



---

<sup>b</sup>  
**UNIVERSITÄT  
BERN**

Universität Bern  
Institut für Psychologie  
Lehrstuhl Kognition, Wahrnehmung und Methodenlehre

**Forschungsatelier**  
Frühlingssemester 2018

# Decrypting imagination

Was Augenbewegungen über wahrgenommene und  
vorgestellte Inhalte verraten

Mirko Bristle  
Betreut von L. Gurtner

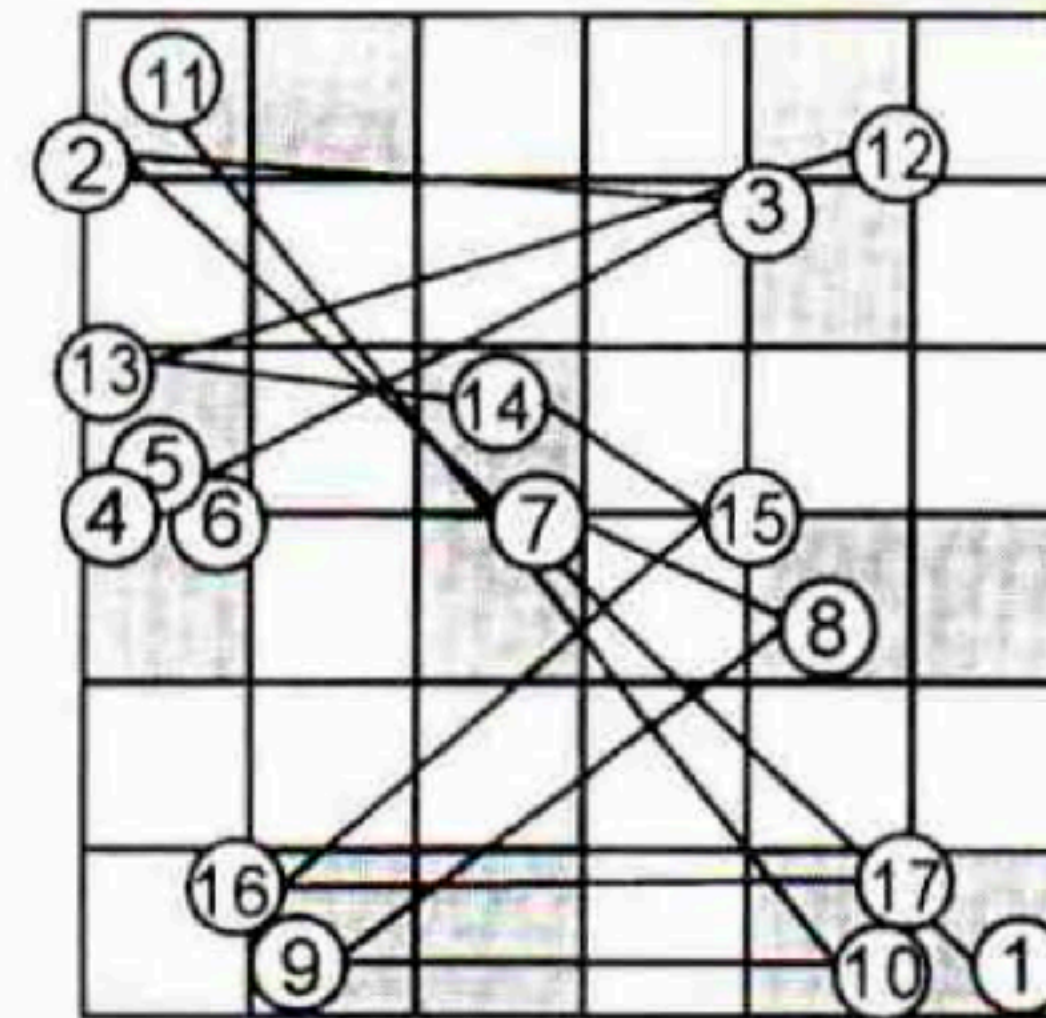
# Überblick

1. Theoretischer Hintergrund
2. Fragestellung
3. Forschungsdesign
4. Analyse
5. Ausblick

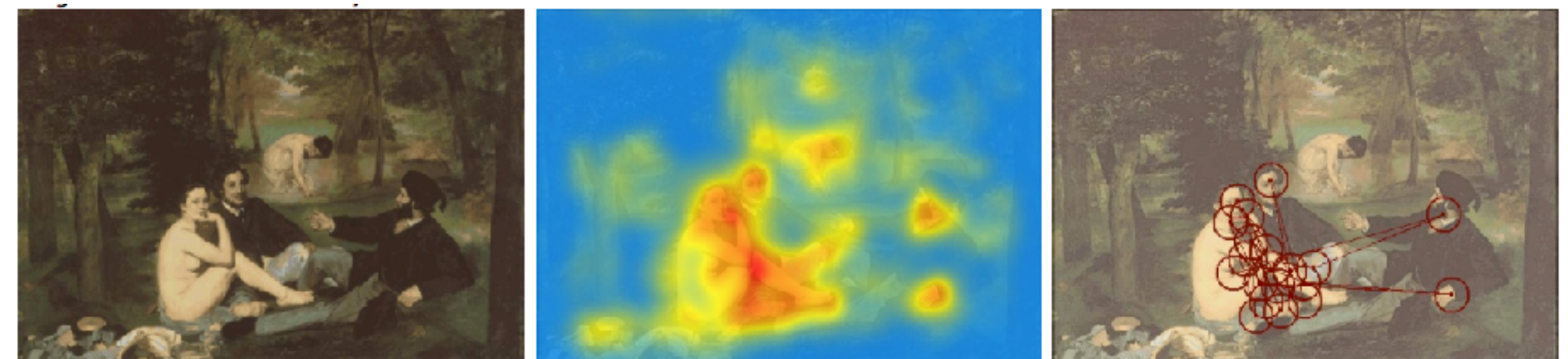
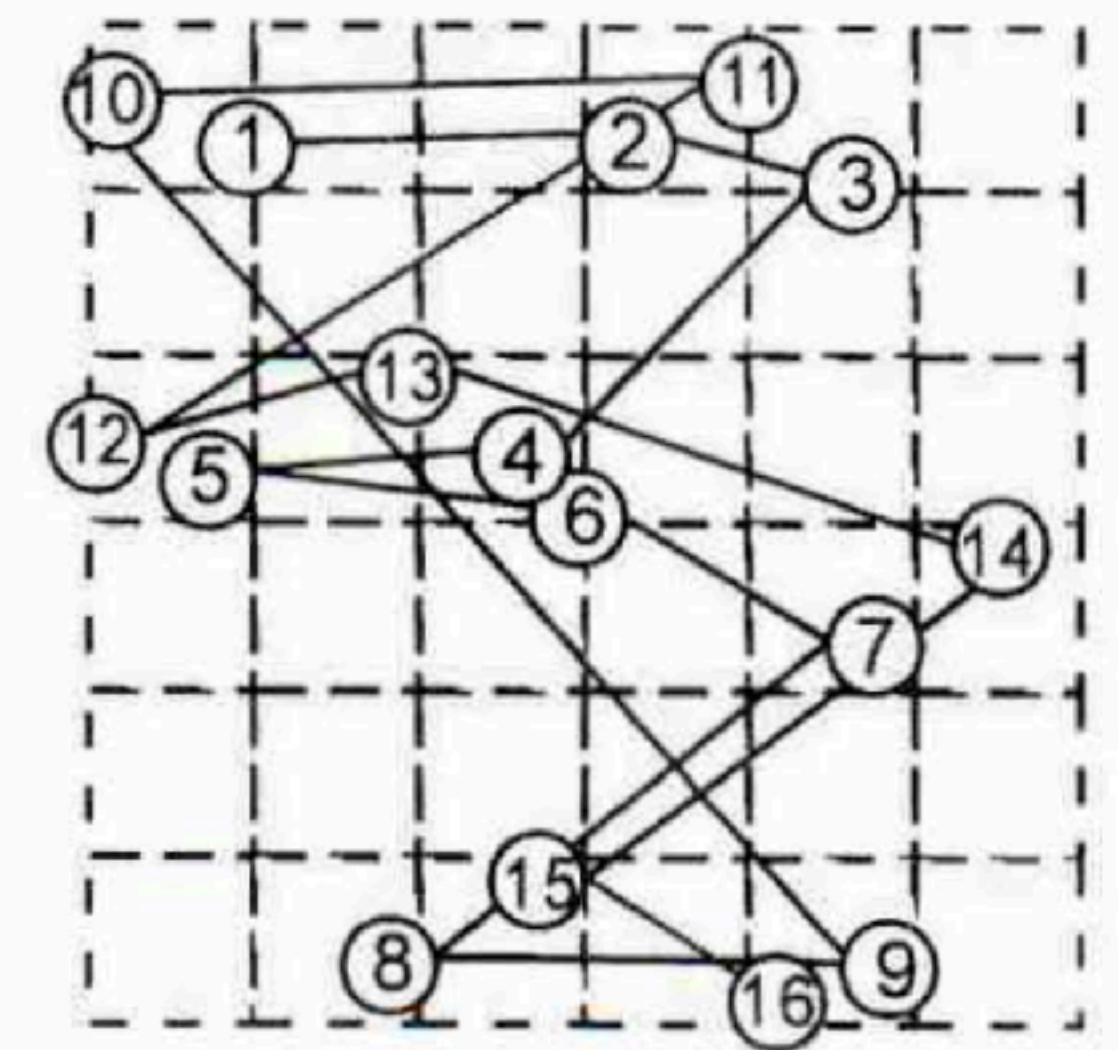
## Theoretischer Hintergrund

- > The Scanpath Theory
  - Komplex
  - Nicht-Zufällig
  - Sequenzen von wiederholten Fixationen
- > Spatial Model
- > Augenbewegungen haben eine funktionale Rolle bei der Encodierung und beim Retrieval
- > Bildliche Abruf-Aufgaben
  - Es wird nicht nur kurzzeitig, sondern auch nach einer Woche an die Stelle des Stimulus zurück geblickt.
  - ScanPath Analysen zeigen, dass das Muster des vorgestellten Bildes dessen Inhalt folgt.

Betrachtung



Vorstellung



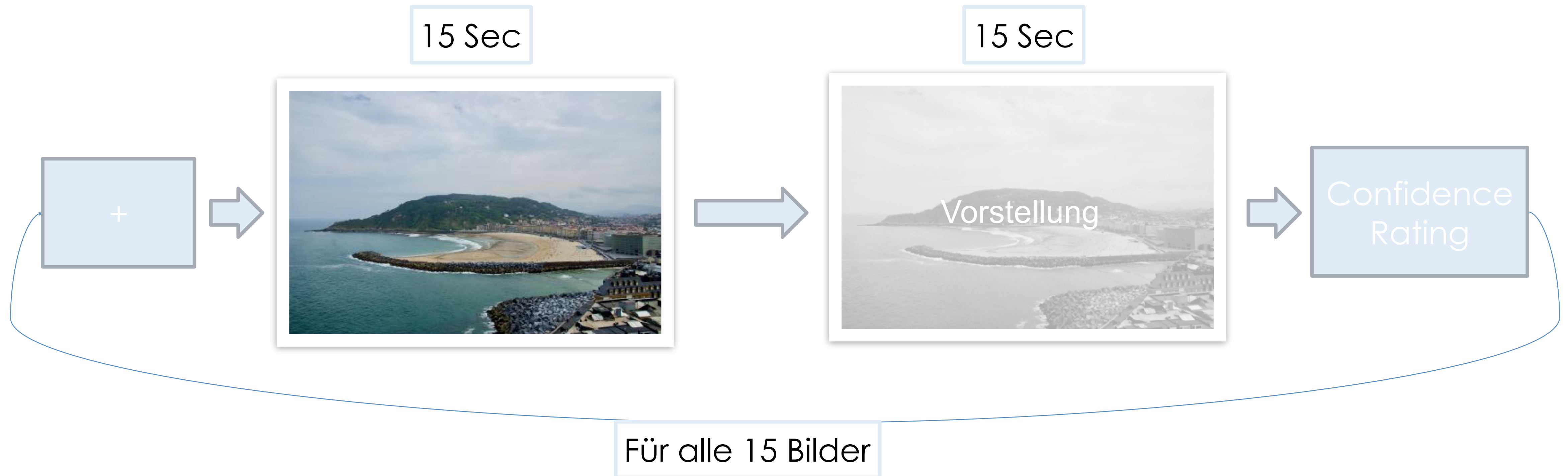
## Fragestellung

1. Kann auf Grund der in der Augenbewegung liegenden Information ein Bild vorher gesagt werden?
  - Kann das Bild für betrachtete Bilder und für vorgestellte Bilder gleich gut vorhergesagt werden?
2. Kann auf Grund der Information der betrachteten Bilder ein vorgestelltes Bild vorhergesagt werden?
3. Welche Feature tragen am meisten zur korrekten Klassifikation bei?



## Forschungsdesign

- > 5 Probanden
- > 15 Bilder geordnet in 3 Kategorien (Kunst, Landschaft, Gesichter)



# Auswertung

$u^b$

<sup>b</sup>  
UNIVERSITÄT  
BERN

Datenerhebung

Aufbereitung

Modellierung

Statistik



Preprocessing

Feature extraction

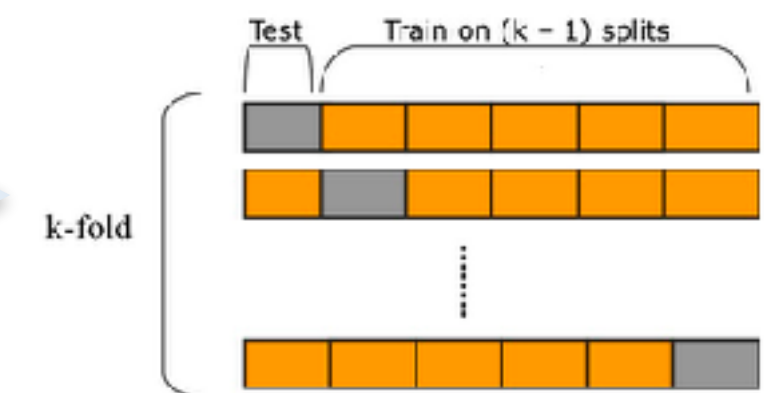
X	y
n x m	Face
n x m	Art
n x m	Landscape
:	:

SVM

$$\begin{aligned} \min_{\gamma, w, b} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i, \quad i = 1, \dots, m \\ & \xi_i \geq 0, \quad i = 1, \dots, m. \end{aligned}$$

Summary - Data

CrossValidation



Feature extraction

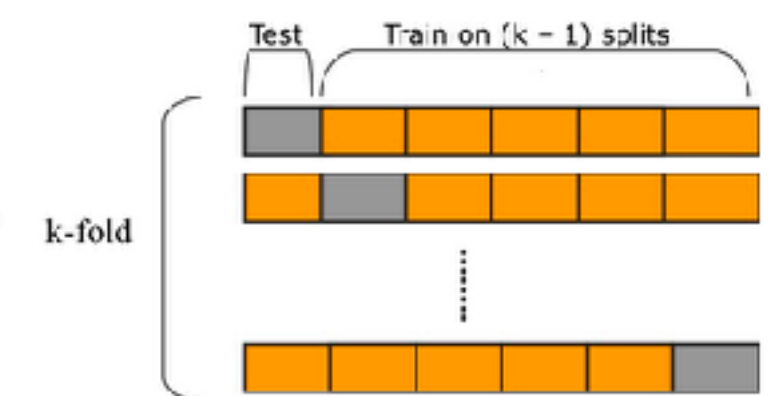
X	y
n x m	Face
n x m	Art
n x m	Landscape
:	:

SVM

$$\begin{aligned} \min_{\gamma, w, b} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i, \quad i = 1, \dots, m \\ & \xi_i \geq 0, \quad i = 1, \dots, m. \end{aligned}$$

Fixation - Map Data

CrossValidation



Kontrollexperiment

Bayesian multilevel model

# Auswertung

Modellierung

Statistik

Post-Hoc Analyse

Interpretation

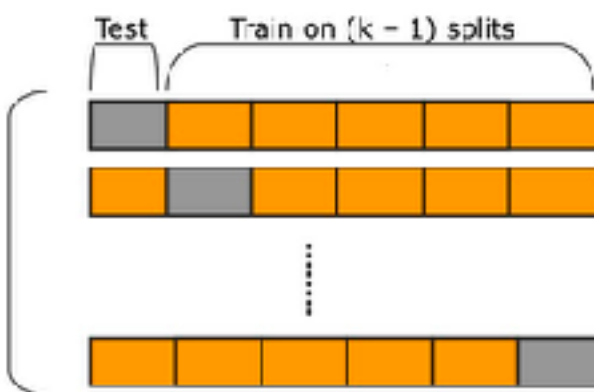
Summary - Data

SVM

$$\begin{aligned} \min_{\gamma, w, b} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i, \quad i = 1, \dots, m \\ & \xi_i \geq 0, \quad i = 1, \dots, m. \end{aligned}$$

k-fold

CrossValidation



Feature Selection  
Feature Ranking

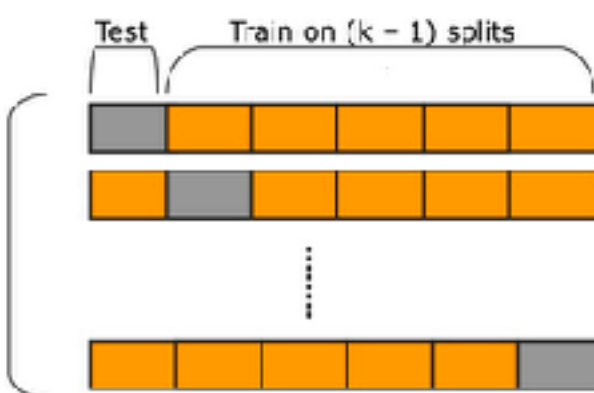
Fixation - Map Data

SVM

$$\begin{aligned} \min_{\gamma, w, b} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i, \quad i = 1, \dots, m \\ & \xi_i \geq 0, \quad i = 1, \dots, m. \end{aligned}$$

k-fold

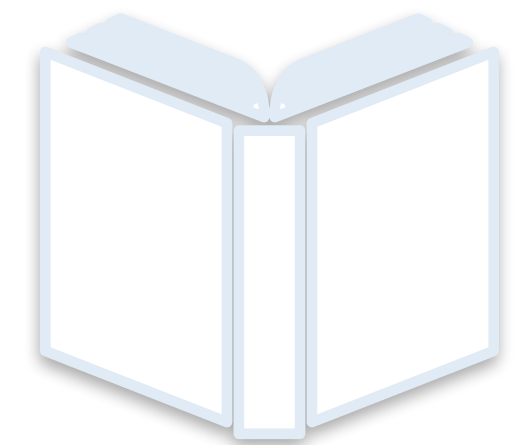
CrossValidation



Feature Selection  
Feature Ranking

Vergleich der Güte

Bayesian multilevel model



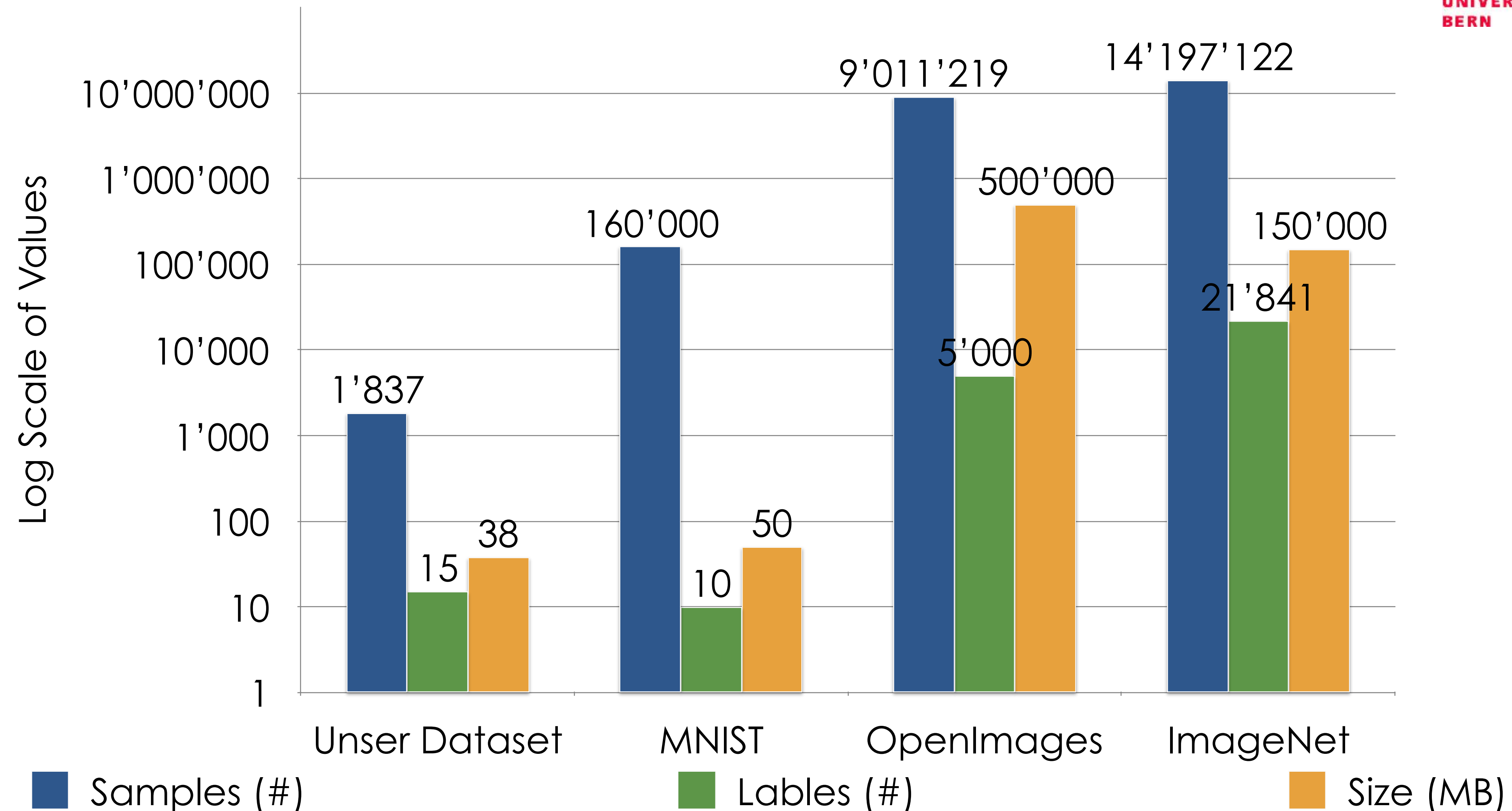
## Feature extraction

- > **Summary Data**
- > **Raw Daten:** [X, Y, Pupillen Dilatation] mit 2000 Hz für 15 Sec x 30 Zyklen
- > **Recurrent Quantification Analysis:** Metrik für zeitlich zusammenhängende Fixationen

Welche Methode ermöglicht eine möglichst gute Vorhersage der Bilder ?

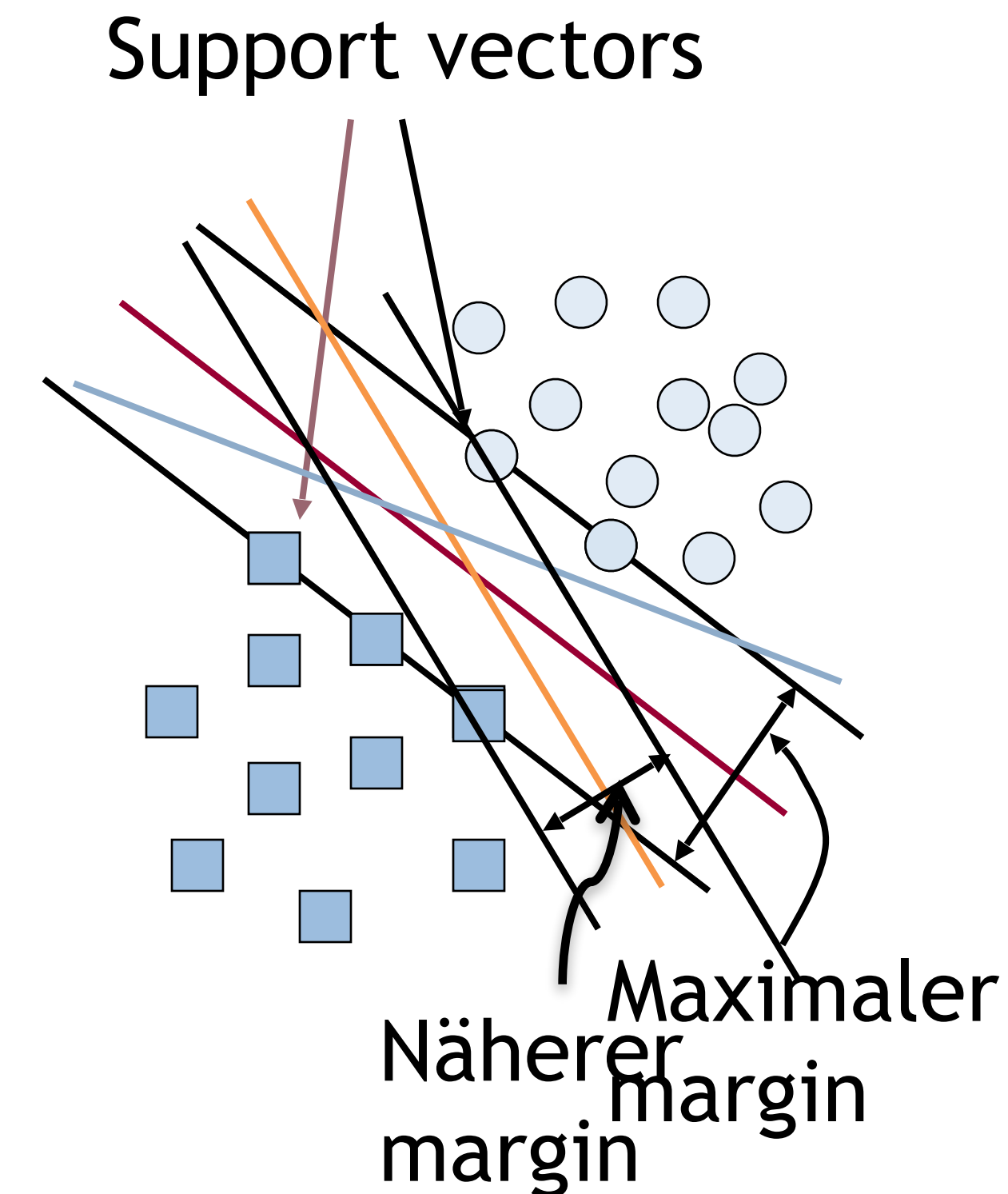


## Methoden Wahl: Ist unser Dataset Big Data?



## Support Vector Machine (SVM)

- > SVMs maximieren den Abstand zu der trennenden Hyperebene.
  - A.k.a. large margin classifiers
- > The decision function ist vollkommen spezifiziert durch ein Subset von training Samples, *den Support Vektoren*.
- > Bis ca. 2012 die beste Methode für Text Klassifikation\*

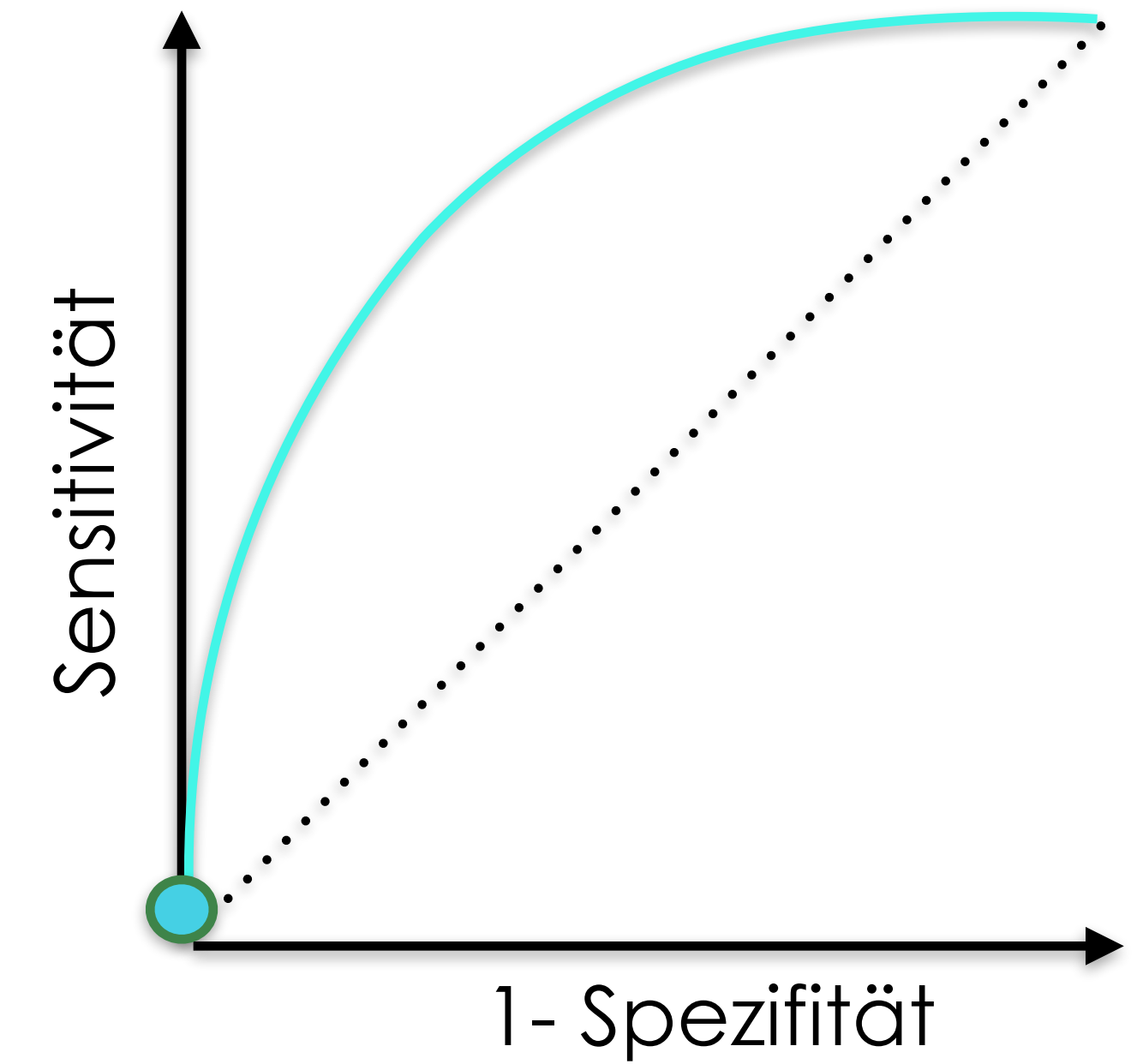
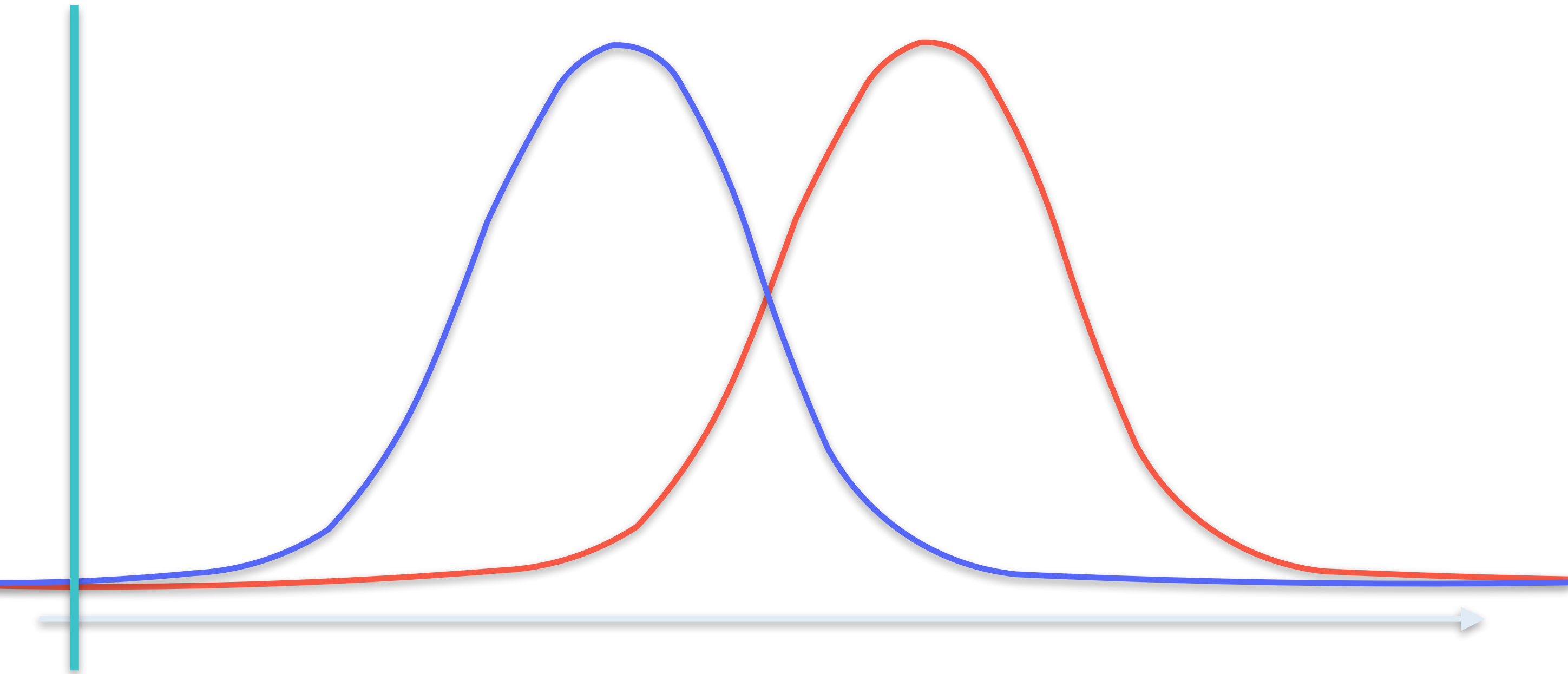


\*aber andere Methoden sind ähnlich gut!

# Receiver Operating Curve

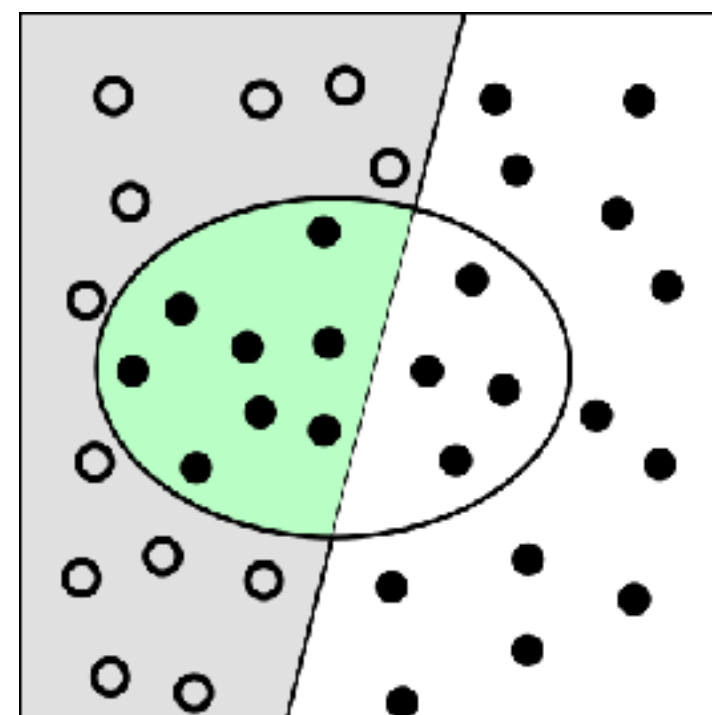
$u^b$

<sup>b</sup>  
UNIVERSITÄT  
BERN



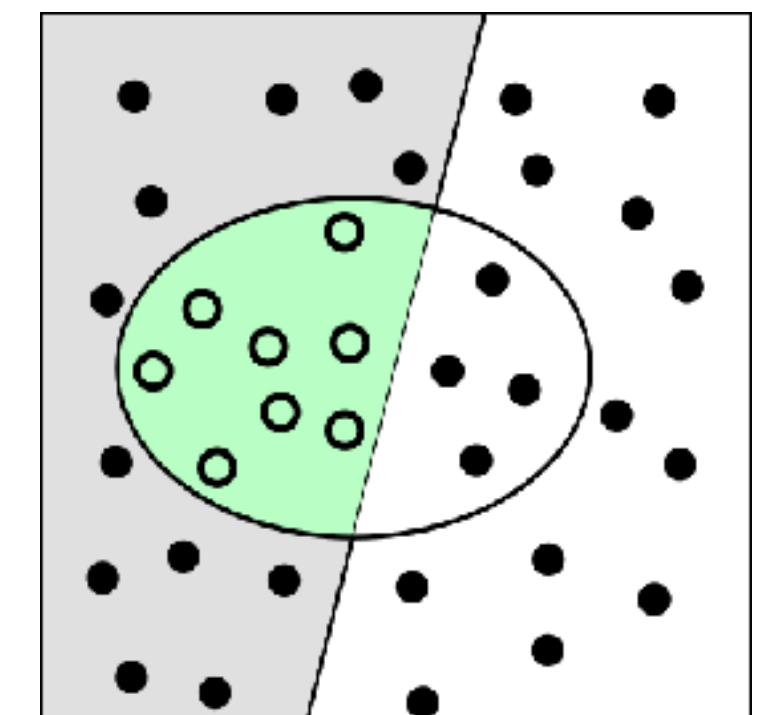
Spezifität

$$P(\text{negatives Testergebnis} | \text{tatsächlich krank}) = \frac{f_n}{r_p + f_n}$$



Sensitivität

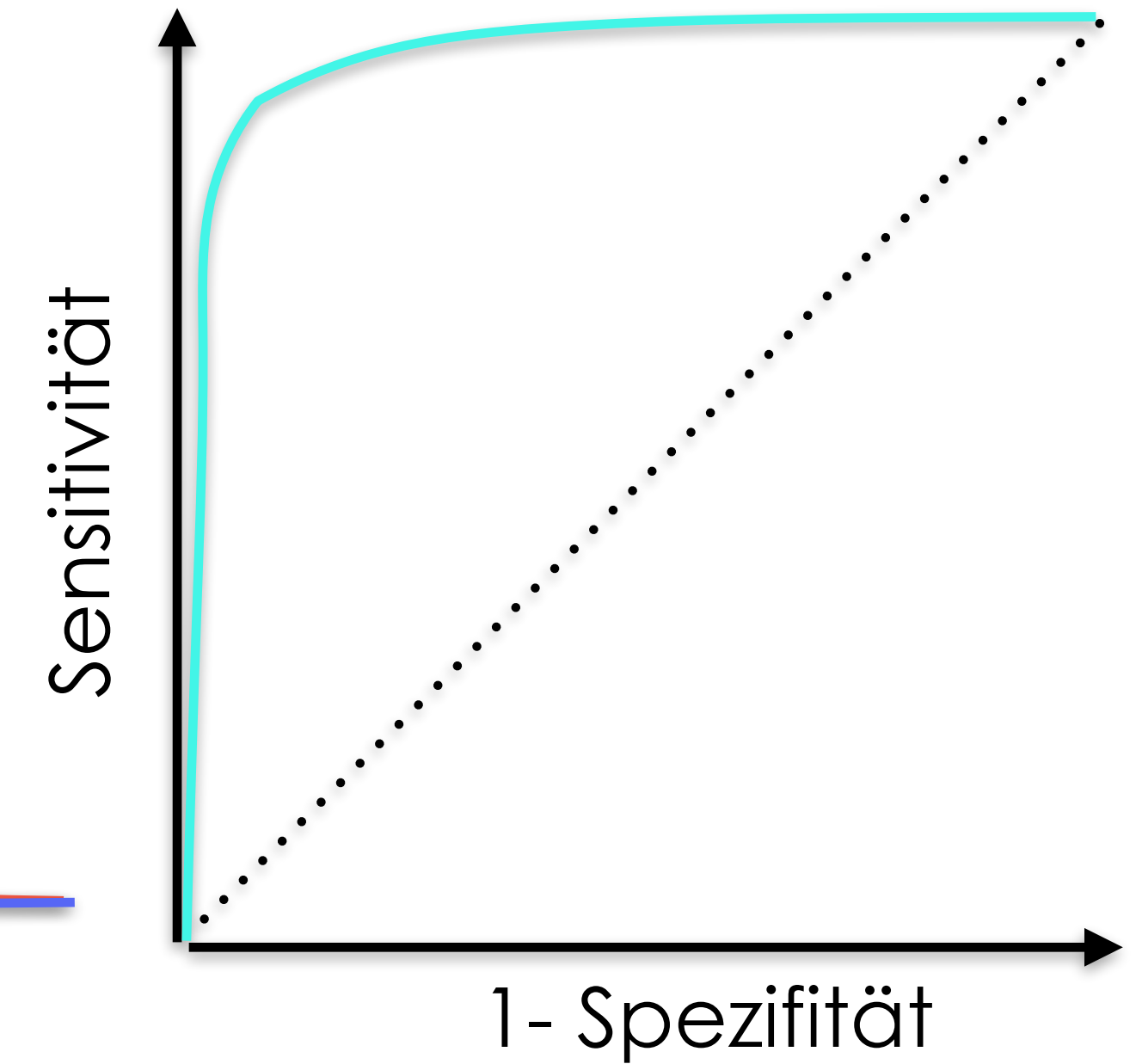
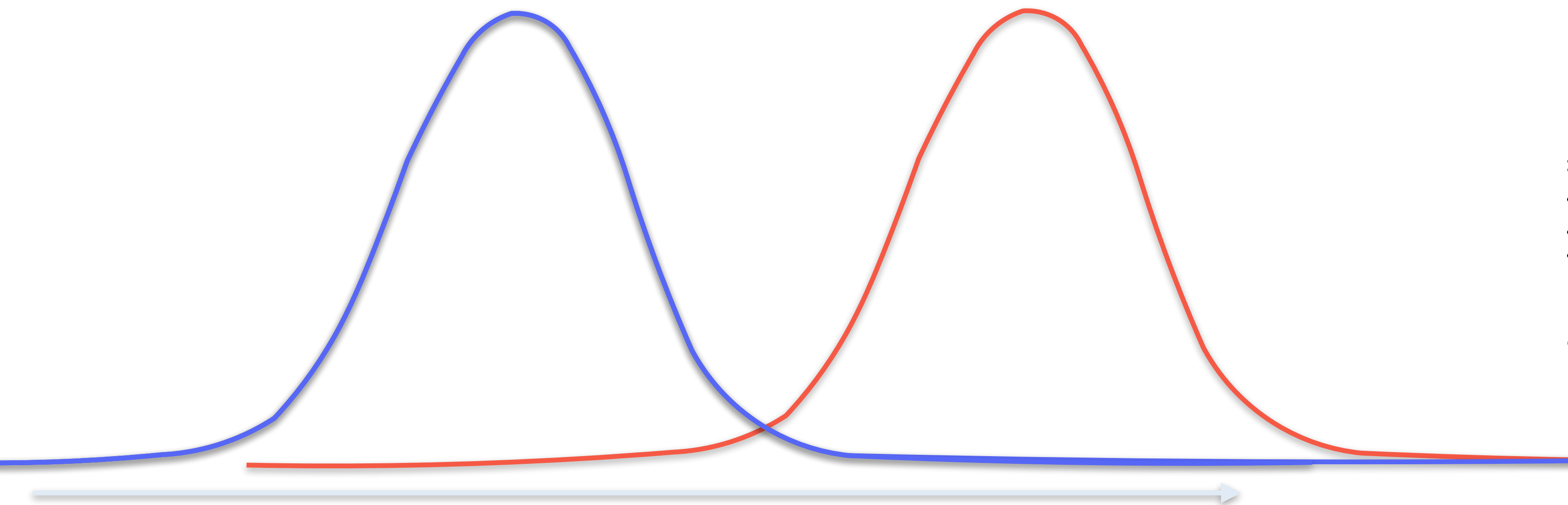
$$P(\text{positives Testergebnis} | \text{tatsächlich krank}) = \frac{r_p}{r_p + f_n}$$



# Receiver Operating Curve

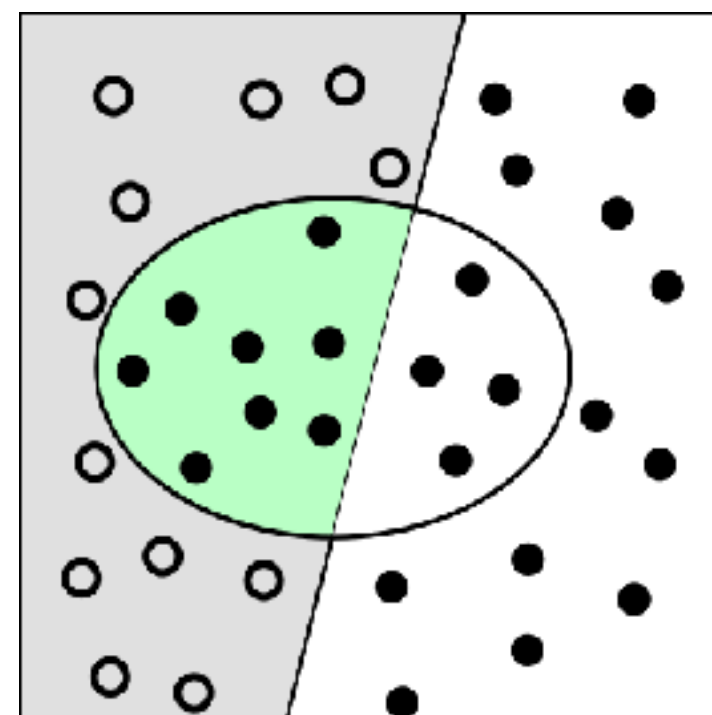
$u^b$

<sup>b</sup>  
UNIVERSITÄT  
BERN



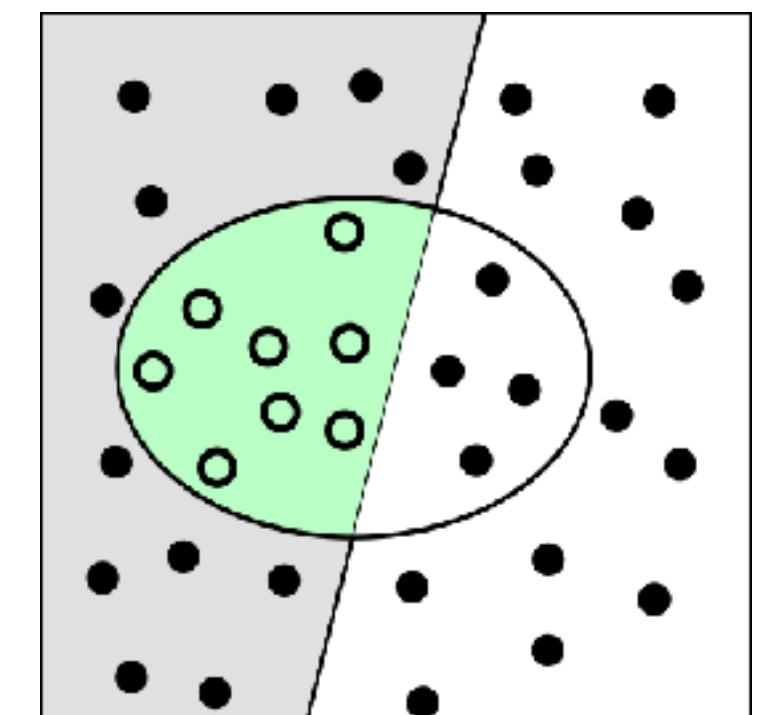
Spezifität

$$P(\text{negatives Testergebnis} | \text{tatsächlich krank}) = \frac{f_n}{r_p + f_n}$$



Sensitivität

$$P(\text{positives Testergebnis} | \text{tatsächlich krank}) = \frac{r_p}{r_p + f_n}$$

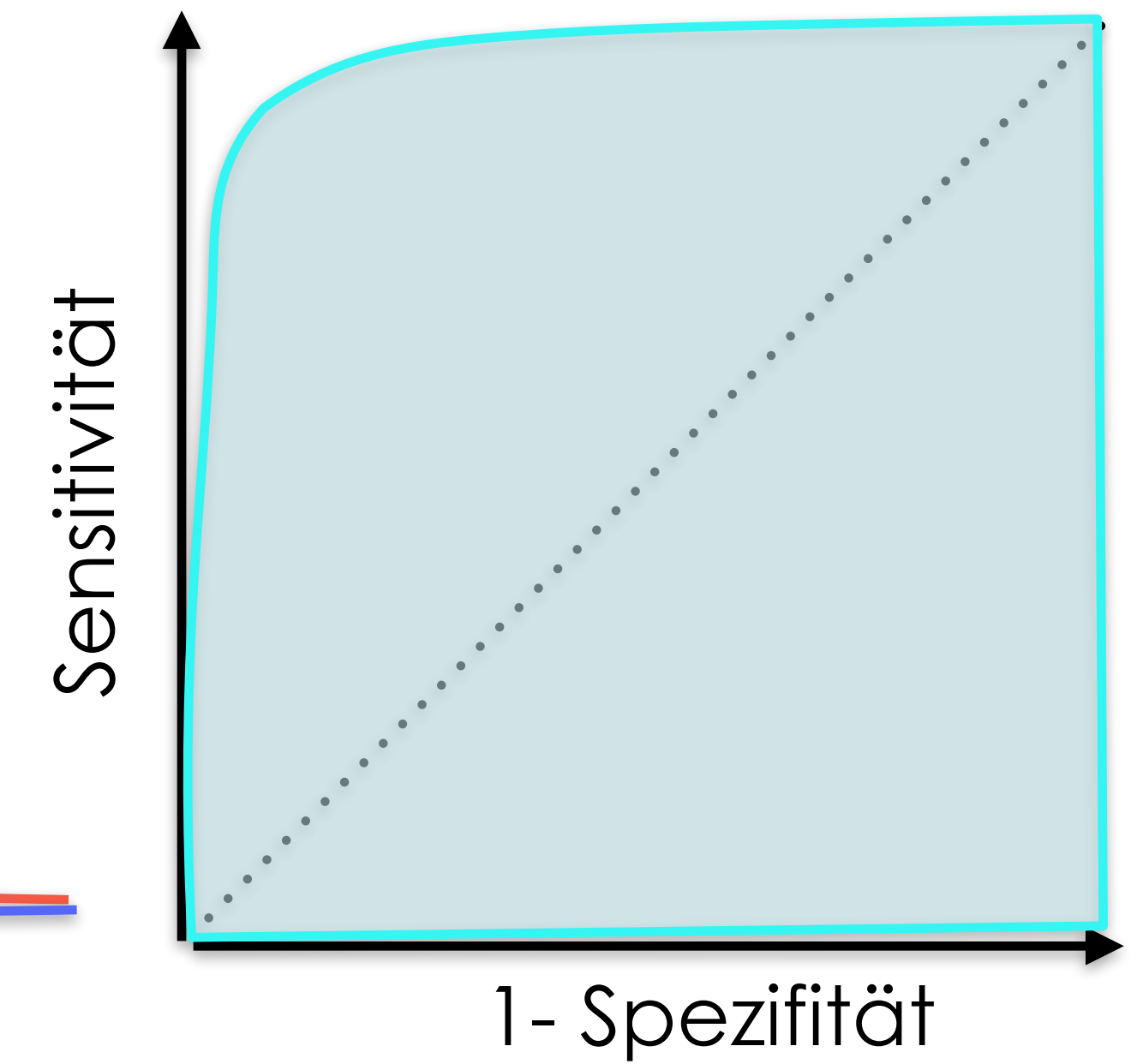
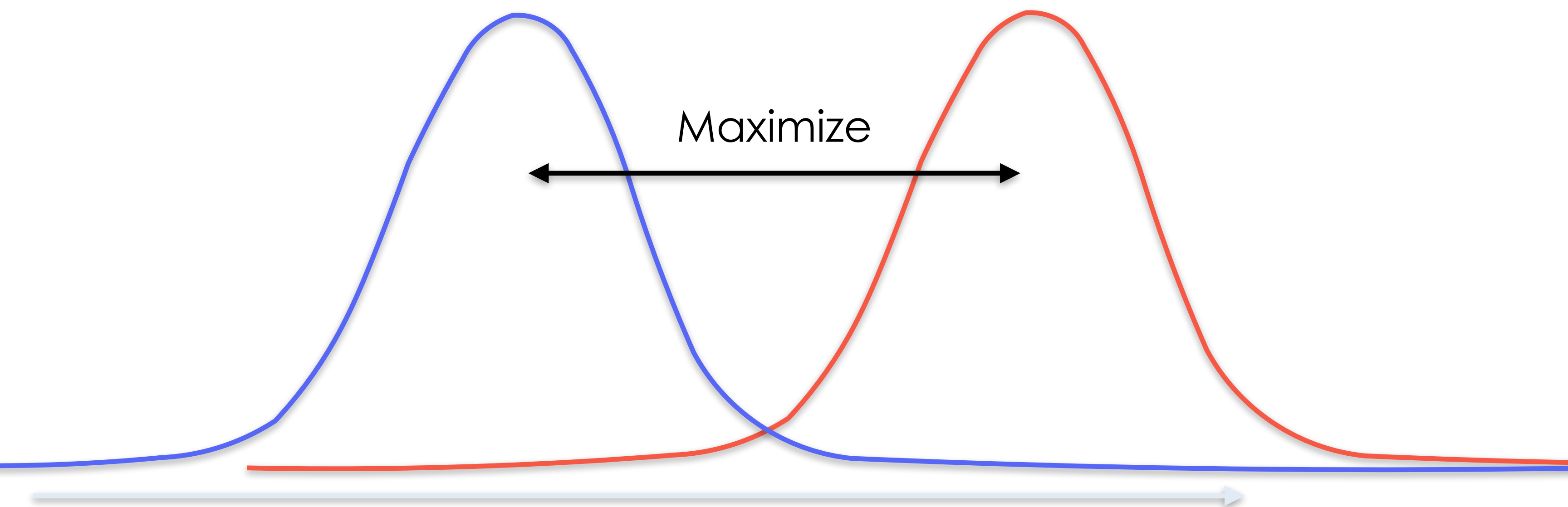




# Receiver Operating Curve

$u^b$

<sup>b</sup>  
UNIVERSITÄT  
BERN



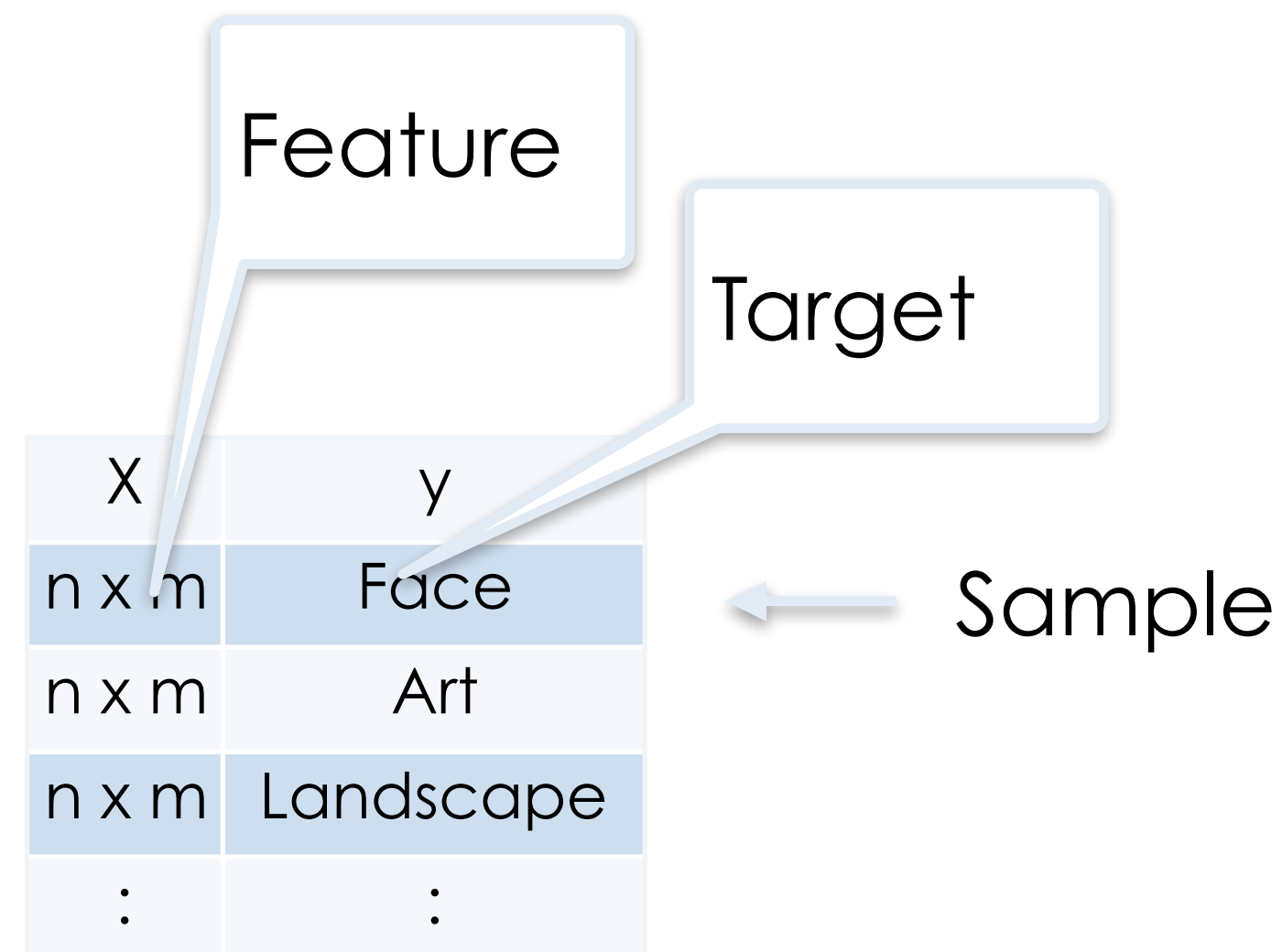
$$0.5 < \text{AUC} < 1$$

# SUMMARY: DATA

Sample = Perception / Imagination von einem Bild

**Problem: Anzahl Fixationen stimmt nicht überein**

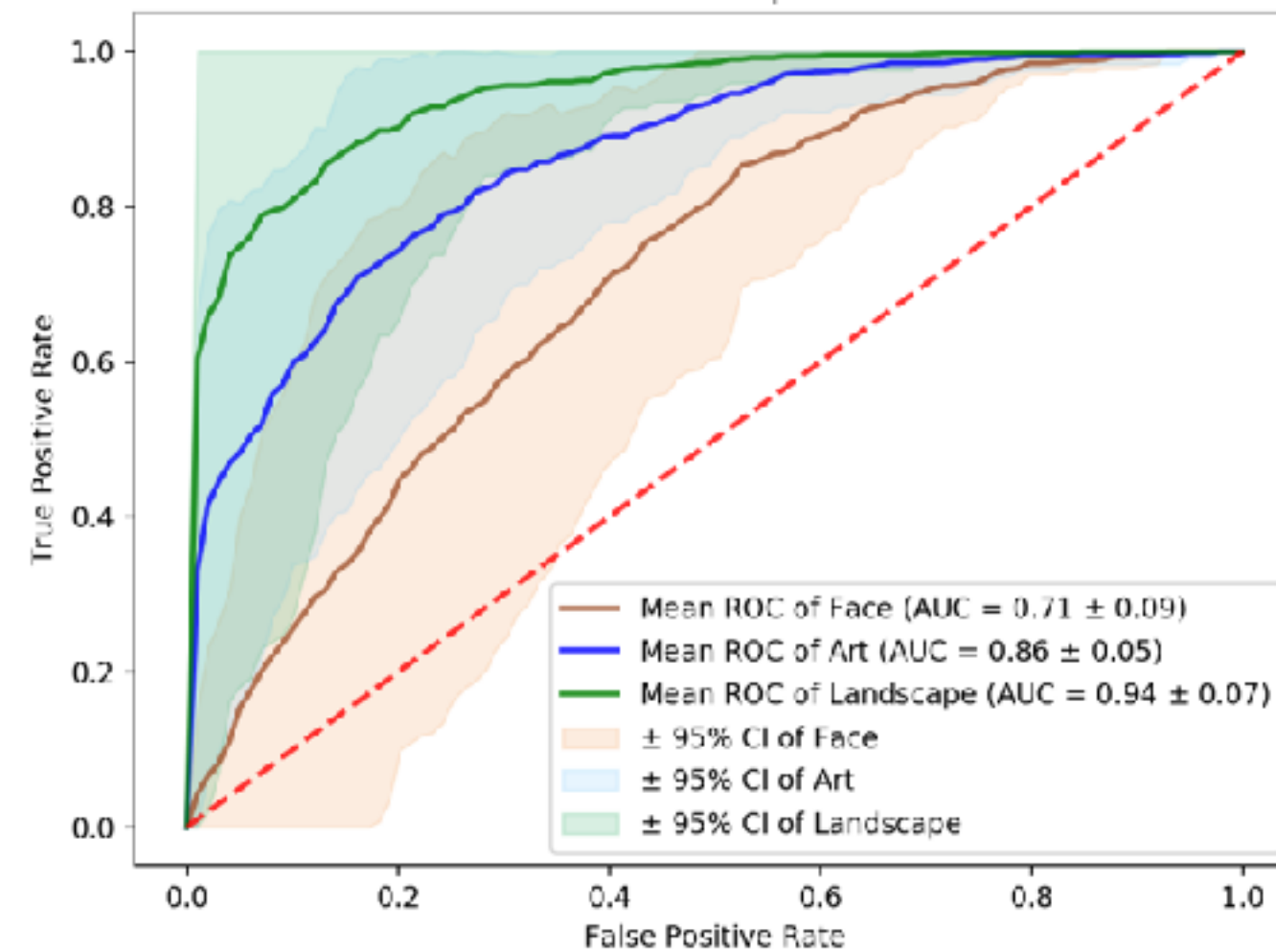
Features of 1 Summary Sample = < mean X<sub>r</sub>,  
mean X<sub>l</sub>,  
mean Y<sub>r</sub>,  
mean Y<sub>l</sub>,  
mean Dauer,  
mean Pupille,  
# Blinks,  
# Fixation >



$$X = 5 \text{ (Vpn)} * 5 \text{ (Session)} * 5 \text{ (Blöcke)} * 15 \text{ (Bilder)} - 38 \text{ (Missing)} = 1837 \text{ features}$$

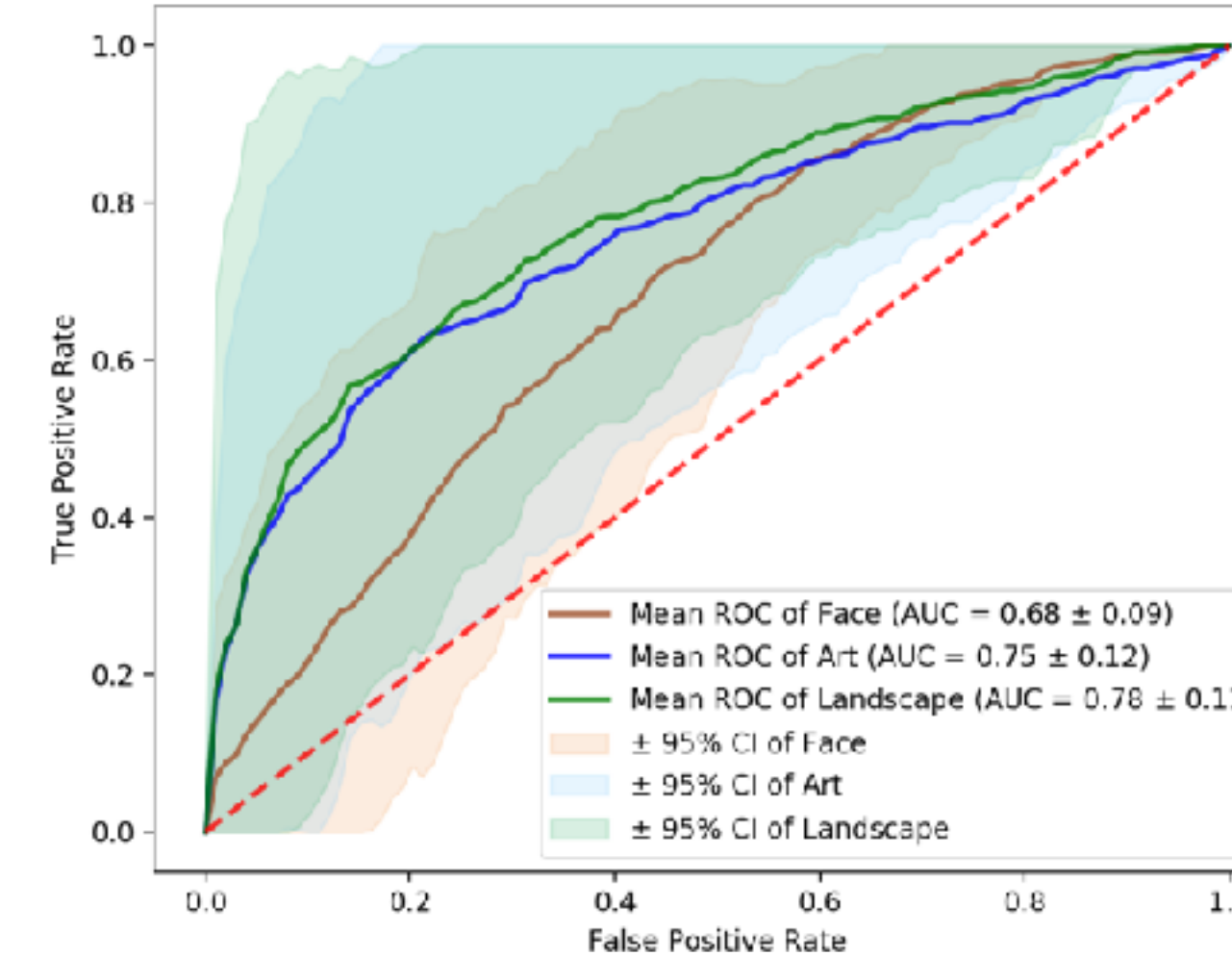
# SUMMARY: PERFORMANCE BY CATEGORY

Perception —> Perception



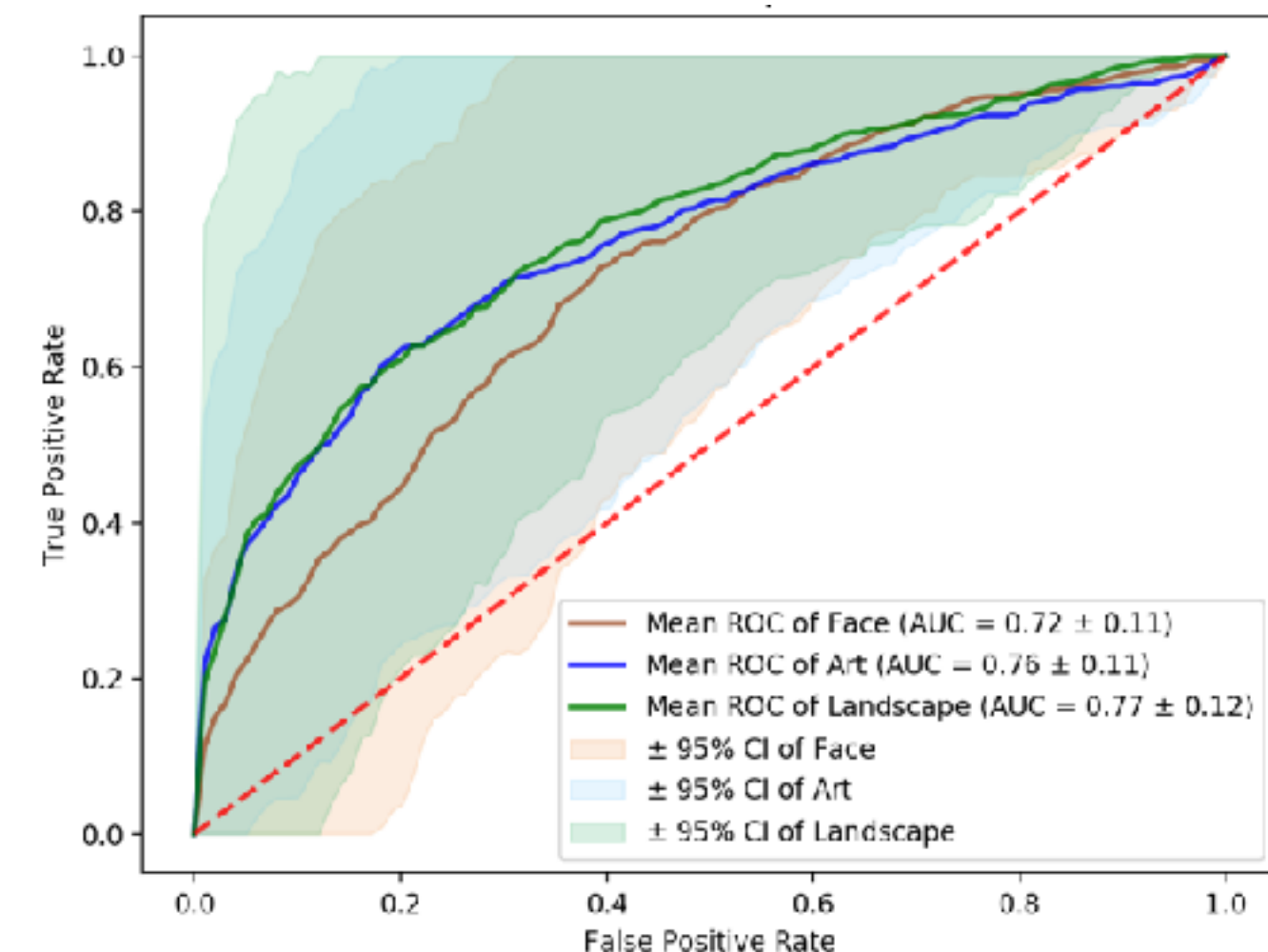
AUC = .87 [.78; .90]

Imagination —> Imagination



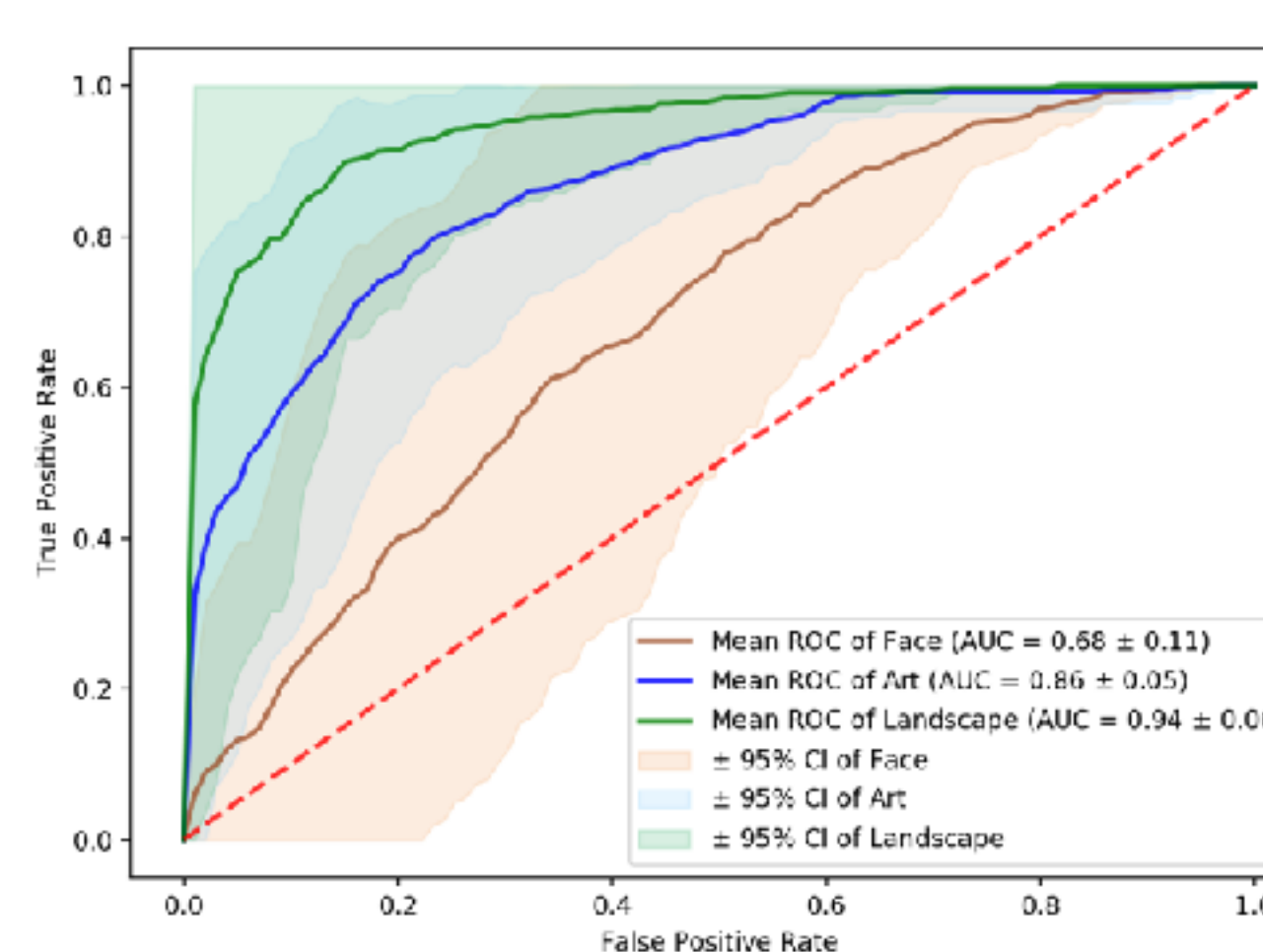
AUC = .77 [.68; .80]

Perception —> Imagination



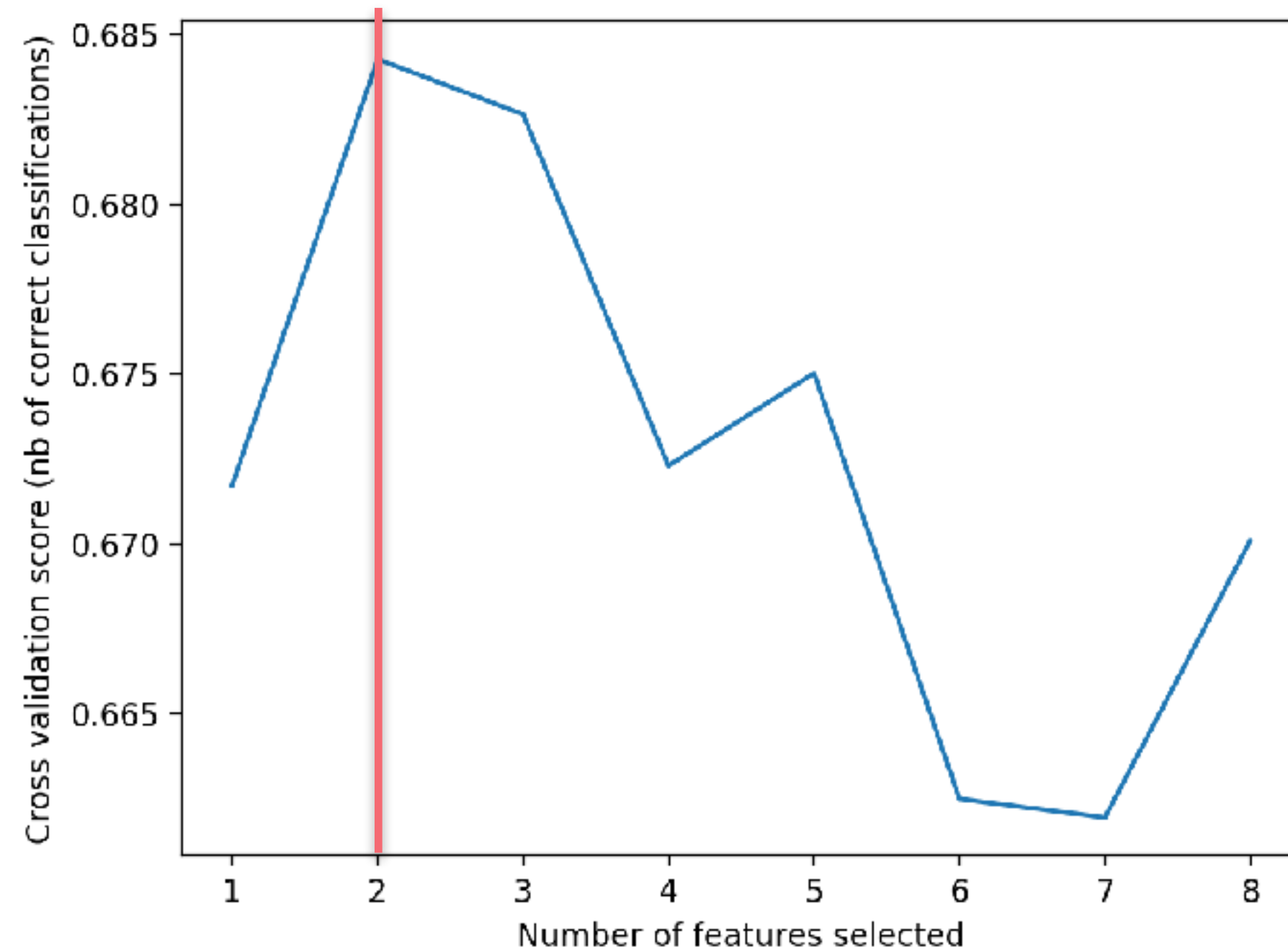
AUC = .77 [.69; .81]

Imagination —> Perception

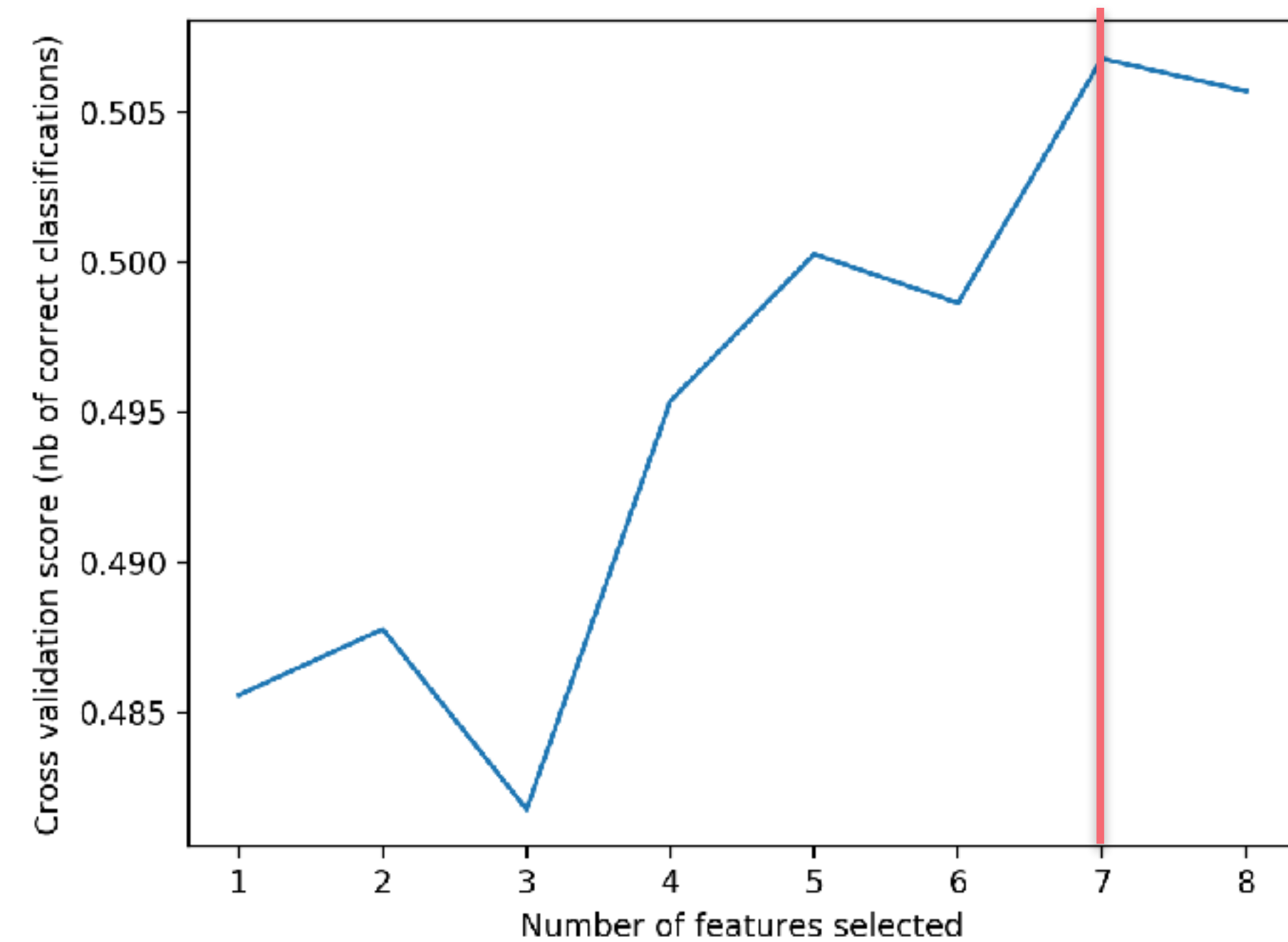


AUC = .84 [.76; .90]

# SUMMARY: RECURSIVE FEATURE ELIMINATION



Perception: **2/8**



Imagination: **7/8**

Optimal number of Features:



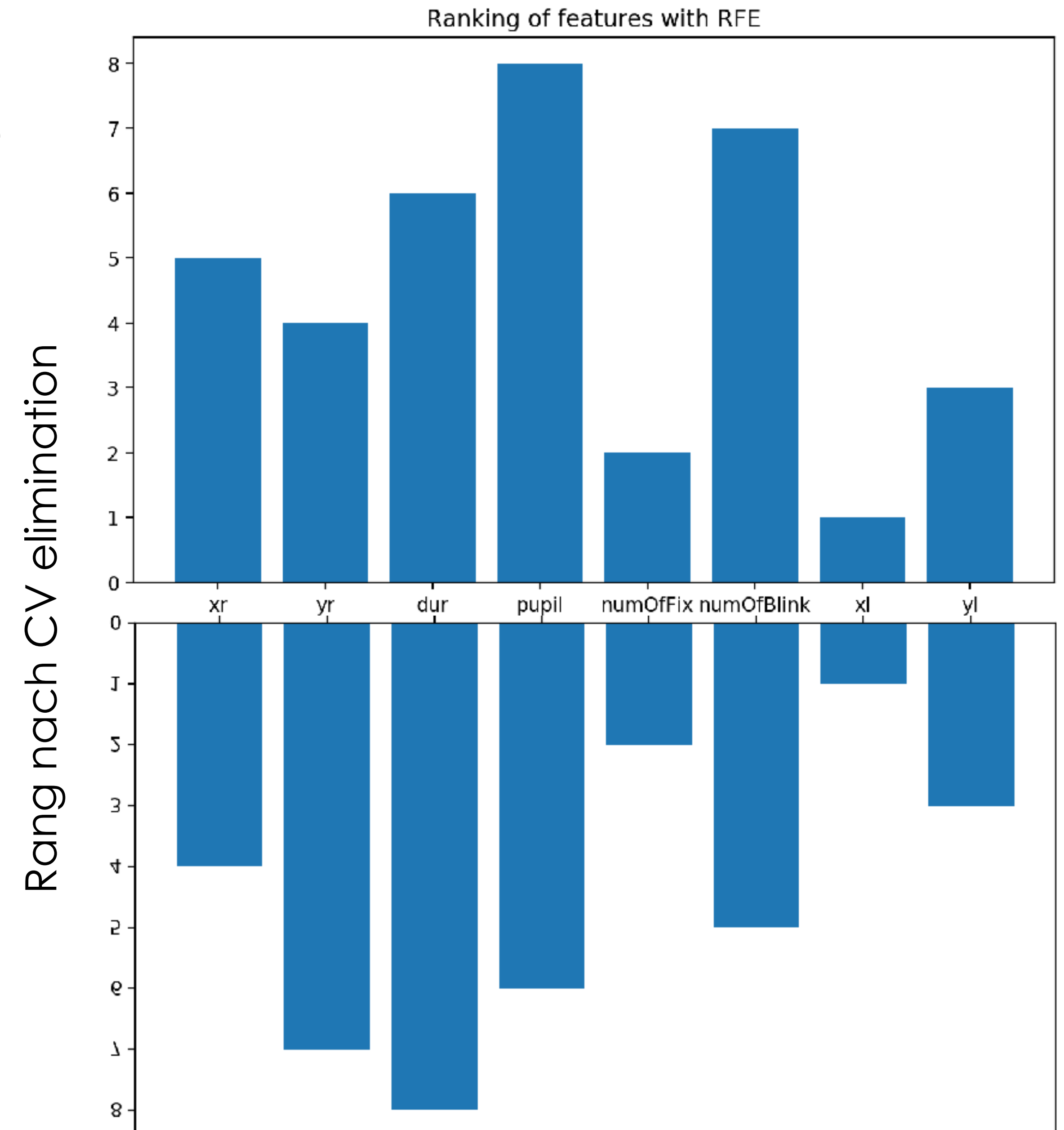
# SUMMARY: FEATURE RANKING

Besonders wichtig:

- Anzahl der Fixationen
- Räumliche Koordinate
- Linkes Auge > Rechtes

weniger wichtig:

- Pupille
- Dauer
- Number of Blinks



Fixation summary = <X\_r, X\_l, Y\_r, Y\_l, Dauer, Pupille, Blink>

Fixation summary = <X\_r, X\_l, Y\_r, Y\_l, Dauer, Pupille, Blink>



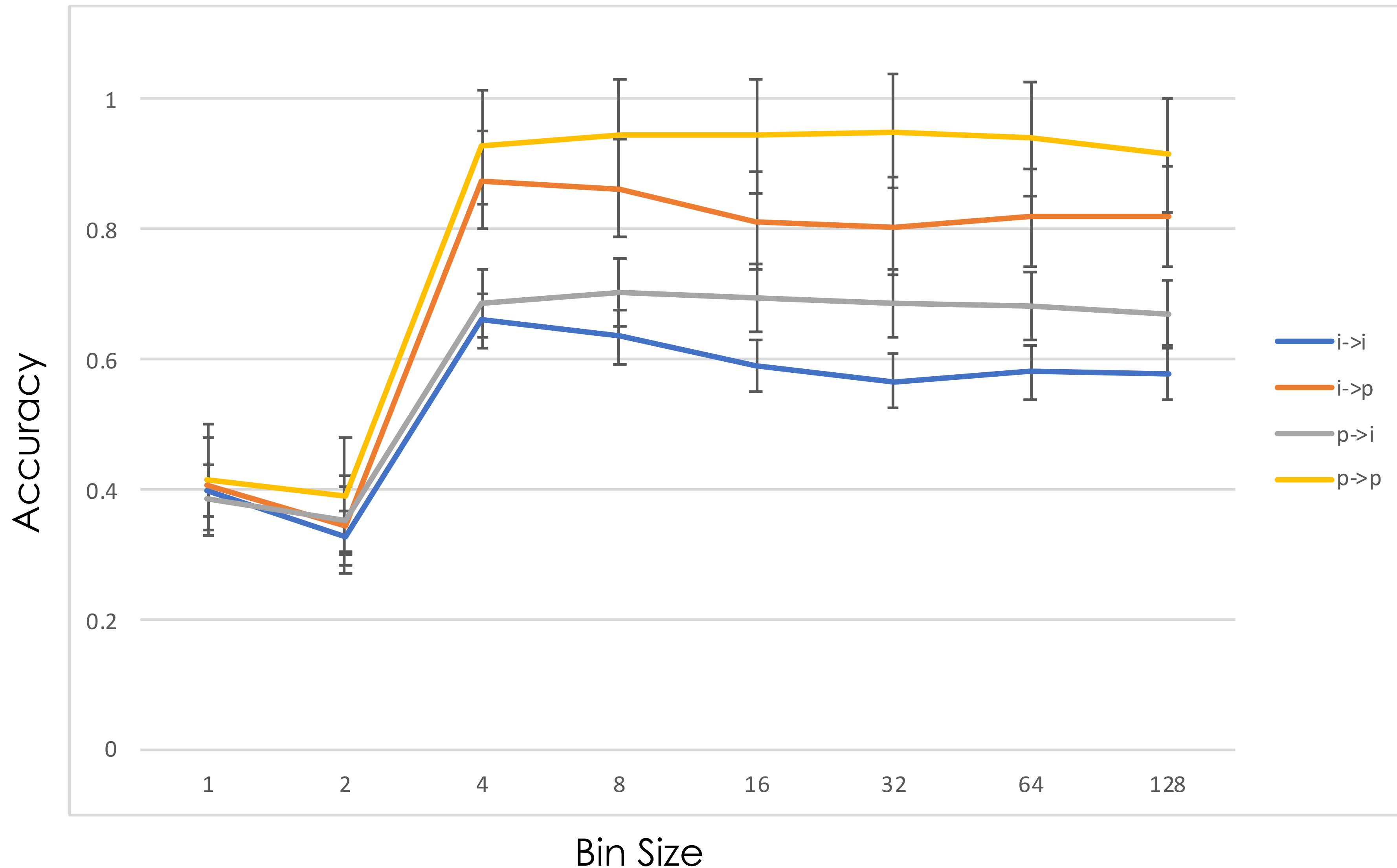
- ## Beispiel:

Wenn Bin Grösse = 10

Sample = 10\*10\*7  
= 700 Features

Ohne Binning: 63'000'000 features! 18

# FIXATION MAP: BIN GRÖSSE

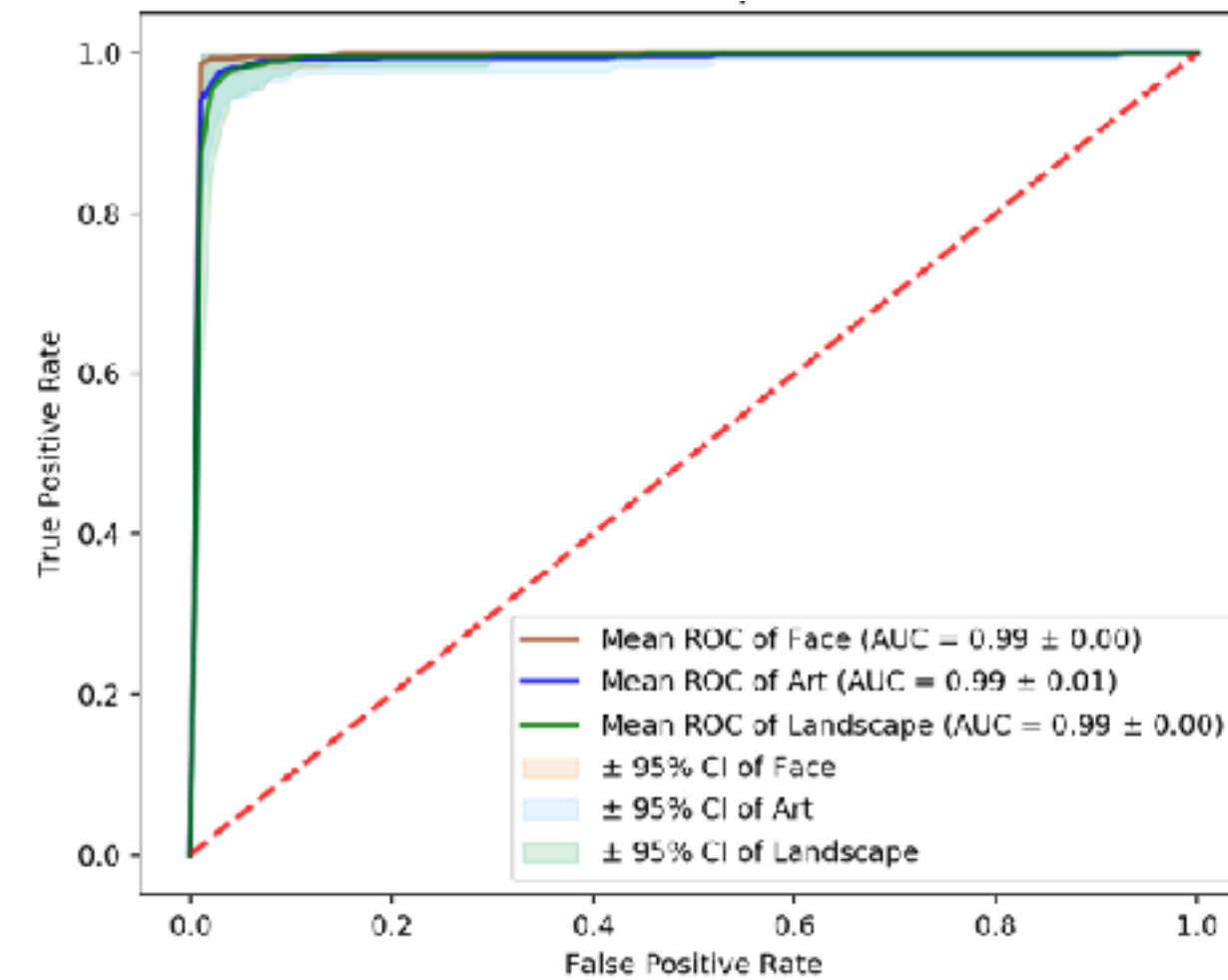


Klassifikationsgüte nach  
Bin Grösse:

- Perception optimal bei etwa **8 (32)** Bins
- Imagination optimal bei etwa **4** Bins

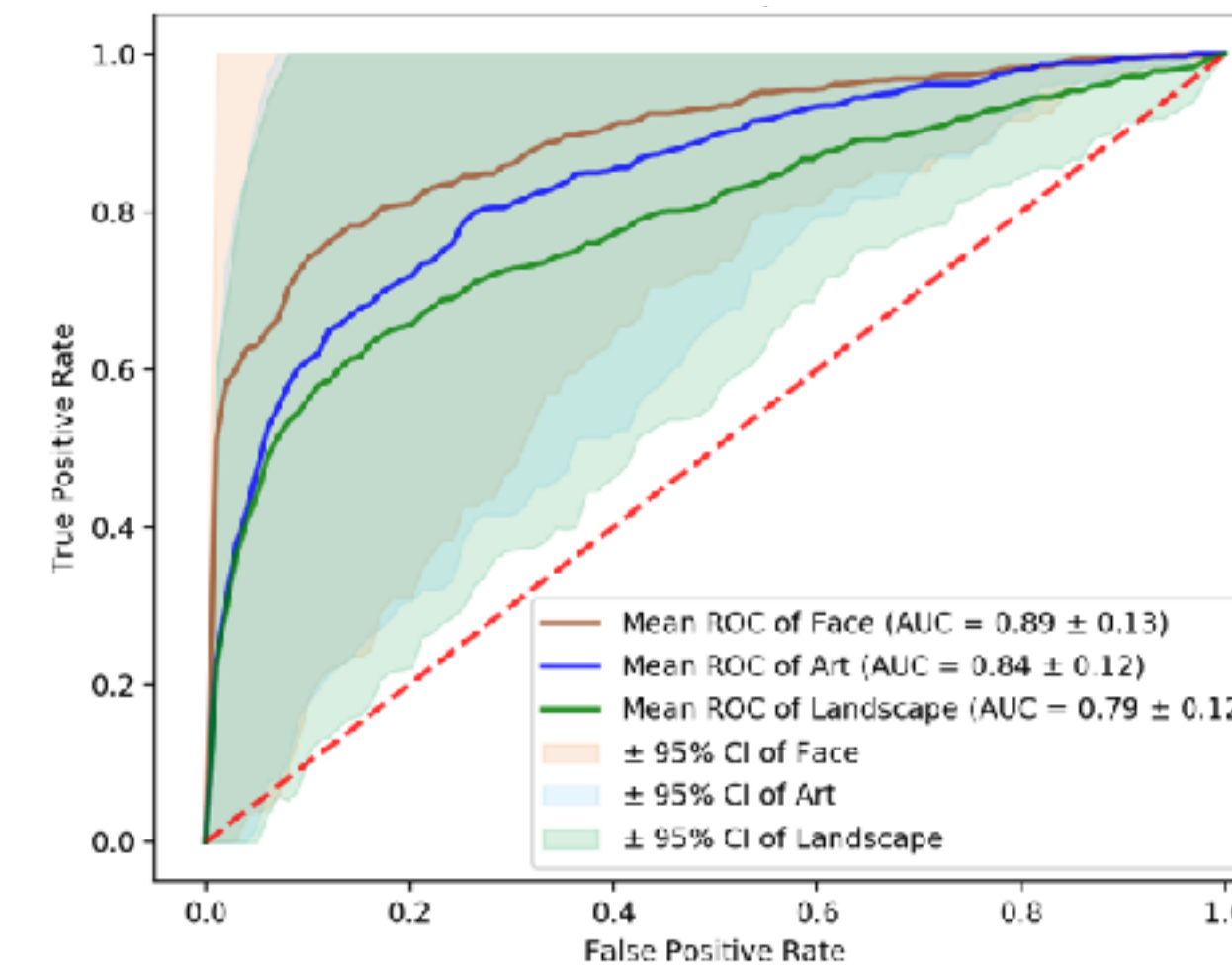
# FIXATION MAP: PERFORMANCE BY CATEGORY

Perception → Perception



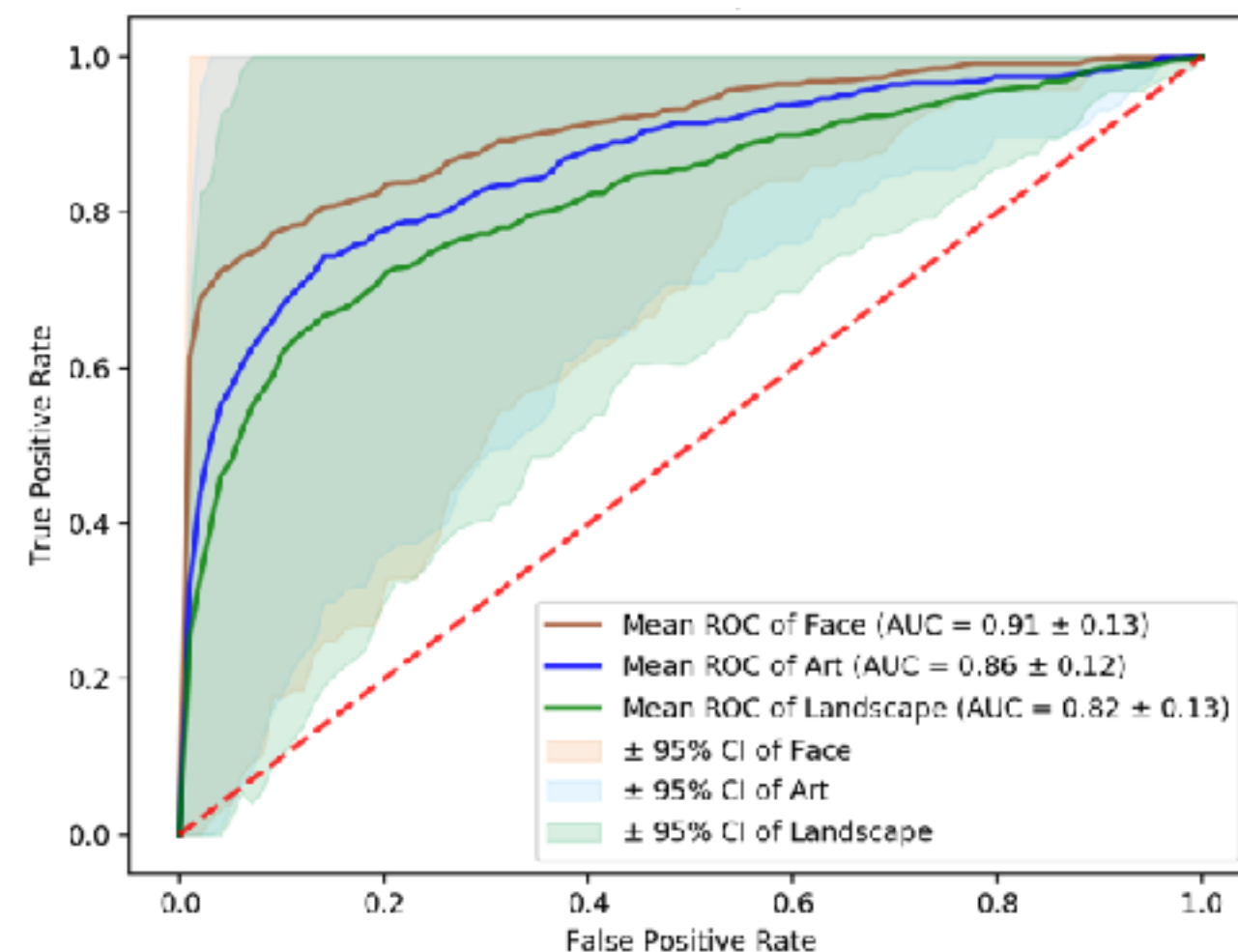
AUC = .99 [.99; 1.00]

Imagination → Imagination



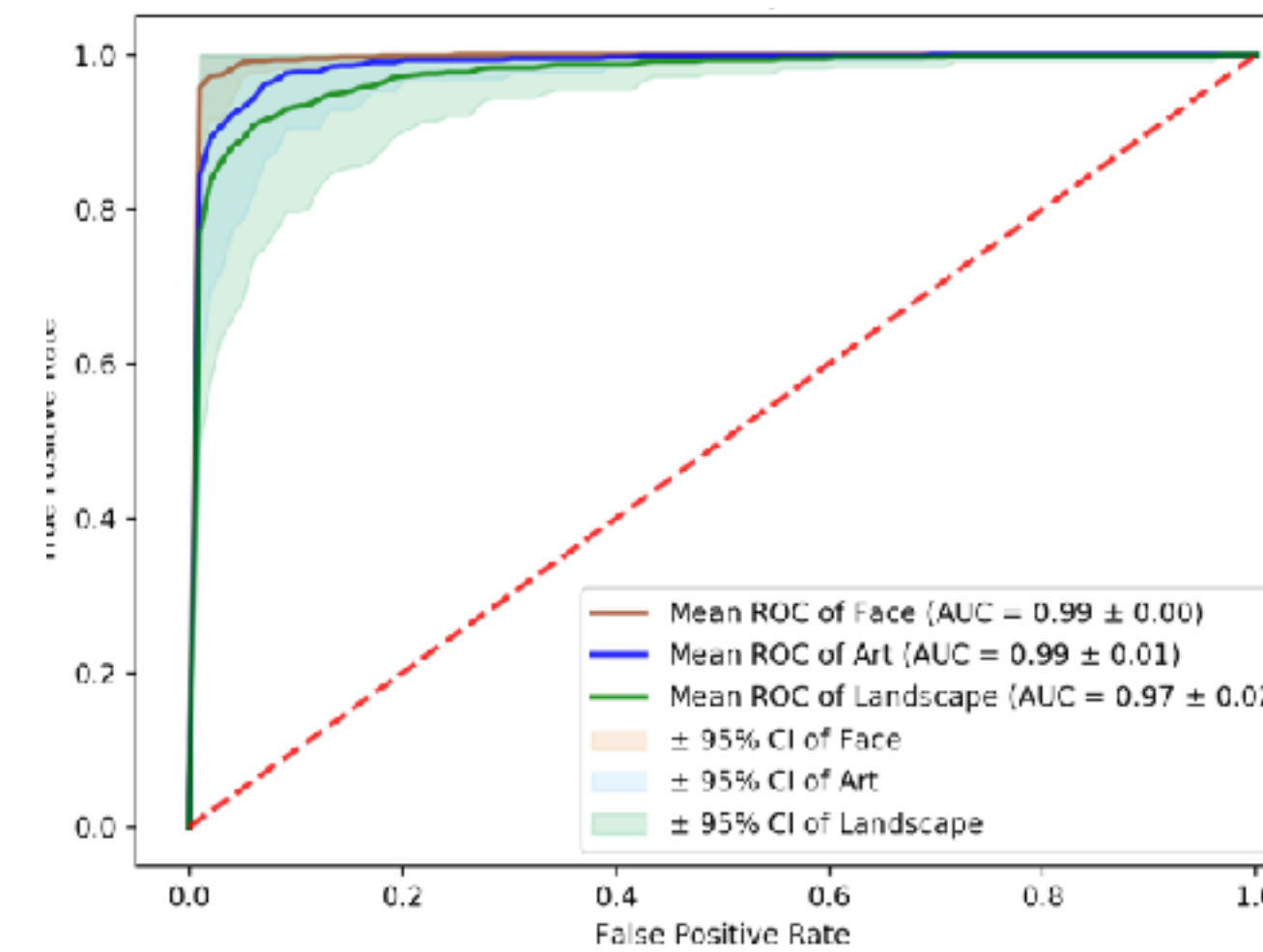
AUC = .80 [.77; .90]

Perception → Imagination



AUC = .89 [.79; .93]

Imagination → Perception



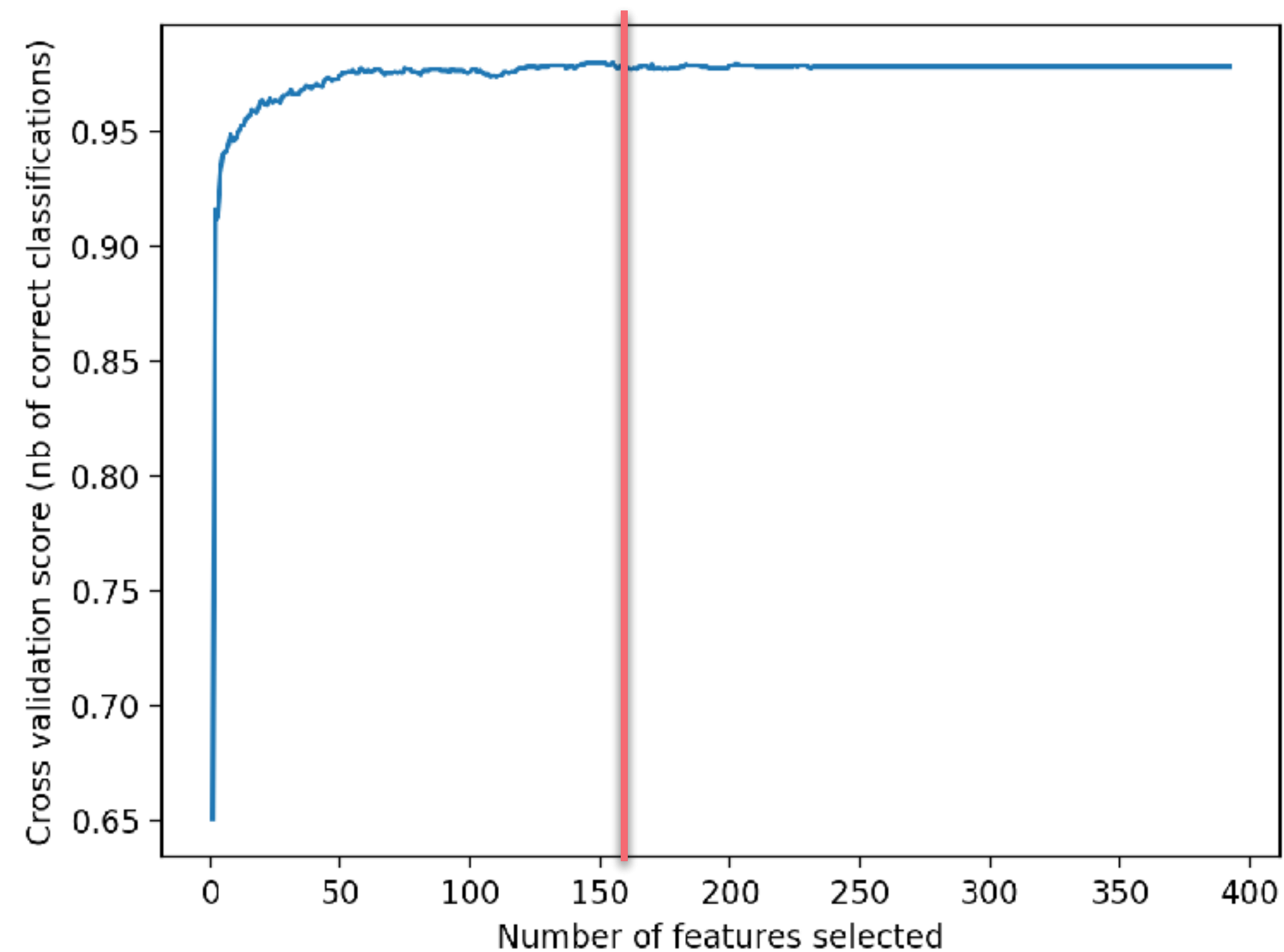
AUC = .98 [.98; .1.0]



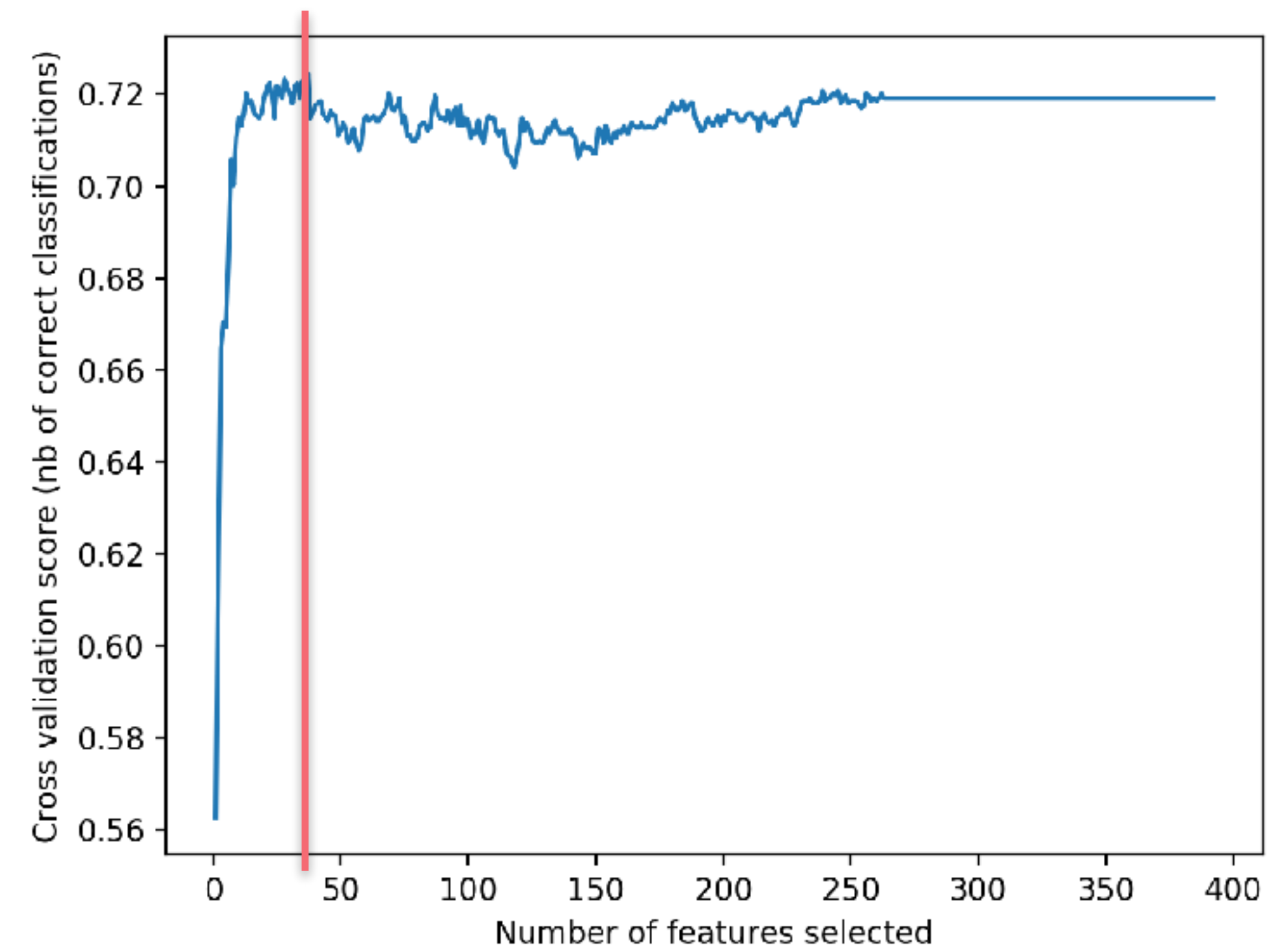
# FIXATION MAP: RECURSIVE FEATURE ELIMINATION

Optimal number of Features:

Perception: **155/392**



Imagination: **37/392**



# FIXATION MAP: FEATURE RANKING

Perception

Guess what... and the best feature-set is?

Bezieht sich jeweils auf jeden Bin während der Vorstellung oder Wahrnehmung eines Bildes

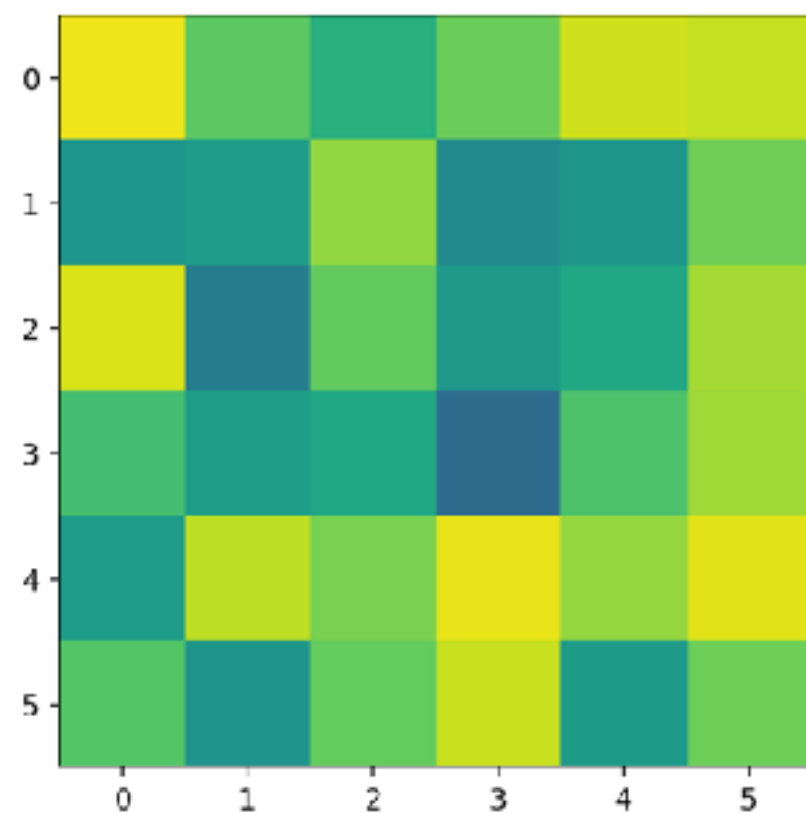
- Relative Hit Rate eines Bins
- Absolute Hit Rate eines Bins
- Relative rechtes Auge
- Relative linkes Auge
- Durchschnittliche Dauer
- Durchschnittliche Pupille
- Durchschnittliche Zwinkern
- Anzahl Zwinkern

# FIXATION MAP: FEATURE RANKING

Perception

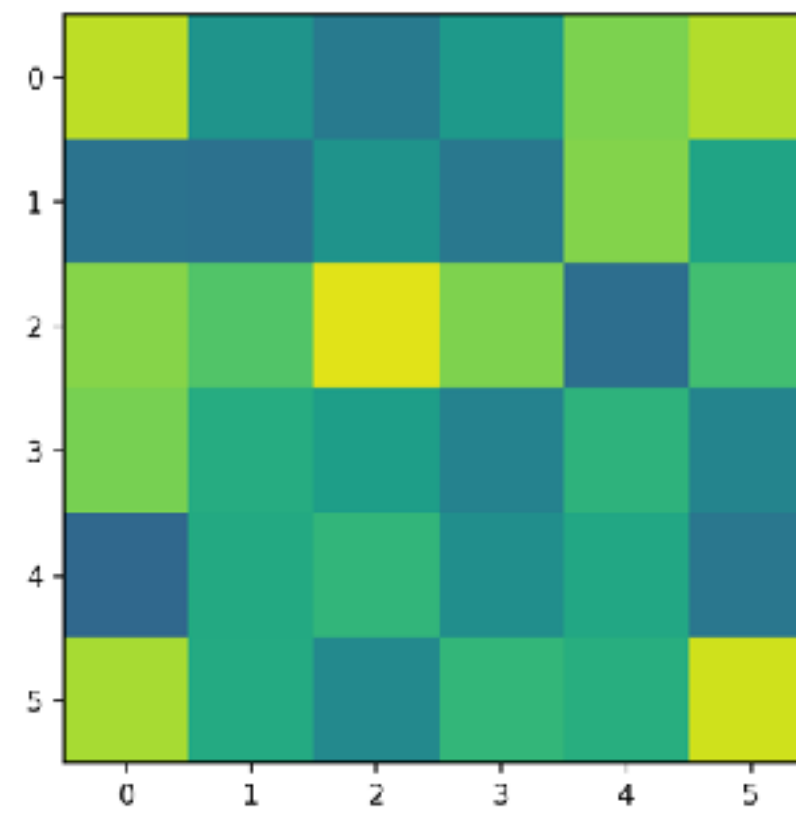
Interpretation: Hell gut,  
dunkel schlecht

Relative Hit Rate eines Bins



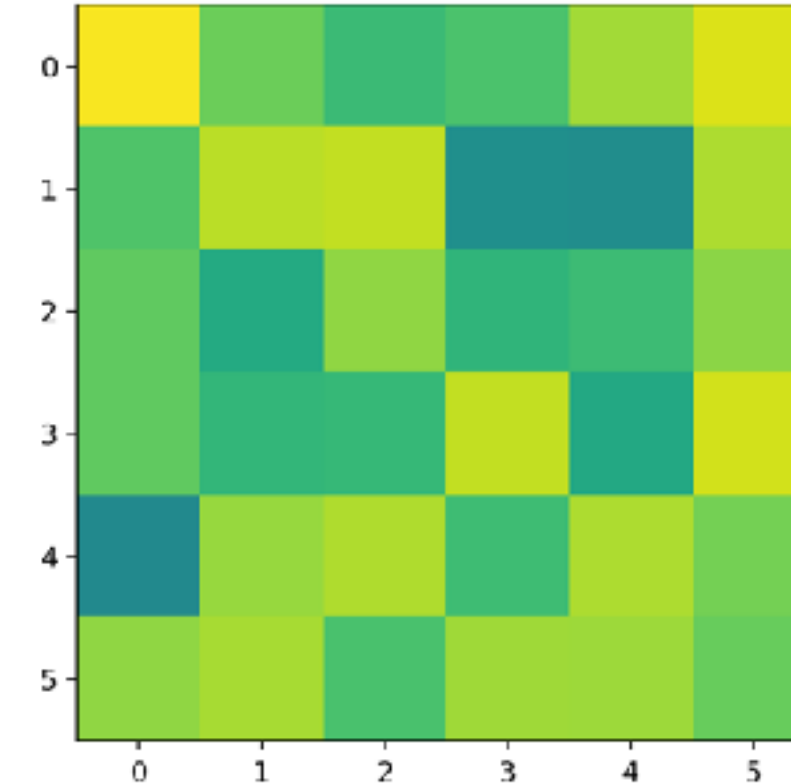
**Median:  
94  
[72, 156]**

Absolute Hit Rate eines Bins



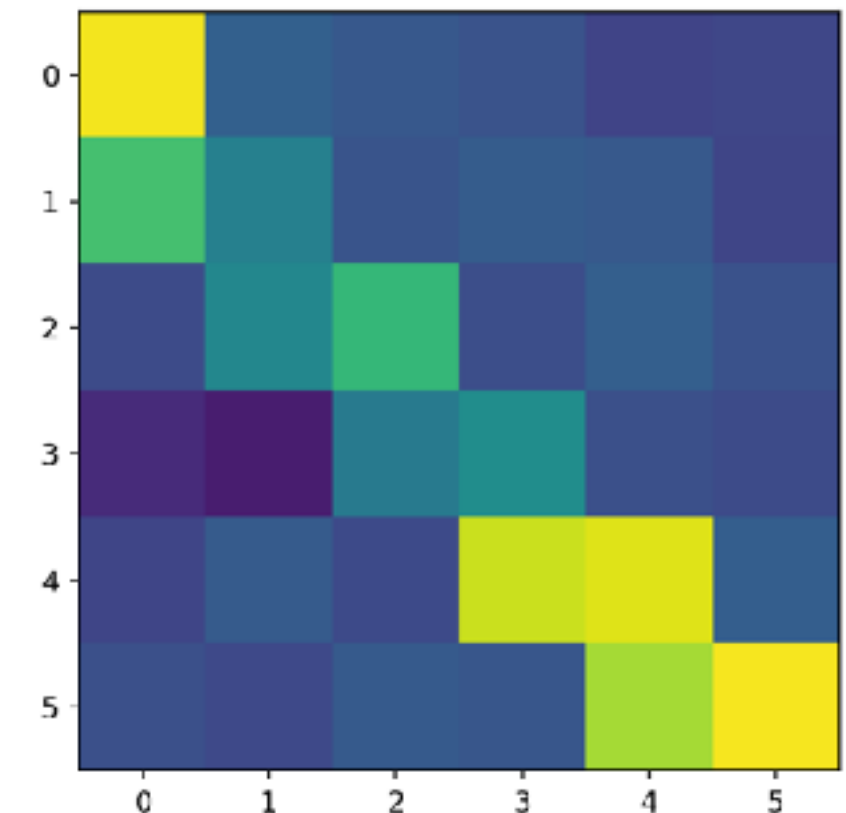
**Median:  
144  
[124, 187]**

Relative rechtes Auge



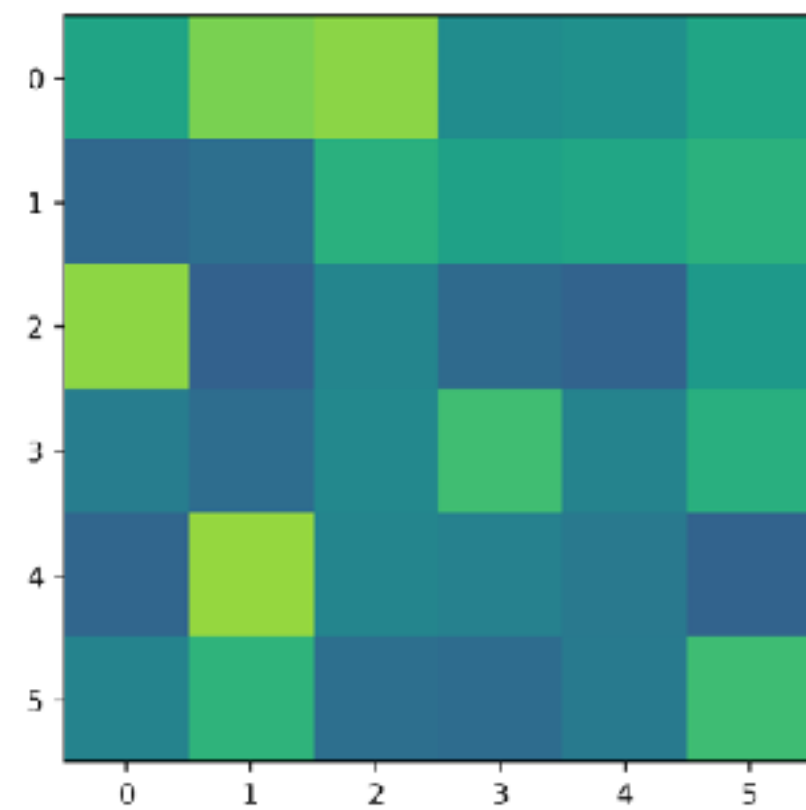
**Median:  
93  
[55, 109]**

Relative linkes Auge



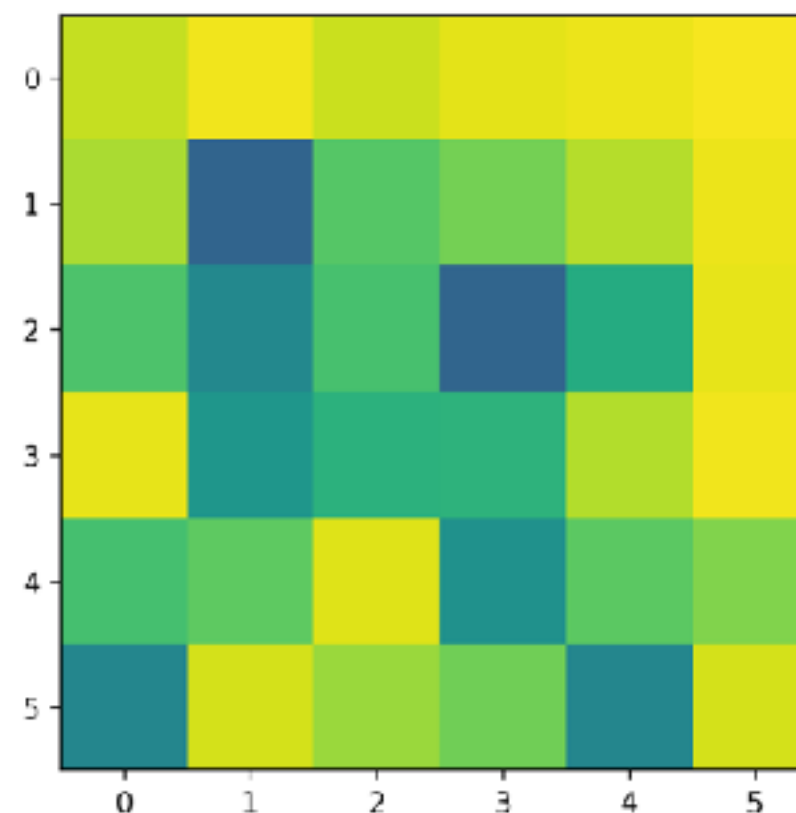
Median:  
279  
[271, 295]

Durchschnittliche Dauer



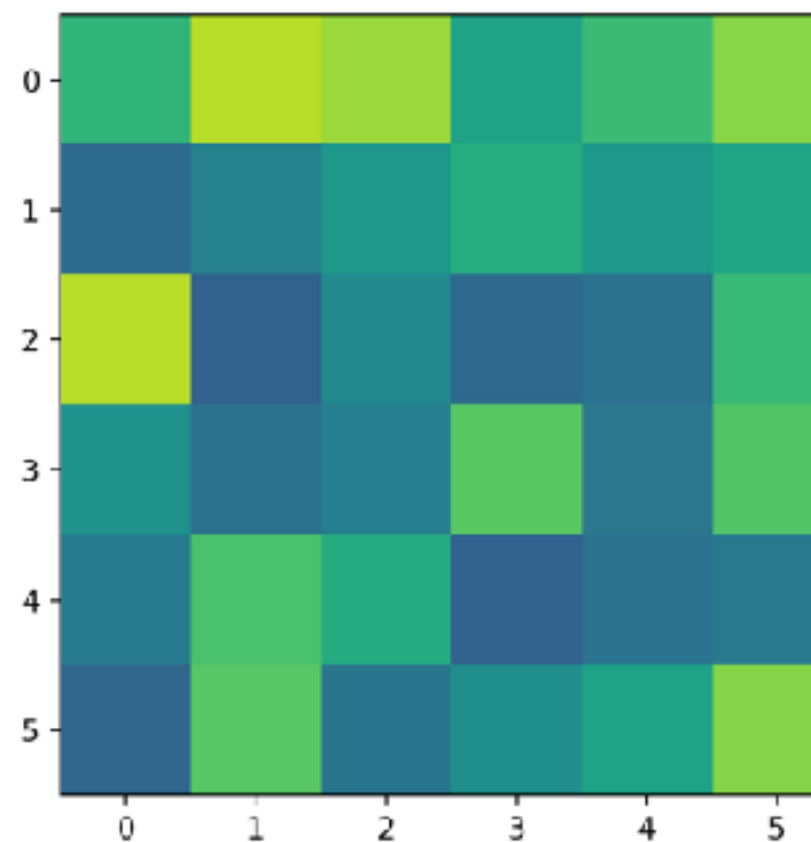
Median:  
215  
[161, 224]

Durchschnittliche Pupille

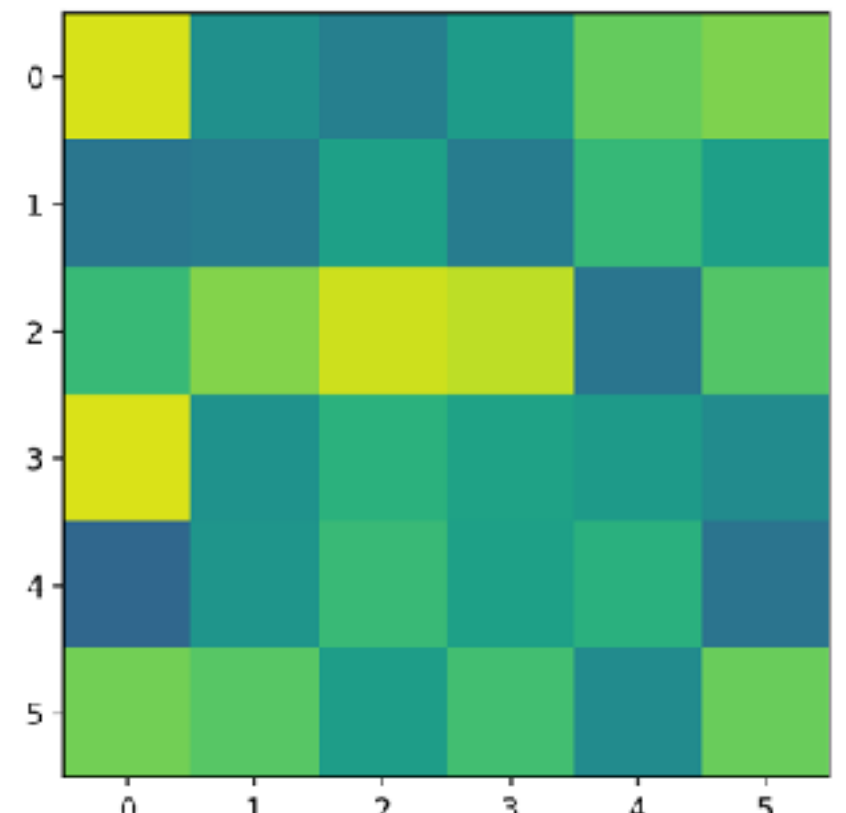


**Median:  
72  
[31, 107]**

Durchschnittliche Zwinkern Anzahl Zwinkern



**Median:  
182  
[145, 225]**



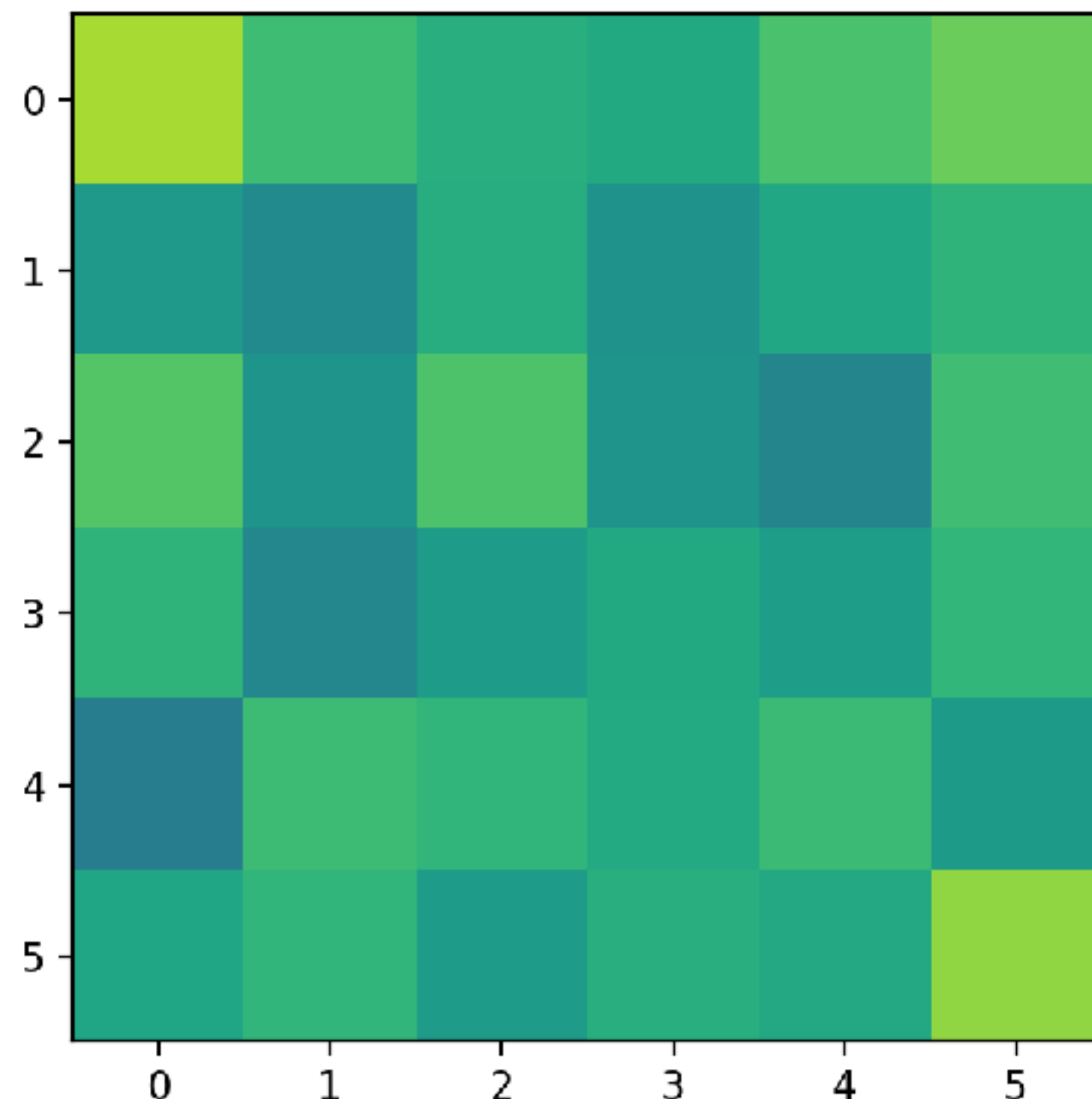
**Median:  
172  
[120, 180]**

# FIXATION MAP: FEATURE RANKING

Perception

Durchschnittlich beste Feature

Interpretation: Hell gut,  
dunkel schlecht



(Mean [ SD] )

**Durchschnittliche Pupille (89.55 [77.36])**

**Relative Hit Rate rechtes Auge (88.80 [50.98])**

**Relative Hit Rate eines Bins (109.34 [67.65])**

Absolute Hit Rate eins Bins (148.69 [70.73])

Anzahl Zwinkern (148.19 [66.43])

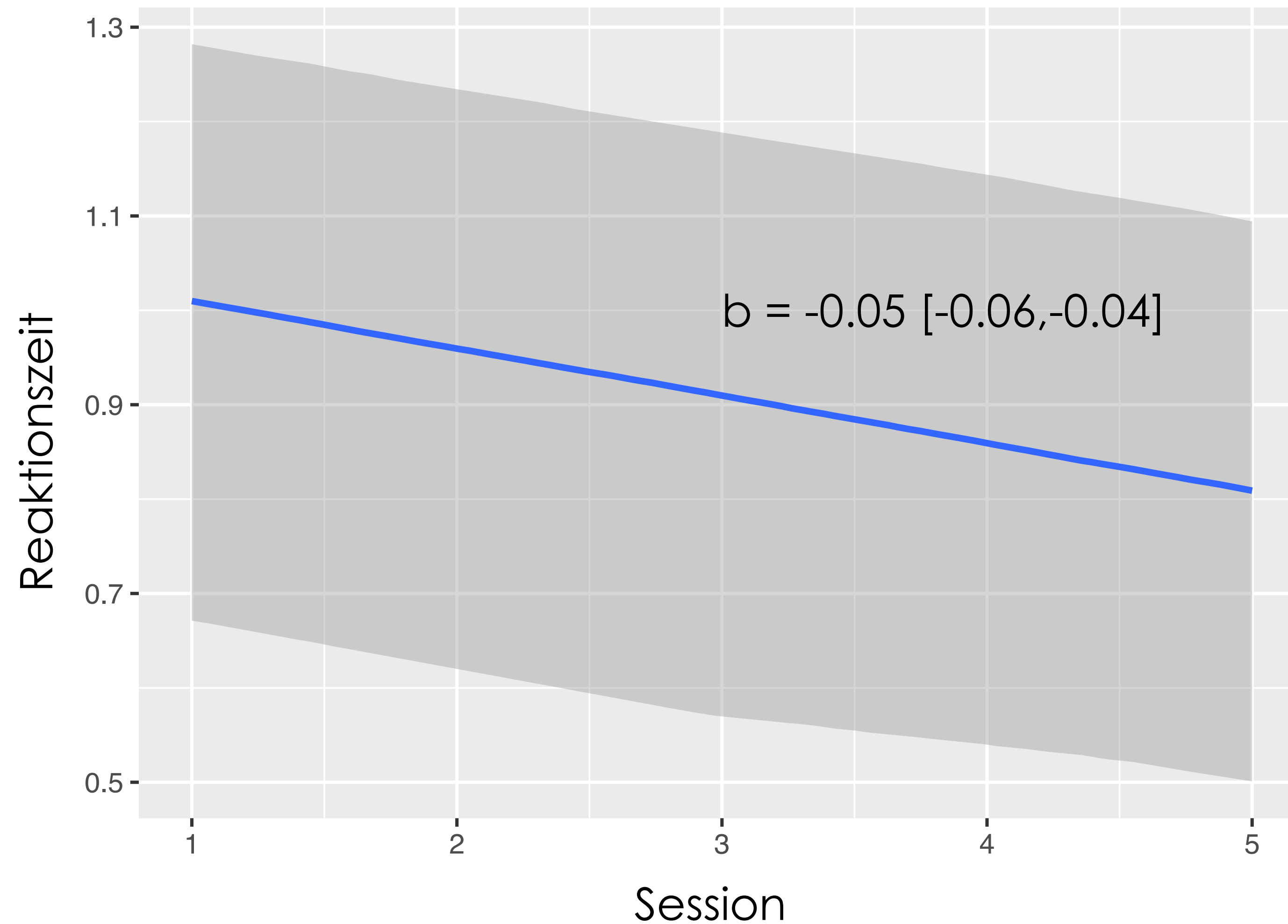
Durchschnittliche Zwinkern (173.58 [69.76])

**Durchschnittliche Dauer (190.52 [62.16])**

**Relative Hit Rate linkes Auge (240.5 [100.63])**



# KONTROLLEXPERIMENT: RESULTS



## Bayesian multi level model

Verteilung: Exgauss

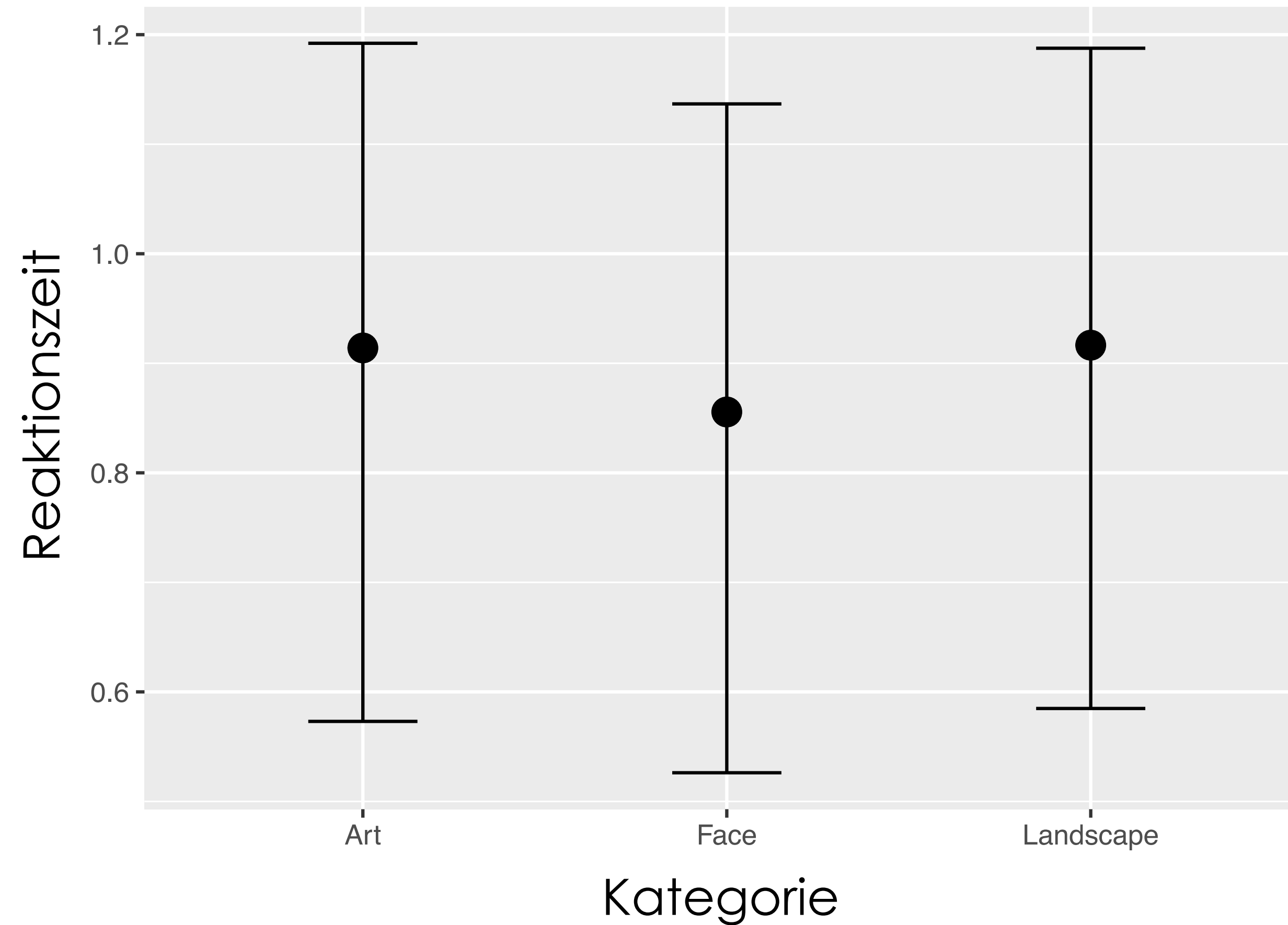
Gruppen level:  
(variierender Intercept):

- *inTraining (Hit/FA)*
- *Subject*
- *Img ID*

Population level:

- *Session*
- *Kategorie*

# KONTROLLEXPERIMENT: RESULTS



## Bayesian multi level model

Verteilung: Exgauss

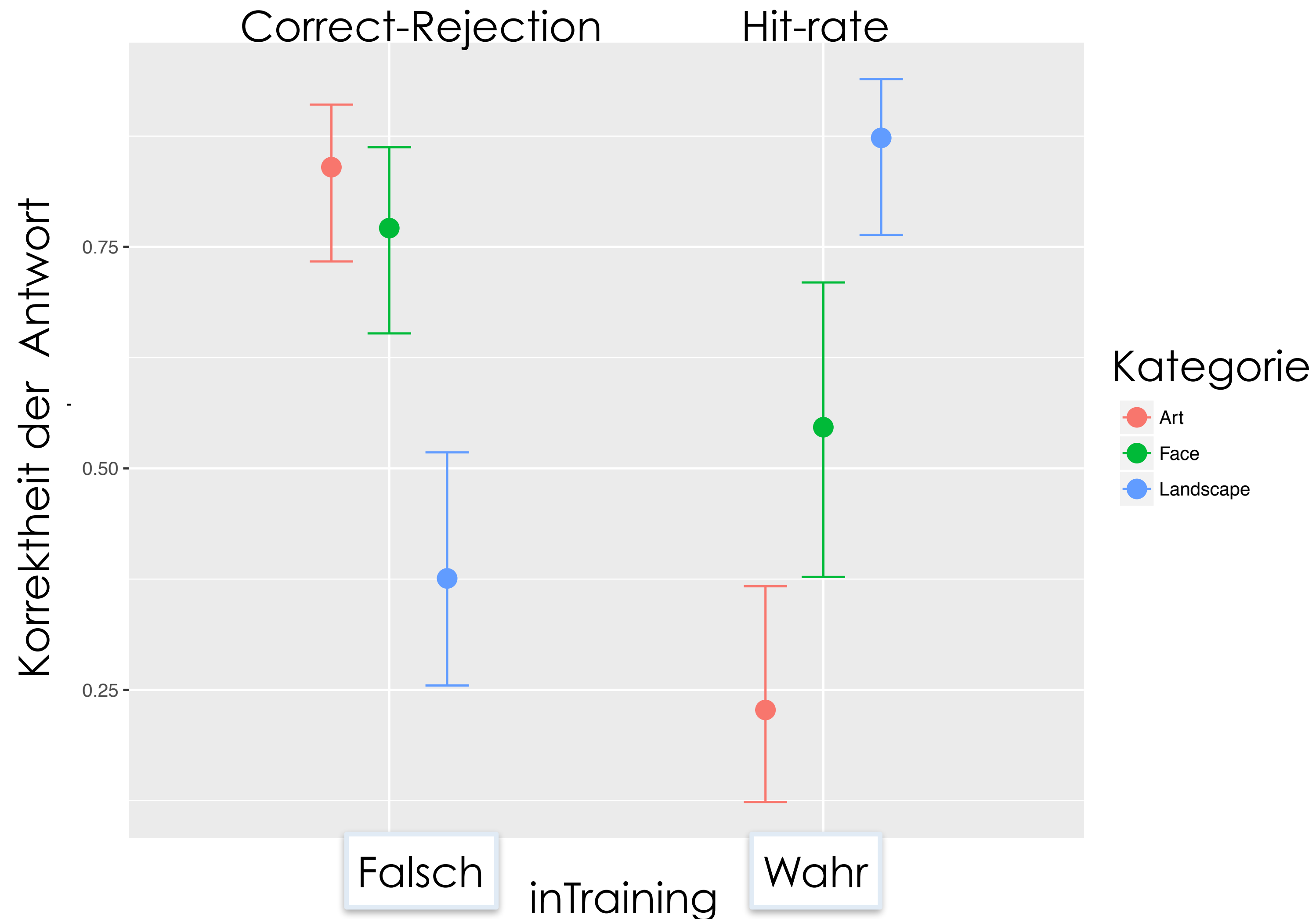
Gruppen level:  
(variierender Intercept):

- *inTraining (Hit/FA)*
- *Subject*
- *Img ID*

Population level:

- *Session*
- *Kategorie*

# KONTROLLEXPERIMENT: RESULTS



## Bayesian multi level model

Verteilung: Bernoulli

Gruppen level:  
(variierender Intercept):

- *Session*
- *Subject*
- *Img ID*

Population level:

- *inTraining* x *Kategorie*

## Résumé & Ausblick

- > **Summary-Statistik** klassifiziert über Zufällig, wird aber von höheren Repräsentation bei Perception als Testset überboten.
- > Perception kann am besten Perception vorhersagen, Imagination klassifiziert jedoch ebenfalls Perception am besten.
- > Dies stimmt mit der Literatur überein, dass Imagination grössere Varianz aufweist. Es zeigt, aber auch, dass ein Grossteil der Information repräsentiert wird. - Wäre ein translator zw. Vorstellung und Wahrnehmung möglich?
- > Featureinterpretationen sprechen für ein Kognitiv - Abtasten des Bildes,
  - > Summary Statistik führt zu einem Vereinfachten Verständnis, welches das **Abtastverhalten** und die **räumliche Auflösung** verdeutlicht.
  - > Die Hit Rate eines Ortes, die Dauer der Betrachtung und das Blinzeln trägt kaum zu einer korrekten Klassifikation zu.
  - > Es sind weniger einzelne Orte, die über die Klassifikation des Bildes informieren, als die
    - > Häufigkeit wie der Ort betrachtet wird und
    - > die Kombination von diesem spezifischen **Ort** und der **kognitiven Reaktion** zu diesem Ort (Pupille) und wie oft dieser im Vergleich zu anderen Orten **besucht** wurde.
  - > Legt Grundlage für Spatial Memory -> Mental Imagery
- > Dies könnte Automatisiert genutzt werden, um in einem natürlichen Umfeld individualisierte Diagnostik zu betreiben.

## Technical Acknowledgement:



**Vielen Dank für Ihre  
Aufmerksamkeit.**



## Literaturverzeichnis

- Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of cognitive neuroscience*, 9(1), 27-38.
- Dewhurst, R., Nyström, M., Jarodzka, H., Foulsham, T., Johansson, R., & Holmqvist, K. (2012). It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach. *Behavior research methods*, 44(4), 1079-1100.
- Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & cognition*, 33(1), 98-106.
- Jarodzka, H., Holmqvist, K., & Nyström, M. (2010, March). A vector-based, multidimensional scanpath similarity measure. In *Proceedings of the 2010 symposium on eye-tracking research & applications*(pp. 211-218). ACM.
- Johansson, R., & Johansson, M. (2014). Look here, eye movements play a functional role in memory retrieval. *Psychological Science*, 25(1), 236-242.
- Le Meur, O., Le Callet, P., Barba, D., & Thoreau, D. (2006). A coherent computational approach to model bottom-up visual attention. *IEEE transactions on pattern analysis and machine intelligence*, 28(5), 802-817.
- Martarelli, C. S., & Mast, F. W. (2013). Eye movements during long-term pictorial recall. *Psychological research*, 77(3), 303-309.