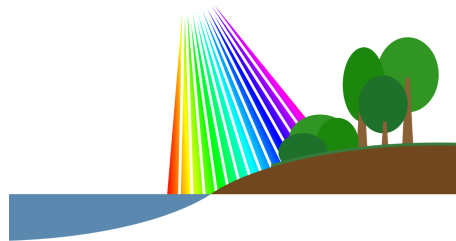


Vegetation Basisklassen Nonnenwerth Sommeraufnahme 2019

mDRONES4rivers



Kontakt

mDORNES4rivers@bafg.de
Bundesanstalt für Gewässerkunde (BfG)
Am Mainzer Tor 1, 56068 Koblenz
www.bafg.de



Bearbeiter

Edvinas Rommel
Laura Giese
Frederik Kathoefer

Projektkoordination

Dr. Björn Baschek
Tel. 0261 / 1306 5395
baschek@bafg.de

Projektpartner

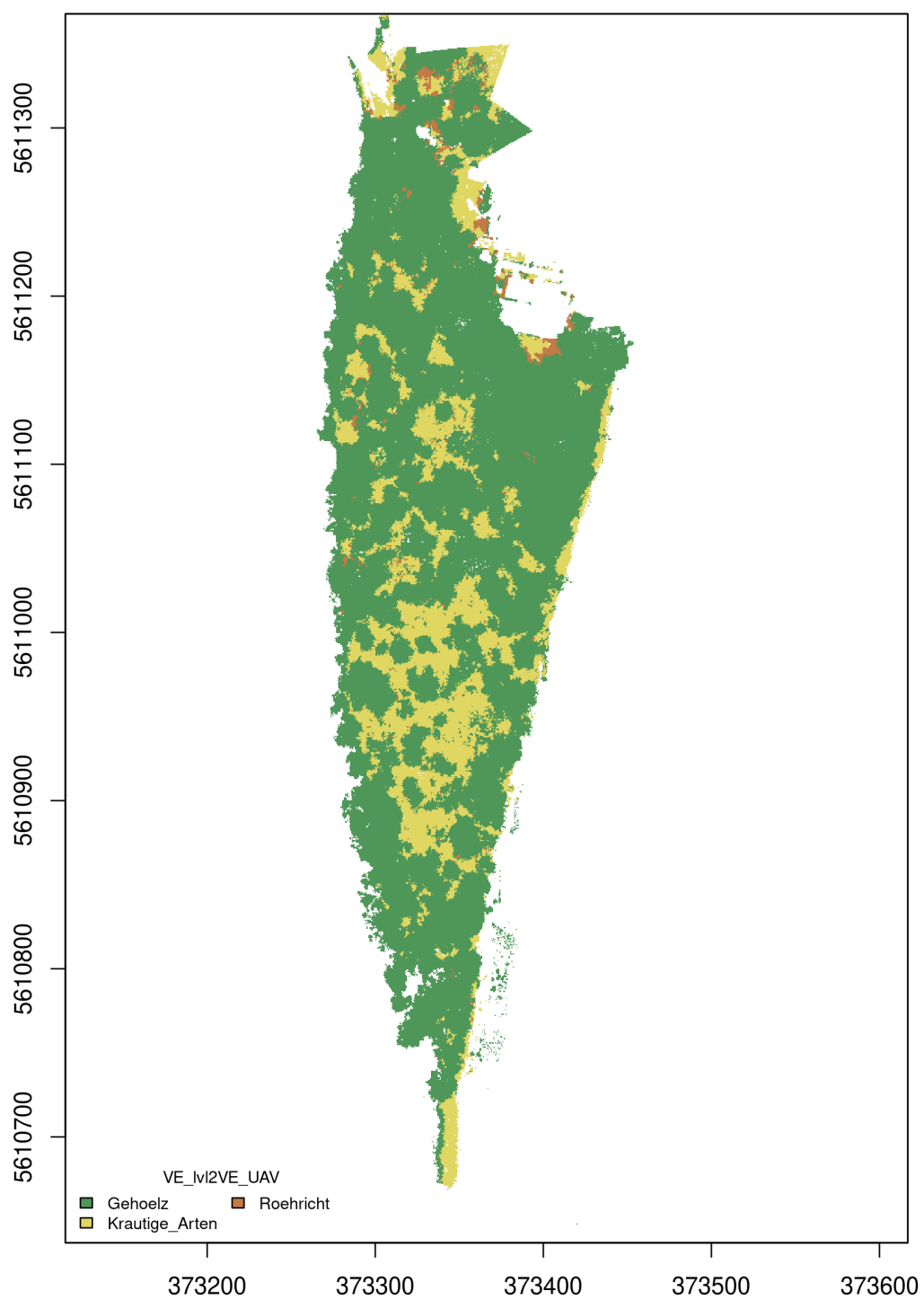


Verwendeter Algorithmus

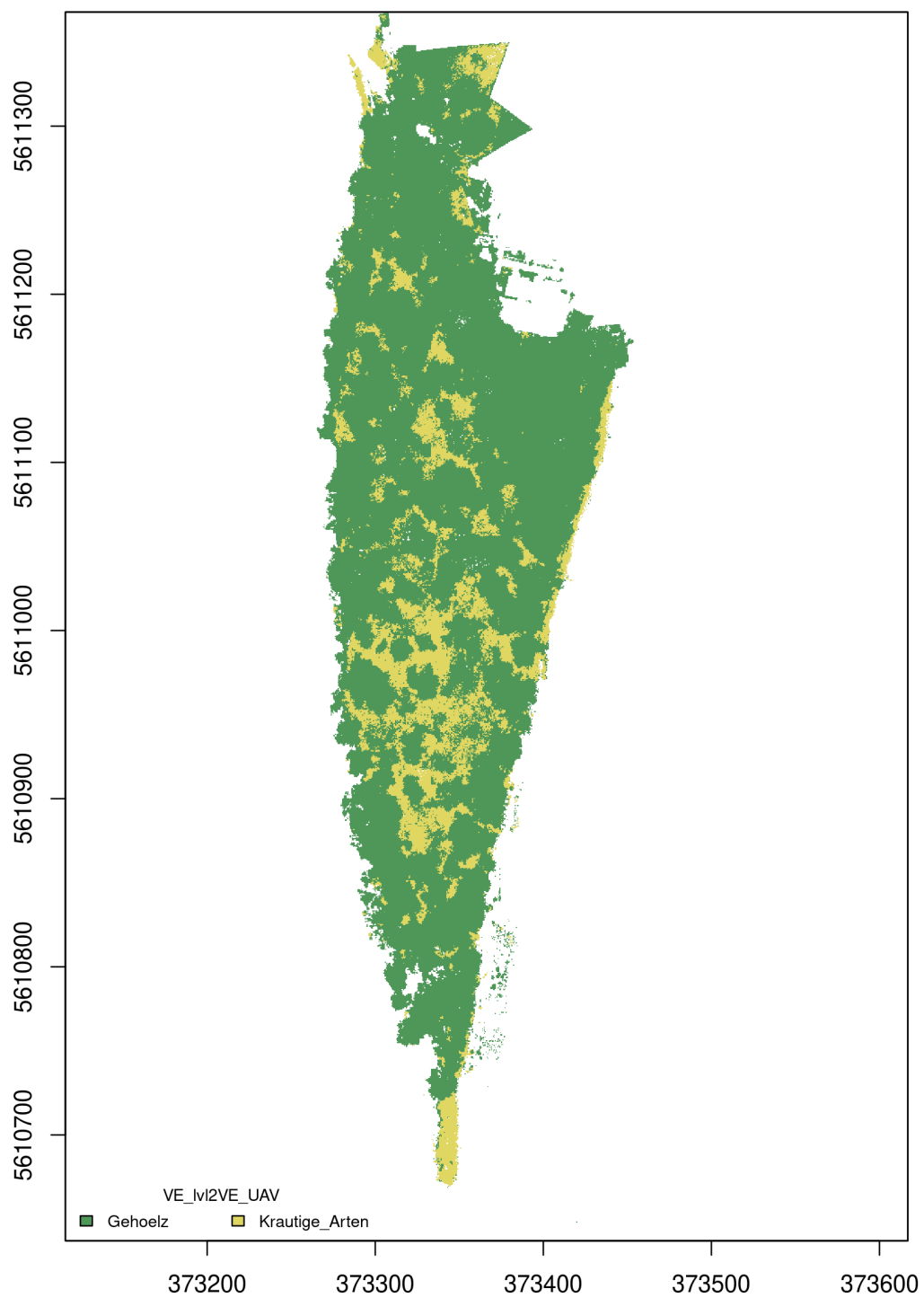
Die Klassifikation wurde mit Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost) durchgeführt

Übersichtskarte

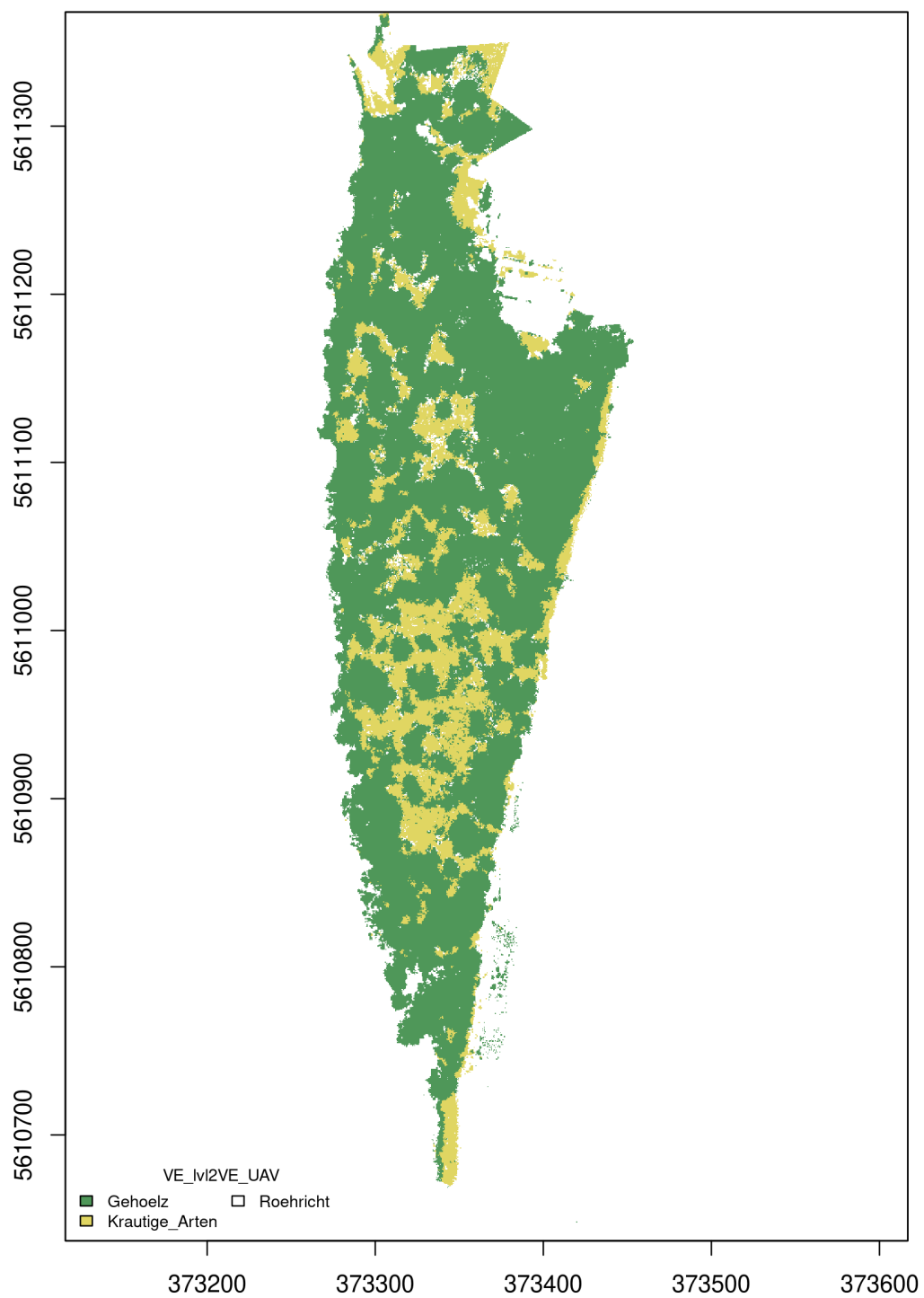
VE_lv12_NoW_2019_3 RF



VE_lvl2_NoW_2019_3 SVM



VE_lvl2_NoW_2019_3 XGBOOST



Klassifikationsgüte

Die Klassifikationsgüte wird anhand einer Konfusionsmatrix und deren abgeleitete Gütemaße dargestellt.

Die Zahlen in der Konfusionsmatrix sind die aufsummierten Werte von 500 Iterationen basierend auf dem Bootstrap Resampling Verfahren. Im Bootstrap Resampling werden die vorhandenen in situ Daten zufällig in Trainings- und Validierungsdaten aufgeteilt. Die tatsächliche Anzahl der in situ Daten pro Klasse ergibt sich aus: $x = \text{Wert in der Konfusionsmatrix} / 500 \text{ Iterationen}$.

Von den aufsummierten Werten wurden die Gütemaße wie User und Producer-Accuracy berechnet.

User's Accuracy (Nutzergenauigkeit):

- Anzahl richtig klassifizierter Objekte der Klasse / Anzahl aller Objekte der Klasse
- fehler der Übererfassung von Objekten
- z.B. zehn Bäume wurden real gezählt, aber nur acht wurden insg. klassifiziert

Producer's Accuracy (Herstellergenauigkeit):

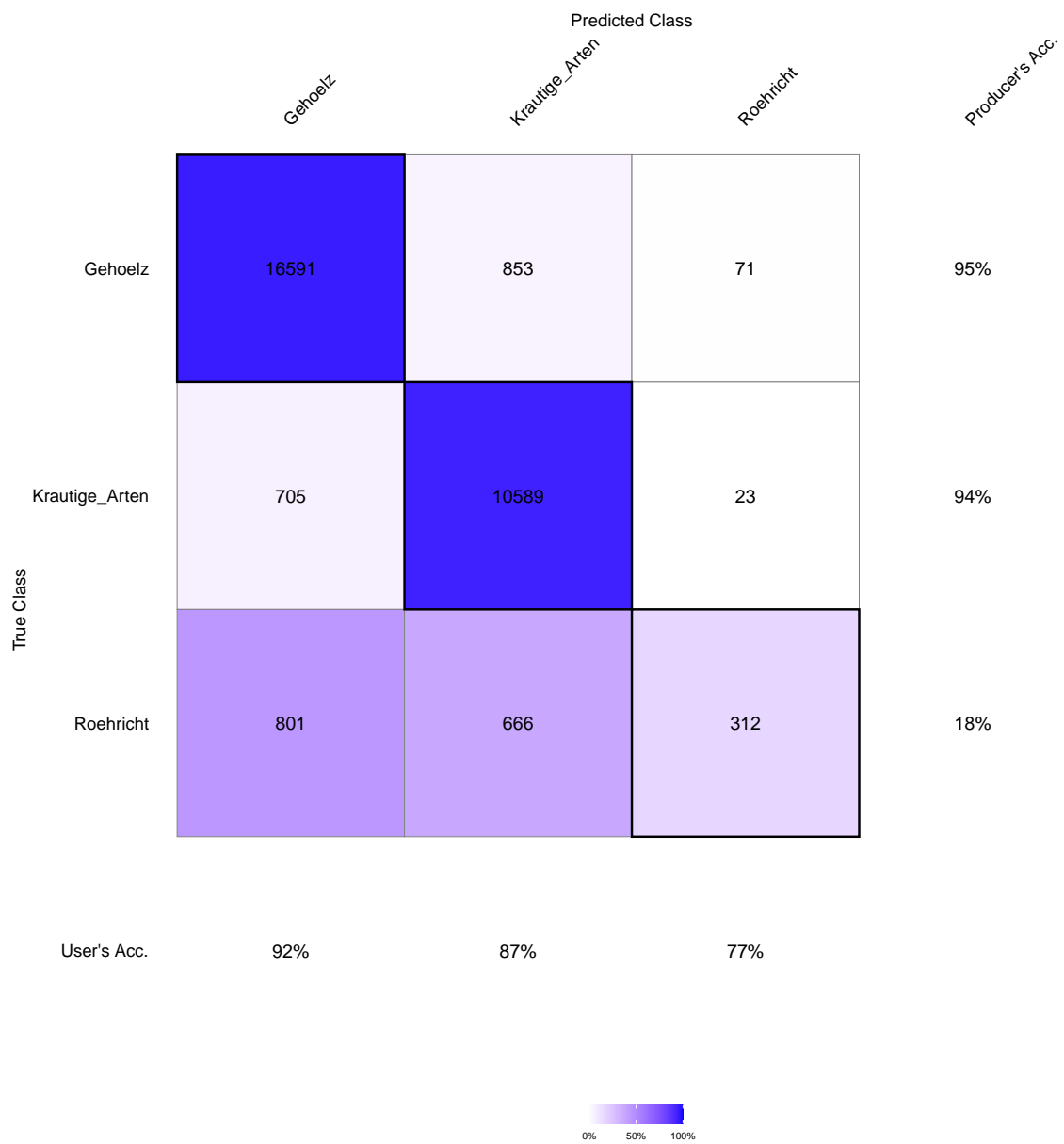
- Anzahl richtig klassifizierter Objekte der Klasse / Anzahl aller klassifizierter Objekte der Klasse
- fehler der Untererfassung von Objekten
- z.B. zwölf Bäume wurden klassifiziert, aber nur zehn wurden real gezählt

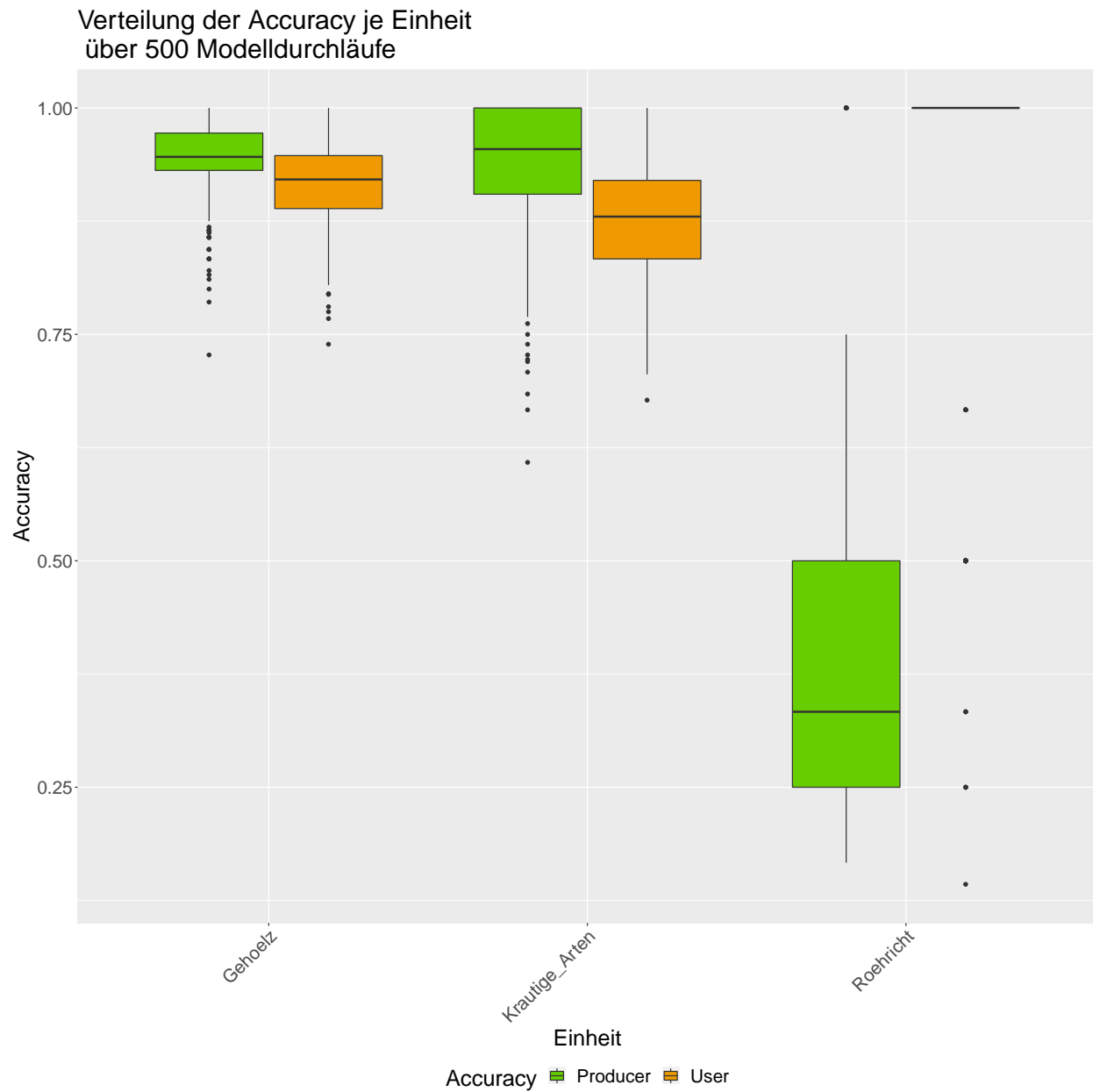
Random Forest (RF)

Overall Accuracy: **89.8 %**

Kappa: **0.8**

Confusion Matrix RF :



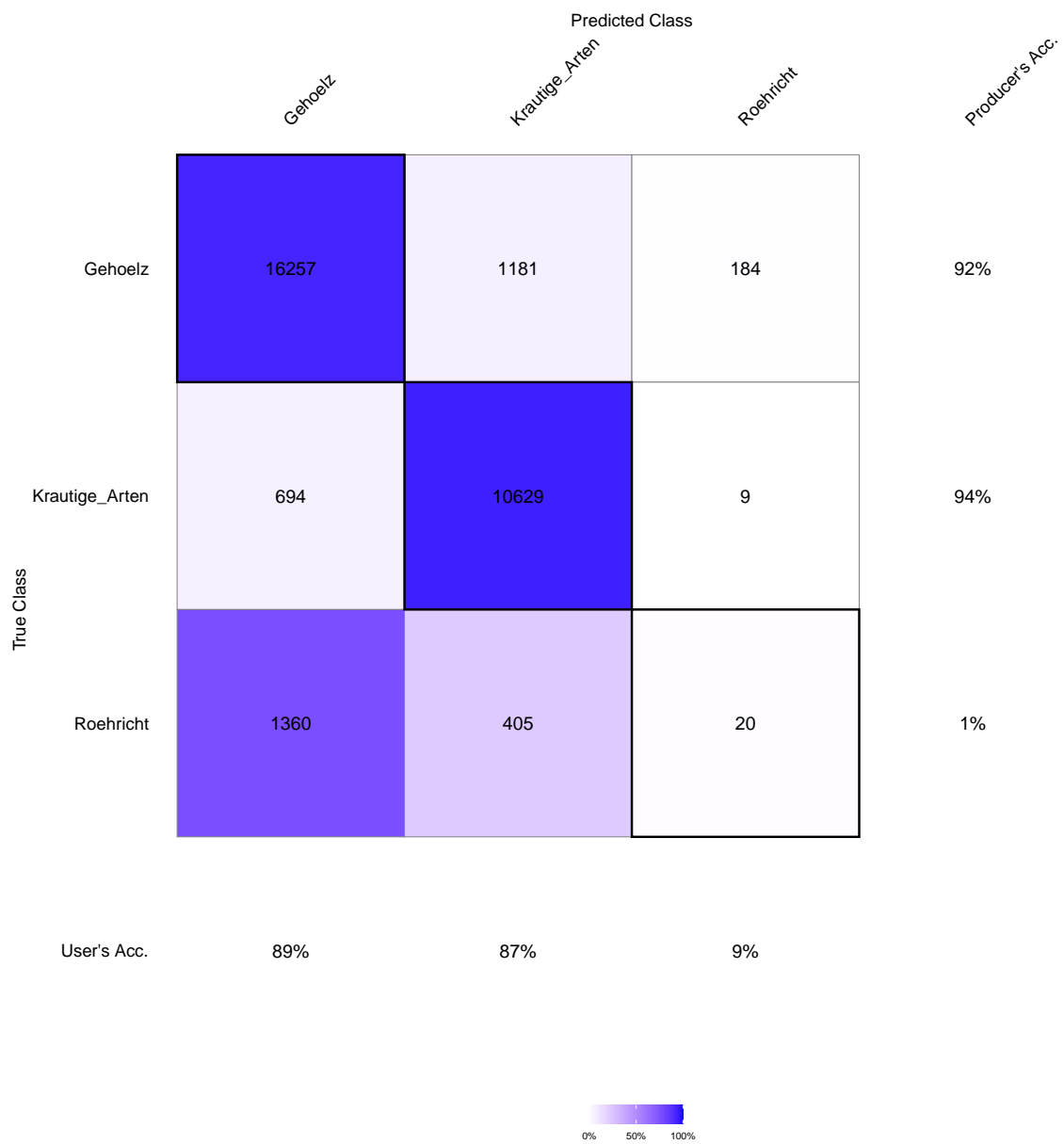


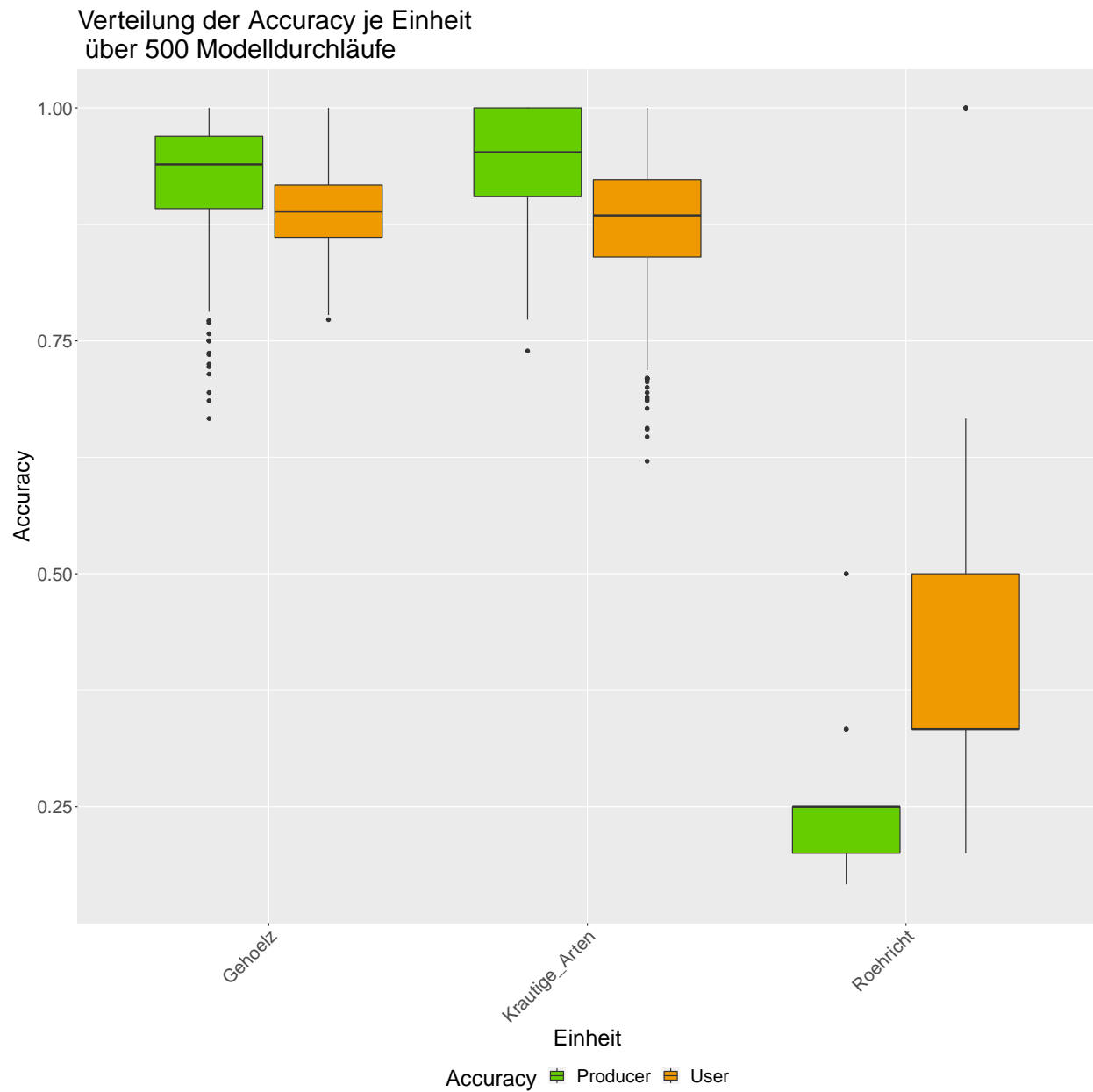
Support Vector Machine (SVM)

Overall Accuracy: **87.5 %**

Kappa: **0.76**

Confusion Matrix SVM :



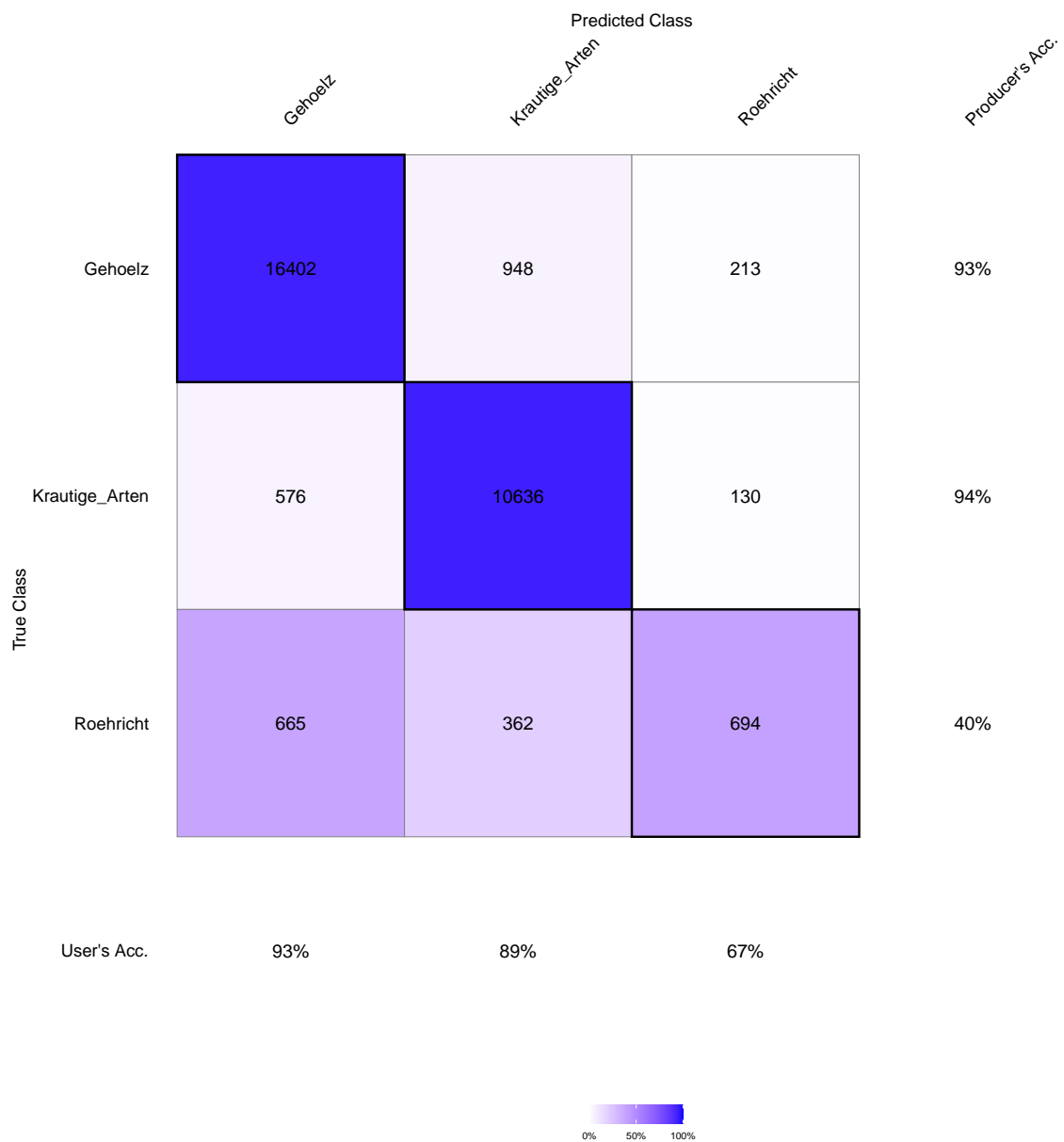


Extreme Gradient Boosting (XGBoost)

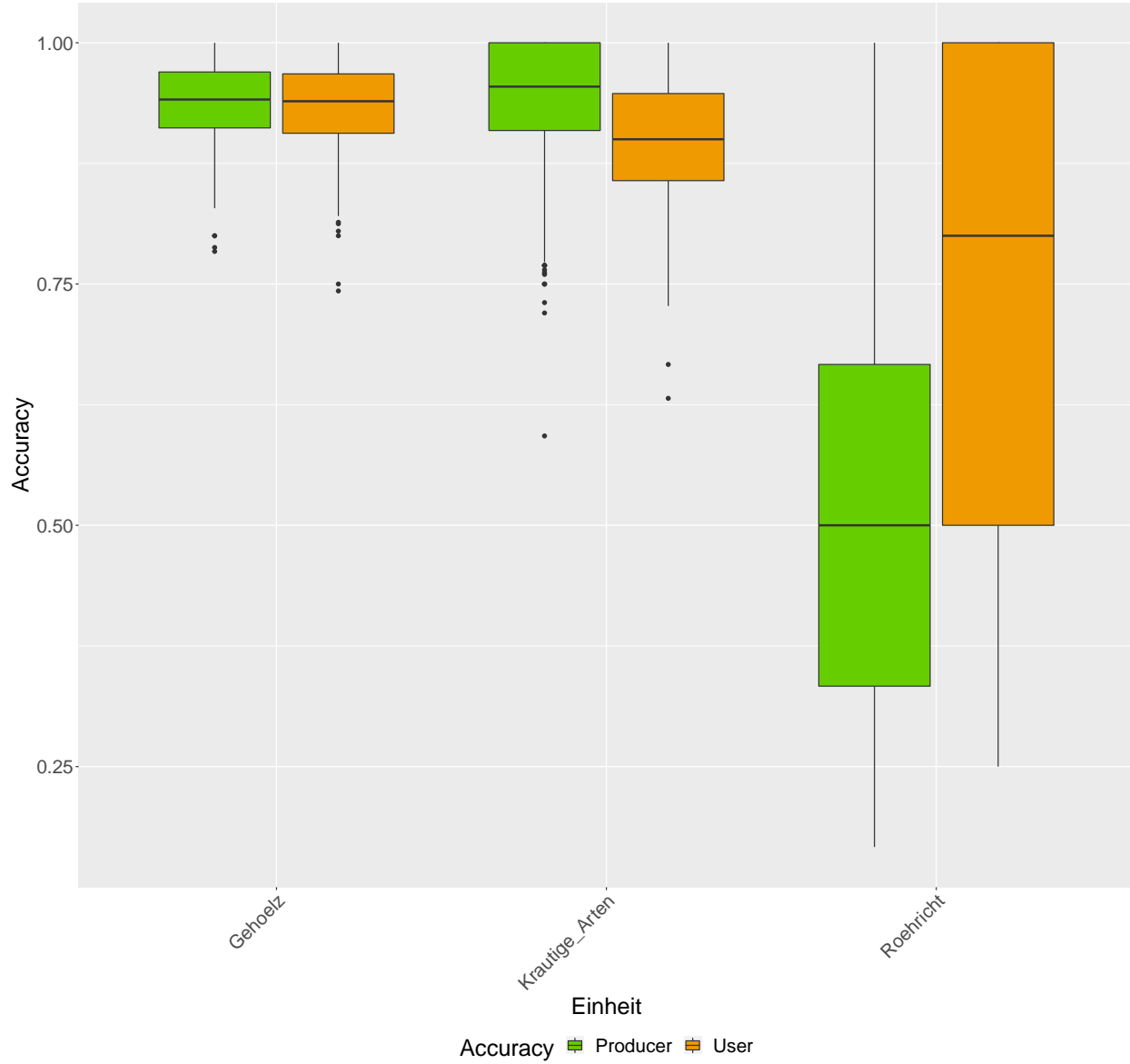
Overall Accuracy: **90.6 %**

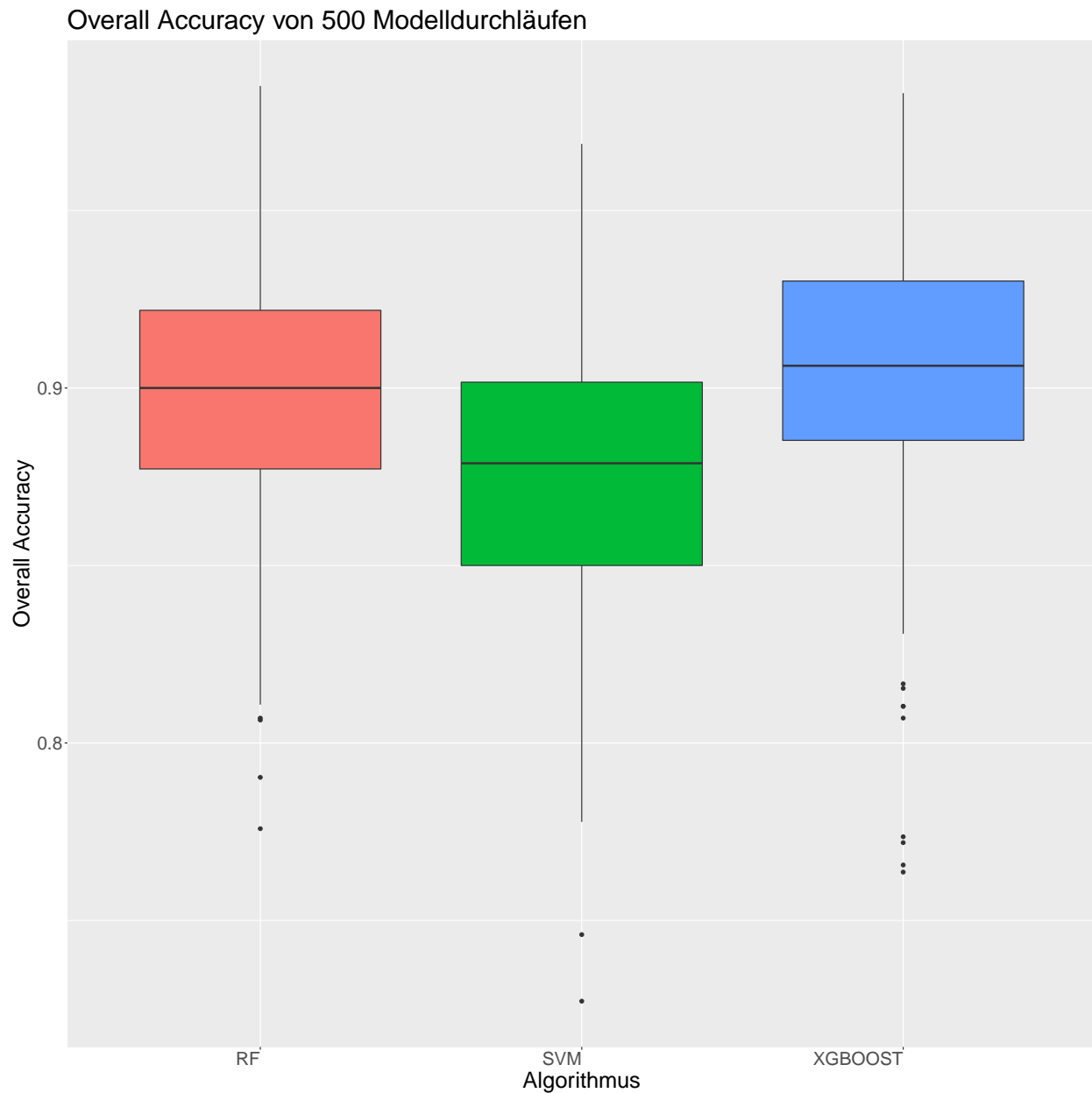
Kappa: **0.82**

Confusion Matrix XGBOOST :



Verteilung der Accuracy je Einheit
über 500 Modelldurchläufe





Modelleigenschaften

Hyperparameter und deren Einstellungen:

Algorithmus: RF

Model for learner.id=classif.ranger; learner.class=classif.ranger Trained on: task.id = impurity; obs = 168; features = 20 Hyperparameters: num.threads=12,verbose=FALSE,respect.unordered.factors=order,num.trees=128,mtry=8

Algorithmus: SVM

Model for learner.id=classif.ksvm; learner.class=classif.ksvm Trained on: task.id = impurity; obs = 168; features = 6 Hyperparameters: fit=FALSE,C=804,sigma=0.00128

Algorithmus: XGBOOST

Model for learner.id=classif.xgboost; learner.class=classif.xgboost Trained on: task.id = impurity; obs = 168; features = 6 Hyperparameters: nrounds=316,verbose=0,nthread=12,max_depth=9,eta=0.322,lambda=0.126

Variablen Selektion

Um die Gefahr einer Überanpassung des Models zu verringern und die Berechnungszeit zu verkürzen erfolgt eine Auswahl von wichtigen Variablen. Als Variablenselektionsmethode wurde die Algorithmunenunabhängige Methode ranger-impurity aus dem R-Paket “ranger” gewählt. Ranger impurity ordnet die Variablen nach deren Wichtigkeit auf Basis der eines Entscheidungsbaumes indem die Knotenreinheit mit dem Gini Index berechnet wird. Je Höher der Wert, desto wichtiger die Variable.

Die optimale Anzahl an Variablen wird mit Hilfe einer “forward feature selection”. Hierbei wird in einem iterativen Verfahren die Anzahl an Variablen schrittweise erhöht und ein Modell erstellt (angefangen bei der wichtigsten Variable). Es werden solange Variablen hinzugefügt bis eine Sättigung der Modelgüte erreicht ist.

Die wichtigsten Variablen sind im Nachfolgenden in absteigender Reihenfolge dargestellt.

Table 1: Ausgewählte Variablen für RF

Variable	Wert
mn_nDOM	14.44
mdiff_neig	11.16
sd_nDOM	9.753
sd_UFD_sd	2.39
mn_ndre	2.263
mn_UFD_mn	2.18
mn_sr	1.959
mn_gvi	1.834
sd_UFD_mn	1.8
mn_NIR	1.755
sd_ndvi	1.703
sd_sr	1.672
sd_gvi	1.558
mn_UFD_sd	1.376
mn_ndwi	1.376
sd_ndwi	1.316
mn_gndvi	1.283
sd_savi	1.221
diss_NIR	1.142
sd_totbr	1.116

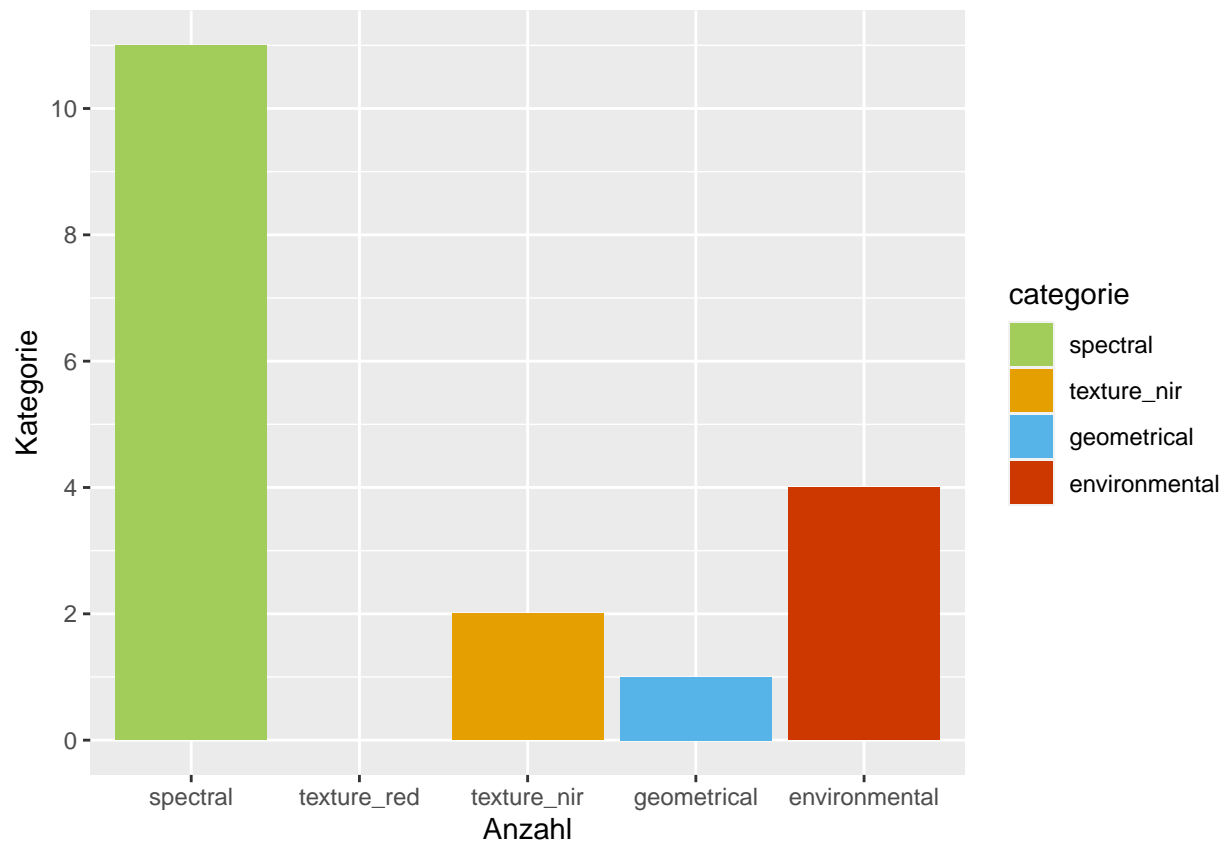


Table 2: Ausgewählte Variablen für SVM

Variable	Wert
mn_nDOM	15.8
mdiff_neig	10.92
sd_nDOM	9.759
sd_UFD_sd	3.016
mn_sr	2.19
mn_ndre	2.021

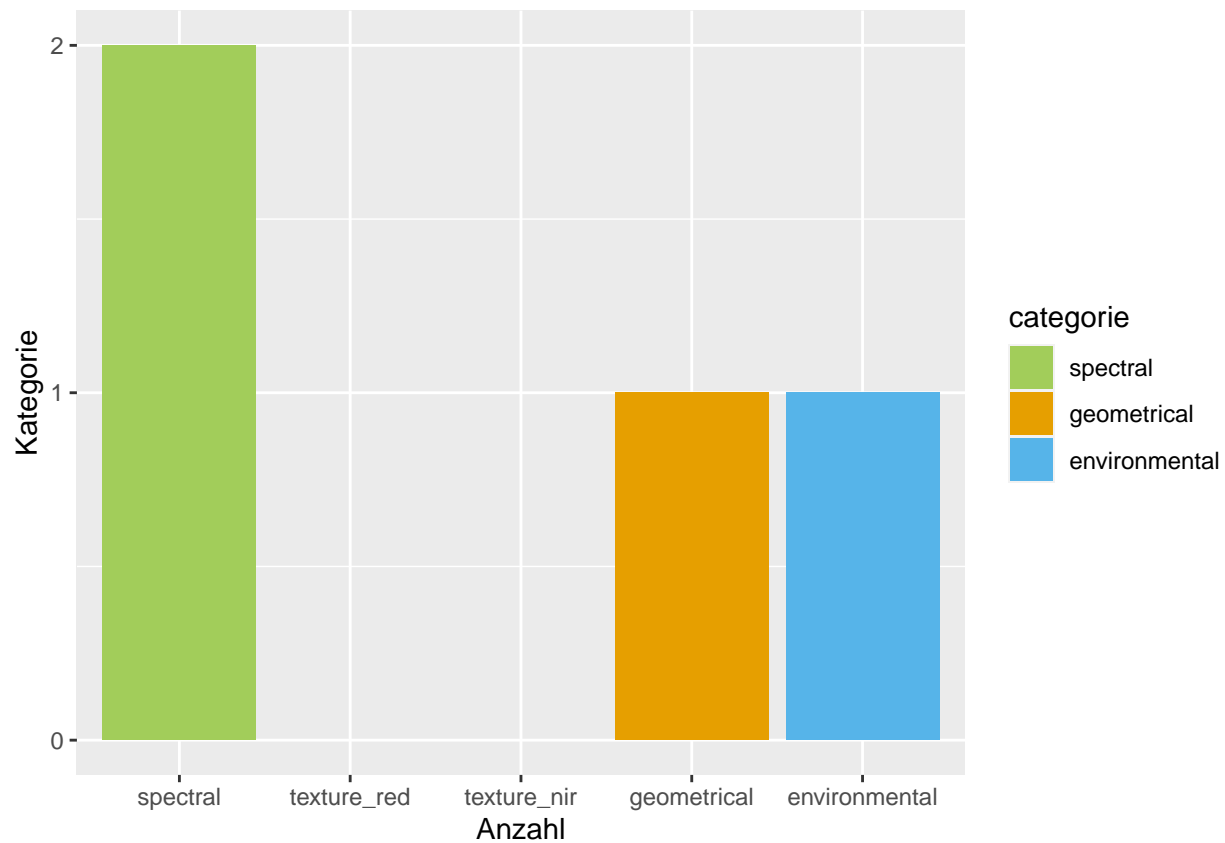
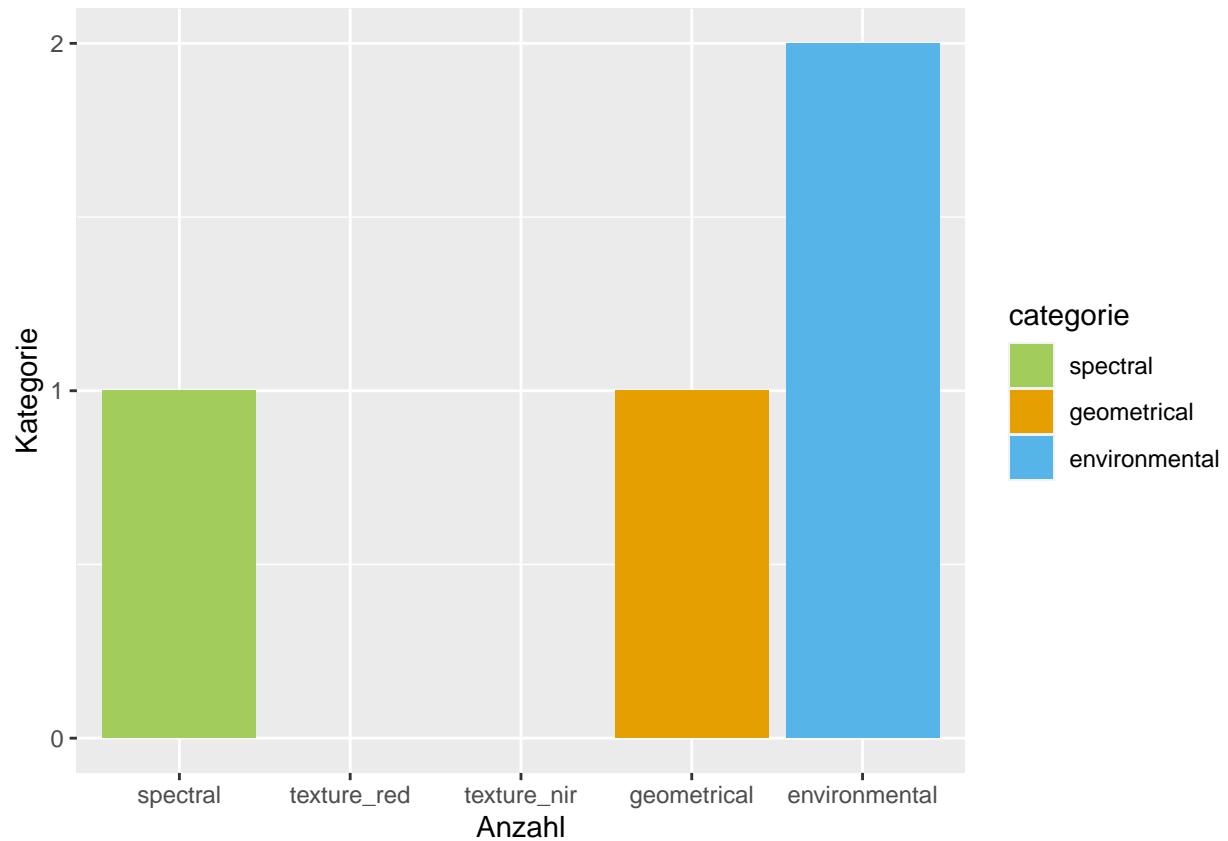


Table 3: Ausgewählte Variablen für XGBOOST

Variable	Wert
mn_nDOM	15.67
mdiff_neig	11.74
sd_nDOM	10.24
sd_UFD_sd	2.618
mn_UFD_mn	2.273
mn_ndre	1.969



Anzahl Trainings- und Validierungsdaten

Verteilung der in situ Daten über die Einheiten, die zum Trainieren und Validieren der Modelle genutzt werden:

Table 4: in situ Punkte: VE_lvl2_NoW_2019_3

Einheit	Anzahl
Gehoelz	96
Krautige_Arten	62
Roehricht	10

Gesamtzahl an in situ Daten: 168