Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen International Center for Information Technology

Institute for Genomic Statistics and Bioinformatics

ukb universitäts klinikumbonn

Quantitative Comparison of Deep Learning Classifiers and Human Attention in Assessing Rare Disorders

March 5, 2024

Anna Rose Johny

Advisors

Prof. Dr.-Ing. Sebastian Houben, Prof. Dr. med. Dipl. Phys. Peter Krawitz, Dr. rer. nat. Tzung-Chien Hsieh

Introduction

- Around 6% of the overall population is affected by genetic conditions.
- The rarity of some genetic conditions makes some diagnoses be missed if a clinician is not experienced enough with them.
- The deep learning classifiers can predict the syndromes depending on the learned facial features.
- We can use Explainable AI (XAI) methods to highlight those regions in a face that a model would have chosen to give the prediction.
- To compare the predictions of XAI, we can use heat maps obtained from an eye-tracker that is used with similar images.

Motivation

- **Rarity of genetic conditions** - makes diagnosis difficult for experienced clinicians
- **Black box between model predictions** - lack explanations
- **Check comparability** - Comparison of clinicians observations to the explanations by Artificial Intelligence (AI) model

Challenges and difficulties

- Lower performance of the classifier
- Smaller dataset and class imbalance
- Lack of methods to validating outputs

Problem statement

RQ1. What are the image saliency methods available to identify the Area Of Interest (AOI) in a given image to make predictions?

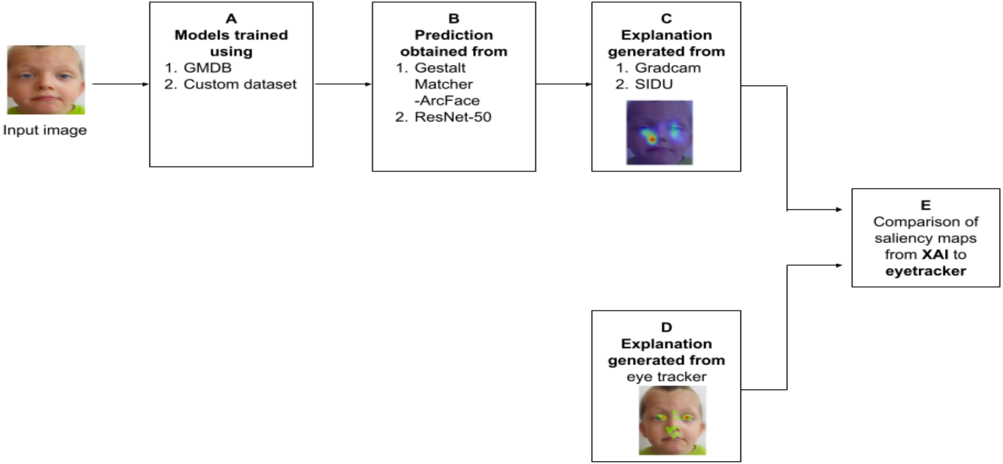
RQ2. What are the important AOI in a face to make predictions related to genetic disorders?

RQ3. Do the saliency maps of clinicians and eye-tracker saliency maps align?

RQ4. Do the clinicians and non-clinicians look at the same facial features?

Methodology

Proposed pipeline



Step A - Datasets

- GestaltMatcher DataBase (GMDB)
- Custom dataset

Figure 1: Few example of images in GMDB.

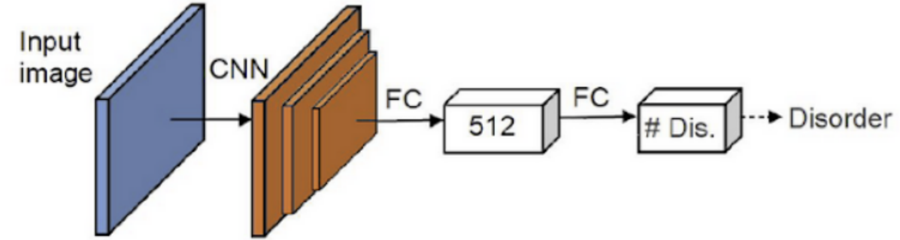
Step B - Rare disorder identification and classification methods

- GestaltMatcher-Arc[1]
- ResNet-50[2]

[1] Hustinx et al., “Improving Deep Facial Phenotyping for Ultra-rare Disorder Verification Using Model Ensembles”.
[2] Mandal, Okeukwu, and Theis, “Masked Face Recognition using ResNet-50”.

GestaltMatcher-Arc

- Extension of **DeepGestalt** approach
- Uses same architecture and face dataset (CASIA) as a base for **transfer learning**
- Each image encoded into **320**-dimensional representation vector
- Representation vectors spanned a **Clinical Face Phenotype Space (CFPS)**
- In the CFPS, patients with **rare disorders** can be matched to other similar patients
- **Clustering analysis** performed to analyze the similarity among different disorders



ResNet-50

- Is a 50 layer convolutional neural network with four main parts
  - Convolutional layers
  - Identity block
  - Convolutional block
  - Fully connected layers

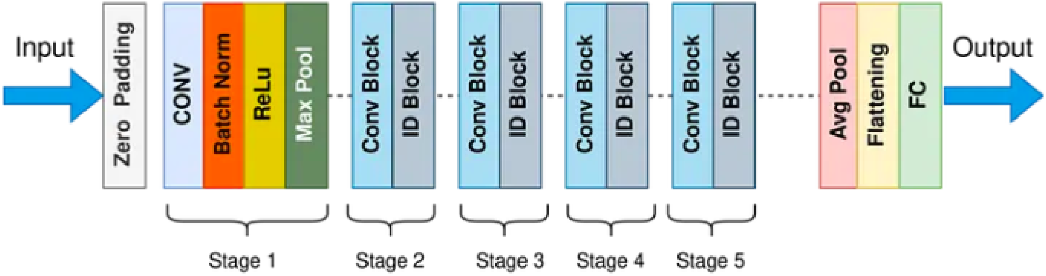


Figure 3: ResNet50 architecture

Experimental setup

Network architectures

Hyperparameters	Values	
	GestaltMatcher-Arc	ResNet-50
Input image size	100×100	100×100
Learning rate	0.001	0.01
Loss Function	Cross-entropy	Categorical cross-entropy
Epochs	50	50
Batch size	32	32

Table 1: Details of training of GestaltMatcher-Arc and ResNet-50.

Step C - Explanation methods

- Gradient-weighted Class Activation Mapping (Grad-CAM)[3]
- Similarity Difference and Uniqueness (SIDU)[4]

[3] Selvaraju et al., **Grad-CAM: Why did you say that?**

[4] Muddamsetty et al., “Visual explanation of black-box model: Similarity Difference and Uniqueness (SIDU) method”.

Grad-CAM

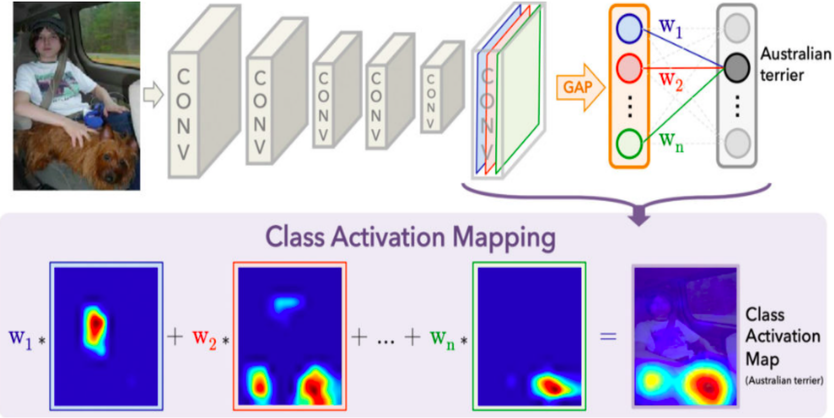


Figure 4: CAM architecture. Image source[5]

[5] Zhou et al., **Learning Deep Features for Discriminative Localization.**

SIDU

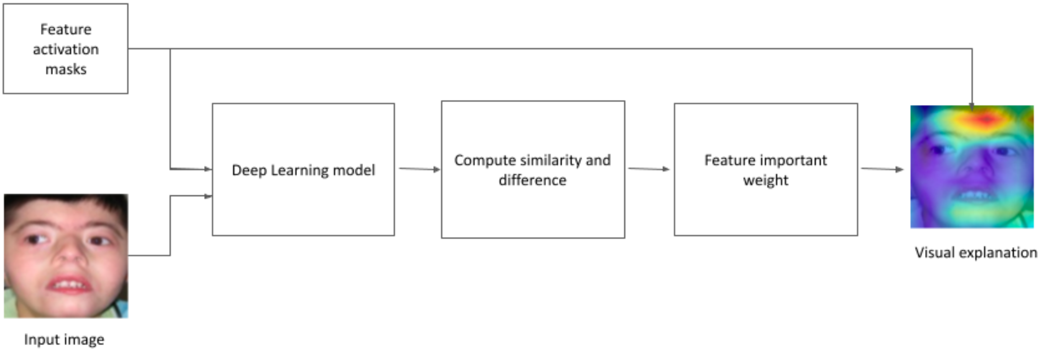


Figure 5: Overall architecture of SIDU in our case.

Implementation details

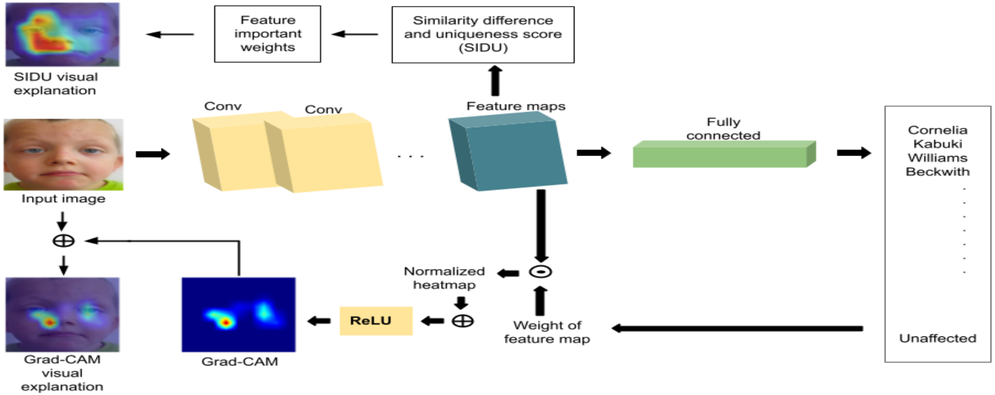


Figure 6: Grad-CAM and SIDU for GestaltMatcher-Arc and ResNet-50 in our research.

Step D - Eye-tracking experiment

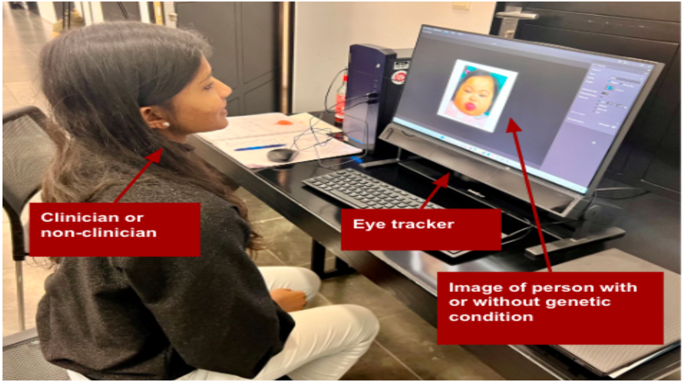


Figure 7: Eye-tracking experiment performed at IGSB Bonn on clinicians and non-clinicians in association with NIH.

Evaluation and results

- Evaluation
  - Experiment 1: Explanation generated by visual explanation methods
  - Experiment 2: Human-grounded evaluation
- Comparison metrics - Evaluation of eye-tracking to XAI methods

Experiment 1: case 1

Explanation generated by visual explanation methods

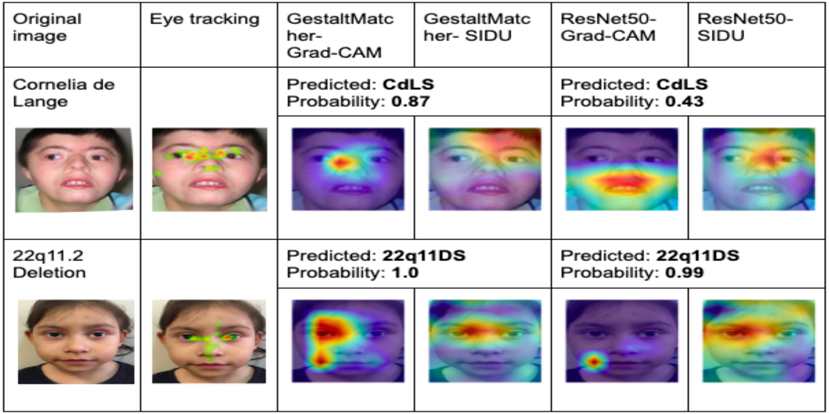


Figure 8: Heat maps of correctly predicted

Experiment 1: case 2

Explanation generated by visual explanation methods

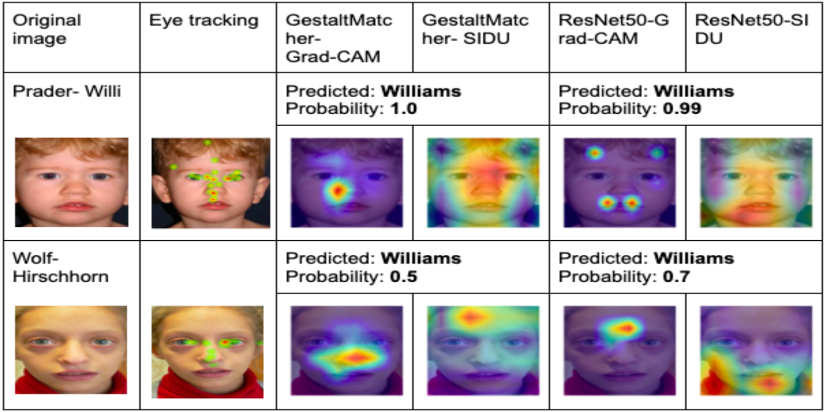


Figure 9: Heat maps of wrongly predicted

Experiment 2

Evaluations from clinicians based on the AOI in face

Condition	Facial features				
	Forehead	Eyes/Periorbital	Nose	Mouth	Ears
Cornelia de Lange		+	+	+	+
22q11.2 deletion			+		
Down		+	+	+	+
Noonan		+			+
Prader-Willi		+			
Wolf-Hirschhorn	+	+	+		
Williams-Beuren		+	+	+	

Table 2: Syndrome-specific facial features identified by a clinician. + sign shows the feature important in a face.

Evaluation: comparison metrics

Syndrome	GestaltMatcher-Grad-CAM	GestaltMatcher-SIDU	ResNet50- Grad-CAM	ResNet50- SIDU
Cornelia de Lange	0.76	0.56	0.30	0.89
22q11.2 deletion	0.78	0.86	0.20	0.67
Down	0.34	0.20	0.34	0.43
Noonan	0.45	0.36	0.56	0.67
Prader-Willi	0.67	0.89	0.56	0.78
Wolf-Hirschhorn	0.56	0.76	0.65	0.10
Williams-Beuren	0.45	0.65	0.78	0.34
Average	**0.58**	**0.62**	**0.48**	**0.55**

Table 3: Comparison table for IoU score. The IoU values are calculated between the eye-tracker heat maps and different XAI heat maps considered.

Revisiting the research questions...

RQ1. What are the image saliency methods available to identify the AOI in a given image to make predictions?
SIDU approach was able to emphasize the crucial regions more accurately than the approach Grad-CAM.
SIDU approach highlights the regions spread over some portion of the face, whereas, Grad-CAM approach highlights small regions in the face.

RQ2. What are the important AOI in a face to make predictions related to genetic disorders?
There is **no genetic disorder specific AOI** identified by the visual explanation method.
SIDU based approach tends to highlight AOI in the images more specifically than the Grad-CAM based approach.

Revisiting the research questions...

RQ3. Do the saliency maps of clinicians and eye-tracker saliency maps align?
There is a huge difference in the features observed by a model and a human while looking at features in an image.

RQ 4 Do the clinicians and non-clinicians look at the same facial features?
Clinicians tend to look at other features **excluding the obvious ones**.
Whereas **non-clinicians** always tend to look at the **eye, nose,** and **mouth** region.

Contributions

- Conducted literature review for rare-disorder identification and classification and also explanation methods.
- Implemented selected classification methods and explanations generated.
- Compared saliency maps from XAI to eye-tracker maps.
- Presented the ideas as a poster at Arbeitsgemeinschaft für Gen-Diagnostik e.V. (AGDev).
- Contributed to recently published paper *"Comparison of clinical geneticist and computer visual attention in assessing genetic conditions"*in PLOS genetics.

Future work

- Improvement to dataset
- Extending eye-tracking survey
- Find approach to quantify results

Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences
universitäts klinikumbonn
Institute for Genomic Statistics and Bioinformatics

Quantitative Comparison of Deep Learning Classifiers and Human Attention in Assessing Rare Disorders

Anna Rose Johny

Motivation and background

- Perform eye tracking analysis and to understand how clinicians diagnose patients.
- Comparing with explainable AI to understand difference between human and AI.

Datasets used

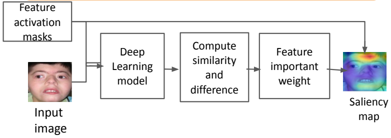
- **GestaltMatcher DataBase** (7 syndromes)
- **Custom dataset** (11 classes -10 syndromes, 1 unaffected

Deep Learning classifiers used

Deep Learning classifiers are used to classify the images into various syndromes. Deep Learning classifiers used in this work,

- **GestaltMatcher ArcFace**
- **ResNet50**

Explainable AI Methods



The regions responsible for prediction are highlighted using explainable AI method (**GradCam**, **SIDU**). The figure above represents the saliency map generation using SIDU.

Eye-tracking data using Tobii-Pro

A Tobii-eye tracker device is used for collecting data from clinicians and non-clinicians provided with syndromic and non-syndromic faces. Heatmaps are generated from the eyetracker based on the time spent on each facial feature.

Comparison to eye-tracking data

The eye tracking heatmaps are compared with the explainable AI heatmaps using KL-Divergence and got 0.31 for ResNet gradcam, 0.51 for ResNet SIDU approach.

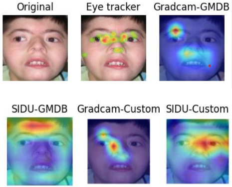


Figure 2: Comparison of eye tracker heatmaps with Gradcam and SIDU.

Conclusion and results

- AI and human look at different features in an image
- Difficult to compare AI and human for different syndromes

References
[1] Satya M. Muddamsetty, Visual explanation of black-box model: Similarity Difference and Uniqueness (SIDU) method, Pattern Recognition,2022

Thank you