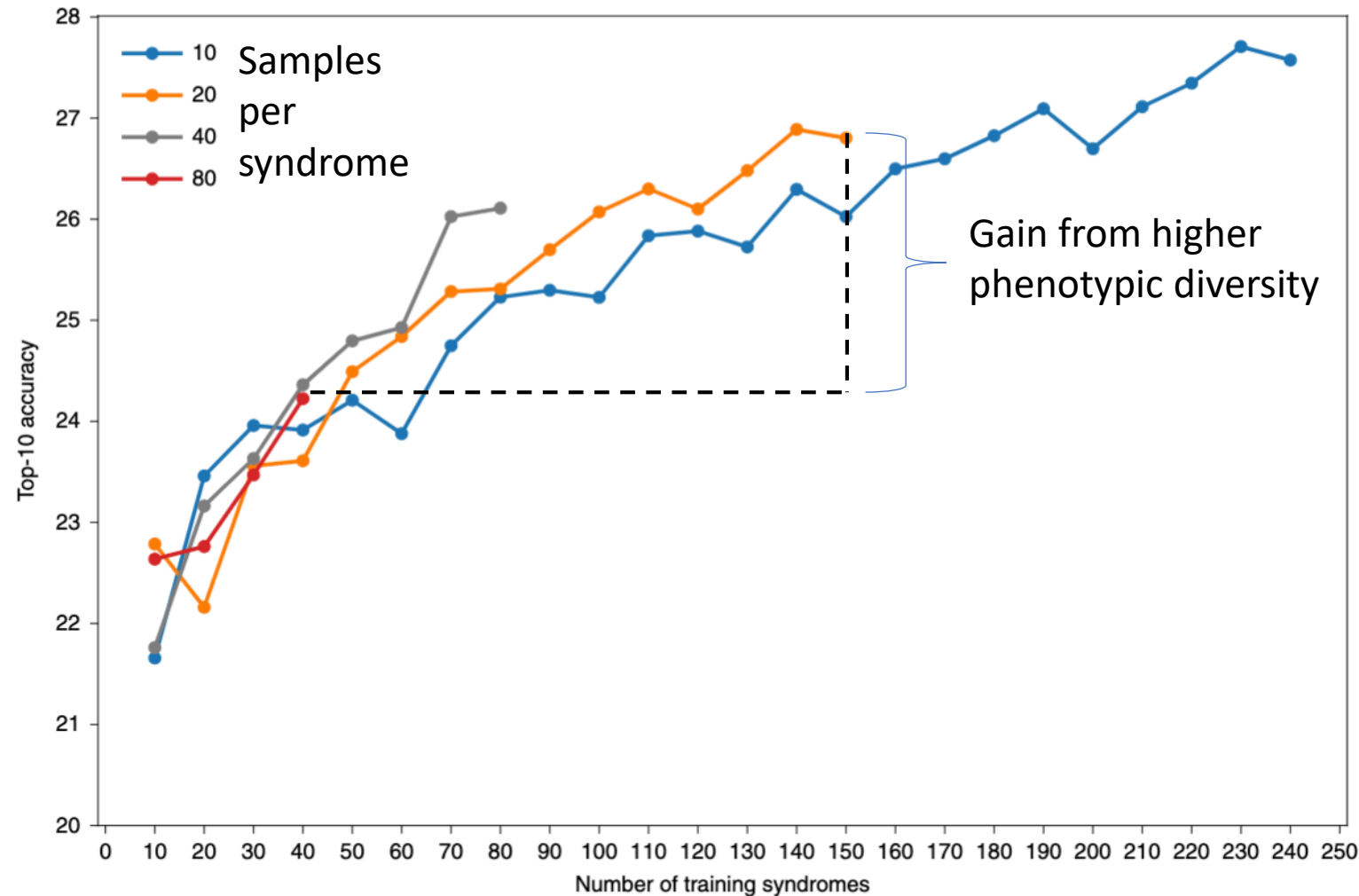
FindMe2care

Diagnosebezogene Medizin bei seltenen Genetischen Erkrankungen

Aufgrund der genetischen Diagnose besteht für den Patienten / die Patientin die Möglichkeit, sich bei der medizinischen Kontaktplattform FindMe2Care zu registrieren. Für die Registrierung benötigt der Patient diesen individuellen QR-Code, in dem die genetische Diagnose ohne die patientenbezogenen Daten verschlüsselt sind.

Registrierung und weitere Information:
<https://www.findme2care.de>

Factors that influence the performance of an AI: diverse data



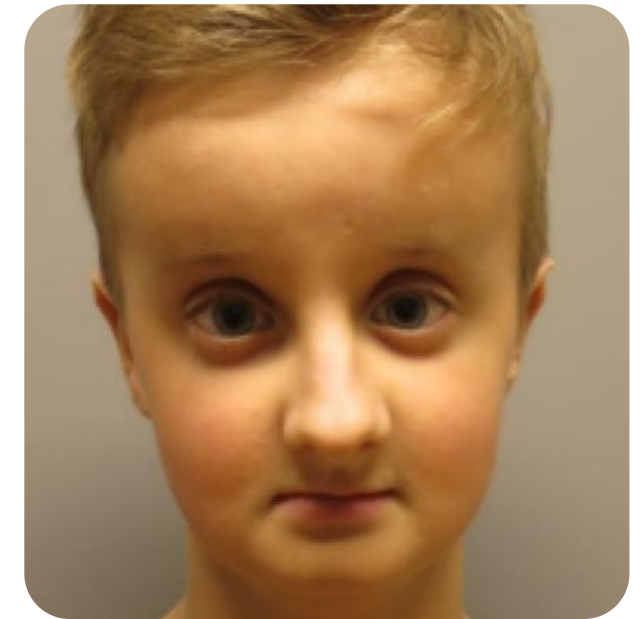
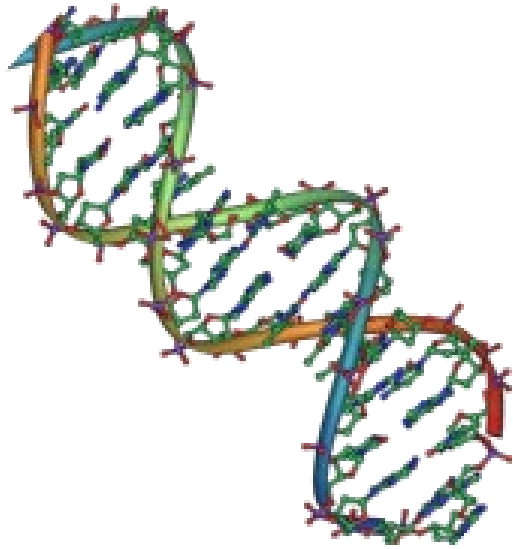
- 40 syndromes x 80 samples
- Has a lower performance than:
 - 160 syndromes x 20 samples
 - Diversity in your data set matters more than sheer size



**form
follows
function**

**Dysfunction
Delineated from
Dysmorphism**





BIP39 Seed Phrase Generator

v1.0.2

Generate a random seed phrase (mnemonic): **GENERATE** 12 words

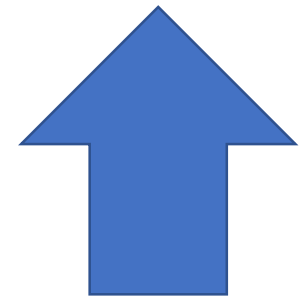
basic lady garden trophy bracket salute hungry any tongue lyrics portion poet



Encodings on different canvases

- Bases in the genome $3 \cdot 10^9$
- Pathogenic variants 10^{10}
- Cells/atoms in body $3 \cdot 10^{15/28}$
- Seed phrases: $5 \cdot 10^{39}$ (= $2048 \cdot 10^{12}$)
- atoms in the universe: 10^{81}
- QR combinations: $8 \cdot 10^{297}$ ($2^{33 \cdot 33}$)
- a 100x100 px portrait: NaN

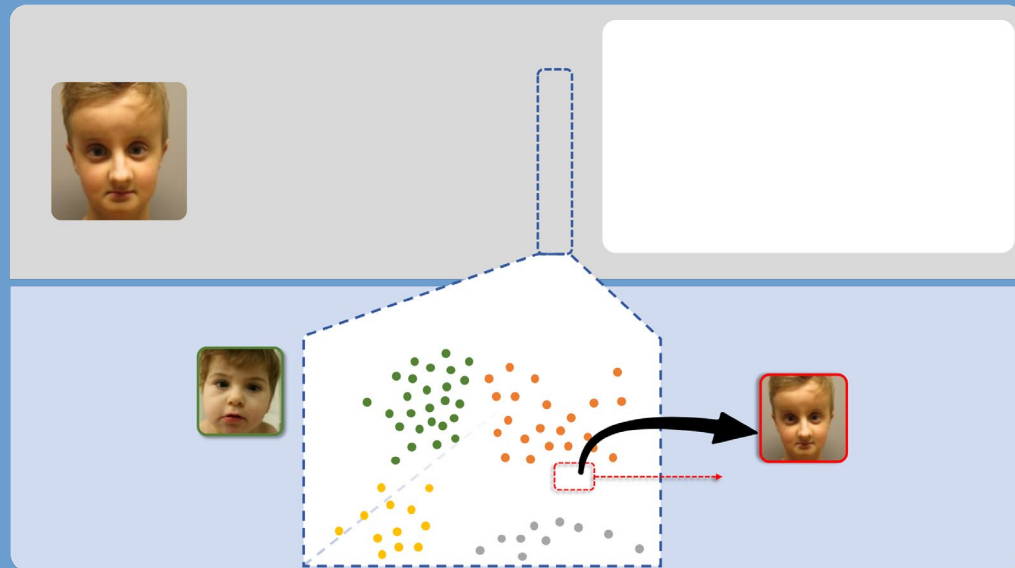
Dysfunction



Dysmorphism

Is there a correct answer to the lumping and splitting question?

Next-generation phenotyping

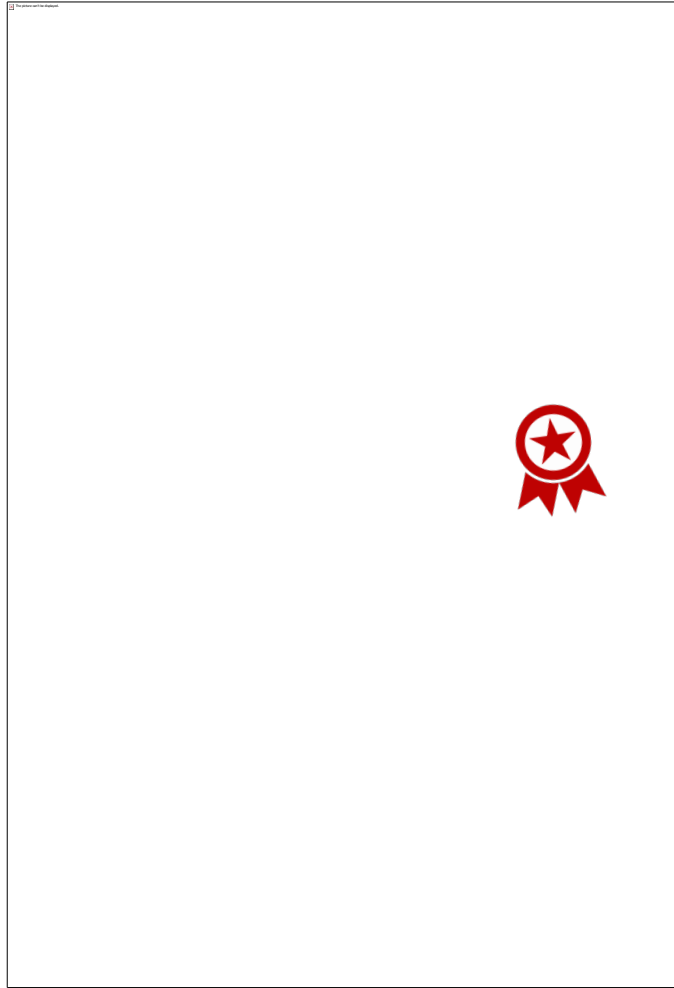


Lumping and splitting





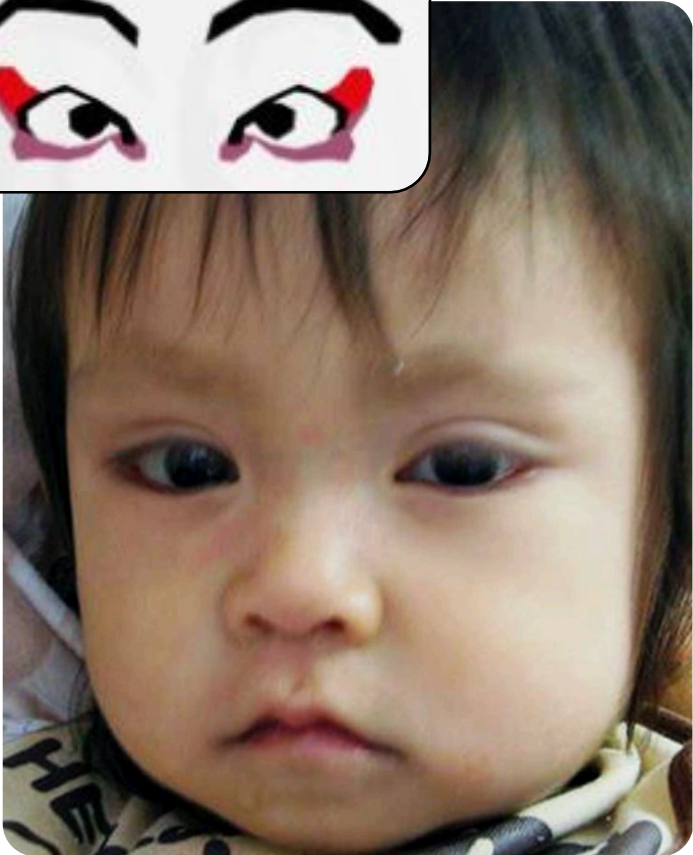
Family similarities



*  = Victor



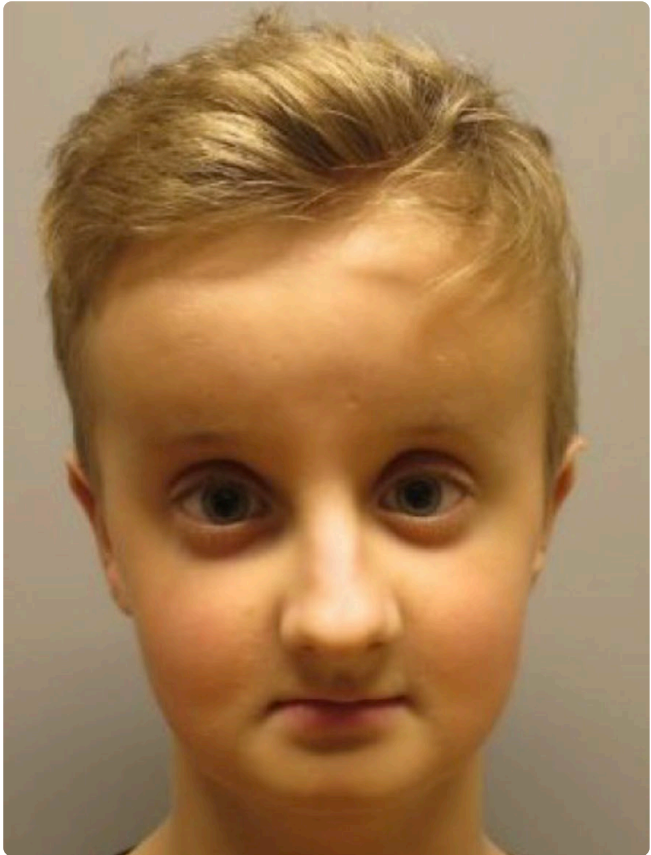
Discuss measurable features?



Kabuki make-up Syndrome



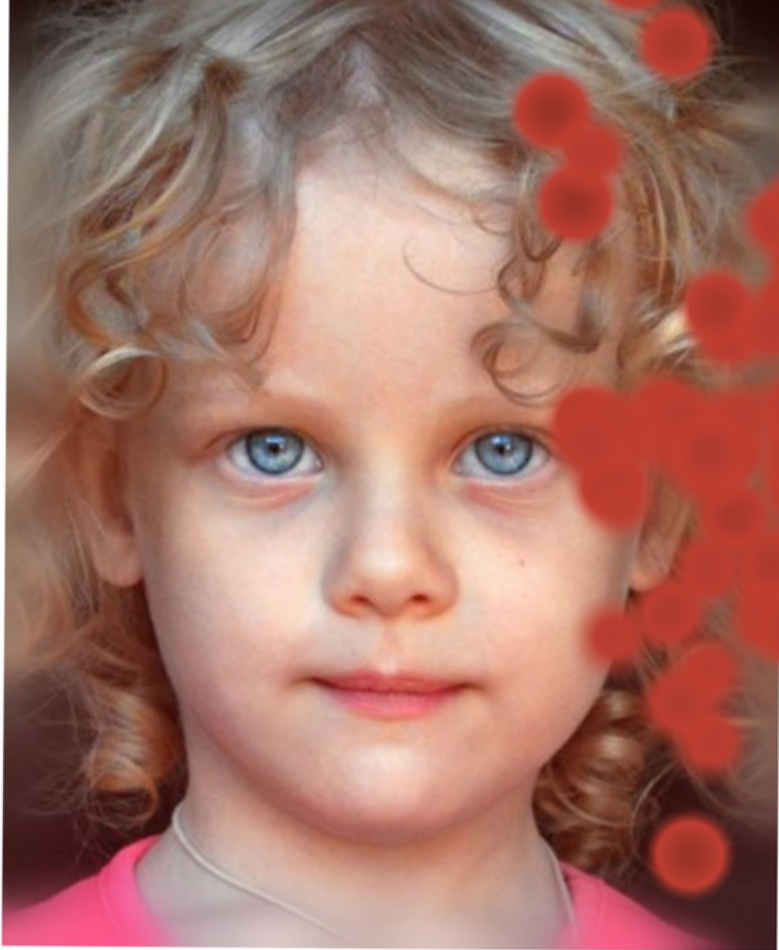
HPMR-Syndrome
(aka Mabry S)



CTRCT46-Syndrome
Cataract 46, Juvenile-Onset, w/wo Arrhythmic
Cardiomyopathy

The (subjective?) answers of Explainable AI (XAI)





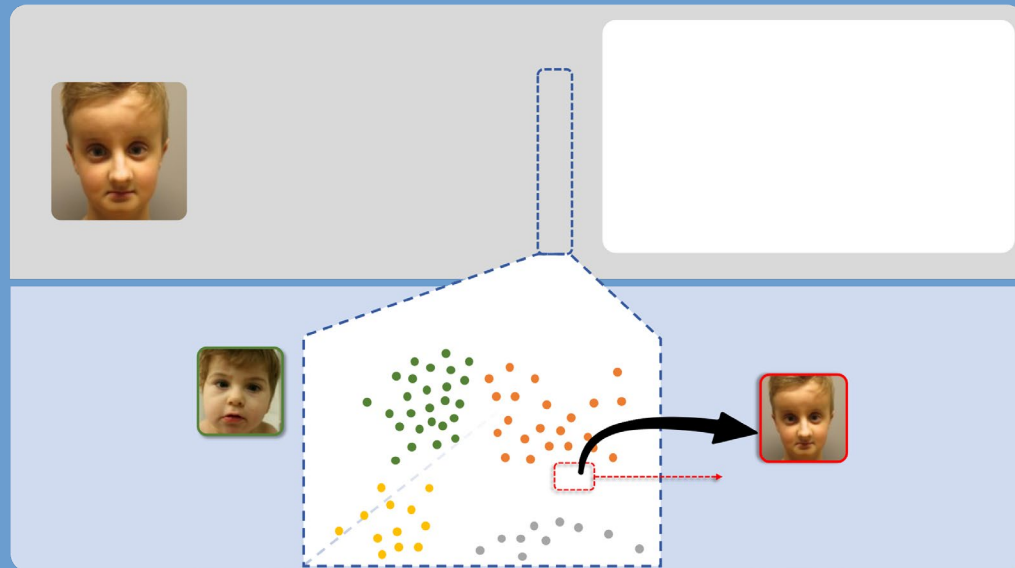


Factors that influence the performance of an AI

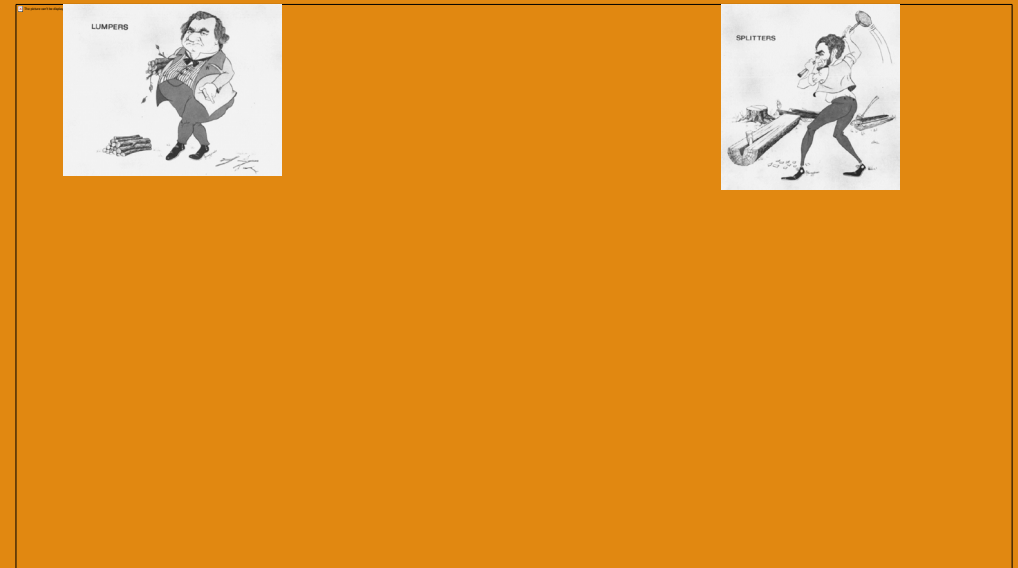
- Architecture of the network
- Learning algorithms
- Training data (prior and post knowledge transfer)

Working on the L&S problem with NGP

Next-generation phenotyping

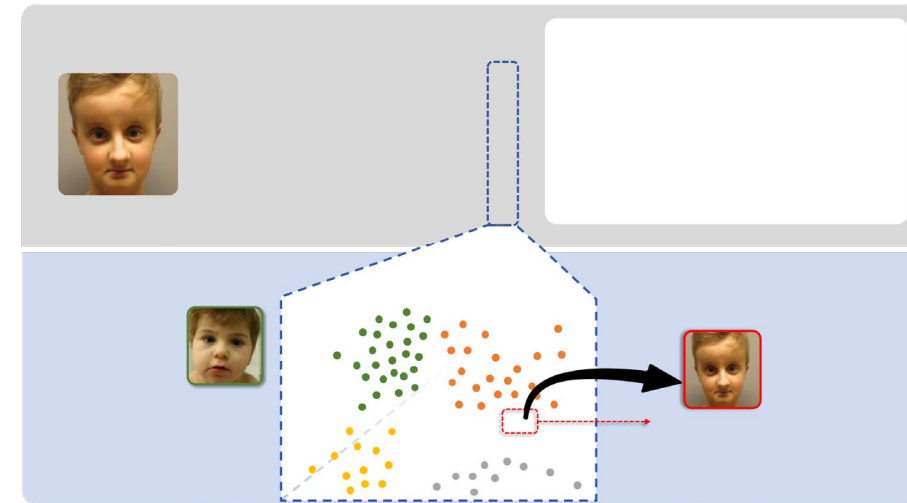


Lumping and splitting



Conclusion

- GestaltMatcher can facilitate the patient matching of ultra-rare disorders.
- GestaltMatcher can contribute to the longstanding discussion of lumping and splitting decisions.



Acknowledgements



IGSB

Prof. Peter Krawitz
Tzung-Chien Hsieh
Hannah Klikhammer
Alexej Knaus
Tom Kamphans
Hellen Lesmann
Jing-Mei Li
Alexander Hustinx
Behnam Javanmardi
Wolfgang Meiswinkel
and all the colleagues



T-NAMSE

Prof. Christian Schaaf
Prof. Regina Betz
Theresa Brunet
Hartmut Engels
Sophia Peters
Axel Schmidt
Matias Wagner
Tobias Haack



CHU de Nantes
Sébastien KÜRY



Charité Berlin

Martin Mensah
Jean Tori Pantel
Stanislav Rosnev



Stellenbosch

Prof. Shahida Moosa
Kimberly Christine Coetzer



Thank you!

Follow us
 [@GestaltMatcher](https://twitter.com/GestaltMatcher)

Acknowledgements

Join FAIR GMDB and
become a coauthor!

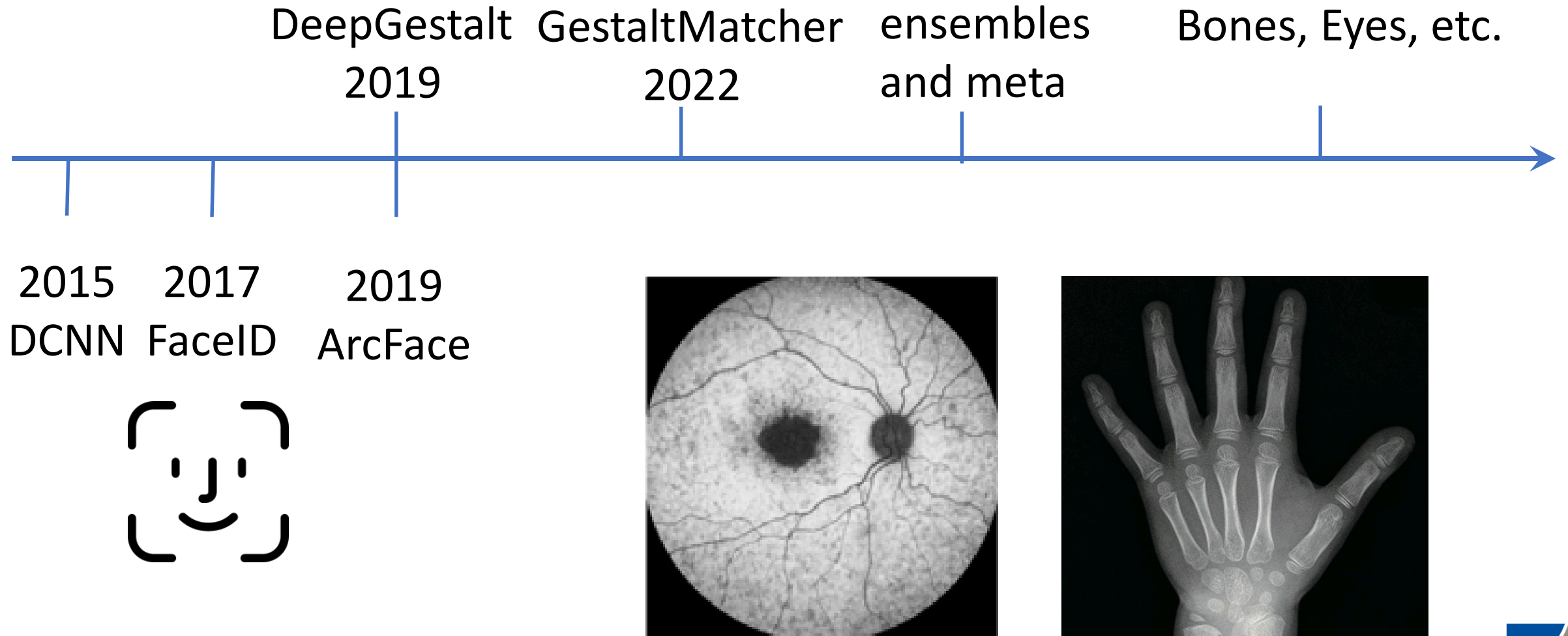


Thank you!

Contact us:
info@gestaltmatcher.org



What to expect in the coming months and years

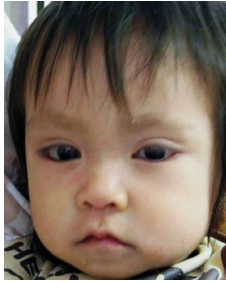


What to expect in the coming months and years

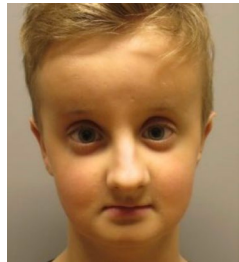
Face2Gene



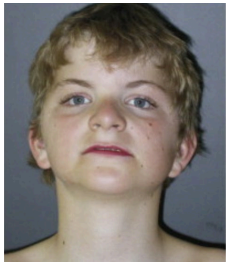
PIGV



KMT2D

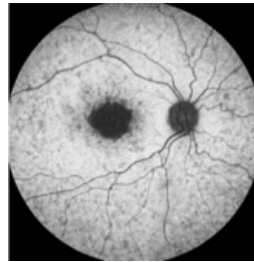


LEMD2

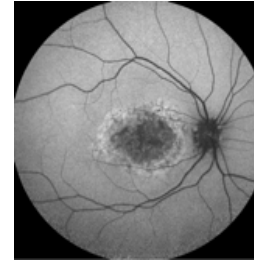


PACS1

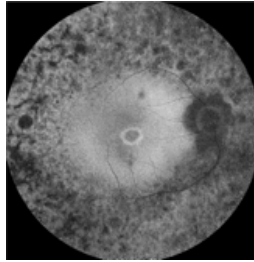
Eye2Gene



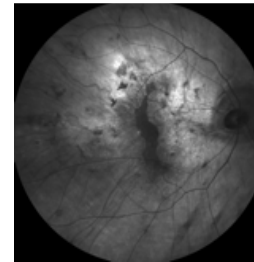
ABCA4



EFEMP1

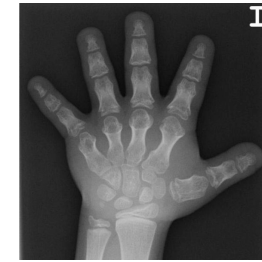


USH2A



CHM

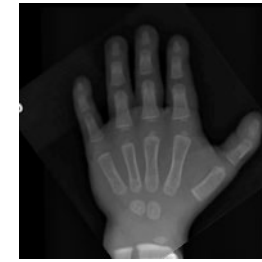
Bone2Gene



PRKAR1A



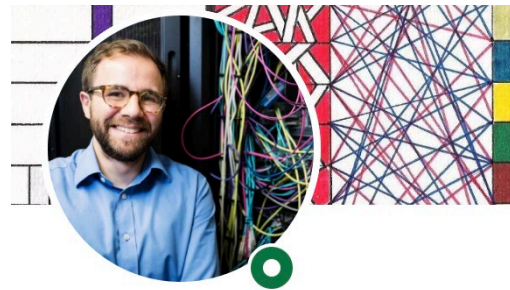
RMRP



PIGV



SHOX



Nikolas Pontikos · 1.



Behnam Javanmardi, PhD · 1.
Artificial Intelligence in Medical Imaging

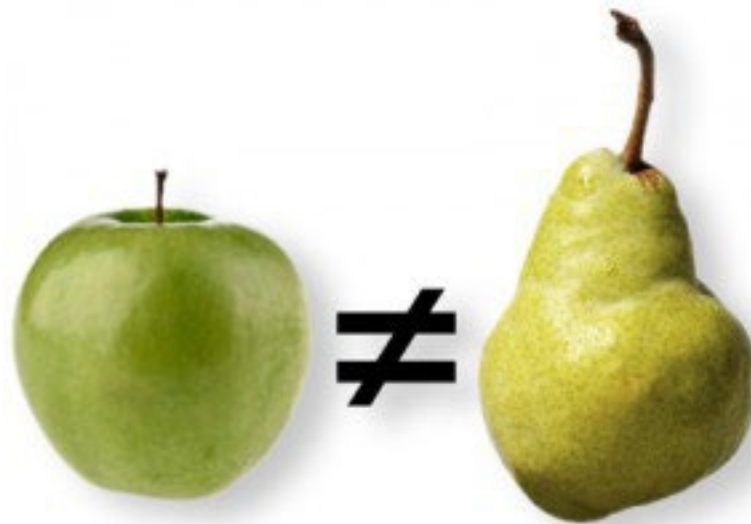


Questions I would like to address in my talk

- How do I measure the performance of an AI?
- What influences the performance of an AI?
- How do I compare the performance of two different AIs?

Performance comparison of AIs

- You need to make sure that the same training and test data is used
- You need to make sure that the same training and test data is used
- You need to make sure that the same training and test data is used



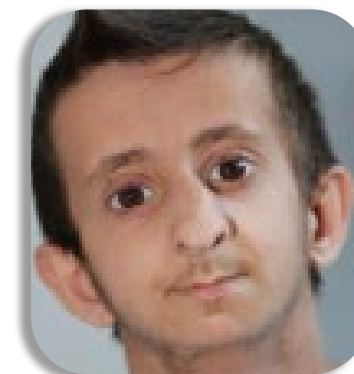
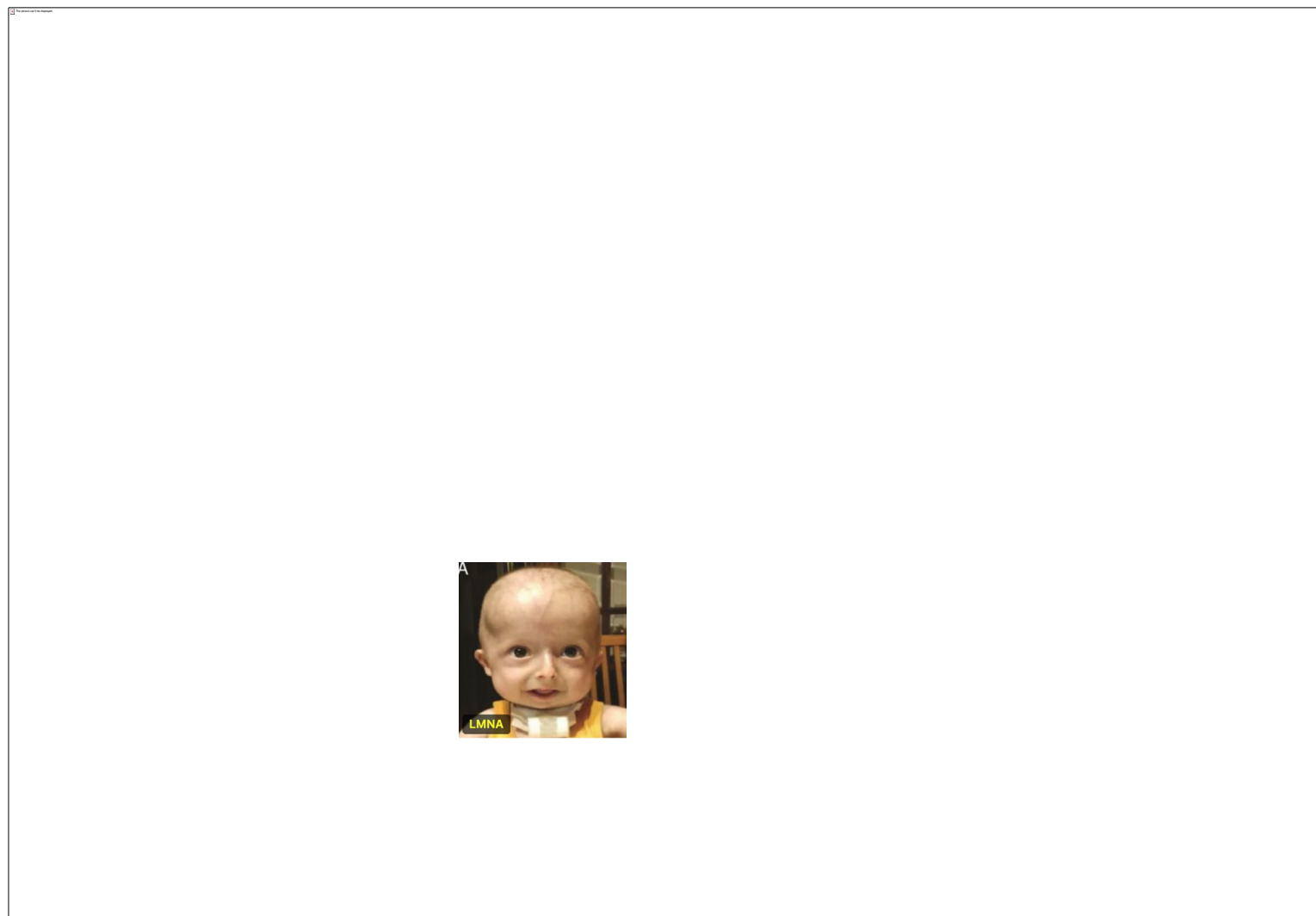


Questions I would like to address in my talk

- How do I measure the performance of an AI?
- What influences the performance of an AI?
- How do I compare the performance of two different AIs?
- What is a FAIR database?



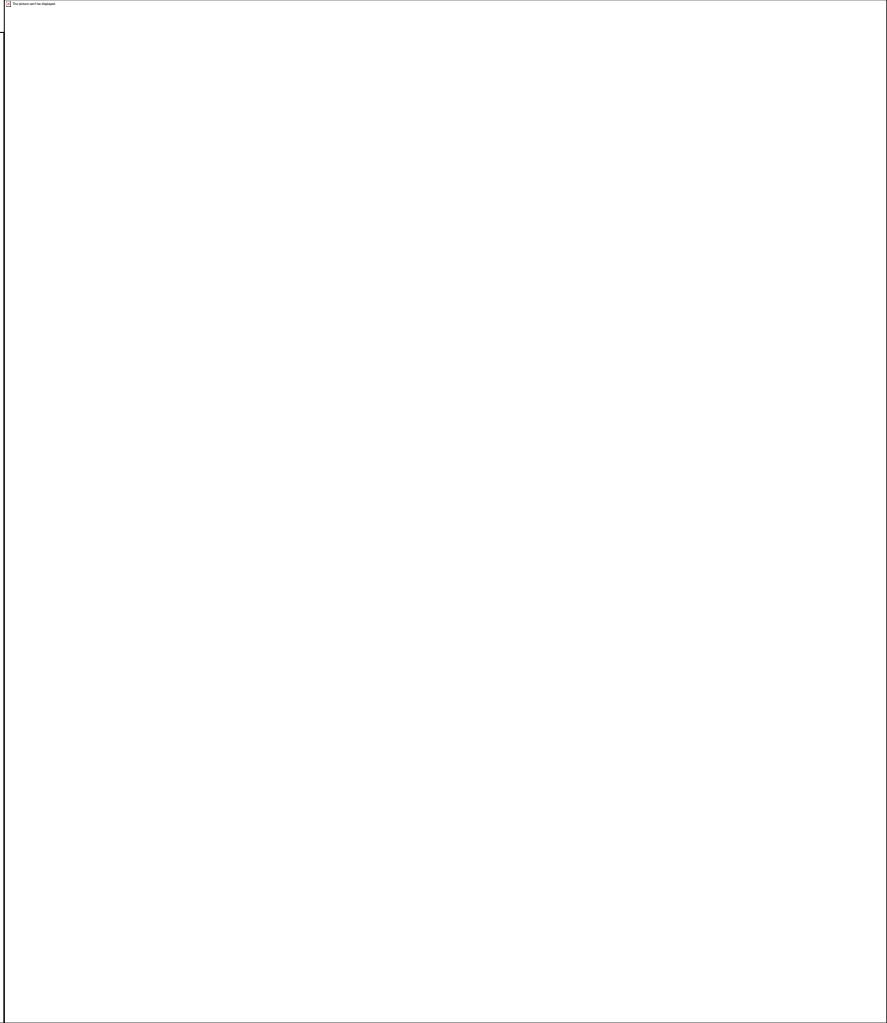
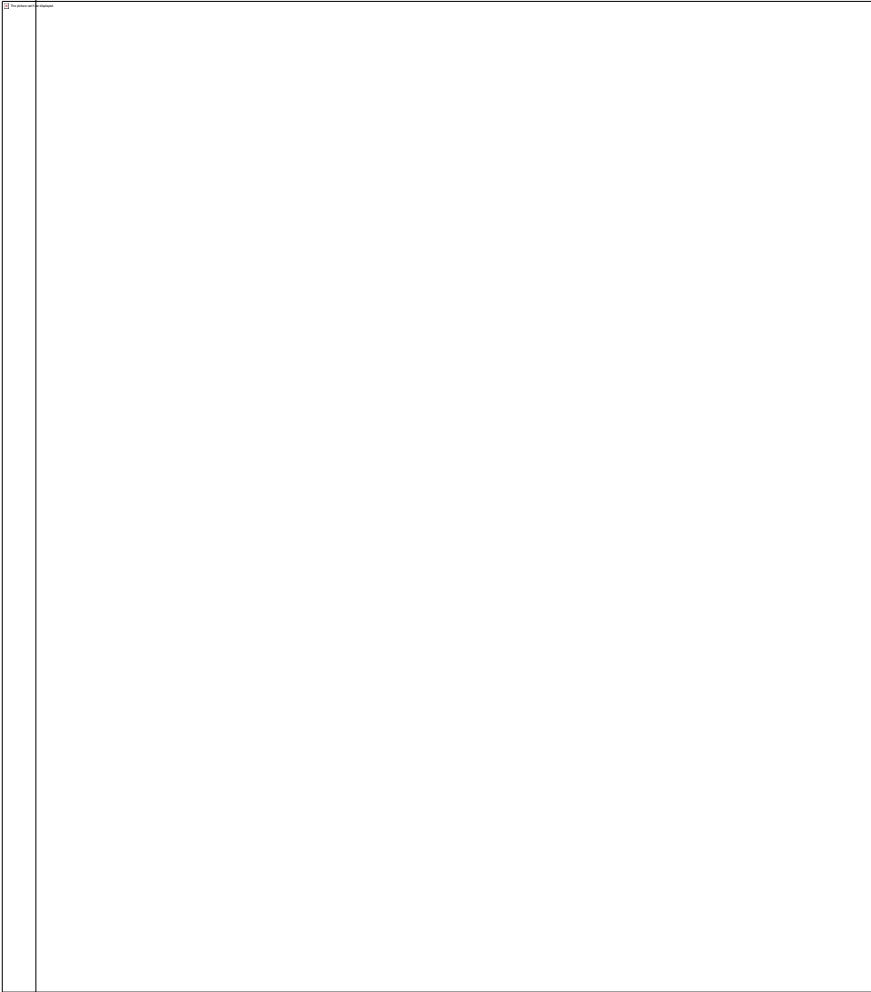
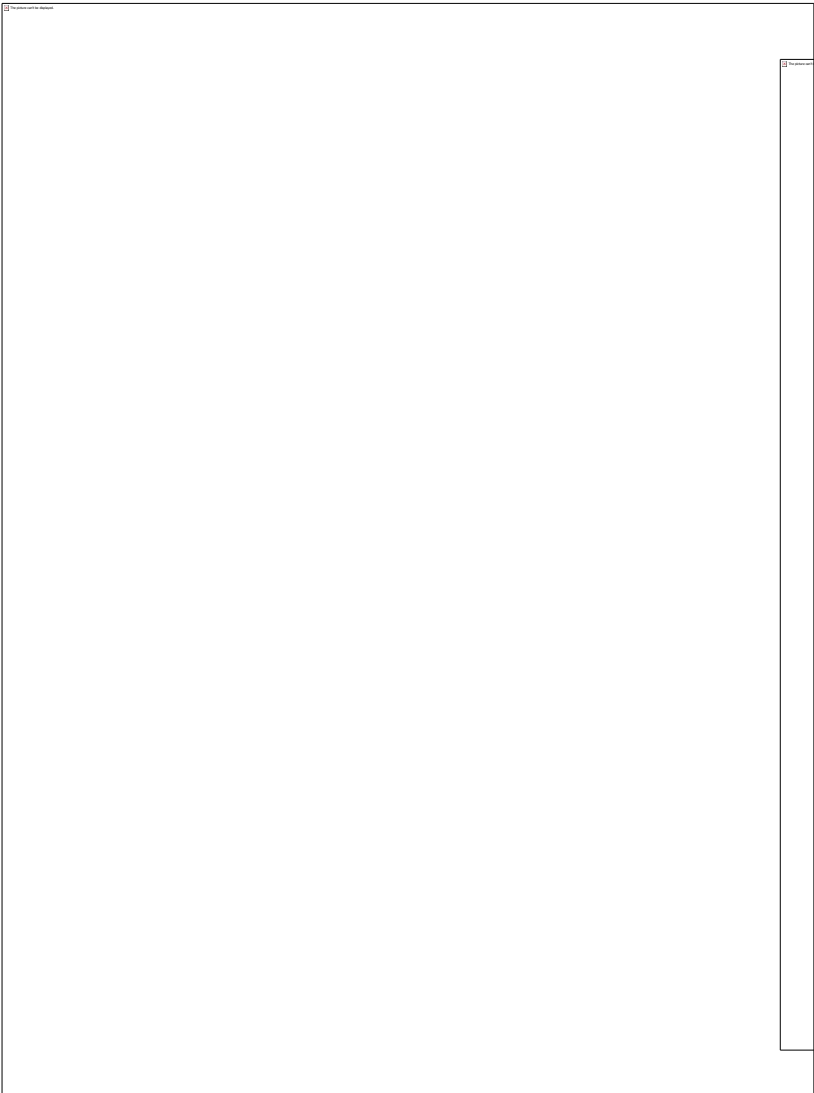
GMDB data is reusable (and therefore suitable for benchmarks)

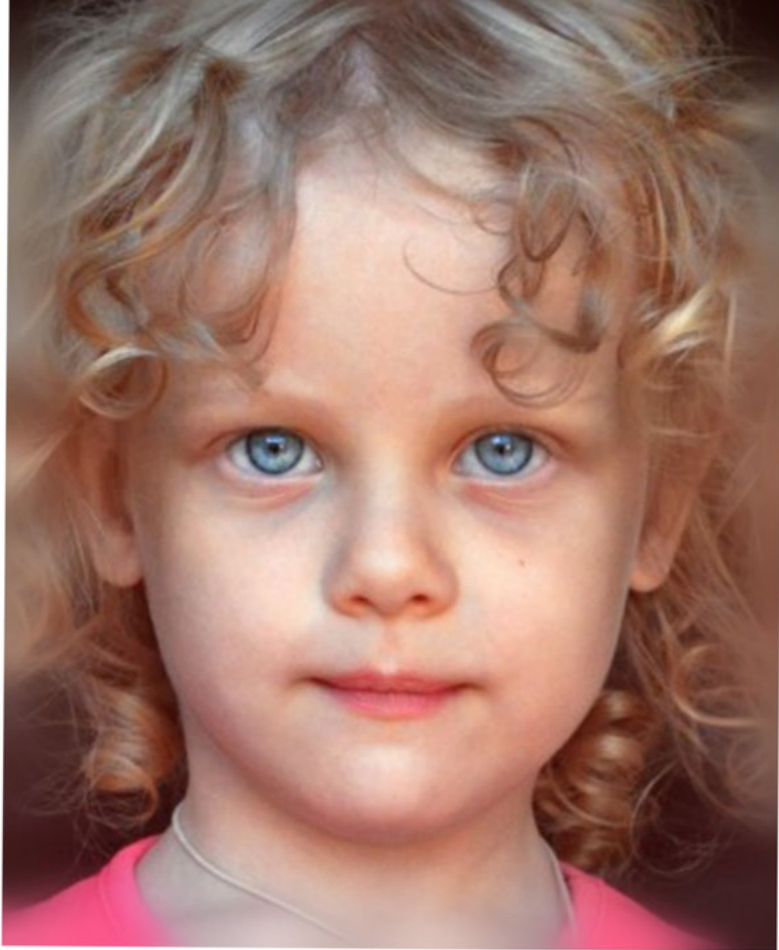


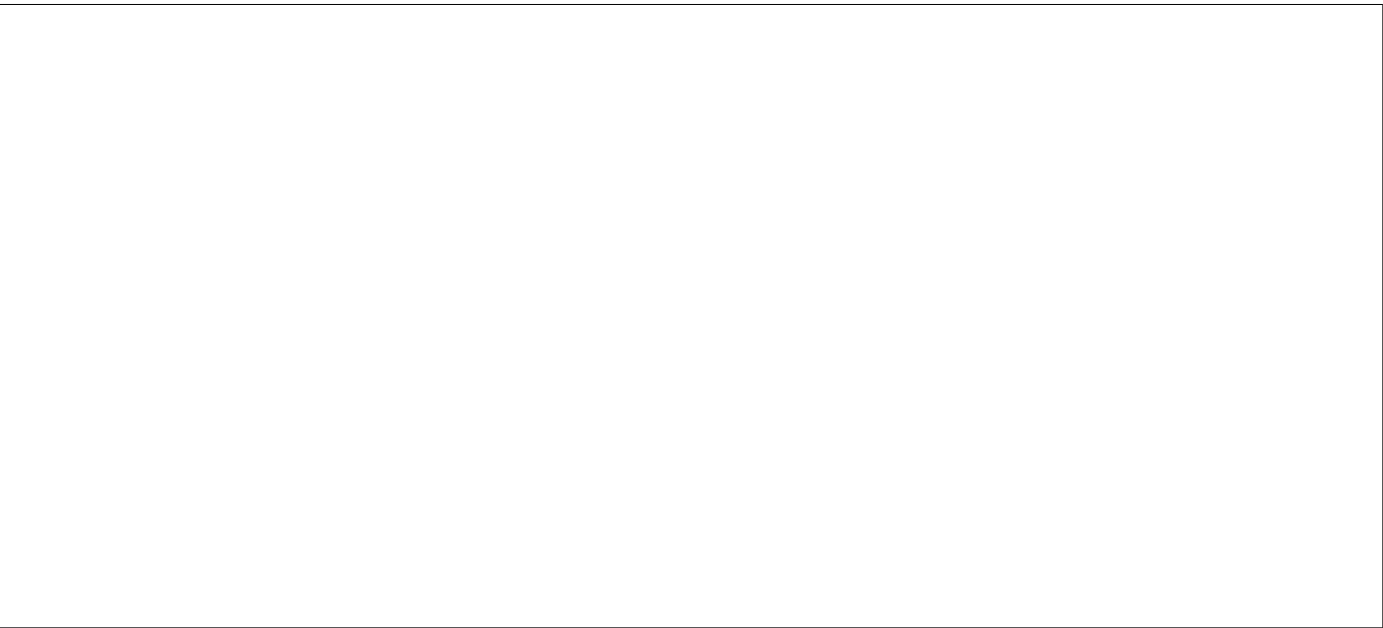


Questions I would like to address in my talk

- How do I measure the performance of an AI?
- What influences the performance of an AI?
- How do I compare the performance of two different AIs?
- What is a FAIR database?
- What can I do with GestaltMatcher in Research?







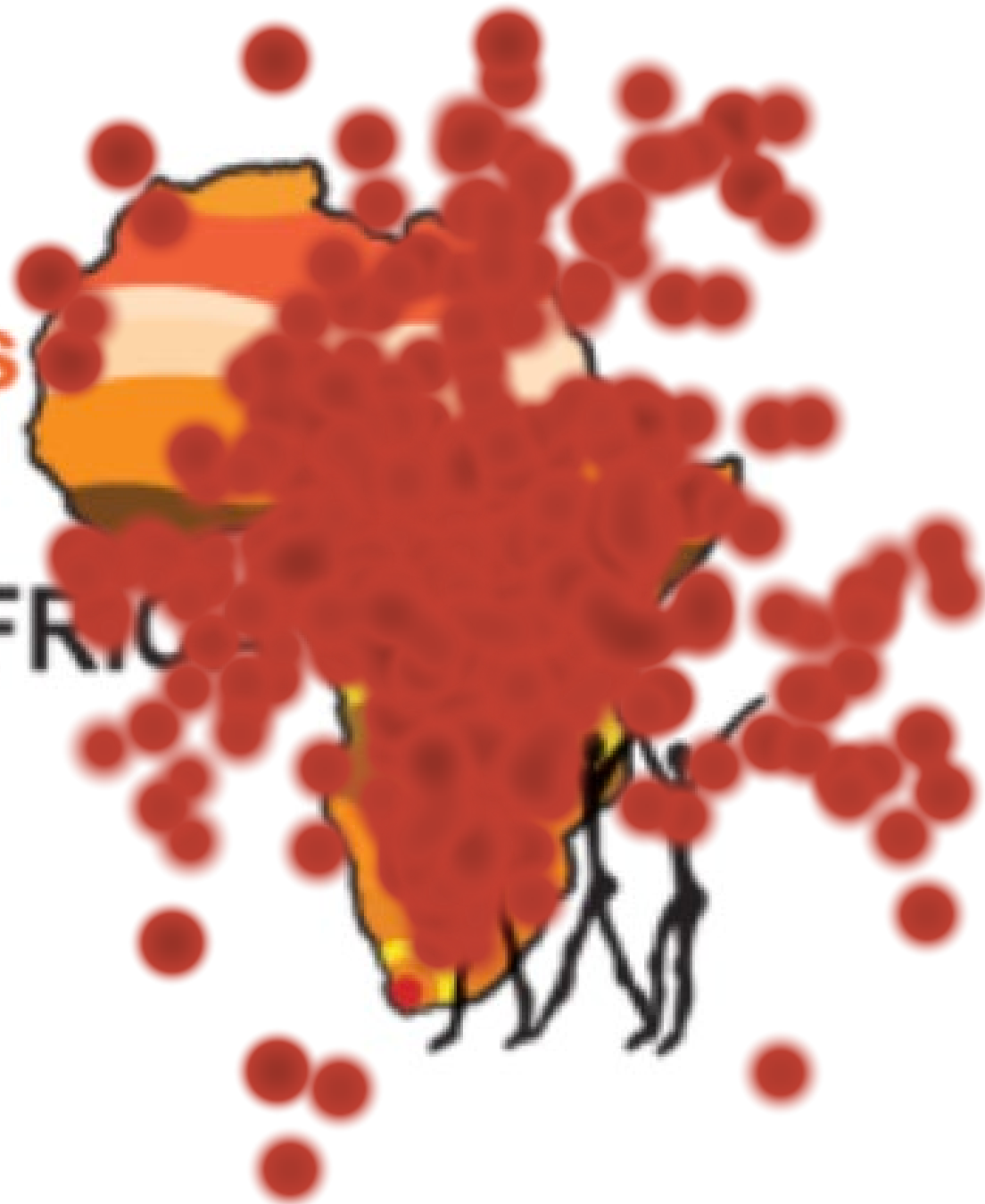
etics

23

| AFRIC

n

1



Training task 1

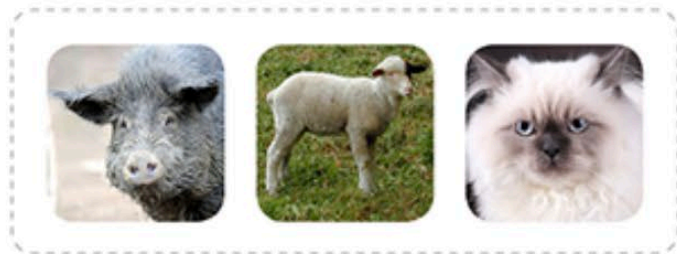
Support set

K=2



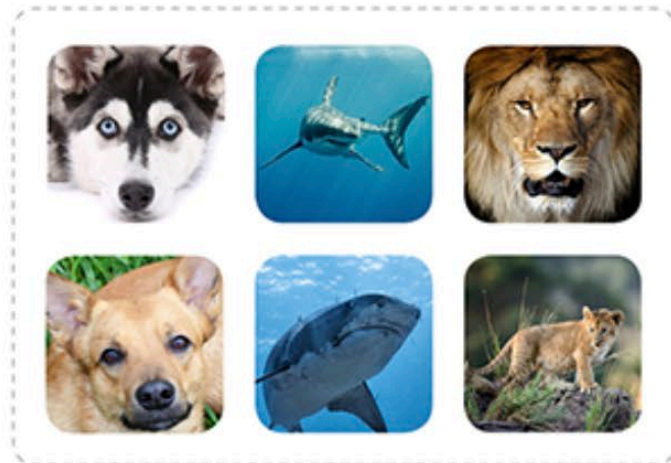
N=3

Query set

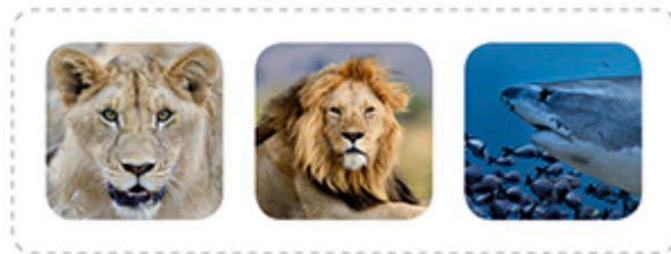


Training task 2 . . .

Support set

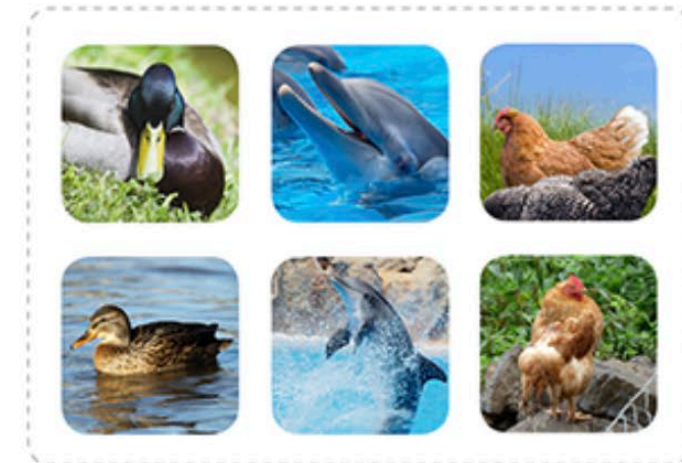


Query set

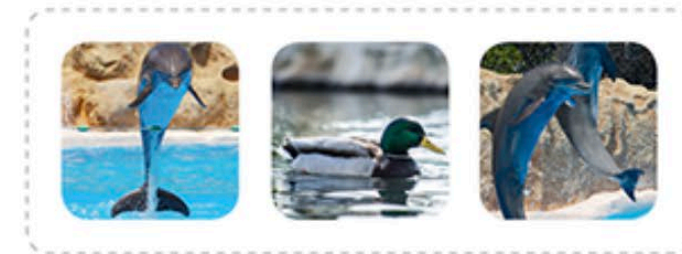


Test task 1 . . .

Support set



Query set



First real world example DeepGestalt

Google's
faux pas

DeepGestalt
2019

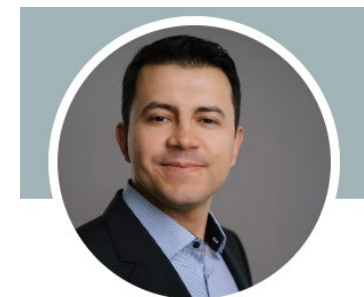
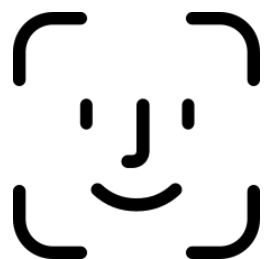
GestaltMatcher
2022

Ensembles & Few Shot
2023

2015
DCNN

2017
FaceID

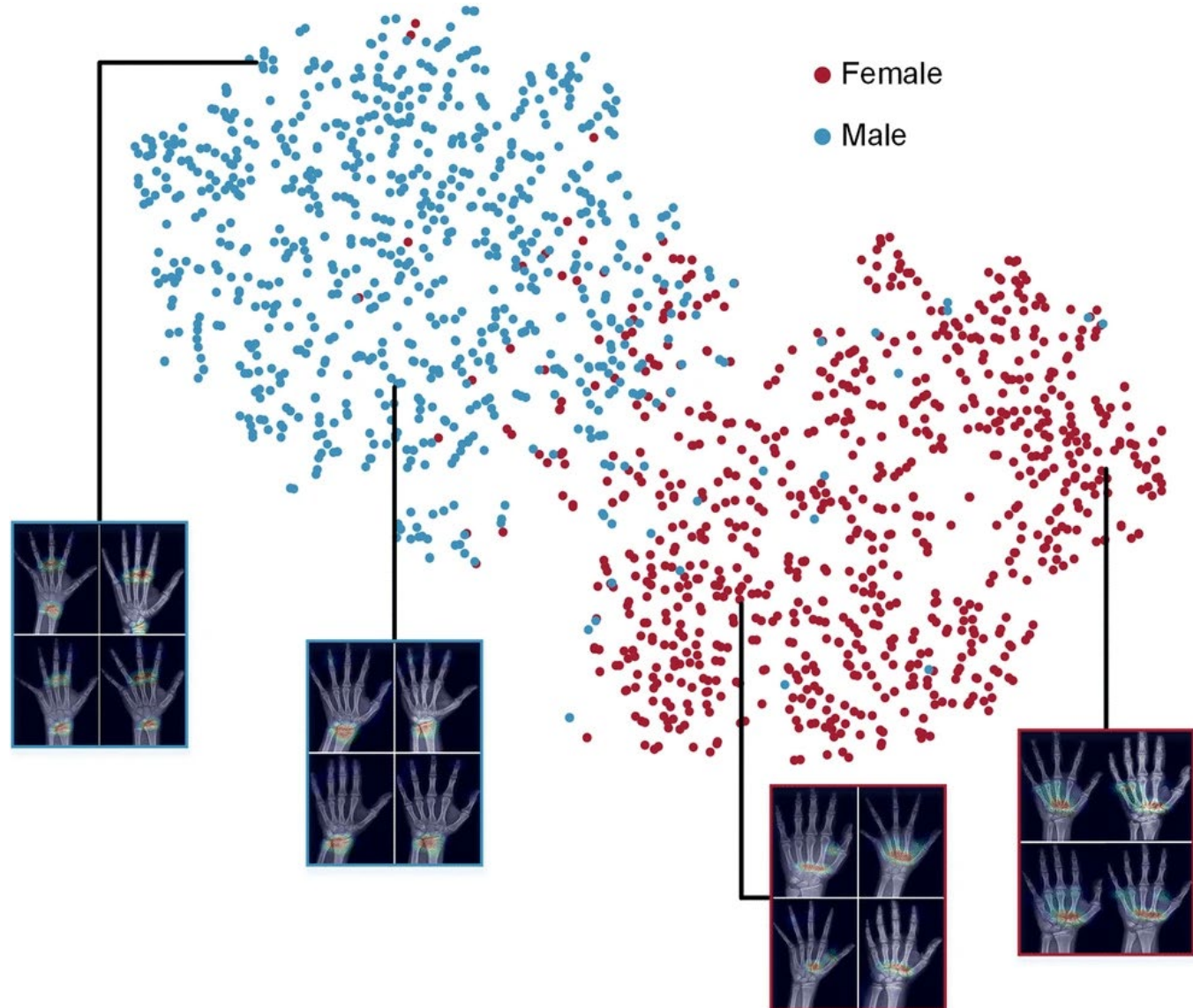
2019
ArcFace



AI can predict sex from hand X-ray

The AI predicted sex based on only hand X-ray with 95% accuracy

Two human radiologists predicted sex based on only hand X-ray with 58% and 46% accuracy.



Prior knowledge about similarity

Learn embedding to discriminate two unseen classes.

Siamese networks (Koch et al., 2015)

Triplet networks (Hoffer & Ailon, 2014)

Learn embedding to discriminate many unseen classes.

Matching networks (Vinyals et al., 2016)

Prototypical networks (Snell et al., 2017)

Relation networks (Santoro et al., 2016)

Prior knowledge about learning

Learn model with parameters that can easily be fine-tuned to give good results.

MAML (Finn & Levine 2017)

FOMAML (Finn & Levine, 2017)

Reptile (Nichol et al., 2018)

Learn update rule that encourages good performance with small datasets.

LSTMs (Ravi & Larochelle, 2016)

Reinforcement learning (Li & Malik, 2016)

Optimization rules (Bello et al., 2017)

Sequence methods. Take entire dataset and test example and predict test label.

Memory-augmented NN (Santoro et al., 2016)

SNAIL (Mishra et al., 2017)

Prior knowledge about the data

Learn generative model for family of data classes.

Pen-stroke model (Lake et al., 2015)

Neural statistician (Edwards & Storkey., 2016)

Learn to synthesize new examples and train with augmented data.

Analogies (Hariharan & Girschick, 2017)

End-to-end (Wang et al., 2018)

Is there a correct answer to the lumping and splitting question?



what is the difference between objective and subjective?



Objective and subjective are two different ways of looking at information or experiences.

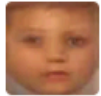


Objective refers to information or experiences that are based on facts, evidence, or observations that can be measured or verified. Objective information is not influenced by personal opinions, biases, or emotions. For example, the temperature of a room, the weight of an object, or the number of people in a crowd are all objective facts.

Subjective, on the other hand, refers to information or experiences that are based on personal opinions, beliefs, feelings, or experiences. Subjective information is influenced by personal perspectives, biases, or emotions. For example, a person's taste in music, their favorite color, or their emotional reaction to a movie are all subjective experiences.

In summary, objective information is based on facts and evidence that can be measured or verified, while subjective information is based on personal perspectives, opinions, or experiences.

Is there a correct answer to the lumping and splitting question?



Are you objective or subjective?



As an artificial intelligence language model, I am programmed to be as objective as possible. I do not have personal opinions or beliefs and I do not experience emotions. My responses are based on the information that I have been trained on and the algorithms that I use to generate them. However, the information that I provide can still be influenced by the quality and accuracy of the data that I have been trained on, so it's important to critically evaluate any information I provide and consider multiple sources.

there a correct answer to the lumping and splitting question?



Physics is generally considered an objective science because it aims to describe the natural world through measurable and verifiable phenomena. Physicists use mathematics, experiments, and observations to make predictions about the behavior of matter and energy in the universe, and they test those predictions through experimentation and observation.

The fundamental laws of physics, such as the laws of motion, the conservation of energy, and the laws of thermodynamics, are based on empirical evidence and have been tested repeatedly under different conditions to confirm their validity. They are considered objective because they apply to all objects and phenomena in the universe, regardless of personal perspectives or biases.

However, it's worth noting that the interpretation of experimental results can sometimes be influenced by personal biases or assumptions, and scientific models are always subject to refinement and revision as new evidence emerges. Nonetheless, physics is generally considered to be an objective science because it seeks to describe the natural world through empirical evidence and logical reasoning.



Can AI be more objective?

Medicine is a field that combines objective scientific knowledge with subjective elements related to patient care and individual experiences.

In terms of diagnosis and treatment, medicine is based on scientific principles and objective evidence. Doctors and other healthcare professionals use a range of objective tests, such as blood tests, imaging studies, and physical examinations, to diagnose medical conditions and develop treatment plans. They also rely on scientific research to guide treatment decisions, with the goal of providing the most effective and evidence-based care possible.

However, medicine also involves subjective elements related to patient care, such as communication, empathy, and cultural sensitivity. Patient experiences, values, and beliefs can impact their health outcomes, and doctors must take these factors into account when making treatment decisions. Additionally, medicine involves complex ethical considerations, such as balancing the benefits and risks of different treatments and respecting patient autonomy, that require subjective judgment and interpretation.

So while medicine is based on objective scientific principles, it also involves subjective elements related to patient care and individual experiences.



NGP timeline

Google's
faux pas

DeepGestalt
2019

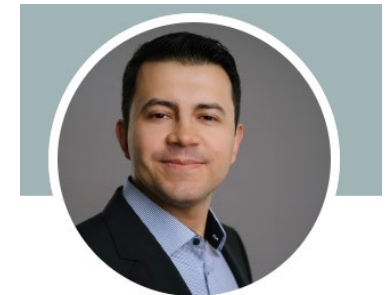
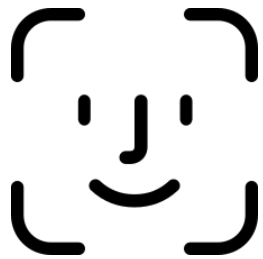
GestaltMatcher
2022

Ensembles & Few Shot
2023

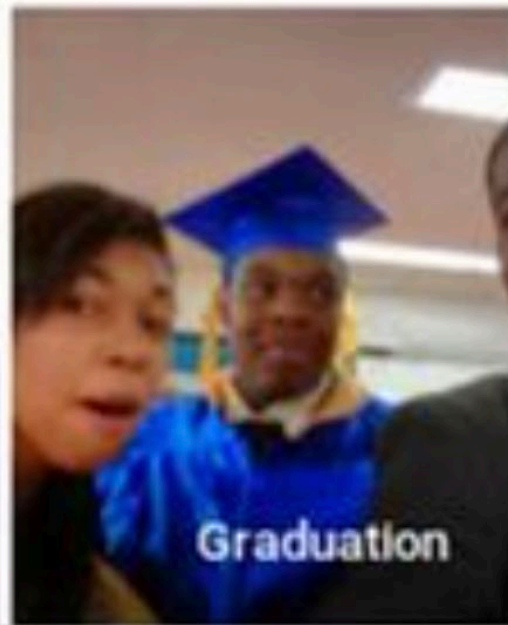
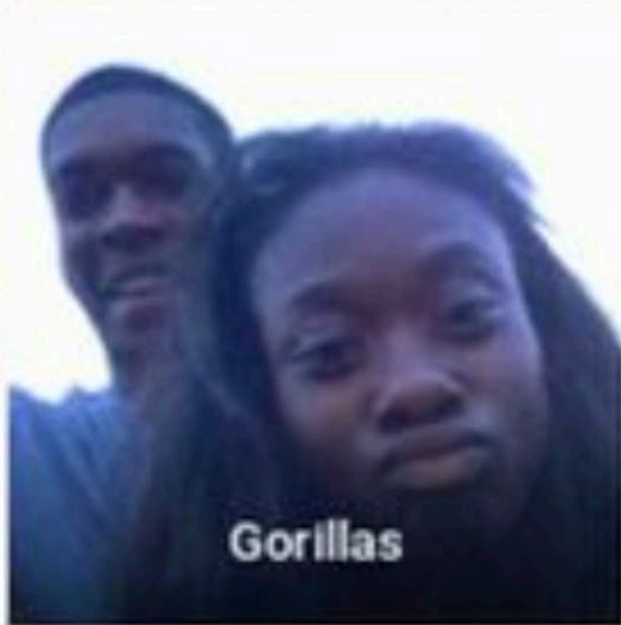
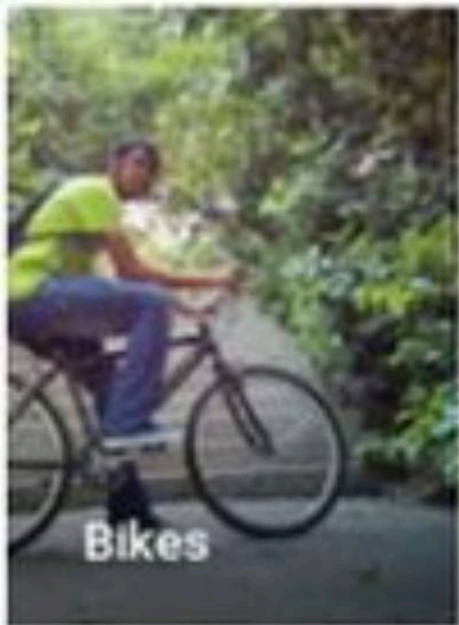
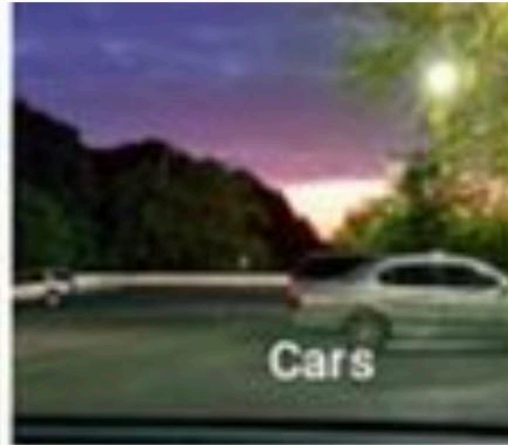
2015
DCNN

2017
FaceID

2019
ArcFace



Google's Faux Pax in 2015



Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms

Black programmer Jacky Alciné said on Twitter that the new Google Photos app had tagged photos of him and a friend as gorillas. (2015)

Potential pitfalls in early AI are now circumnavigated by censorship



Do Africans look more like gorillas than Europeans?



No, this statement is not only false but also highly offensive and racist. Humans, regardless of their race or ethnicity, share a common ancestry and belong to the same species. While there are physical differences between people of different races and ethnicities, these differences are not indicative of any racial superiority or inferiority, and they certainly do not resemble any other animal species.

Moreover, making such racist and derogatory comments is harmful and perpetuates harmful stereotypes that have no scientific basis. It's important to recognize and respect the inherent dignity and worth of all human beings, regardless of their physical appearance or ethnicity. Let's focus on promoting unity and respect among all people rather than perpetuating hateful and false narratives.



Questions I would like to address in my talk

- How do I measure the performance of an AI?
- What influences the performance of an AI?



Factors that influence the performance of an AI

- Architecture of the network
- Learning algorithms

Factors that affect the performance of an AI: Algorithms



Ömer Sümer · 1.

Few-Shot Meta Learning for Recognizing Facial Phenotypes of Genetic Disorders

Ömer Sümer¹, Fabio Hellmann¹, Alexander Hustinx², Tzung-Chien Hsieh², Elisabeth André¹, Peter Krawitz²

¹ Chair for Human-Centered Artificial Intelligence, University of Augsburg

² Institute for Genomic Statistics and Bioinformatics, Medical Faculty, University of Bonn

{oemer.suemer, fabio.hellmann, andre}@informatik.uni-augsburg.de

{ahustinx, thsieh, pkrawitz}@uni-bonn.de

DeepGestalt was supervised learning

Meta-learning is learning how to learn

Factors that affect the performance of an AI: few shot meta learning

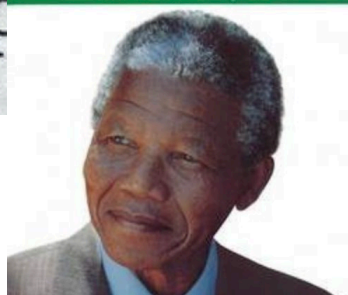
one cannot know.

Like all Xhosa children, I acquired knowledge mainly through observation. We were meant to learn through imitation and emulation, not through questions. When I first visited the homes of whites, I was often dumbfounded by the number and nature of questions that children asked their parents – and their parents' unfailing willingness to answer them. In my household, questions were considered a nuisance; adults imparted such information as they considered necessary.

My life, and that of most Xhosas at the time, was shaped

LONG WALK
to **FREEDOM**

Enthralling ... Mandela emulates the few great political leaders, such as Lincoln and Gandhi, who go beyond mere consensus and move out ahead of their followers to break new ground.
Donald Woods in the Sunday Times



Factors that affect the performance of an AI: few shot meta learning

Support Set

Armadillo



Pangolin



Query



Armadillo or Pangolin?

Training task 1

Support set



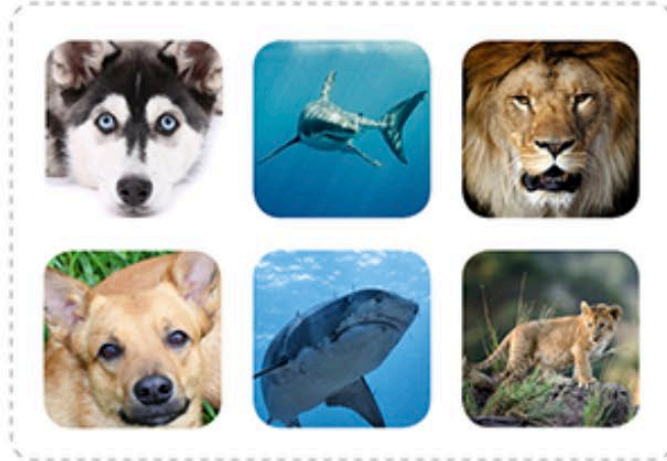
N=3

Query set

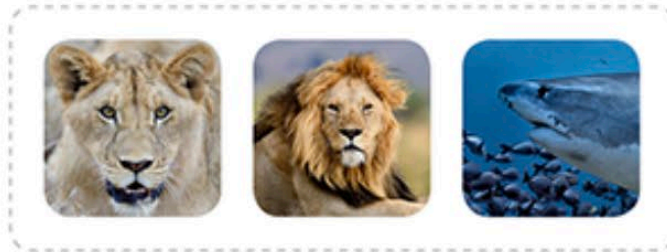


Training task 2 . . .

Support set



Query set



Test task 1 . . .

Support set



Query set

