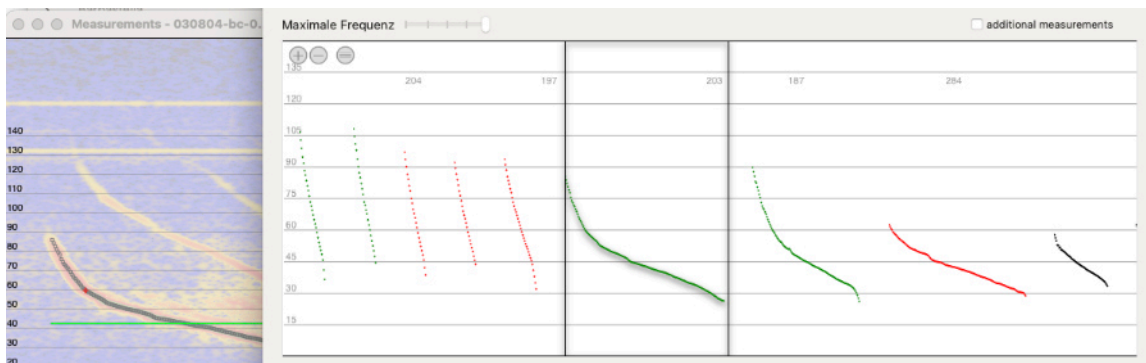


Rufbestimmung 2024

Ein bereits länger andauerndes Projekt ist die Neuauflage von batIdent ergänzt durch neue Referenzaufnahmen und neue Rufmesswerte. Seit Ende 2023 liegt ein neuer Klassifizierungsalgorithmus vor, der bereits einen Teil der neuen Aufnahmen enthält und die Klassifizierung mittels zusätzlicher, neuer Messwerte durchführt. Wir wollen Ihnen im folgenden die Datengrundlage und das Vorgehen zur Erstellung solcher neuer Klassifizierer vorstellen. Weiterhin stellen wir die Ergebnisse und die Qualitätsbeurteilung dieser Klassifizierer, die auf verschiedenen taxonomischen Ebenen arbeiten, vor. Anwender von bcAdmin mit einer Lizenz, die wenigstens Updates für das Jahr 2024 beinhaltet, können diese neuen Klassifizierer bereits vorab verwenden und mit den Ergebnissen von batIdent direkt vergleichen.

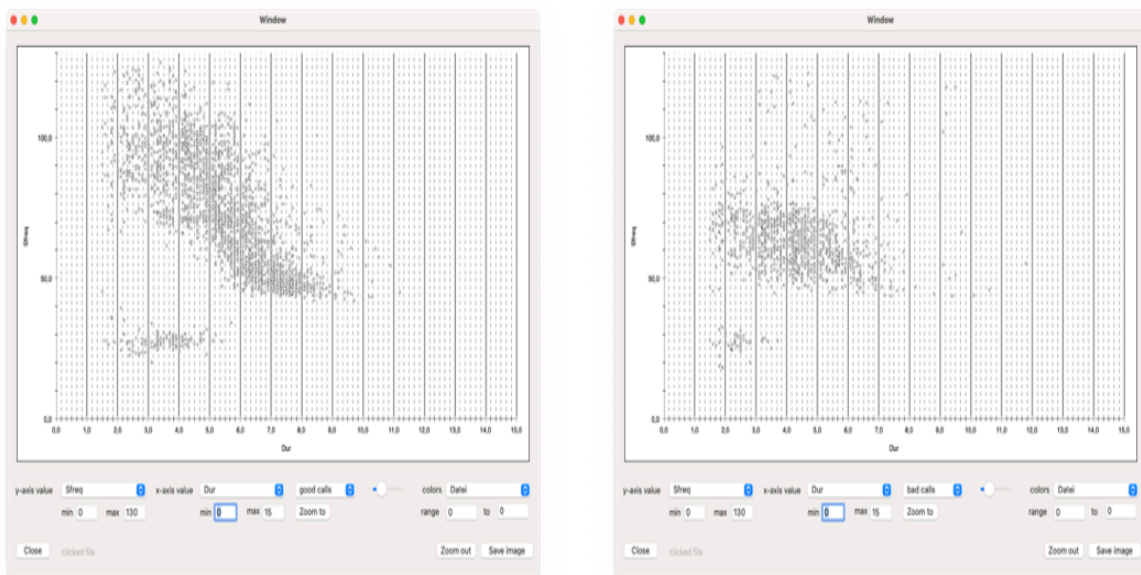
Datengrundlage

Die der Modellerstellung zugrunde liegende Datenbank enthält Referenzaufnahmen aus verschiedensten Flugsituationen. Die Aufnahmen wurden zum Großteil mit hochauflösender Aufnahmetechnik (zumeist batCorder) erstellt. Alle Aufnahmen, die als verlässliche Referenz in der Datenbank markiert sind, wurden Rufweise manuell geprüft. Alle Rufe wurden klassifiziert als „gute“, „schlechte“ oder „andere“ Rufe. Als gut wird ein Ruf dann eingetragen, wenn er möglichst vollständig aufgezeichnet wurde und die Vermessung weitestgehend vollständig ist. Schlechte Rufe sind solche, die nicht ganz aufgezeichnet wurden oder nicht vollständig vermessen wurden. Nicht klassifiziert werden solche Rufe, die nur noch als Bruchstücke vorhanden sind oder zu anderen Arten gehören. Im folgenden Screenshot aus bcRefCalls 3.0 ist ein Beispiel gezeigt mit grün = gutem Ruf und rot = schlechtem Ruf:



Beim Export der Rufmesswerte aus bcRefCalls können gezielt solche ausgegeben werden, die eine gewünschte Klassifizierung haben. Dies ermöglicht die Verwendung von Rufen basierend auf der Aufnahme- und Rufqualität.

Die Daten können auch direkt in bcRefCalls für einfache Plots verwendet werden. Auch hier lassen sich die Rufe entsprechend der Klassifizierung auswählen. Die folgenden Plots zeigen für die Referenz-Daten der Zwergfledermaus die Startfrequenz über Rufdauer (beides stark abhängig von der Rufqualität). Die linke Abbildung beinhaltet nur gute Rufe, die rechte Abbildung nur schlechte Rufe. Deutlich werden hier die Unterschiede sichtbar:

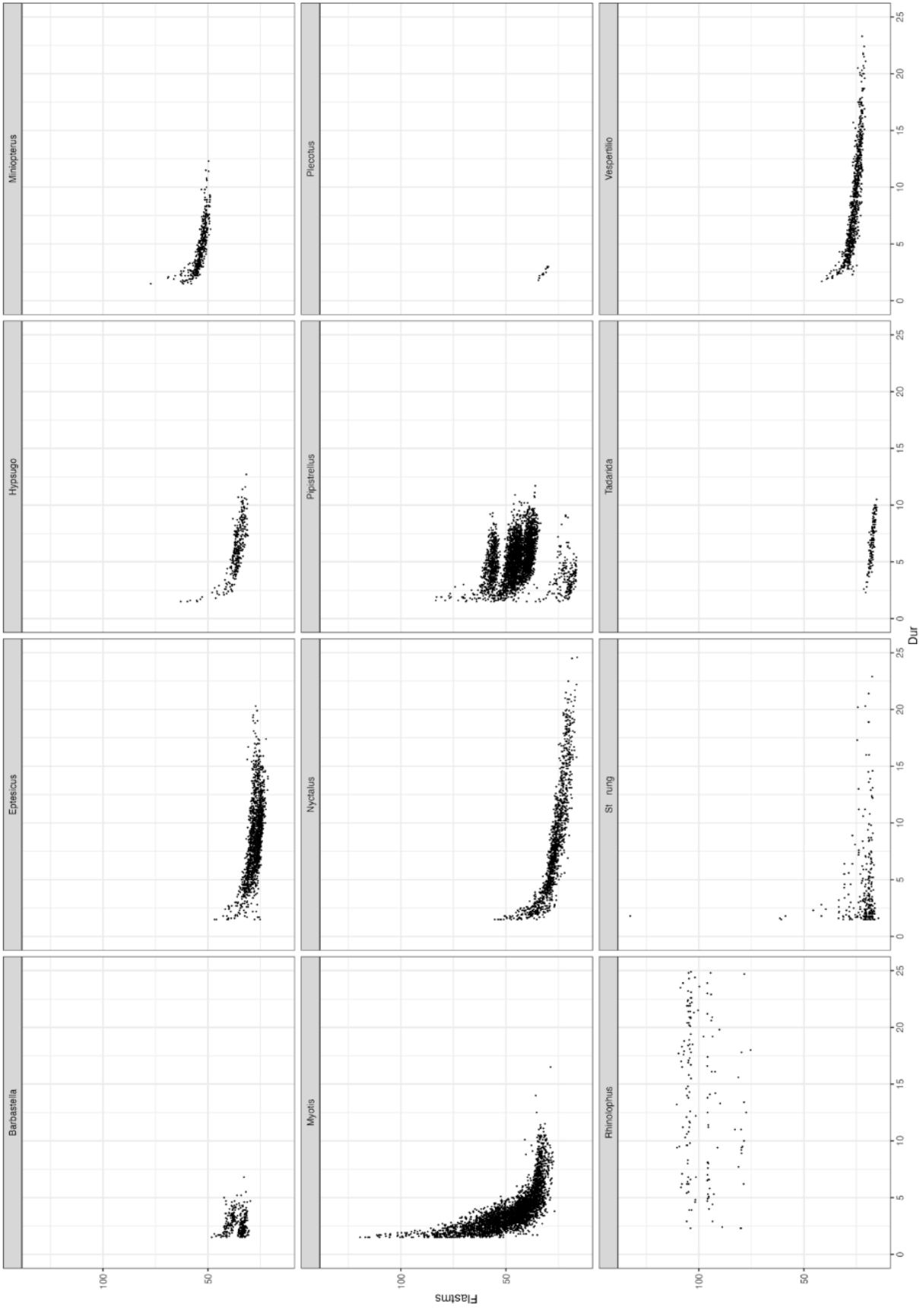


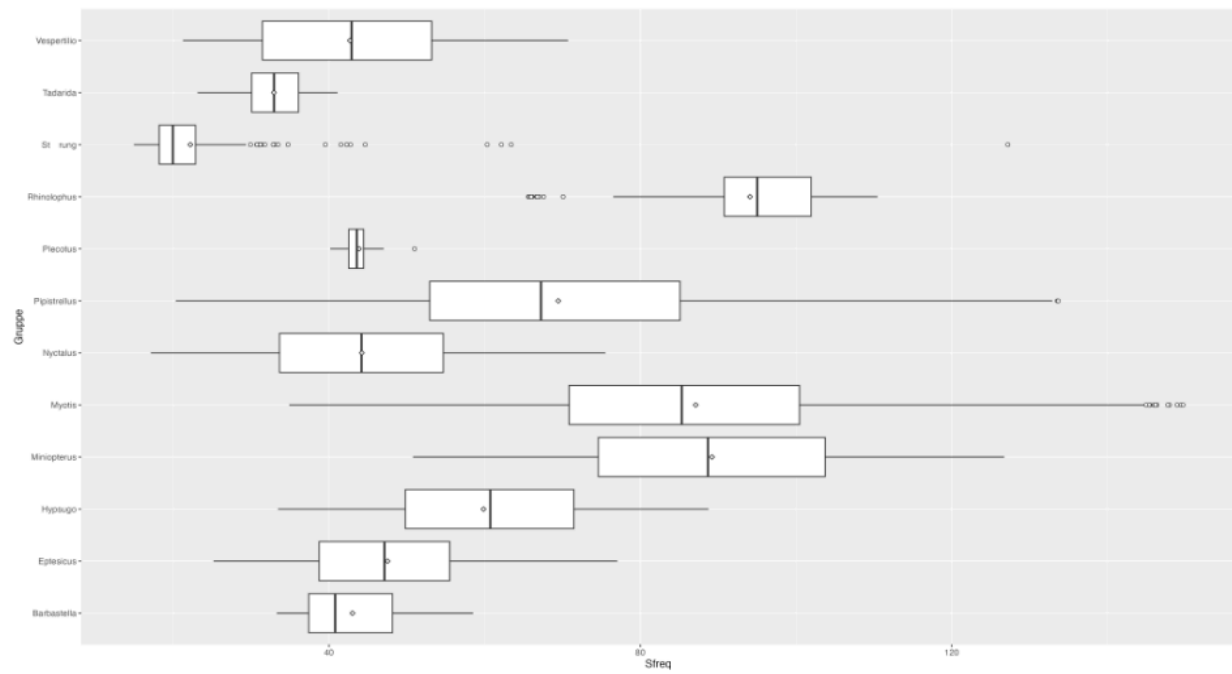
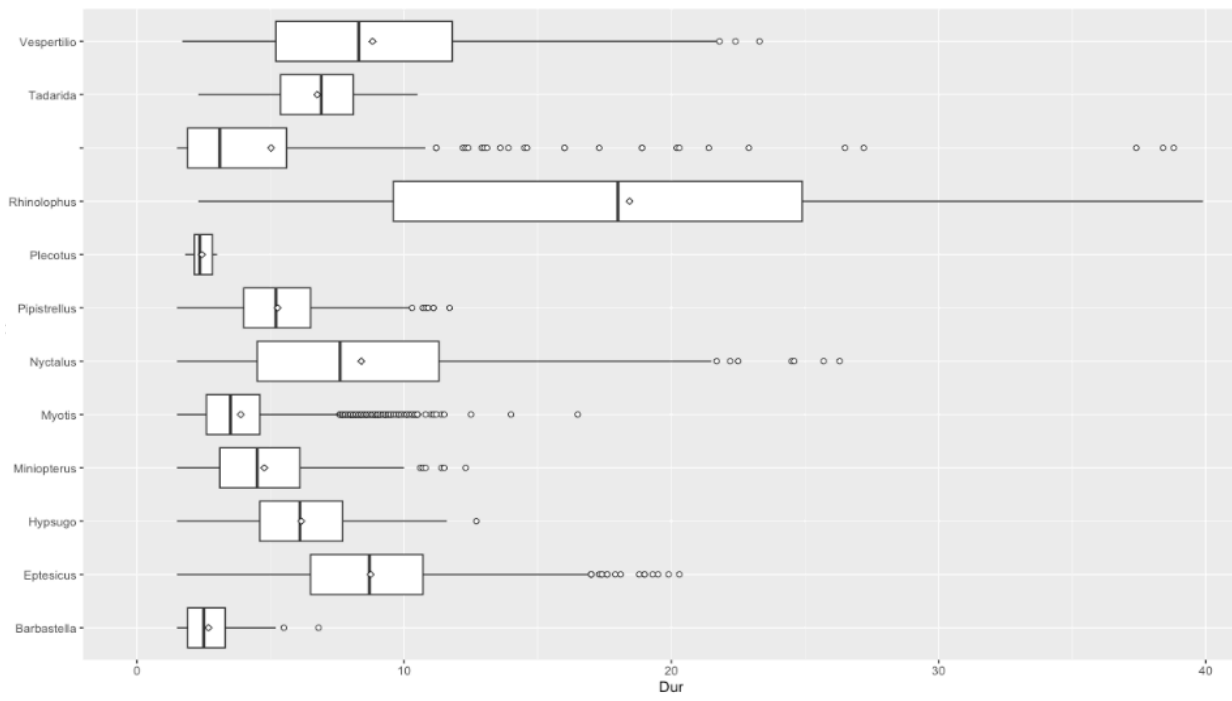
Solche Grafiken sind wichtig, um basierend auf unterschiedlichen Parameter-Paarungen zu ermitteln, in wie weit die Referenzdaten vollständig sind oder ob es noch Datenlücken gibt, die bei einer späteren Klassifizierung von Rufen zu Fehlern führen können.

Für die im späteren beschriebenen Klassifizierer wurden die in der bcRefCalls-Datenbank gespeicherten Referenz-Rufe entsprechend der gewünschten Ruftypen exportiert. Basierend auf den Messwerten je Ruf können auch in Programmen wie zB R (Cran) weitere Grafiken erstellt und Analysen durchgeführt werden. Die Grafik auf der nächsten Seite zeigt als Beispiel einen Plot des Messwerts Flastms aufgetragen gegen die Rufdauer. Je Gattung/Gruppe wurde dazu automatisch eine Grafik erstellt.

Auch ohne zwingend Details in der Grafik erkennen zu können wird deutlich, dass es sich um homogene Rufdaten handelt, die den typischen Rufbereich der jeweiligen Gattungen gut abdecken. Die Daten sind somit geeignet für die Verwendung in einem Klassifizierer. Ca. 30 Parameter haben eine höhere Relevanz und es ergeben sich dann bei paarweisen Plots je Gattung mehrere Hundert Grafiken. Die grobe Sichtung dieser ist relativ aufwendig, jedoch hilfreich, um ein Verständnis für die Daten zu erhalten. Hier im Dokument ist eine Abbildung sinnvoll jedoch nicht möglich. Weiterhin sind je Klassifizierer, also zum Beispiel auch für die Bestimmung innerhalb von Gattungen, solche Analysen notwendig. Dadurch erhöht man die Stabilität der Bestimmung, der Aufwand ist selbst bei den wenigen europäischen Arten dadurch jedoch recht hoch.

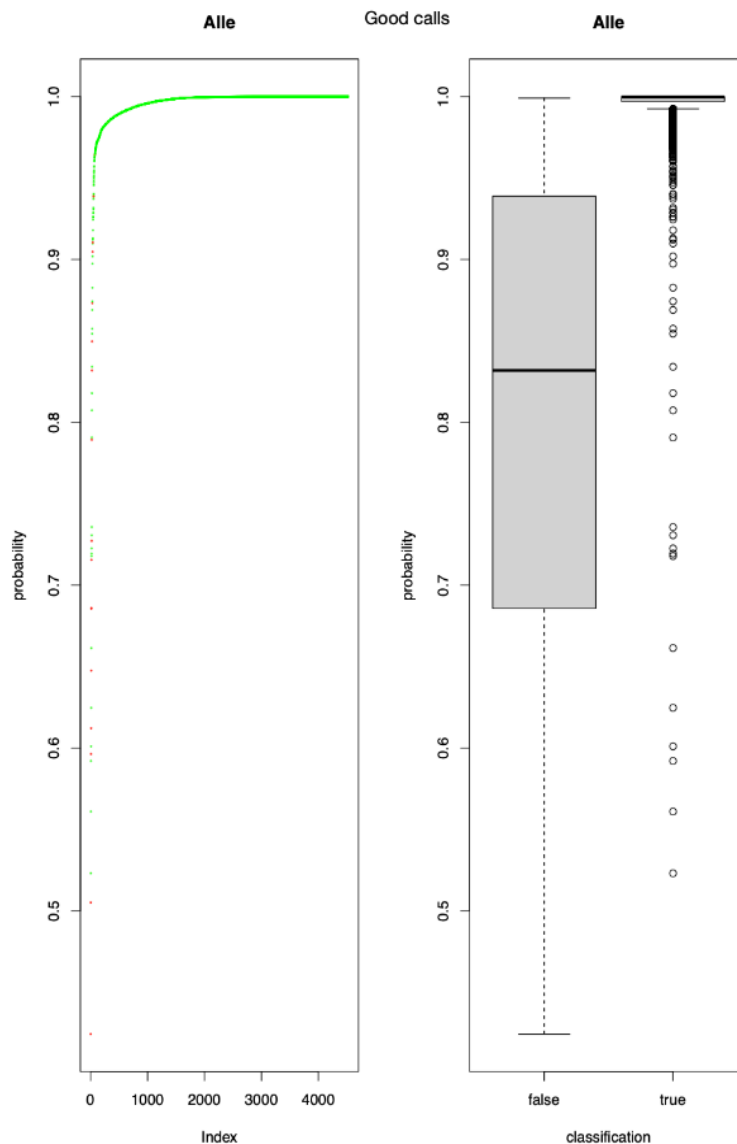
Zur Verdeutlichung der Verteilung der Daten werden in zwei weiteren Grafiken noch Boxplots dargestellt, die die Verteilung der Rufdauern und der Startfrequenzen zeigen.



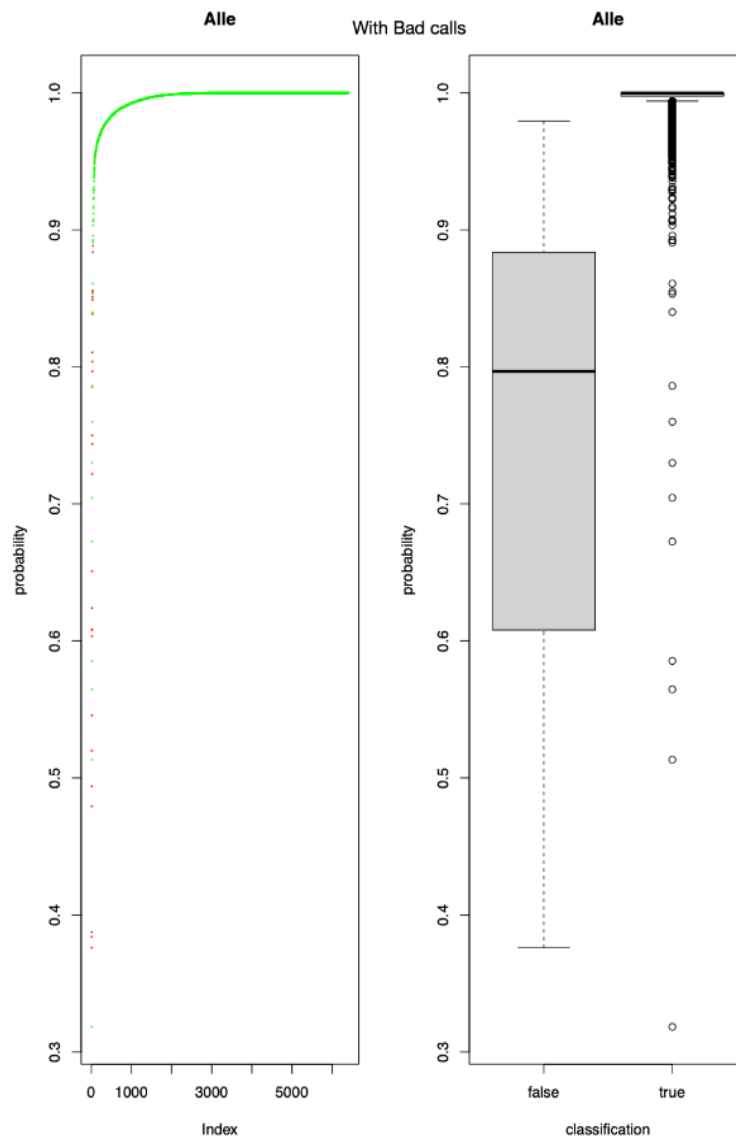


Klassifizierer - Gruppen/Gattungen

Nach kompletter Prüfung der vorhandenen Referenzen aller Arten sowie der Hinzunahme neuer Störungen wurde ein Export von Referenzrufen der geprüften Gattungen als CSV durchgeführt (nur gute Rufe). Der Trainingsdatensatz besteht aus jedem zweiten Ruf (alle Gattungen, sortiert nach Ruflänge). Für den Validierungsdatensatz wurde beginnend beim zweiten Ruf jeder vierte Ruf ausgewählt. So sind Trainings- und Validierungsdatensatz aus dem selben Datenpool ohne Überlappung einzelner Rufe extrahiert worden. Die Datensätze wurden in CreateML (Apple) für das Training und die Evaluation eines Klassifizierungsmodells verwendet. Dazu wurde ein randomForest basiertes Training durchgeführt. Der Parameter maxIterations wurde auf 15 gesetzt (Standard ist 10) und maxDepth auf 9 (Standard ist 6). Die Bestimmungssicherheit wurde ebenso wie die korrekte Bestimmung für den Validierungsdatensatz ermittelt und geplottet. Dazu wird die Wahrscheinlichkeit der Bestimmung über dem Rufindex aufgetragen (aufsteigende Sortierung nach Wahrscheinlichkeit vor Plot). Korrekte Klassifizierungen sind mit einem grünen Punkt dargestellt, falsche Klassifizierungen mit einem roten Punkt. Im zweiten Plot sind für die falschen und korrekten Bestimmungen die Wahrscheinlichkeitsverteilungen gezeigt.



Entsprechend des obigen Vorgehens wurde ein weiterer Gattungs-Klassifizier erstellt, der neben den guten auch die als schlecht markierten Rufe beinhaltet hat. Die nach dem Training und der Validierung in CreateML erhaltenen Ergebnisse sind ähnlich gut, wie beim ersten Klassifizierer. Die Fehlerrate der Evaluation mit dem -Valid-Datensatz ist minimal größer. Auch für diesen Klassifizierer wurde basierend auf den Validierungsdaten eine Übersicht der Bestimmungssicherheiten erstellt.



Der Datensatz „gute Rufe“ beinhaltet 9049 Rufe im Training und 4525 Rufe für die Validierung. Der Datensatz mit „guten und schlechten Rufen“ ist etwas größer und beinhaltet 12813 respektive 6407 Rufe.

Bewertung der neuen Klassifizierer

Basierend auf den bestehenden Referenzdaten wurden Modelle mittels CoreML trainiert, die die Rufe zwischen den Gattungen / Gruppen (Nyctaloid) als auch innerhalb der Gruppen und Gattungen im Idealfall auf Artniveau bestimmen können.

Die Qualität des Trainings kann als precision oder recall quantifiziert werden. Weiterhin ist es möglich aus diesen beiden Maßzahlen einen kombinierten, sogenannten F1-score, zu ermitteln. Für die Berechnung werden die true-positives (tp; korrekte Bestimmungen) und die false-positives (fp; falsch-positives Bestimmungsergebnis) verwendet. Analog gibt es noch true-negatives und false-negatives.

Die folgenden Auswertungen beziehen sich immer auf einzelne Rufe im Gegensatz zum normalen Bestimmen im Arbeitsalltag. Hier werden dann alle Rufe einer Aufnahme herangezogen, um ein Ergebnis für die jeweilige Aufnahme zu ermitteln. Da eine Aufnahmen in der Regel mehr als nur einen Ruf enthält, ist das Ergebnis auf Ebene der Aufnahmen zumeist genauer, als dies bei einzelnen Rufen der Fall ist.

Die precision gibt an, wie häufig ein Modell bei der Bestimmung einer Art korrekt ist. Der recall hingegen zeigt, wie gut das Modell darin ist, möglichst alle korrekten Arten in den Daten zu finden. Die precision und der recall werden dabei wie folgt ermittelt:

$$\text{precision} = \text{tp} / (\text{tp} + \text{fp})$$

$$\text{recall} = \text{tp} / (\text{tp} + \text{fn})$$

Der F1-score ist das harmonische Mittel von precision und recall:

$$\text{F1-score} = (2 * \text{precision} * \text{recall}) / (\text{precision} + \text{recall})$$

Betrachtet man das Modell für alle Gattungen auf auf der Ebene von einzelnen Rufen (also nicht normale Aufnahmen aus mehreren Rufen bestehend), dann werden Klassifizierungsergebnisse von wenigstens 99% erreicht. Das bedeutet, das Modell kann beinahe absolut sicher und ohne Fehler alle einzelnen Rufe der korrekten Art zu ordnen.

Es gilt zu beachten, dass in diesem Fall für den Test ausschliesslich solche Rufe verwendet wurden, die in der Referenzrufdatenbak vorhanden und nach Qualitätskriterien ausgesucht waren. Bei der Entwicklung von batldent konnten wir in der Situation des Tests mit bekannten Rufen auch deutlich bessere Ergebnisse erzielen, als dann mit normalen Aufnahmen aus dem Habitat. Insofern muss das Ergebnis der neuen Klassifizierer, das nochmals besser als bei batldent ausfällt, nicht unbedingt auf Habitataufnahmen übertragbar sein. Ein Test mit regulären Habitataufnahmen erreichte, wie später beschrieben, dennoch ähnlich gute Ergebnisse.

Das Modell zur Klassifizierung von Einzel-Rufen innerhalb der Gruppe Nyctaloid erweist sich mit einer Genauigkeit von 81% als wenigstens so gut, wie batldent. Ein direkter Vergleich mit batldent ist jedoch nicht möglich, da dort Art-übergreifende Untergruppen bestehen. Am sichersten kann der Klassifizierer die Gattung Eptesicus bestimmen, die geringste Sicherheit erreicht Vespertilio.

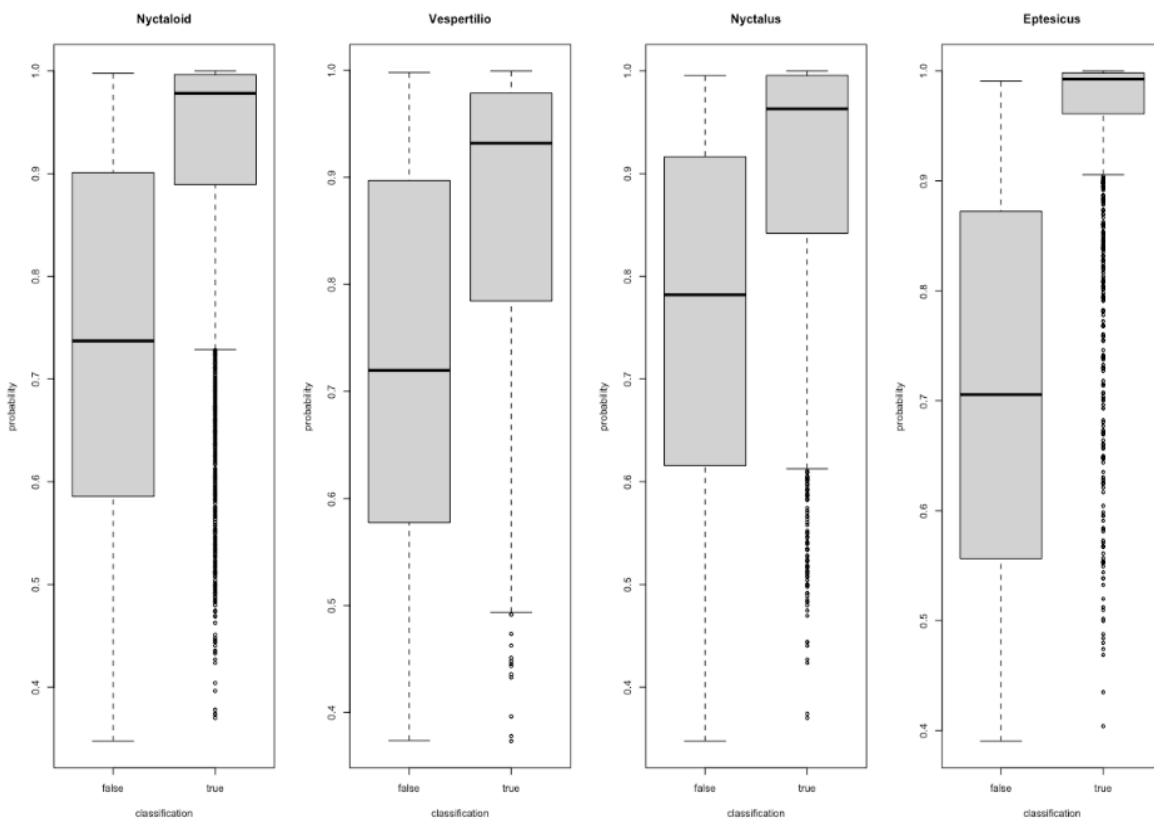
Precision

Predicted	Nyctalus	Eptesicus	Vespertilio	
Korrekt				
Nyctalus	81 %			
Eptesicus		84 %		
Vespertilio			75 %	
			Mittelwert	80 %

Recall

Predicted	Nyctalus	Eptesicus	Vespertilio	
Korrekt				
Nyctalus	76 %	12 %	11 %	
Eptesicus	7 %	89 %	4 %	
Vespertilio	13 %	13 %	74 %	
			Mittelwert	79 %

Die beiden Werte precision und recall haben eine ähnliche Größenordnung und deuten so auf ein stabiles Modell hin. Für Nyctalus und Vespertilio werden immerhin 3/4 aller Rufe korrekt gefunden, bei Eptesicus sogar 9 von 10 Rufen (recall).



Die Boxplots zeigen für die korrekt und die falsch zugeordneten Rufe jeweils die Verteilung der Wahrscheinlichkeit der Artzugehörigkeit. Analog zu precision und recall erkennt man auch hier die größte Streuung bei *Vespertilio* und die geringste Streuung bei *Eptesicus*.

Innerhalb der Gattung *Eptesicus* werden precision und recall Werte von 96% bis 98% erreicht. Die Quote an Fehlbestimmungen ist somit minimal.

Eptesicus recall

Korrekt <-> ID	serotinus	nilssonii
serotinus	97 %	3 %
nilssonii	3 %	97 %

Eptesicus precision

Korrekt <-> ID	serotinus	nilssonii
serotinus	98 %	
nilssonii		96 %

Die Bestimmung von Rufen innerhalb der Gattung *Nyctalus* ist, wenn man sich auf die beiden in Deutschland häufigen Arten Großer Abendsegler und Kleinabendsegler fokussiert, nicht ganz so gut wie bei *Eptesicus*. Insbesondere die Bestimmung des Kleinabendseglers ist relativ unsicher und wirkt sich negativ auf den recall des Großen Abendseglers aus. Korrekte Rufe des Kleinabendseglers werden jedoch mit hoher Sicherheit auch richtig erkannt.

Nyctalus recall

Korrekt <-> ID	noctula	leisleri
noctula	71 %	29 %
leisleri	7 %	93 %

Nyctalus precision

Korrekt <-> ID	noctula	leisleri
noctula	95 %	
leisleri		64 %

Innerhalb der Gattung *Pipistrellus* werden Ergebnisse in etwa auf dem Niveau von *batIdent* erreicht. Die beiden Arten Weißrand- und Flughautfledermaus werden anhand ihrer Ortungsrufe weniger sicher erkannt. Die Arten Zwerg- und Mückenfledermaus hingegen sehr sicher bestimmt. Dementsprechend gut fallen die recall und precision Werte aus. Aber auch Flughaut- und Weißrandfledermaus-Rufe werden in ca. 80% aller Fälle richtig bestimmt.

Pipistrellus recall

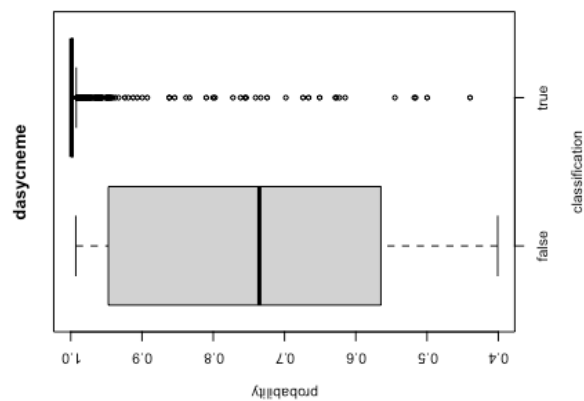
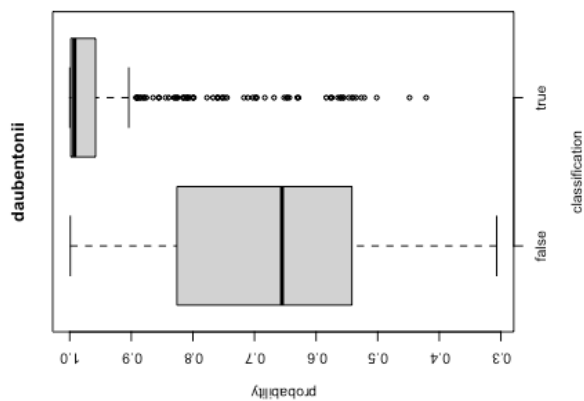
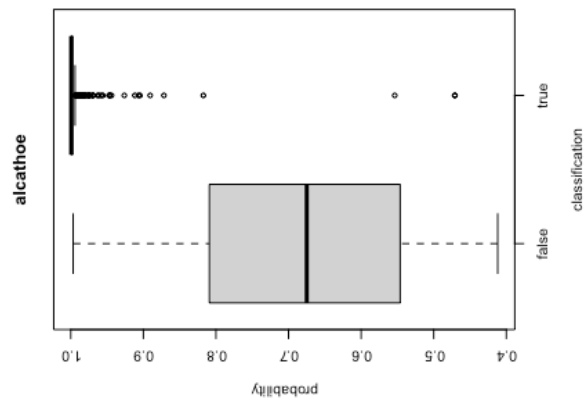
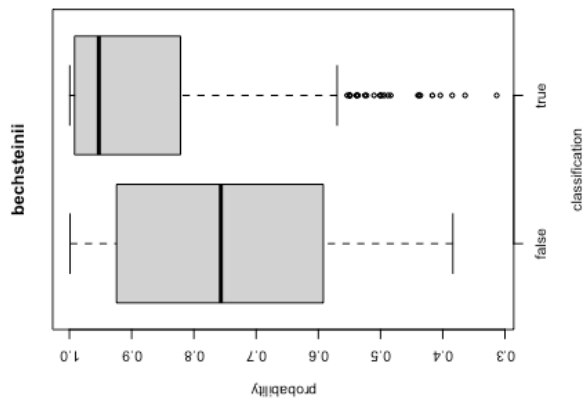
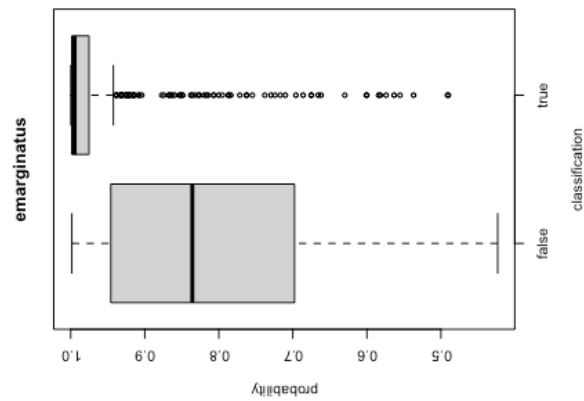
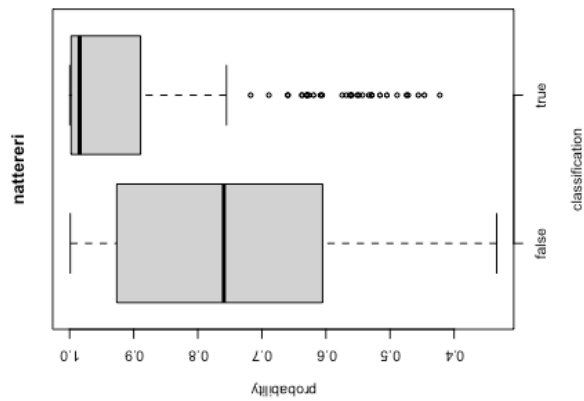
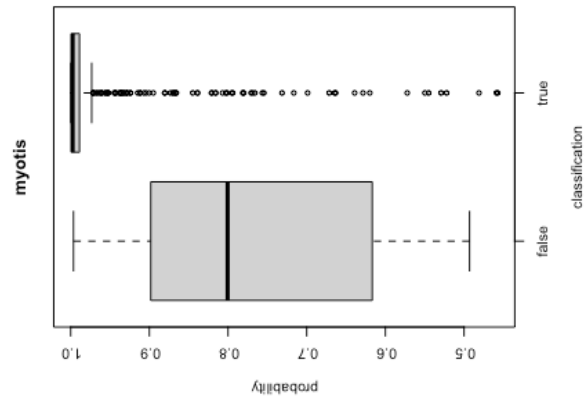
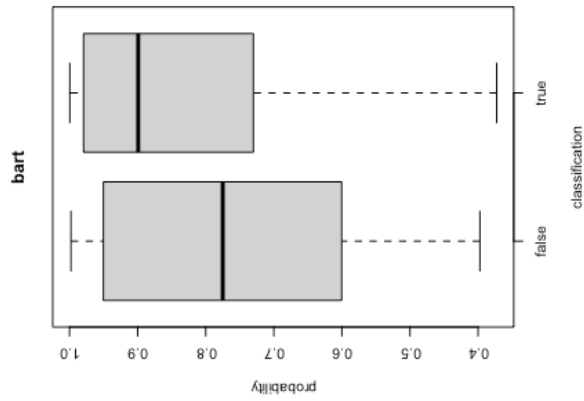
Predicted	pygmaeus	pipistrellus	nathusii	kuhlii
Korrekt				
pygmaeus	99 %			
pipistrellus		99 %		
nathusii			80 %	
kuhlii				75 %

Pipistrellus precision

Predicted	pygmaeus	pipistrellus	nathusii	kuhlii
Korrekt				
pygmaeus	98 %			
pipistrellus		98 %		
nathusii			80 %	
kuhlii				79 %

Für die Gattung Myotis sind die Ergebnisse durchwegs schlechter. Dies ist zum einen der hohen Überlappung im Rufrepertoire geschuldet, zum anderen ist diese Gattung auch deutlich artenreicher. Je größer die Anzahl an Klassen, desto schwieriger ist in der Regel ein gutes Modell zu finden. Wir haben uns für die erste Version für ein Modell entschieden, dass im Gegensatz zu batIdent sofort alle Arten auftrennen soll. Die Boxplots zeigen die Wahrscheinlichkeitsverteilung der falschen und korrekten Artzuweisung.

Nur Arten wie Teichfledermaus, Nymphenfledermaus, Wimpernfledermaus oder das Große Mausohr lassen sich recht gut bestimmen. Die Genauigkeit des Modells liegt immerhin bei ca. 85% und auch precision und recall erreichen im Mittel Werte um 85%. Entsprechend der dargestellten Boxplots gibt es dabei deutliche Unterschiede zwischen den Bestimmungsergebnissen der einzelnen Arten. Während einzelne Arten recall-Werte von über 90% erreichen und somit recht sicher auch korrekt bestimmt werden, liegen die anderer Arten bei deutlich unter 80%.



Myotis recall

Predicted	alcatthoe	bart	bechsteinii	daubentonii	dasycneme	emarginatus	myotis	nattereri
Korrekt								
alcatthoe	95 %							
bart		63 %						
bechsteinii			73 %					
daubentonii				91 %				
dasycneme					97 %			
emarginatus						88 %		
myotis							93 %	
nattereri								75 %

Myotis precision

Predicted	alcatthoe	bart	bechsteinii	daubentonii	dasycneme	emarginatus	myotis	nattereri
Korrekt								
alcatthoe	96 %							
bart		77 %						
bechsteinii			76 %					
daubentonii				81 %				
dasycneme					94 %			
emarginatus						83 %		
myotis							92 %	
nattereri								86 %

Validierung mit naiven Aufnahmen

Wie beschrieben, wurden im Rahmen der Erstellung der neuen Klassifizierer deren Qualität mit einem Teildatensatz der Rufe ermittelt, die vom Trainingsatz abgespalten wurden. Insofern ist davon auszugehen, dass einzelne Rufe einer Sequenz zum Training und die verbleibenden zur Validierung verwendet wurden. Dies ist ein normales Vorgehen bei der Modellierung. Wir haben anschließend weitere Aufnahmen unserer Referenzdatenbank verwendet, um eine zusätzliche Validierung zu erstellen. Diese Aufnahmen waren nicht im Trainingsdatensatz enthalten, sondern zusätzliche, sicher bestimmte Aufnahmen der unterschiedlichen Arten. Sie können daher als naiv angesehen werden. Auch für diese Aufnahmen wurde eine Prüfung aller Rufe durchgeführt und solche eliminiert, die nur unzureichend aufgezeichnet oder vermessen wurden.

	Predicted	Barbastella	Hypsugo	Miniopterus	Myotis	Nyctaloid	Pipistrellus	Plecotus	Rhinolophus	Stoerung	Tadarida
Correct											
Barbastella		99	0	0	0	0	0	0	0	0	0
		100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Hypsugo		0	87	0	0	3	26	0	0	0	0
		0.0%	75.0%	0.0%	0.0%	2.6%	22.4%	0.0%	0.0%	0.0%	0.0%
Miniopterus		0	0	44	0	0	0	0	0	0	0
		0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Myotis		7	2	2	868	2	8	0	0	0	0
		0.8%	0.2%	0.2%	97.6%	0.2%	0.9%	0.0%	0.0%	0.0%	0.0%
Nyctaloid		4	0	1	37	932	1	2	0	10	0
		0.4%	0.0%	0.1%	3.7%	94.4%	0.1%	0.2%	0.0%	1.0%	0.0%
Pipistrellus		1	0	8	1	4	742	1	0	7	1
		0.1%	0.0%	1.0%	0.1%	0.5%	97.0%	0.1%	0.0%	0.9%	0.1%
Plecotus		1	0	0	0	13	17	44	0	0	0
		1.3%	0.0%	0.0%	0.0%	17.3%	22.7%	58.7%	0.0%	0.0%	0.0%
Rhinolophus		0	0	0	0	0	0	0	27	0	0
		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%
Stoerung		0	0	0	0	4	46	0	0	21	0
		0.0%	0.0%	0.0%	0.0%	5.6%	64.8%	0.0%	0.0%	29.6%	0.0%

Die Ergebnisse der Klassifizierung von 3073 naiven Einzelrufen ergibt die obige Konfidenztabelle. Die Ergebnisse sind auf einem ähnlichem Niveau, wie bei dem Validierungsdatensatz. Nur einzelne Gattungen wurden etwas schlechter bestimmt. Dabei handelt es sich um Plecotus-Rufe, hier kommt es zu Verwechslungen mit Nyctaloiden und mit Pipistrelliden (Sozialrufe) Rufen. Weiterhin sind Verwechslungen von Rufen der Alpenfledermaus mit solchen der Gattung Pipistrellus häufig. Die Erkennung von Störungen ist ebenso schlechter.

Auch bei Betrachtung der Ergebnisse innerhalb einer Gattung zeigen sich schlechtere Ergebnisse im Vergleich mit den Validierungen basierend auf einem Teil der ursprünglichen Referenzdaten. Die folgenden Tabellen geben eine Übersicht über recall und precision für 889 Rufe der Gattung Myotis. Man erkennt, dass Rufe der Arten Wasserfledermaus, Bartfledermäuse und Bechsteinfledermaus zum Verwechseln neigen. Hier finden sich die größten Fehler oder Unsicherheiten. Jedoch ist dieses Ergebnis beinahe zu erwarten, denn die Rufe überlappen stark. Eine sichere Unterscheidung durch den Klassifizierer hätte dann unter Umständen auf einen Fehler oder Bias in den Daten hingewiesen.

Myotis recall

	predicted	alcahthoe	bart	bechsteinii	dasychneme	daubentonii	emarginatus	myotis	nattereri
correct									
alcahthoe		25	0	0	0	0	0	0	0
		100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
bart		2	64	27	3	36	5	0	4
		1.4%	45.4%	19.1%	2.1%	25.5%	3.5%	0.0%	2.8%
bechsteinii		6	2	58	0	3	3	0	2
		8.1%	2.7%	78.4%	0.0%	4.1%	4.1%	0.0%	2.7%
dasychneme		1	4	2	177	1	0	0	0
		0.5%	2.2%	1.1%	95.7%	0.5%	0.0%	0.0%	0.0%
daubentonii		1	4	19	4	152	1	0	5
		0.5%	2.2%	10.2%	2.2%	81.7%	0.5%	0.0%	2.7%
emarginatus		0	0	2	0	0	83	0	5
		0.0%	0.0%	2.2%	0.0%	0.0%	92.2%	0.0%	5.6%
myotis		0	0	0	1	0	0	91	13
		0.0%	0.0%	0.0%	1.0%	0.0%	0.0%	86.7%	12.4%
nattereri		0	0	0	0	0	2	3	78
		0.0%	0.0%	0.0%	0.0%	0.0%	2.4%	3.6%	94.0%

Myotis precision

	predicted	alcahthoe	bart	bechsteinii	dasychneme	daubentonii	emarginatus	myotis	nattereri
correct									
alcahthoe		25	0	0	0	0	0	0	0
		71.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
bart		2	64	27	3	36	5	0	4
		5.7%	86.5%	25.0%	1.6%	18.8%	5.3%	0.0%	3.7%
bechsteinii		6	2	58	0	3	3	0	2
		17.1%	2.7%	53.7%	0.0%	1.6%	3.2%	0.0%	1.9%
dasychneme		1	4	2	177	1	0	0	0
		2.9%	5.4%	1.9%	95.7%	0.5%	0.0%	0.0%	0.0%
daubentonii		1	4	19	4	152	1	0	5
		2.9%	5.4%	17.6%	2.2%	79.2%	1.1%	0.0%	4.7%
emarginatus		0	0	2	0	0	83	0	5
		0.0%	0.0%	1.9%	0.0%	0.0%	88.3%	0.0%	4.7%
myotis		0	0	0	1	0	0	91	13
		0.0%	0.0%	0.0%	0.5%	0.0%	0.0%	96.8%	12.1%
nattereri		0	0	0	0	0	2	3	78
		0.0%	0.0%	0.0%	0.0%	0.0%	2.1%	3.2%	72.9%

Für die dargestellten Tabellen wurden nur die Rufe verwendet, die anhand der Aufnahme sicher der Art zugeordnet wurden und durch eine manuelle Kontrolle als Gute oder schlechte Rufe eingeteilt wurden. In der Regel enthalten die Aufnahmen noch zahlreiche weitere Rufe. Dabei handelt es sich um solche, die im Rahmen der Kontrolle als zu schlecht vermessen oder aufgezeichnet erkannt und sich damit nicht für die weitere Verwendung geeignet haben. Weiterhin sind in einigen Aufnahmen auch Ortungsrufe anderer Arten, die ebenso bei diesem Vergleich nicht berücksichtigt wurden.

In der täglichen Praxis wird bei der automatischen Artbestimmung ein solch aufwendiger Prozess und die Auswahl einzelner Rufe für die Bestimmung jedoch nicht durchgeführt. Daher dienen die

vorhergehenden Tabellen rein der Beurteilung der Klassifizierer im Hinblick auf die generellen Qualität der zugrunde liegenden Statistik. Sie lassen sich noch nicht zwingend Aussagen für die tägliche Arbeit ableiten. In der Regel werden die Rufe vor der Bestimmung nicht aufwendig kontrolliert, dies ist bei den üblichen Datenmengen der automatischen Erfassung nicht möglich und auch nicht sinnvoll. Weiterhin liefern die obigen Tabellen die Ergebnisse je Einzelruf. Eine Aufnahme besteht aus mehreren Rufen und somit kann das Ergebnis für die Aufnahme dann auch abweichen.

Aus diesem Grund wurde der obige Datensatz in bcAdmin 4 eingelesen und dort sowohl die Bestimmung mit batldent als auch mit den CoreML-basierten Funktionen durchgeführt. Es lässt sich festhalten, dass CoreML um ca. den Faktor 10 schneller ein Ergebnis ermittelt hat. Auf Ebene der Gattungen/Gruppen, also die erste Bestimmungsebene in batldent ebenso wie im CoreML-basierten Modell, wird eine Rate von ca. 93% / 96% erreicht (batldent/CoreML). Eine korrekte Bestimmung wurde dann gezählt, wenn wenigstens eine der - in batldent möglichen - drei Arten des Ergebnisses zur zugehörigen Gruppe/Gattung gehört. Ebenso wurden die Aufnahmen im Hinblick auf die Bestimmung der Art überprüft. Wie zu erwarten, sind die Ergebnisse in diesem Fall generell schlechter. Die Überlappung der Rufe nahe verwandter Arten lässt dies erwarten. So werden aber dennoch 69% (batldent) und 87% (CoreML) korrekte Bestimmungen erreicht. Die Tabellen zeigen die Unterschiede und Arten mit guter/schlechter Bestimmung auf Ebene von Aufnahmen.

Korrekte Bestimmung auf Gattungs/Gruppenebene

	batldent	CoreML
Barbastella	91 %	100 %
Eptesicus	89 %	100 %
Hypsugo	100 %	100 %
Miniopterus	100 %	100 %
Myotis	93 %	99 %
Nyctalus	90 %	97 %
Pipistrellus	100 %	100 %
Plecotus	67 %	42 %
Rhinolophus	100 %	100 %
Vespertilio	100 %	100 %

Korrekte Bestimmung auf Artniveau

	batldent	CoreML
B. barbastellus	91 %	100 %
E. nilssonii	100 %	100 %
E. serotinus	15 %	85 %
H. savii	100 %	100 %
M. schreibersii	100 %	100 %
M. alcaethoe	25 %	75 %
M. bart	54 %	38 %
M. bechsteinii	0 %	91 %
M. dasycneme	88 %	100 %
M. daubentonii	76 %	94 %
M. emarginatus	43 %	100 %
M. myotis	70 %	100 %
M. nattereri	38 %	100 %
N. leisleri	45 %	45 %
N. noctula	72 %	78 %
P. nathusii	72 %	89 %
P. pipistrellus	95 %	100 %
P. pygmaeus	100 %	100 %
Plecotus	67 %	42 %
Rhinolophus	100 %	100 %
V. murinus	63 %	88 %

Vergleich mit batldent

Für einen Vergleich mit batldent wurden Habitataufnahmen aus ca. 15 Jahren akustischer Erfassung ausgewählt. Diese wurden in bcAdmin 4 vermessen und anschliessend eine Bestimmung sowohl mit batldent als auch mit dem neuen Klassifizierer auf Gattungsniveau durchgeführt.

Es wurden auf Gruppen/Gattungsniveau 34 von 1428 Aufnahmen mit dem neuen Klassifizierer falsch bestimmt. Das sind 2% der Aufnahmen, somit wurde bei 98% der Aufnahmen die Gattung/Gruppe korrekt ermittelt. Der neue Klassifizierer bestimmt beinahe 100% aller Myotis-Aufnahmen korrekt im Gegensatz zu batldent, dass hier eine Fehlerquote von 35% aufweist. batldent weist 15% falsche Bestimmungen von Rufen der Gruppe Nyctaloid auf, wohingegen mit dem neuen Klassifizierer keine falschen Bestimmungsraten erhalten wurden. Die Bestimmungsgenauigkeit von batldent lag bei diesem Datensatz bei 89%.

Genauigkeit neuer Klassifizierer

Vorhersage	Myotis	Pipistrellus	Nyctaloid
Korrekt			
Myotis	95 %		
Pipistrellus		98 %	
Nyctaloid			100 %

Genauigkeit batldent

Vorhersage	Myotis	Pipistrellus	Nyctaloid
Korrekt			
Myotis	65 %		
Pipistrellus		99 %	
Nyctaloid			85 %

Anwendung in bcAdmin 4

bcAdmin 4 in der Version 1.4.0 (2024) verfügt über eine erste Implementierung der beschriebenen Klassifizierer. Ausgewählte Aufnahmen können mittels des Menü Aufnahmen auf Gattungs- oder Artniveau bestimmt werden. Die Ergebnisse werden vorerst im Kommentar gespeichert und an den bestehende Kommentar angefügt. Für optimale Ergebnisse empfehlen wir die Experteneinstellungen anzupassen, um schlechte Rufe in Aufnahmen zu eliminieren. Der folgende Screenshot zeigt den von uns momentan empfohlenen Ansatz.

