



Abschlusspräsentation - Data Challenge

Von Chantal Klemm, Emmanuela Georgoula und Robin Krause

Gliederung

1. Einleitung
2. Grundlagen - Ensemble Learning
3. Umsetzungen
4. Ergebnisse
5. Ausblick
6. Quellen

Einleitung

- Verschiedene CNNs zur Klassifizierung der Münzstätten trainieren (Base Models)
 - Architekturen: VGG16, ResNet50 und ResNet101
 - Transfer Learning
 - Trainingsdatensätze (17.821 Münzbilder):
 - Nur die Vorderseite
 - Nur die Rückseite
 - Vorder- und Rückseite
 - Anzahl der Klassen: 92
- Methoden des Ensemble Learnings verwenden, um mit den Outputs der Base Models einen finalen Output zu generieren

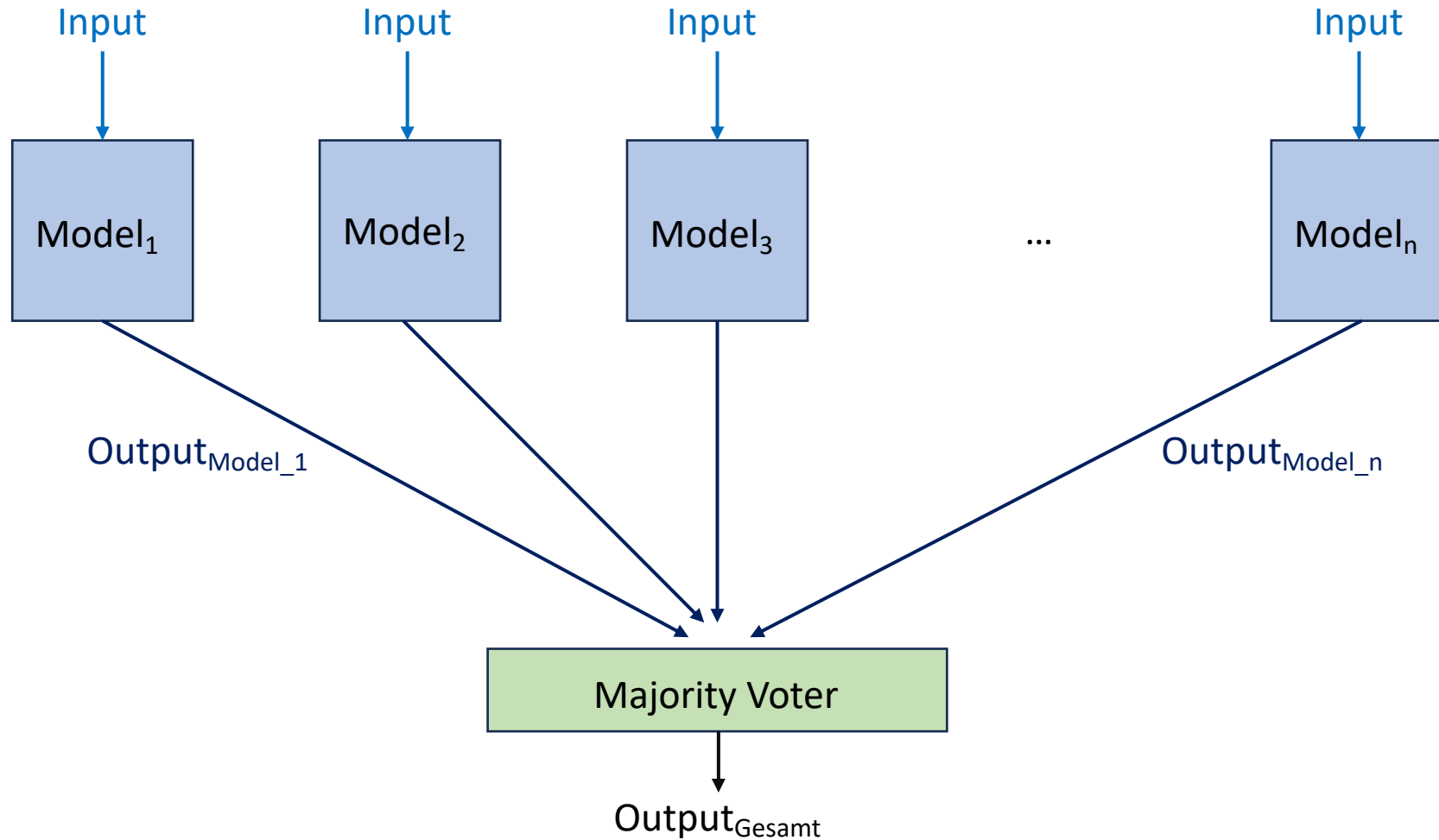


Grundlagen – Ensemble Learning

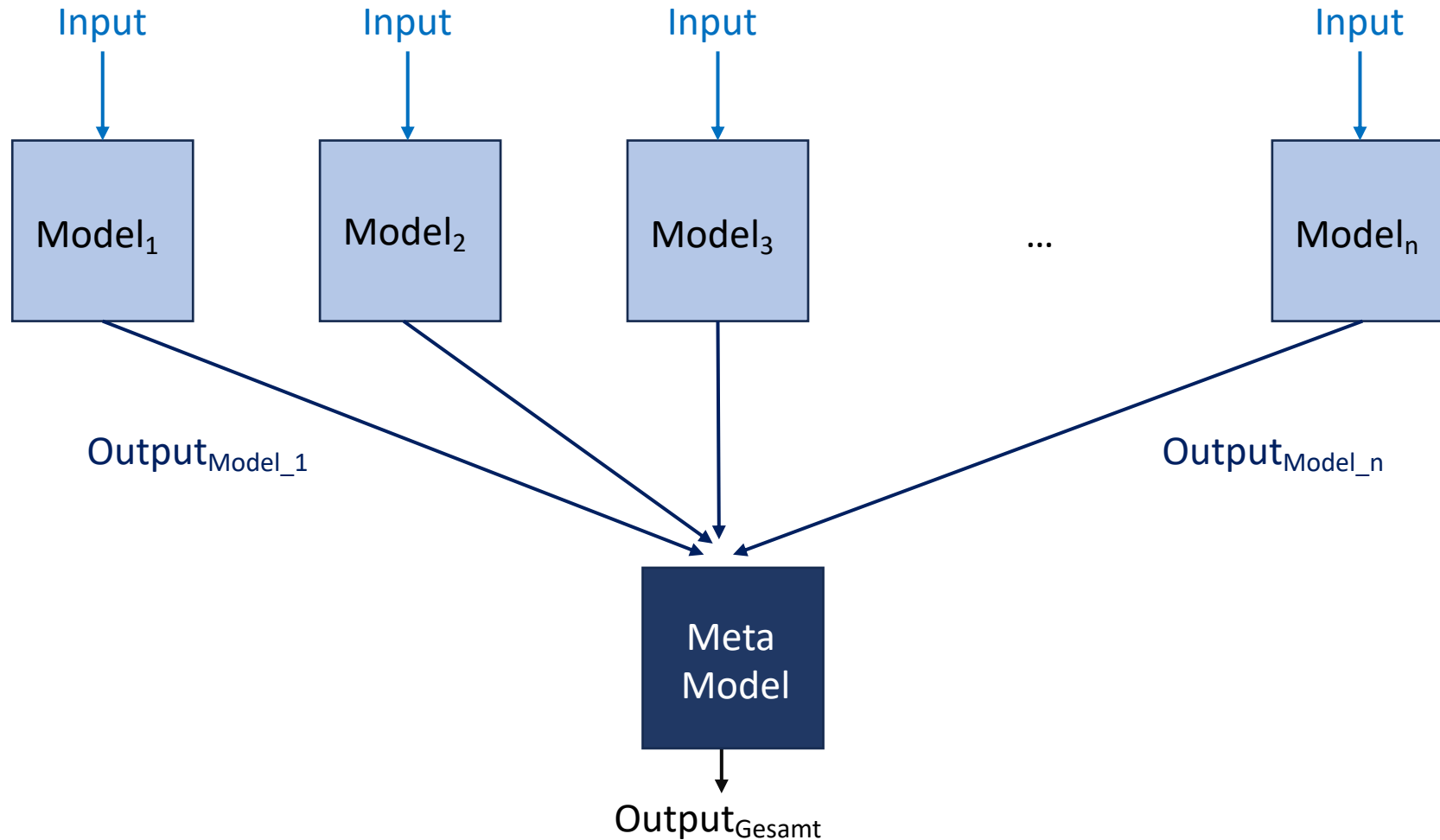
Allgemein

- Kombination mehrerer Einzelmodelle (Base Models), um bessere Ergebnisse zu erzielen.
- Methoden:
 - Decision Rules
 - Majority Voting
 - Stacked Generalization (Stacking)
 - Bootstrap aggregating (Bagging)
 - Boosting
 - ...

Voting (Decision Rule)



Stacking (Stacked Generalization)

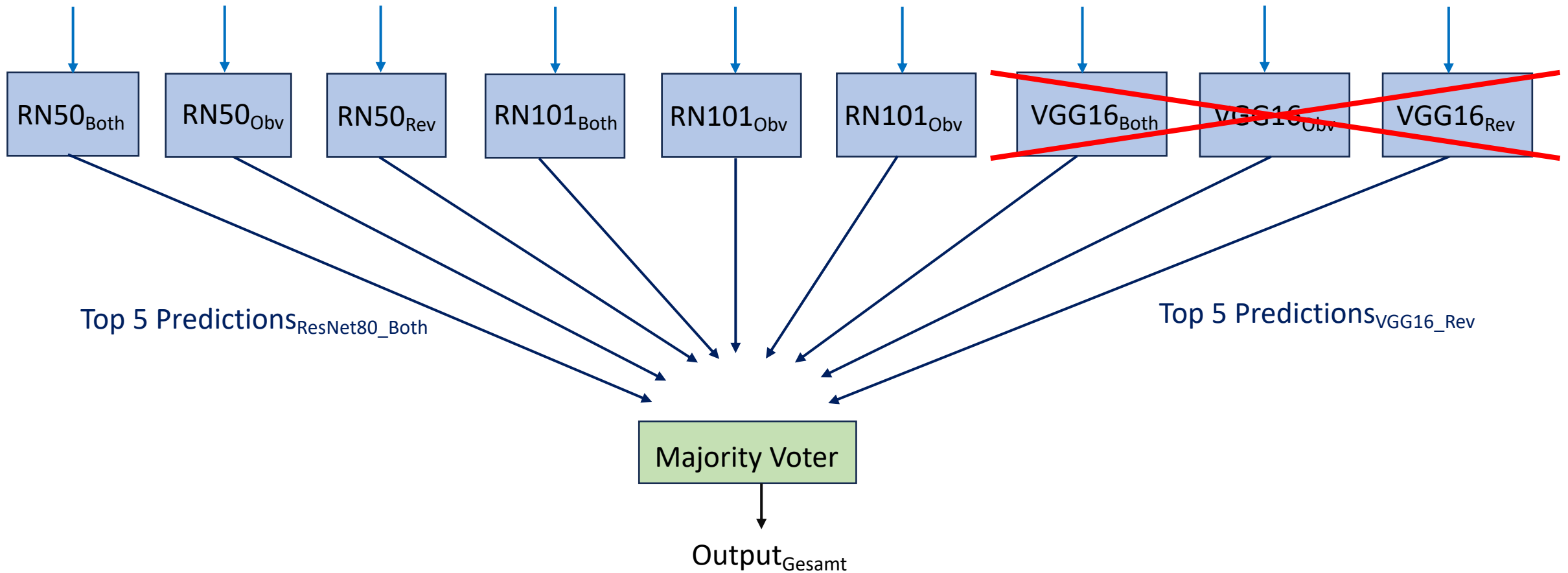




Umsetzungen

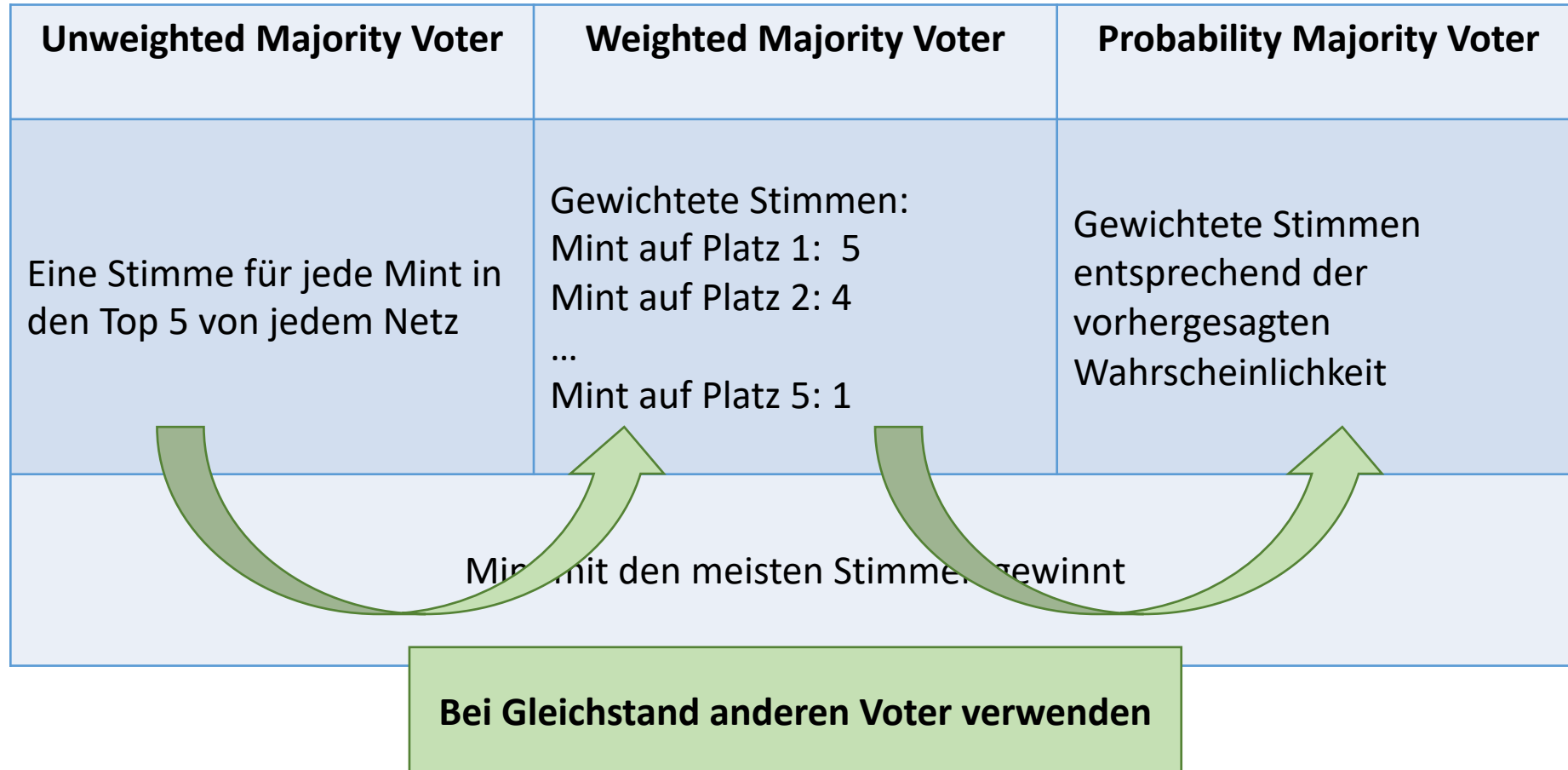
Voting

Münzbilder des Testsets



Anmerkung: RN = ResNet

Voting: Majority Voter



Voting Beispiel

RN50_{Both} Top 3

1. Mint A 0.5
2. Mint B 0.4
3. Mint C 0.01

RN50_{Rev} Top 3

1. Mint B 0.95
2. Mint D 0.03
3. Mint C 0.01

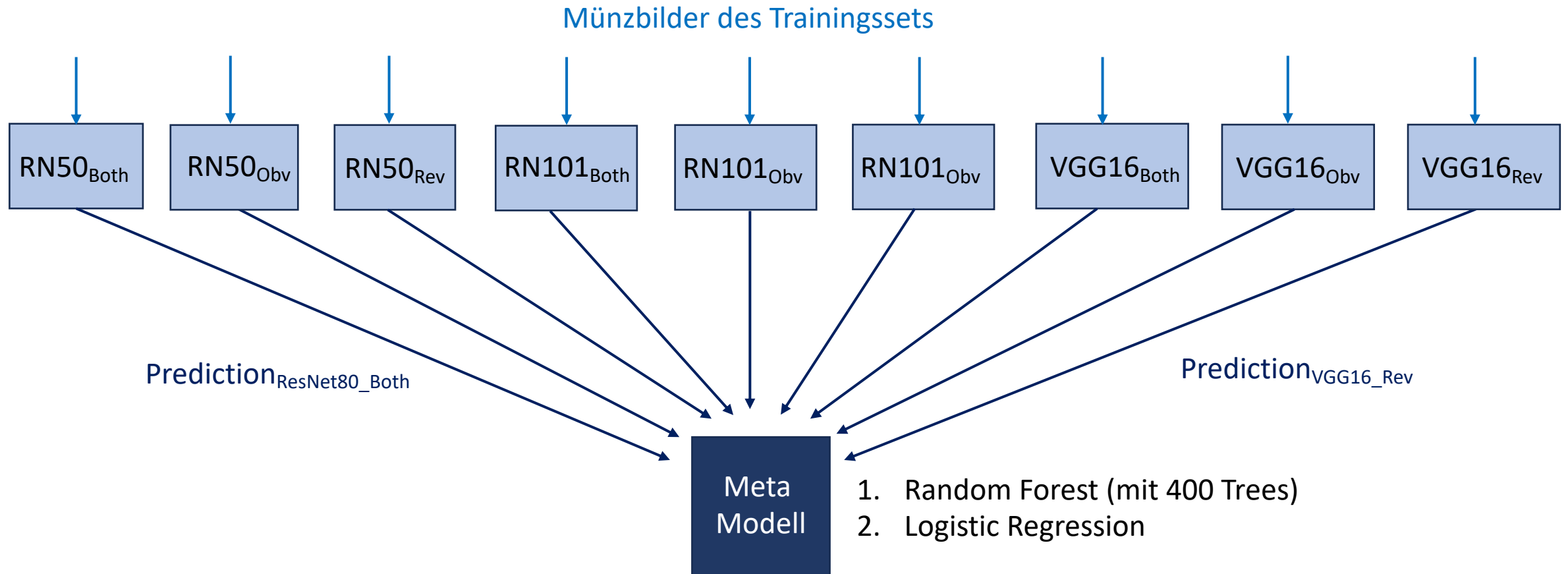
VGG16_{Both} Top 3

1. Mint A 0.6
2. Mint E 0.11
3. Mint C 0.1

	Mint A	Mint B	Mint C	Mint D	Mint E
Unweighted	2	2	3	1	1
Weighted	6	5	3	2	1
Probability	1.1	1.35	0.12	0.03	0.11

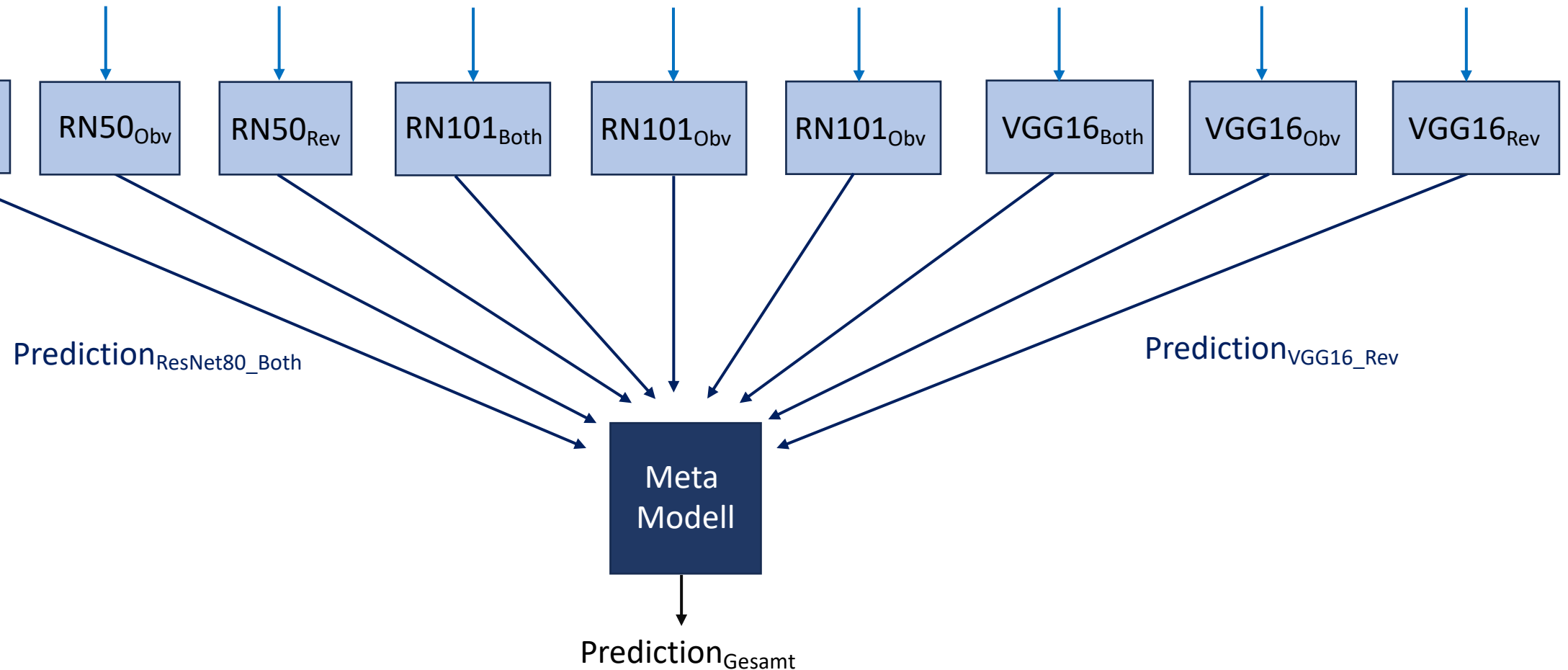
Bei Gleichstand: Anderen Voter verwenden

Stacking - Training des Meta Models



Stacking - Evaluation des Meta Models

Münzbilder des Testsets

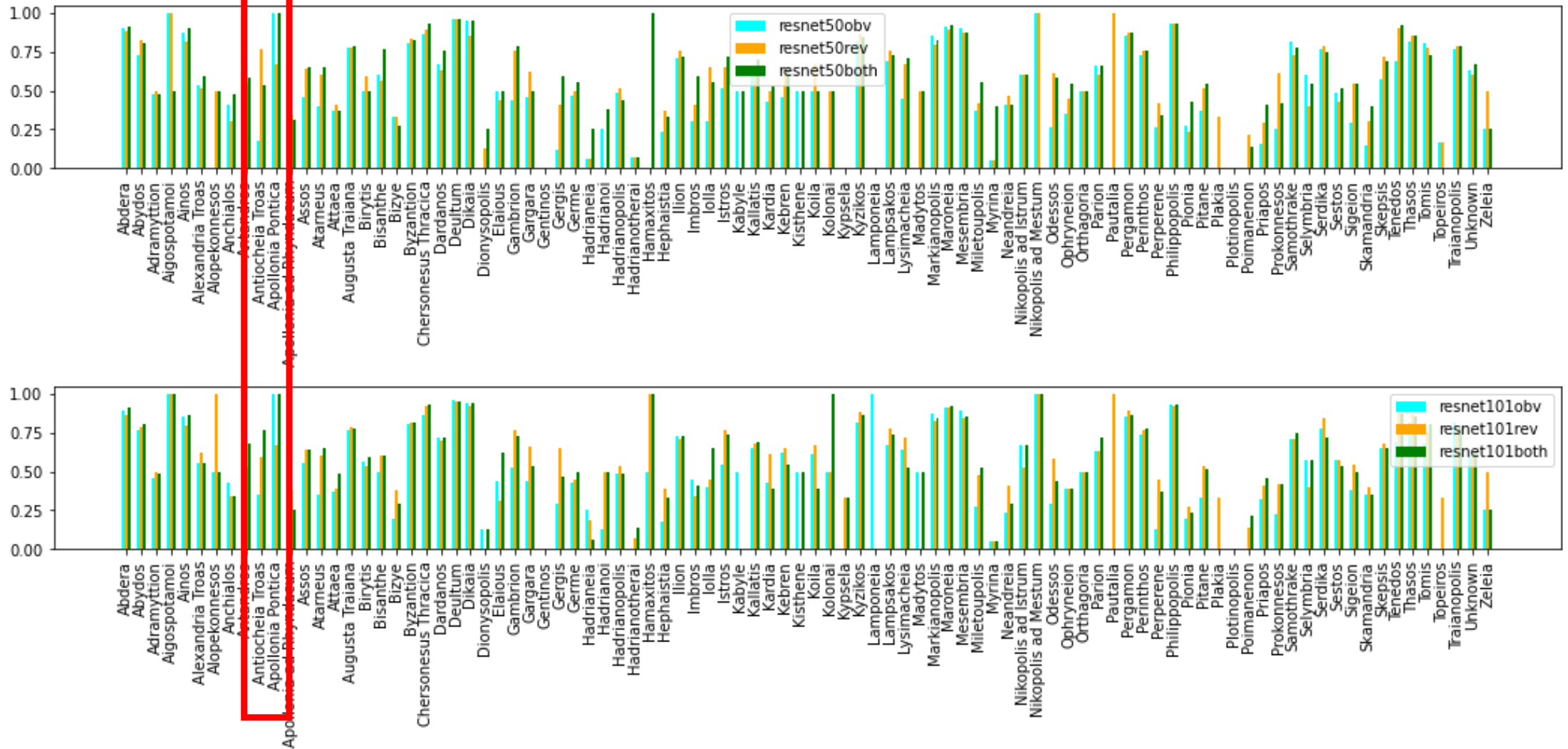


Anmerkung: RN = ResNet

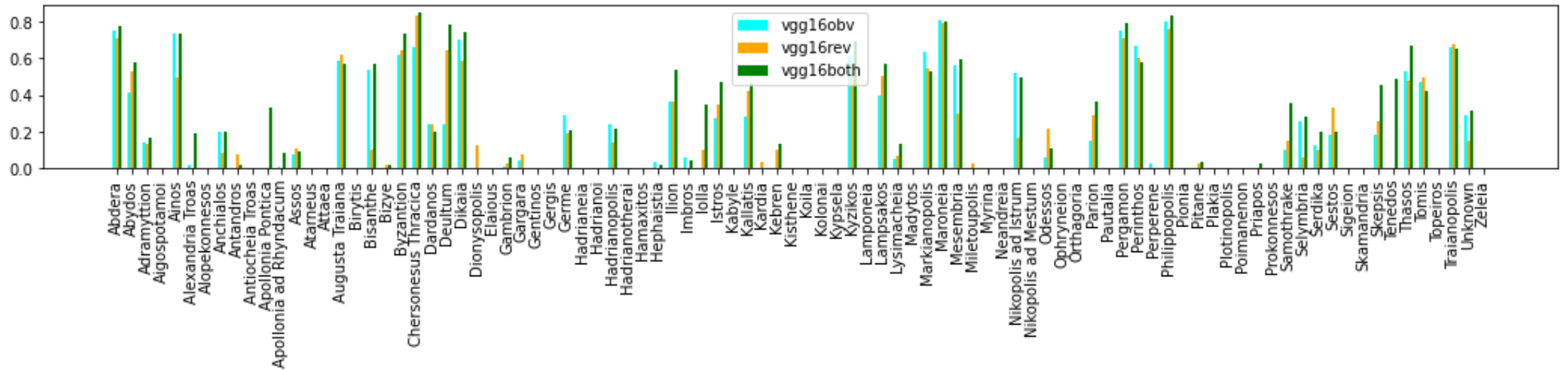
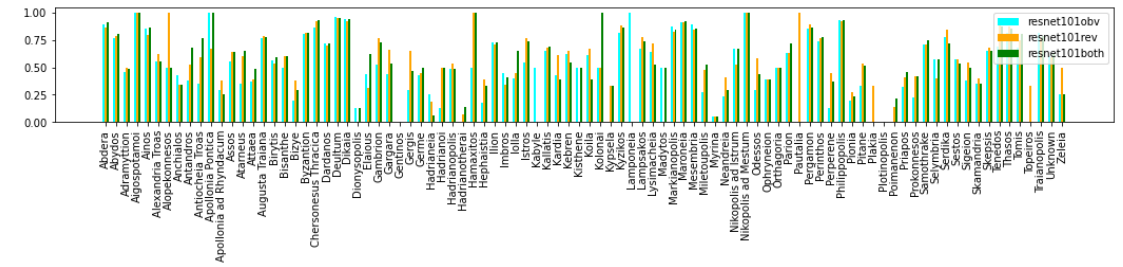
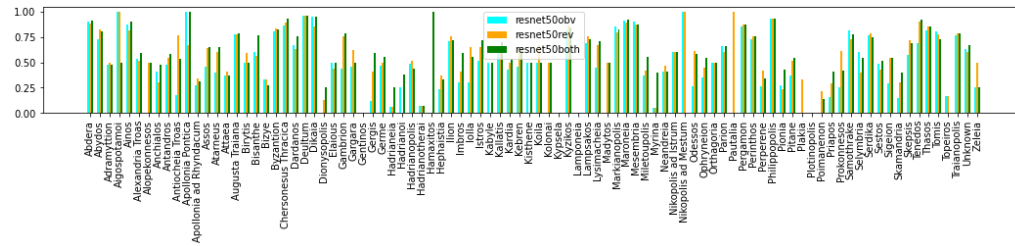


Ergebnisse

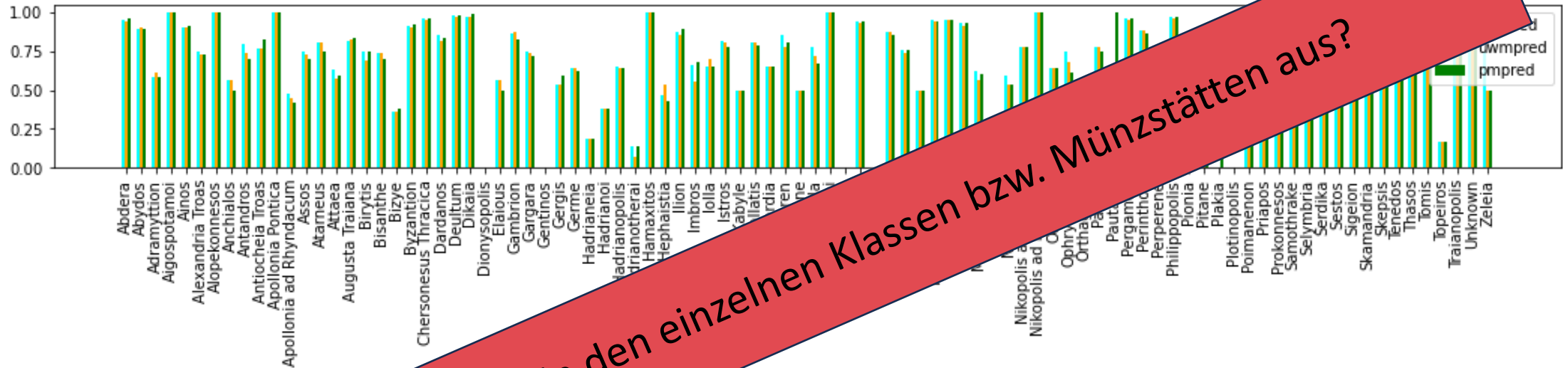
Vergleich der Base Models - ResNet



Vergleich der Base Models – ResNet und VGG



Vergleich der Voter – ohne VGG16



- Voting auf 11.8
- Richtige Antwort: 11.8

Majority: 10.248 Münzbilder
 Simple Majority: 10.151 Münzbilder
 Qualified Majority: 10.187 Münzbilder



Weighted Majority Voter insgesamt am besten!
Accuracy: ca. 86,44%

Vergleich der Voter in den einzelnen Klassen

- **Frage:**

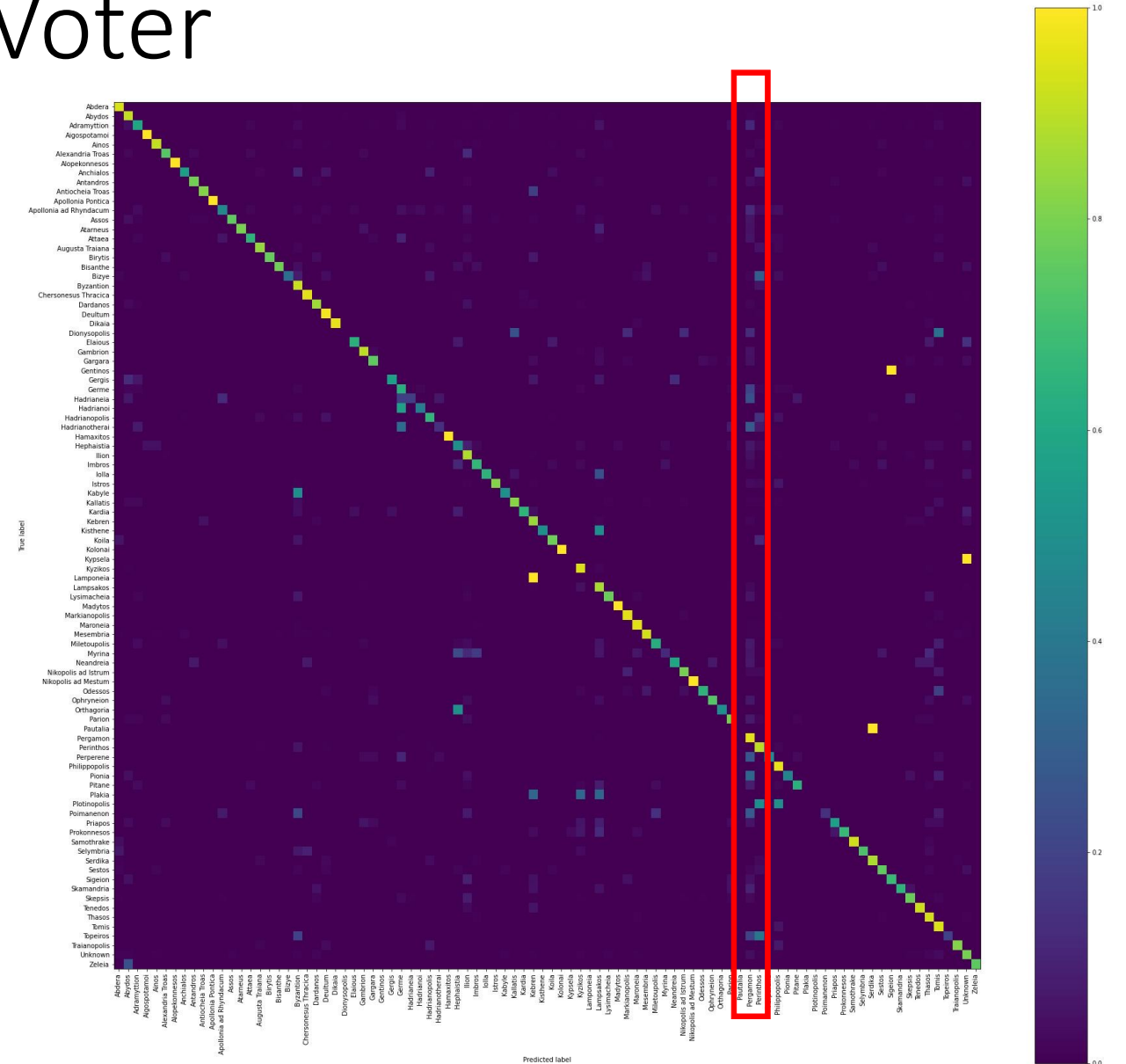
Wie viele Bilder wurden in den unterschiedlich großen Klassen richtig erkannt?

	≤ 10 Bilder	≤ 100 Bilder	≥ 100 Bilder	≥ 200 Bilder	≥ 1000 Bilder
Unweighted	38,60%	65,05%	89,95%	91,65%	95,39%
Weighted	40,35%	66,94%	90,54%	92,26%	96,05%
Probability	42,11%	64,71%	90,31%	92,35%	96,13%

Weighted Majority Voter für die Klassen mit bis zu 199 Bildern am besten und Probability Majority Voter für die Klassen mit weniger als 10 Bildern und mehr als 200 Bildern am besten!

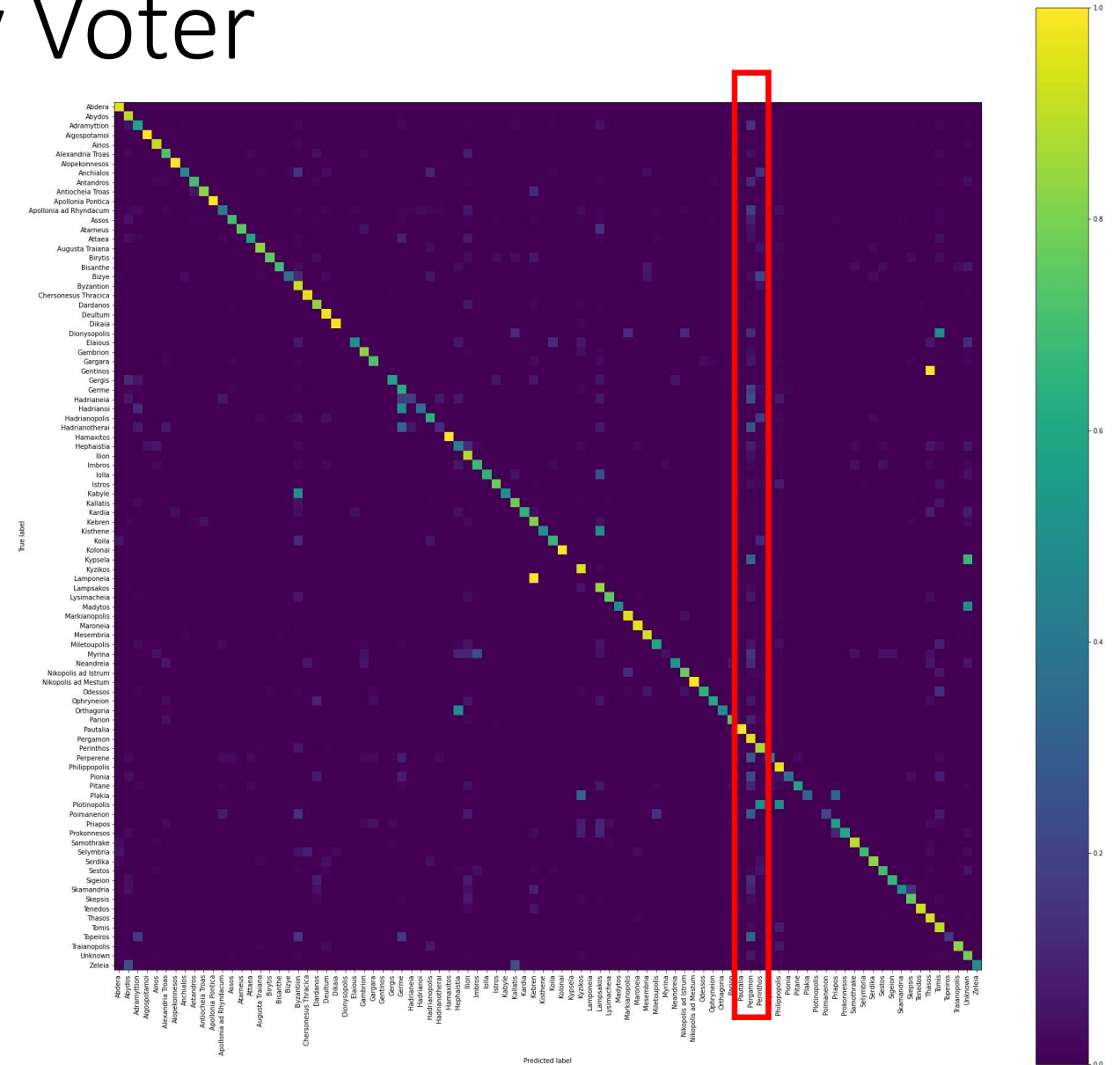
Weighted Majority Voter

Overfitting Problem besteht weiterhin!



Probability Majority Voter

Overfitting Problem besteht weiterhin!



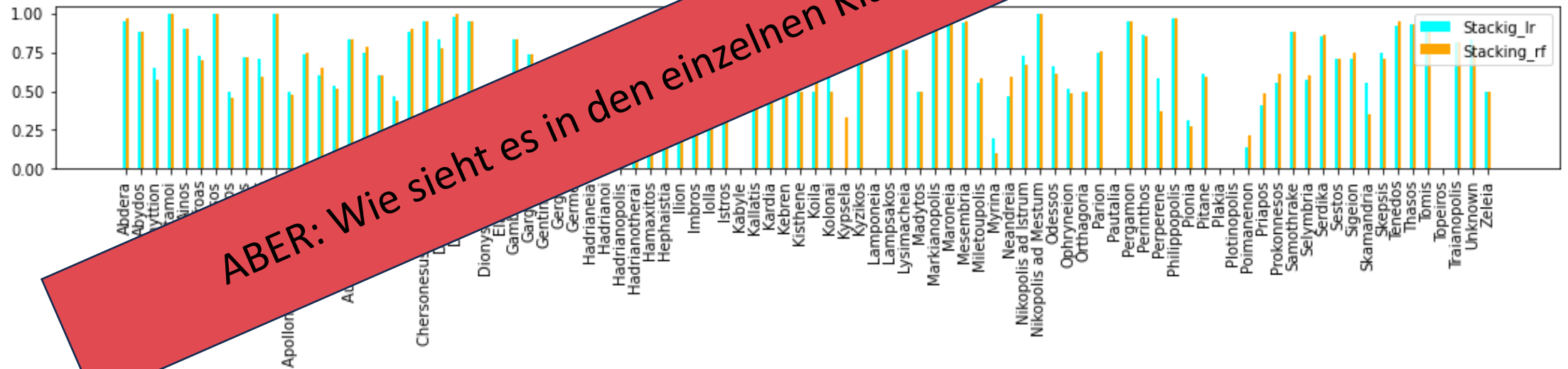
Stacking – Vergleich der Meta Models

Logistic Regression (LR):

- Erzeugt immer eine Prediction!
 - Richtig: 10.144 Münzbilder
- Accuracy: 85.56%

Random Forest (RF):

- Bei 3.305 Bildern aus der Münzstättendatenbank wird eine Coin-Probability erzeugt
- ⇒ Wählst du dir ein Münzbild aus? → Coin-Probability Prediction!



mer eine Prediction!

: 10.144 Münzbilder

85.56%

Bei 3.305 Bildern aus der Testmenge wird eine Prediction erzeugt.

⇒ Wählst du die richtige Münzstätte?

97.26%

ABER: Wie sieht es in den einzelnen Klassen bzw. Münzstätten aus?

A bar chart titled 'Accuracy by Coin Type' comparing predicted vs. actual coin types. The x-axis lists various ancient Greek city-states (e.g., Athens, Corinth, Sparta). The y-axis shows accuracy percentages from 0% to 100%. Each city-state has two bars: a blue one for 'Predicted' and an orange one for 'Actual'. Most accuracies are between 80% and 100%, with some outliers like 'Sparta' at 100% and 'Corinth' at approximately 95%.

Logistic Regression ist minimal besser!

Vergleich der Meta Models in den Klassen

- **Frage:**

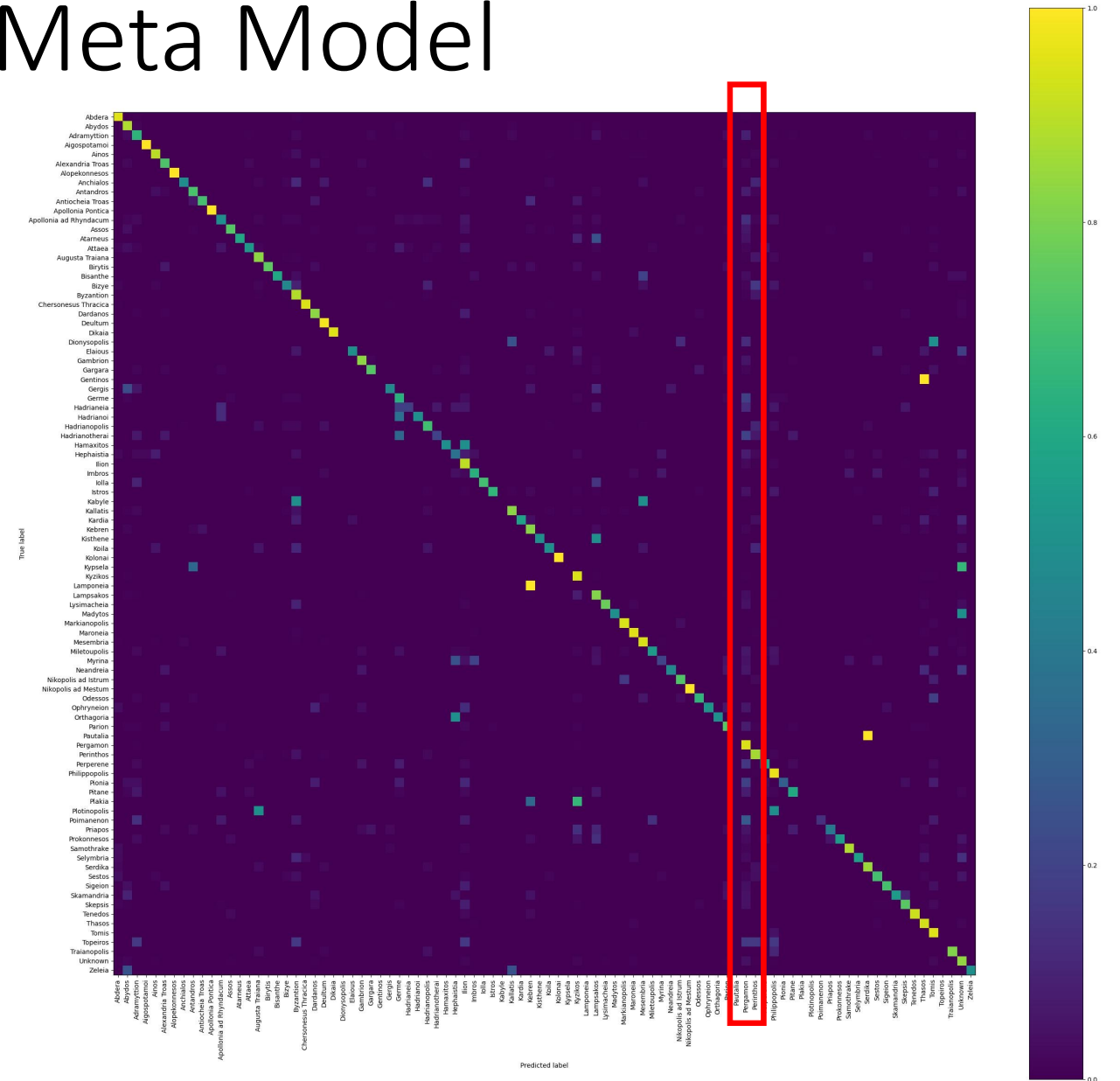
Wie viele Bilder wurden in den unterschiedlich großen Klassen richtig erkannt?

	≤ 10 Bilder	≤ 100 Bilder	≥ 100 Bilder	≥ 200 Bilder	≥ 1000 Bilder
LR	35,09%	63,89%	90,31%	91,89%	95,70%
RF	36,84%	62,53%	90,05%	92,14%	95,55%

Logistic Regression ist bei bis zu 199 Bildern und ab 1000 Bildern minimal besser und Random Forest ist bei weniger als 10 Bildern und bis 999 Bildern minimal besser!

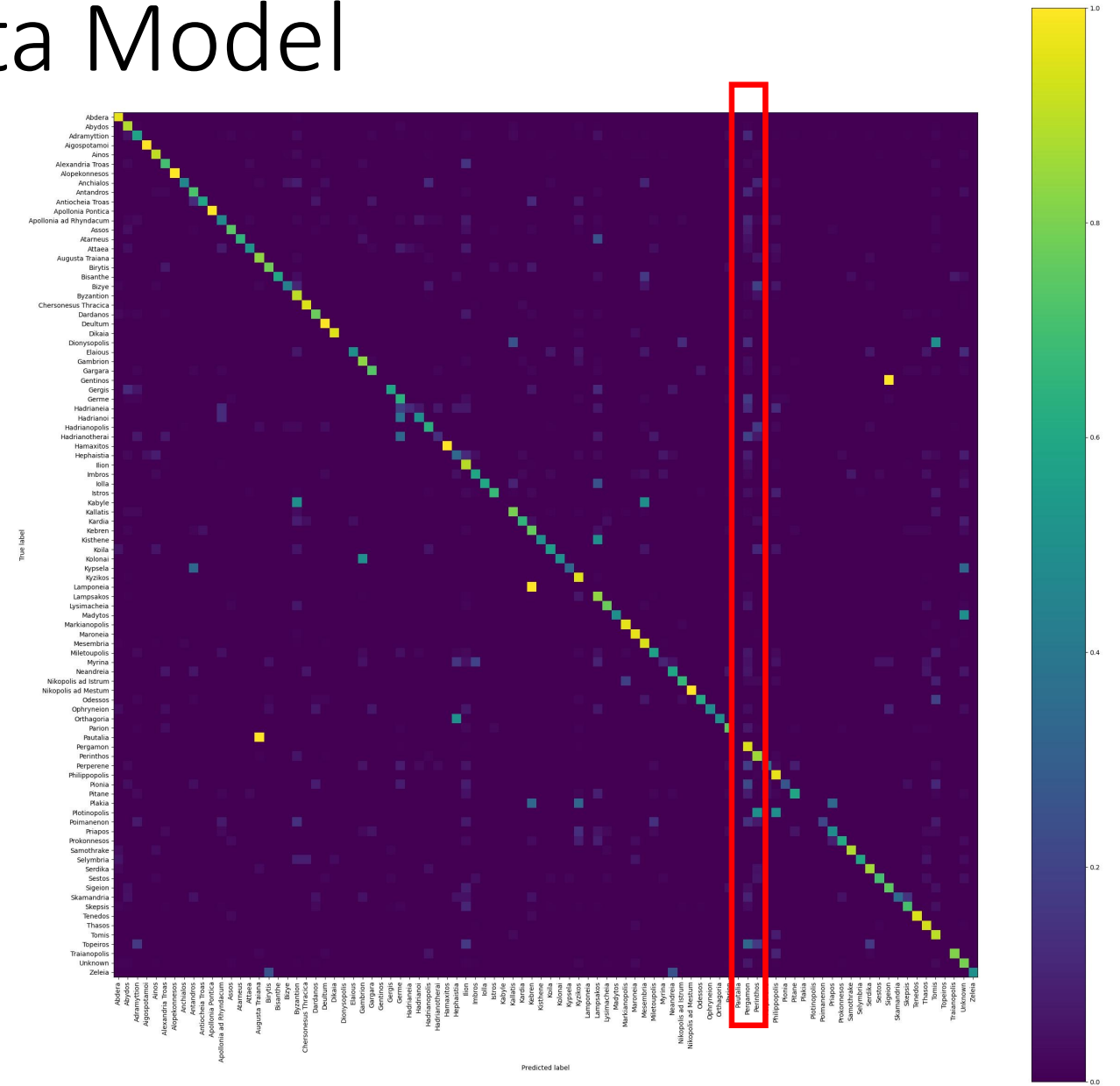
Logistic Regression Meta Model

Overfitting Problem besteht weiterhin!



Random Forest Meta Model

Overfitting Problem besteht weiterhin!



Einzelfallanalyse von Apollonia ad Rhyndacum

- 37 von 74 Münzbilder wurden beim LR-Stacking falsch erkannt
 - Vorderseite: meist ein Kopf (im Profil)
 - Rückseite: unterschiedlich mit schlechter Qualität
 - Viele wurden Pergamon zugeordnet
 - I.d.R. war die Abbildung auf der Rückseite nicht erkennbar
 - Gipsmünzen wurden prozentual öfter richtig zugeordnet
 - Aber: Verteilung im Trainingsset gleich

Einzelfallanalyse von Apollonia ad Rhyndacum

Münzen aus dem Testset (falsch zugeordnet):



Münzen aus dem Trainingsset:



Einzelfallanalyse von Apollonia ad Rhyndacum

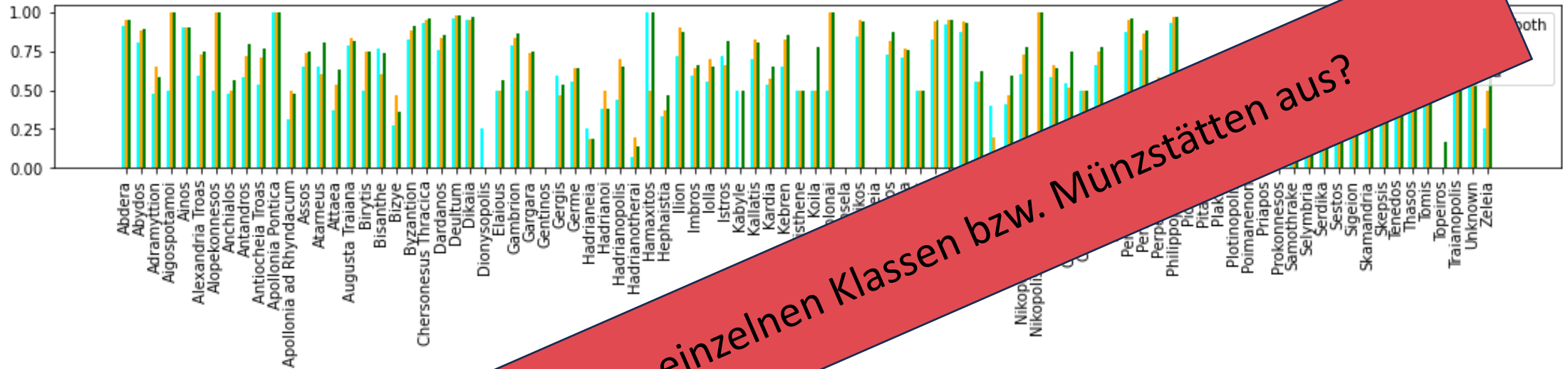
Münzen aus dem Testset (falsch zugeordnet):



Münzen aus dem Trainingsset:



Voting vs. Stacking



- Voting und Stachel
- Richtige Parameterwahl (Trainset, Testset)

ABER: Münzbilder

Majority: 10.248 Münzbilder

Regression: 10.144 Münzbilder



Weighted Majority Voter insgesamt am besten!
Accuracy: ca. 86,44%

Voting vs. Stacking

- **Frage:**

Wie viele Bilder wurden in den unterschiedlich großen Klassen richtig erkannt?

	≤ 10 Bilder	≤ 100 Bilder	≥ 100 Bilder	≥ 200 Bilder	≥ 1000 Bilder
Weighted	40,35%	66,94%	90,54%	92,26%	96,05%
LR	35,09%	63,89%	90,31%	91,89%	95,70%

Voting ist minimal besser als Stacking!

Zusammenfassung

- Voting und Stacking sind besser als die einzelnen Base Models
 - Bestes Base Model (ResNet50_{Both}) bei ca. 77% Accuracy
 - Voting und Stacking bei ca. 85-86% Accuracy
- Voting ist besser bei Münzstätten mit maximal 100 Bildern
 - Probability Voting erkennt ca. 42% richtig
 - Ab 100 Bildern pro Münzstätte ist Voting nur minimal besser als Stacking
- Problem des Overfittings besteht weiterhin!



Ausblick

Ausblick

- Mehr Base Models trainieren
- Base Models unterschiedlich gewichten beim Voting
 - Wie gut sind die einzelnen Base Models in den einzelnen Klassen?
⇒ Gute Vorhersagewahrscheinlichkeit wird stärker gewichtet
- Bei mehr Bildern pro Klasse: Stacking auf dem Validationset
 - Base Models und Meta Model würden auf unterschiedliche Münzbildern trainiert werden
 - Mögliche Verbesserung der Accuracy
 - Base Models werden auf das Trainingsset trainiert
 - Meta Model wird auf das Validationset trainiert
 - Evaluation auf das Testset

Quellen

1. Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115, 105151.
2. Sarkar, D., & Natarajan, V. (2019). *Ensemble Machine Learning Cookbook: Over 35 practical recipes to explore ensemble machine learning techniques using Python*. Packt Publishing Ltd.
3. Wikipedia contributors. (2023, June 23). Ensemble learning. In *Wikipedia, The Free Encyclopedia*. Retrieved 11:01, July 20, 2023, from https://en.wikipedia.org/w/index.php?title=Ensemble_learning&oldid=1161590114