NTNU
Norwegian University of Science and Technology

Improving tools for citizen scientists

Wouter Koch

Digital transformation projects

02.12.2022

Transforming Citizen Science for Biodiversity

Citizen science

A person who...

Citizen science

A person who...

is there with the species...

Citizen science

A person who...

is there with the species...

observes it...

Citizen science

A person who...

is there with the species...

observes it...

recognizes it...

Citizen science

A person who...

is there with the species...

observes it...

recognizes it...

and reports it

Citizen science

A person who...

is there with the species...

observes it...

recognizes it...

and reports it

All biases, especially in CS

What is lacking in the available citizen science data? What can be done to collect better data?

Paper II

Identification is fundamental

AI is helping more and more

Data are taxonomically biased

Maximizing citizen scientists' contribution to automated species recognition

Wouter Koch, Laurens Hogeweg, Erlend B. Nilsen & Anders G. Finstad

Scientific Reports, 2022:
doi:10.1038/S41598-022-11257-X

Paper II

Identification is fundamental

AI is helping more and more

Data are taxonomically biased

How does that affect image data?

Which images would help AI models the most?

Maximizing citizen scientists' contribution to automated species recognition

Wouter Koch, Laurens Hogeweg, Erlend B. Nilsen & Anders G. Finstad

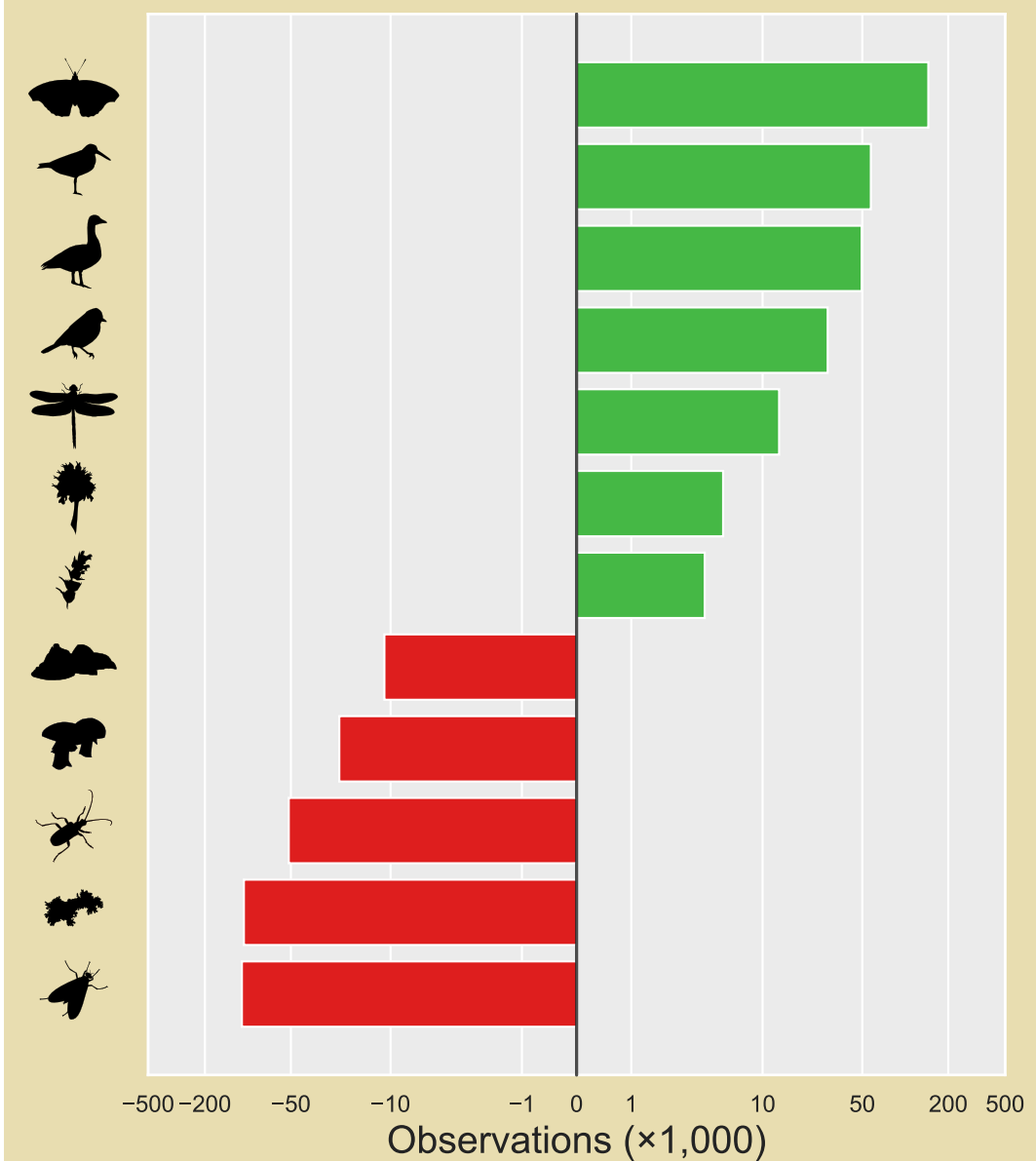
Scientific Reports, 2022:
doi:10.1038/S41598-022-11257-X

Taxonomic bias

Known issue, e.g. Troudet *et al*.

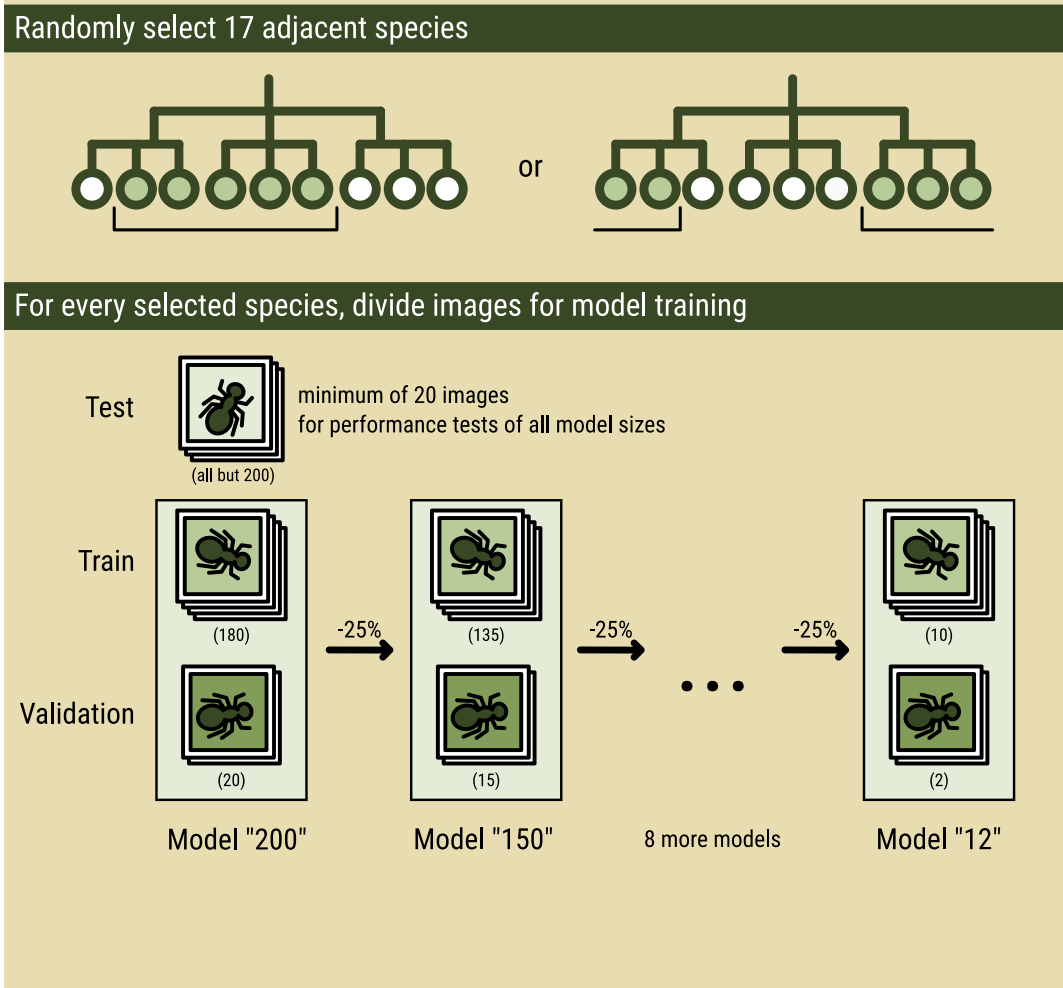
We tested and confirm this within Norwegian Citizen Science image data

Selection of 12 taxonomic orders



Training models

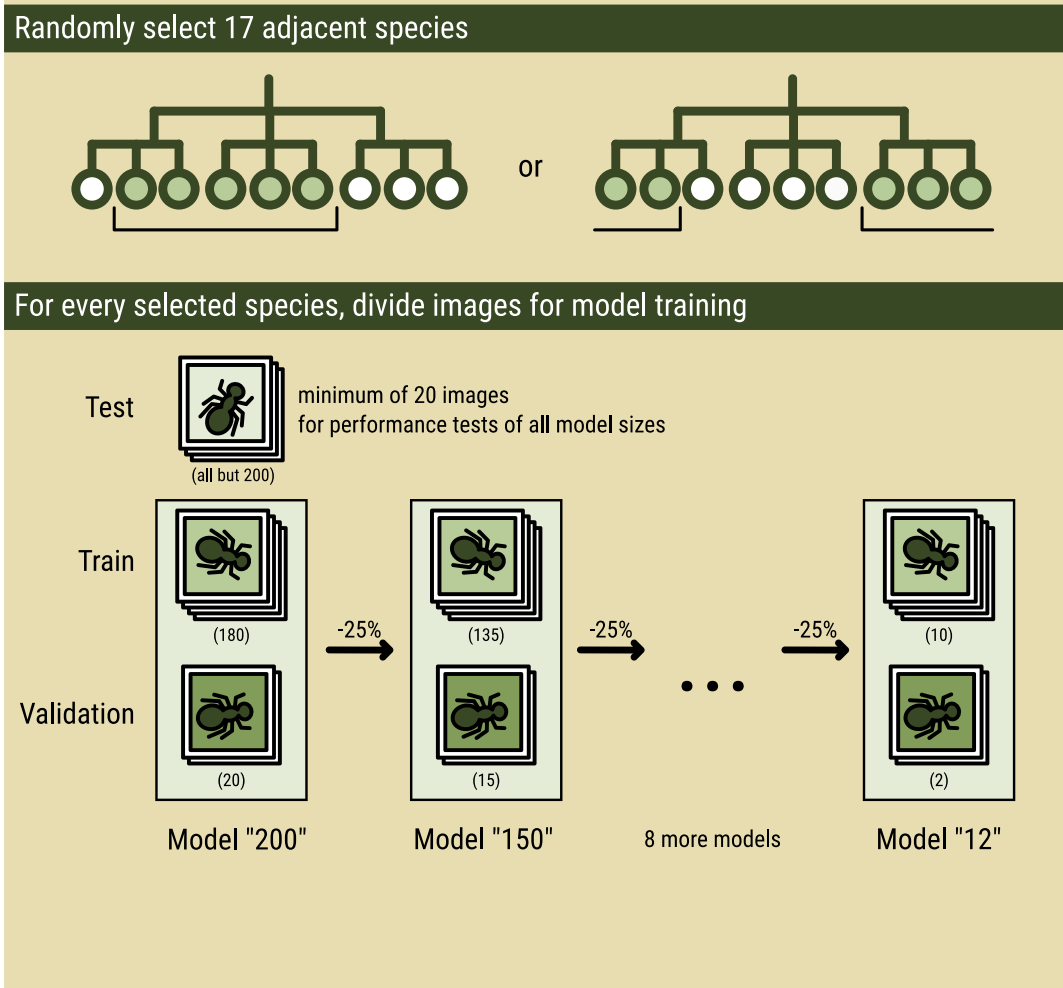
Goal: find the *Value of Information* of adding images per taxon



Training models

Goal: find the *Value of Information* of adding images per taxon

Take 17 species for one order

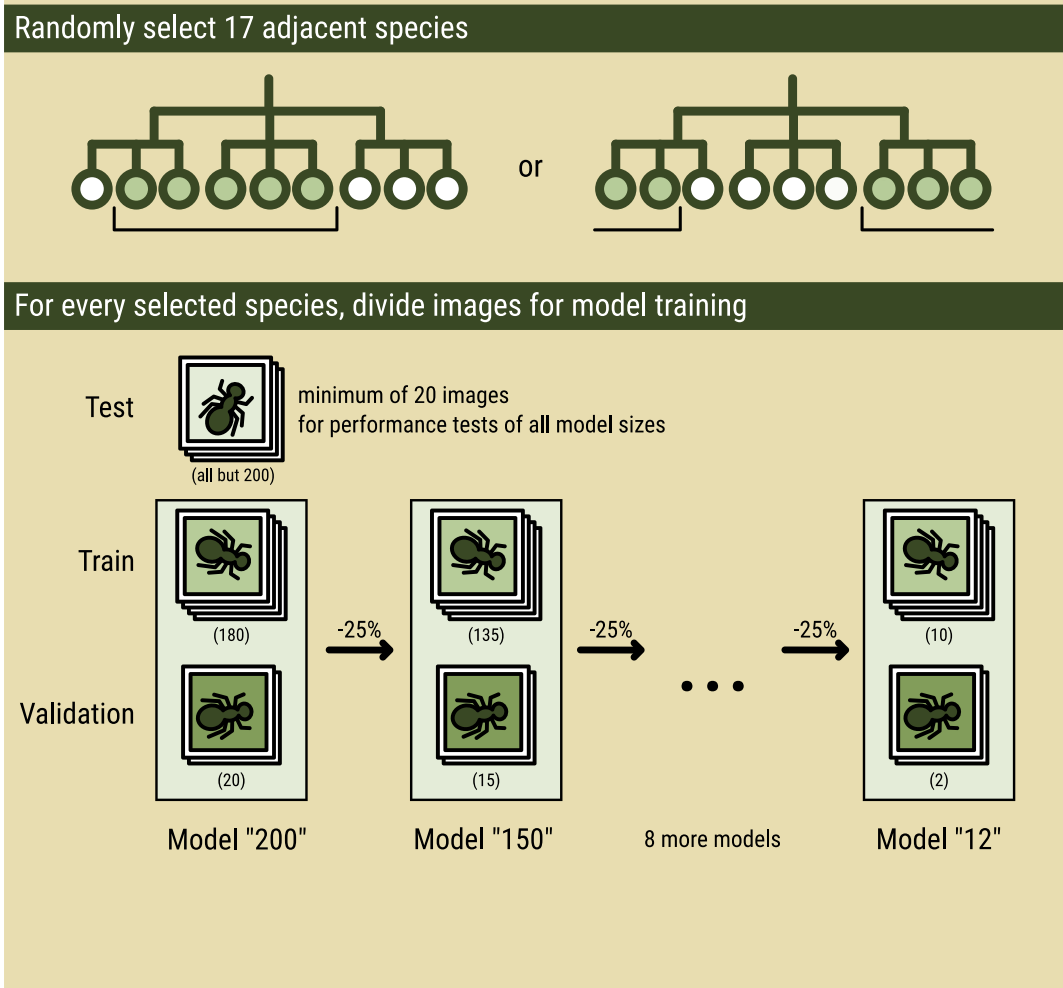


Training models

Goal: find the *Value of Information* of adding images per taxon

Take 17 species for one order

Train on 200, 150 ... 12 images per species (25% reduction)

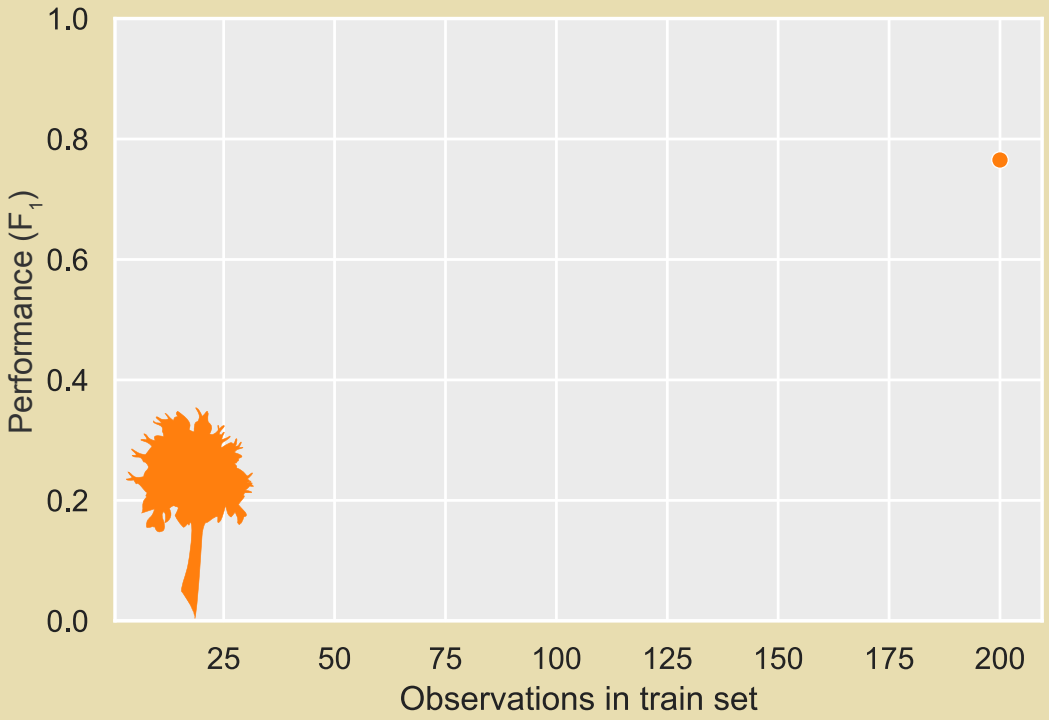


Training models

Goal: find the *Value of Information* of adding images per taxon

Take 17 species for one order

Train on 200, 150 ... 12 images per species (25% reduction)

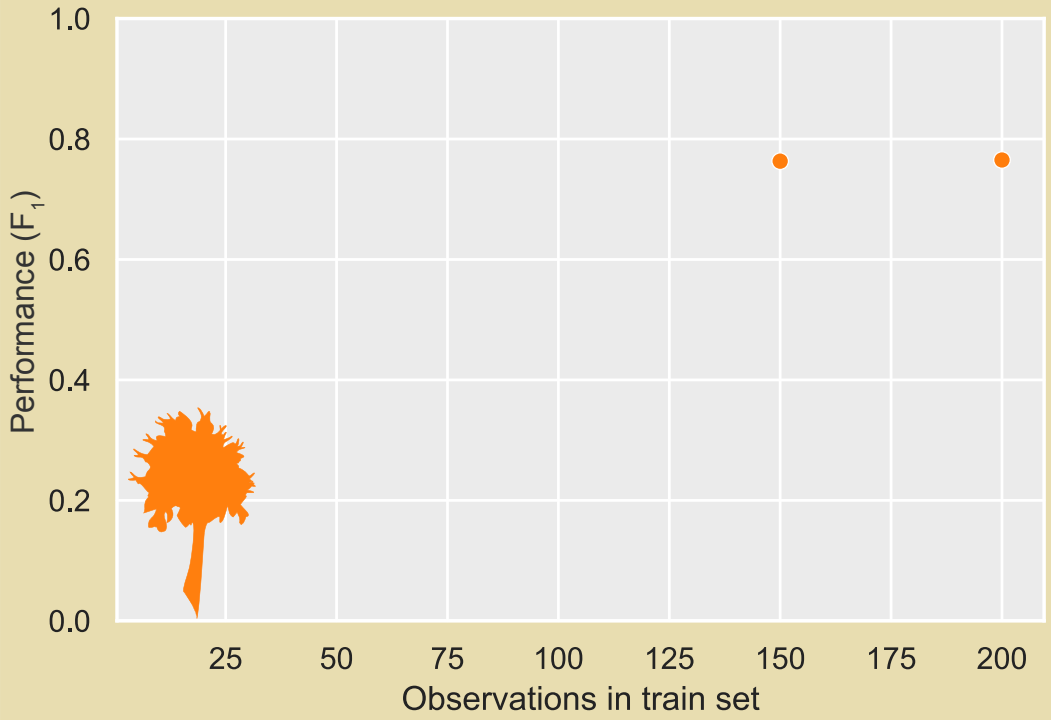


Training models

Goal: find the *Value of Information* of adding images per taxon

Take 17 species for one order

Train on 200, 150 ... 12 images per species (25% reduction)

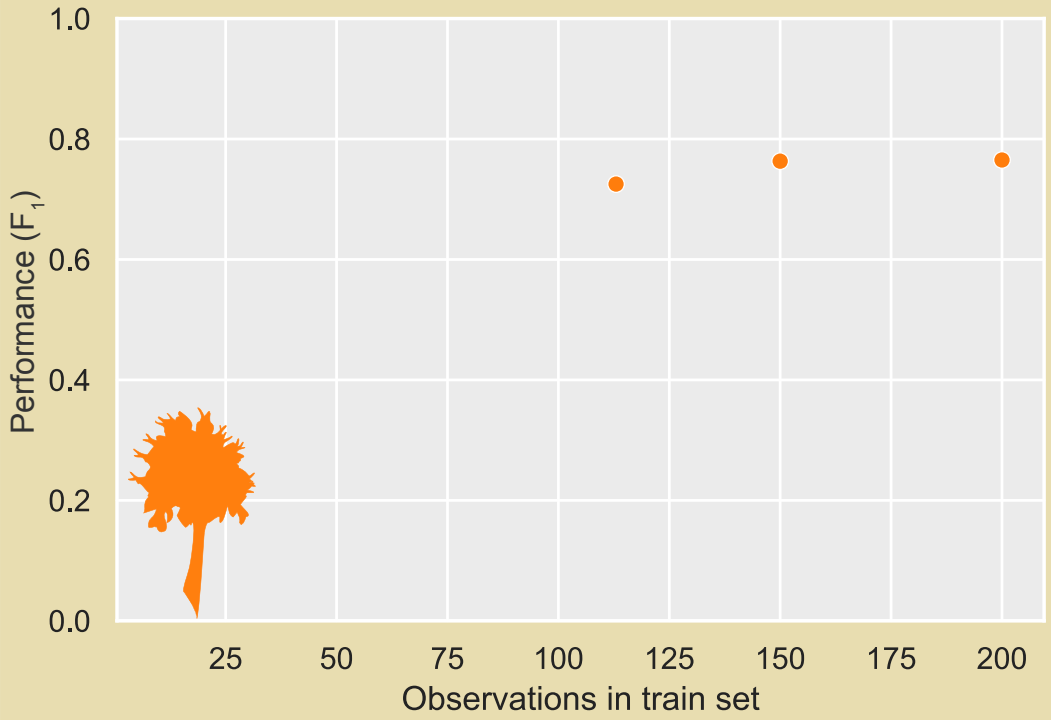


Training models

Goal: find the *Value of Information* of adding images per taxon

Take 17 species for one order

Train on 200, 150 ... 12 images per species (25% reduction)

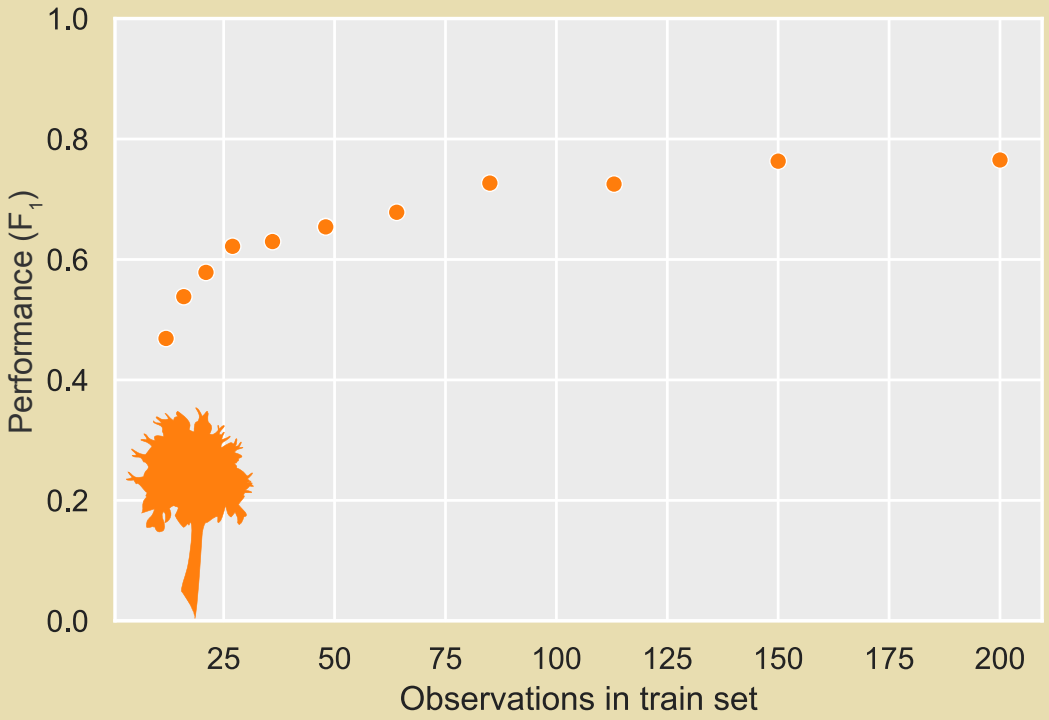


Training models

Goal: find the *Value of Information* of adding images per taxon

Take 17 species for one order

Train on 200, 150 ... 12 images per species (25% reduction)



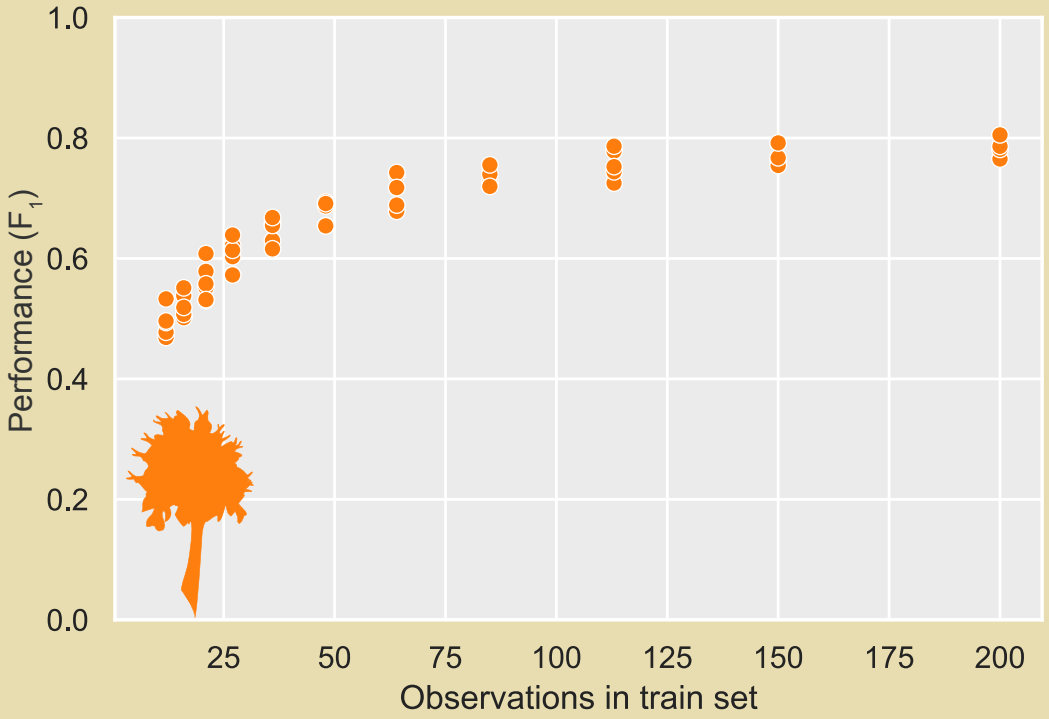
Training models

Goal: find the *Value of Information* of adding images per taxon

Take 17 species for one order

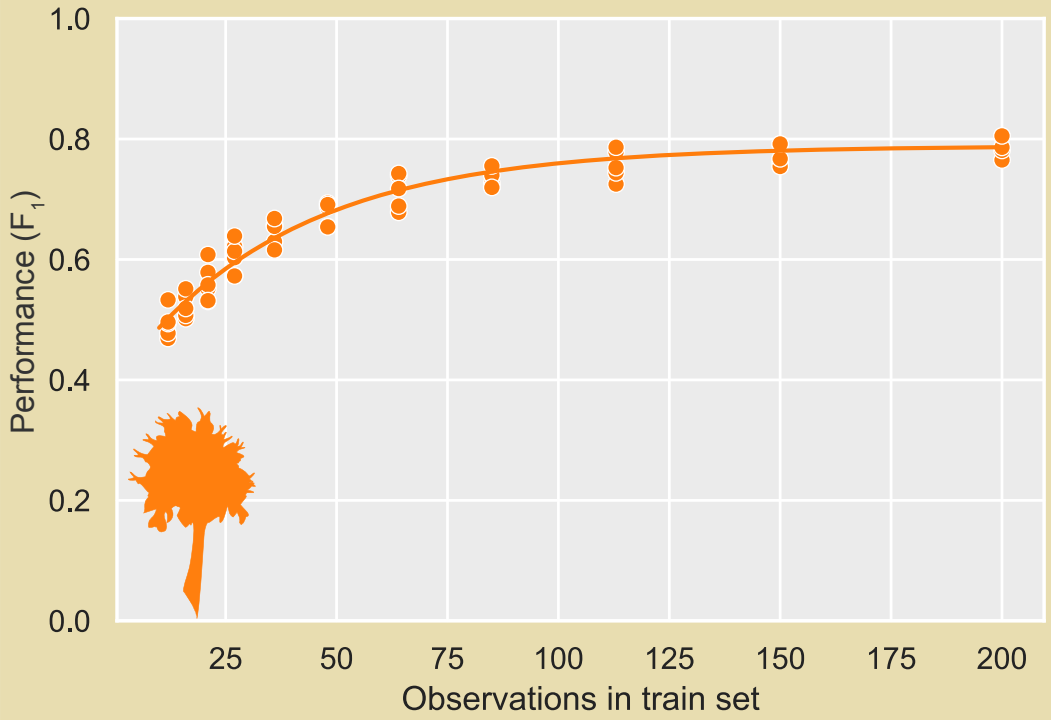
Train on 200, 150 ... 12 images per species (25% reduction)

5 times per order



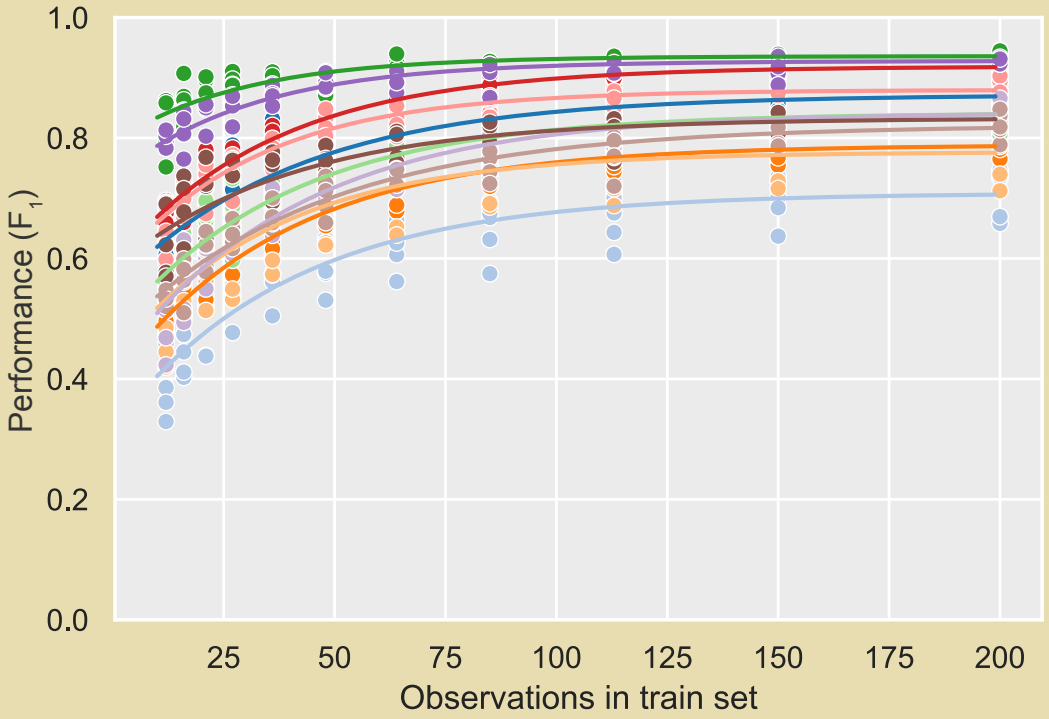
Value of Information

Gives performance curves over images per species



Value of Information

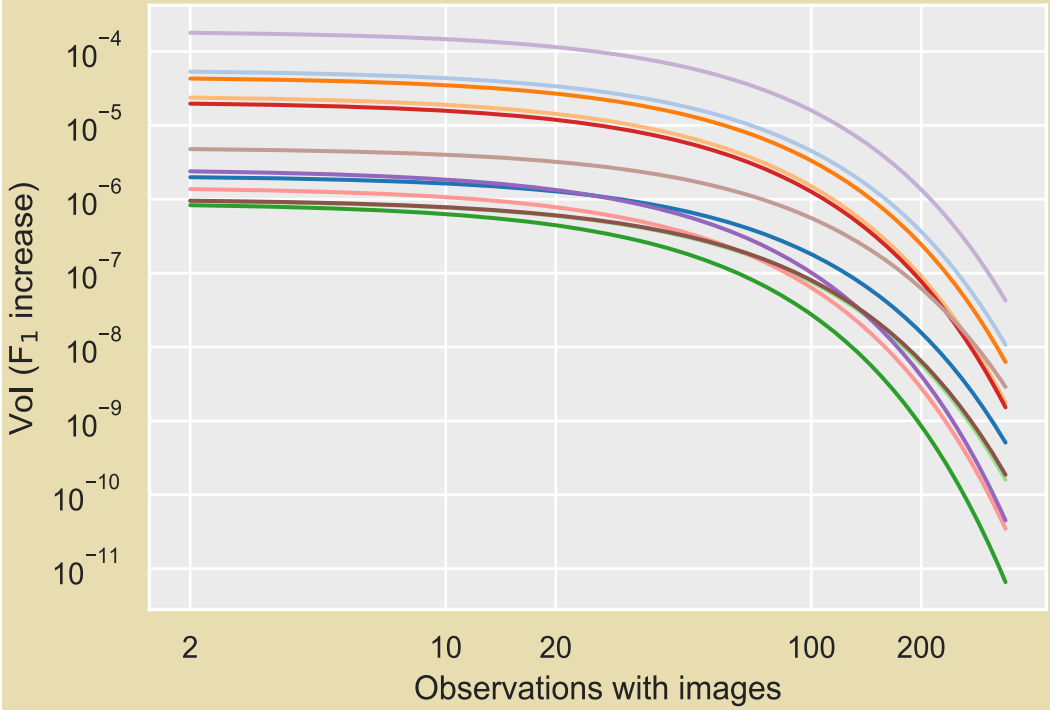
Gives performance curves over images per species



Value of Information

Gives performance curves over images per species

Get the slope of the fitted curves

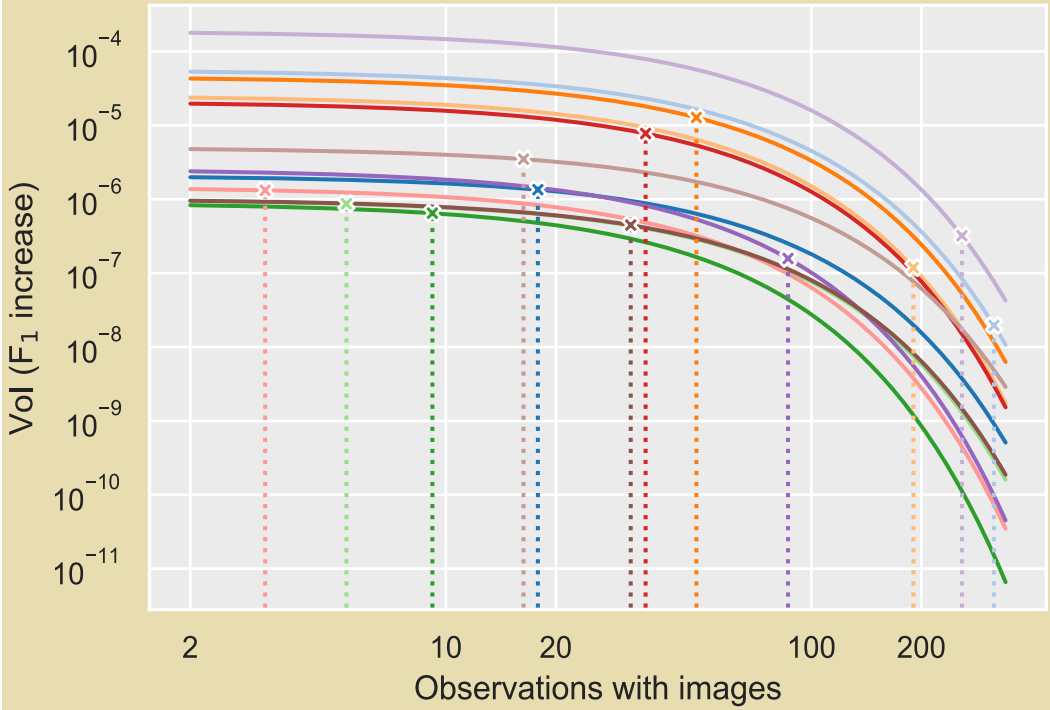


Value of Information

Gives performance curves over images per species

Get the slope of the fitted curves

Look up the Value of Information at the current amount

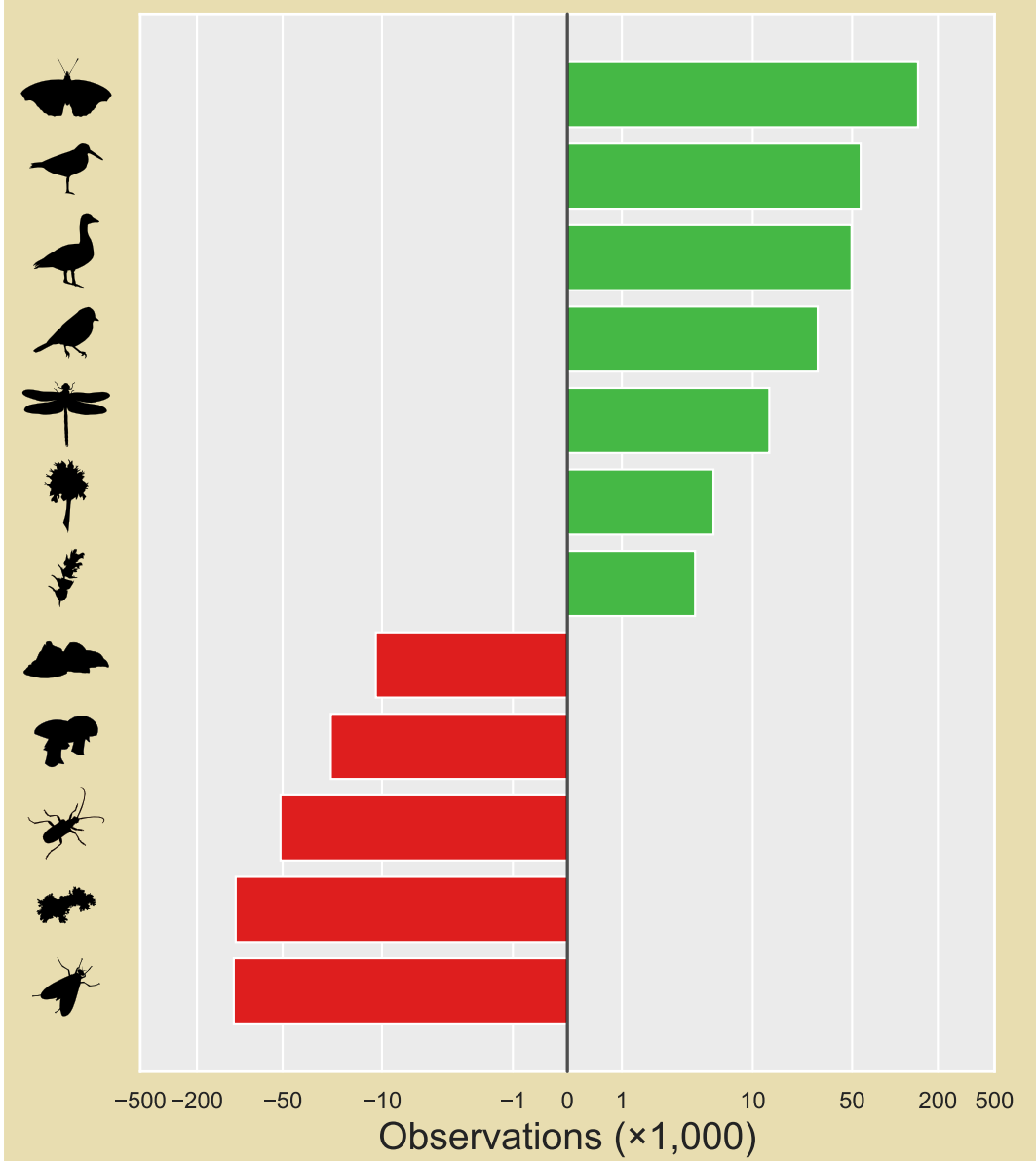


Value of Information

Gives performance curves over images per species

Get the slope of the fitted curves

Look up the Value of Information at the current amount



Value of Information

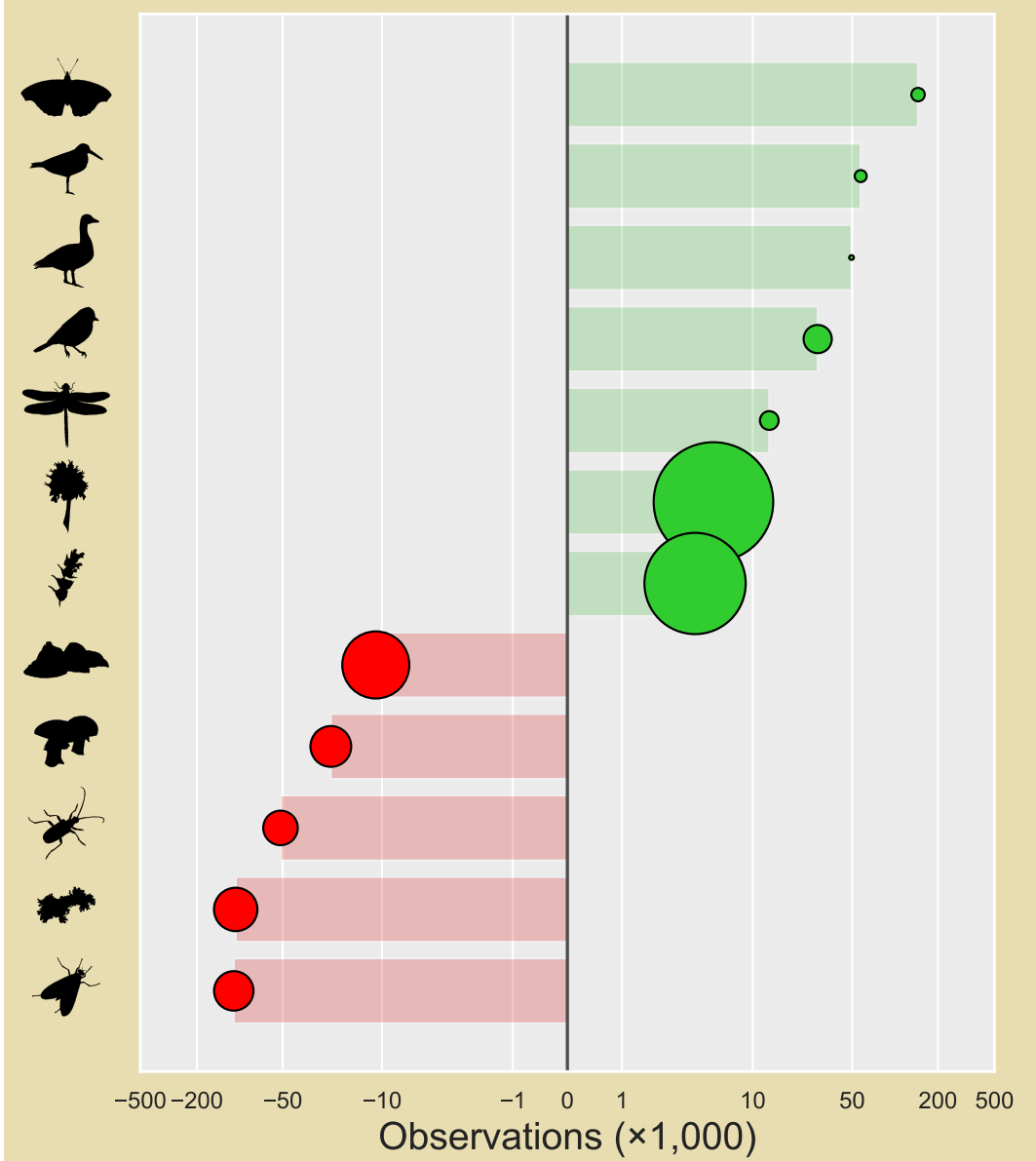
Gives performance curves over images per species

Get the slope of the fitted curves

Look up the Value of Information at the current amount

Value not just in most scarce

Opportunity to focus resources



Paper III

Taxonomic bias is equated to popularity

Recognizability bias in citizen science photographs

Wouter Koch, Laurens Hogeweg, Erlend B. Nilsen, Robert B. O'Hara & Anders G. Finstad

Preprint on bioRχiv:
doi:10.1101/2022.06.25.497604

Paper III

Taxonomic bias is equated to popularity

What about recognizability?

Is AI biased in the same way as humans?

What does that mean for training AI models?

Recognizability bias in citizen science photographs

Wouter Koch, Laurens Hogeweg, Erlend B. Nilsen, Robert B. O'Hara & Anders G. Finstad

Preprint on bioRχiv:
doi:10.1101/2022.06.25.497604

Performance vs popularity

Same 12 orders as in paper II

Deep dive into birds: most prevalent and good standardized trait data

Train AI models with 200 images per species: model does not “know” which are reported more

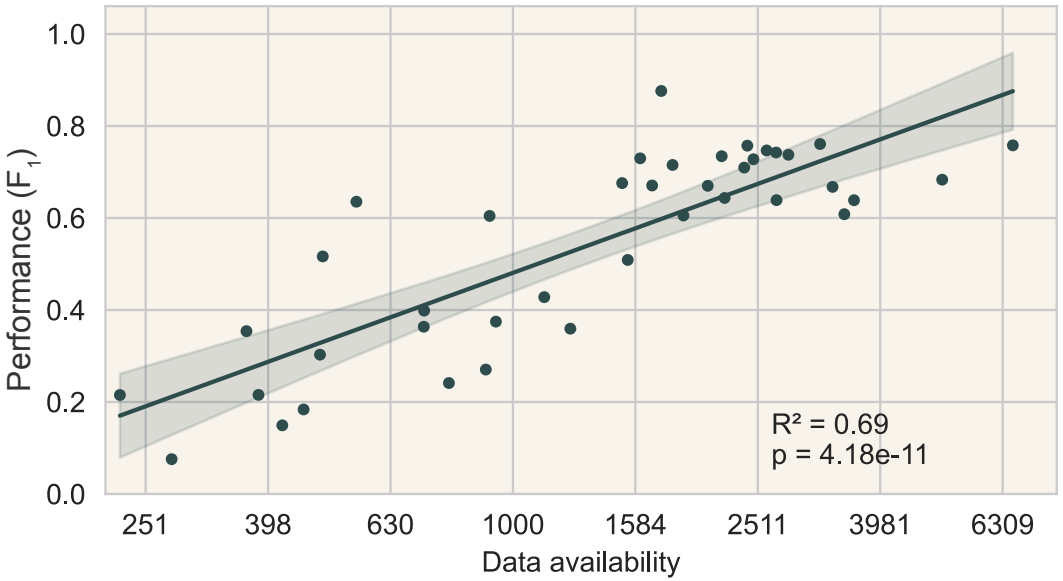
Performance vs popularity

Same 12 orders as in paper II

Deep dive into birds: most prevalent and good standardized trait data

Train AI models with 200 images per species: model does not “know” which are reported more

Still, model does better on more popular species. Why?



Are the pictures better?

Close by or better zoom?

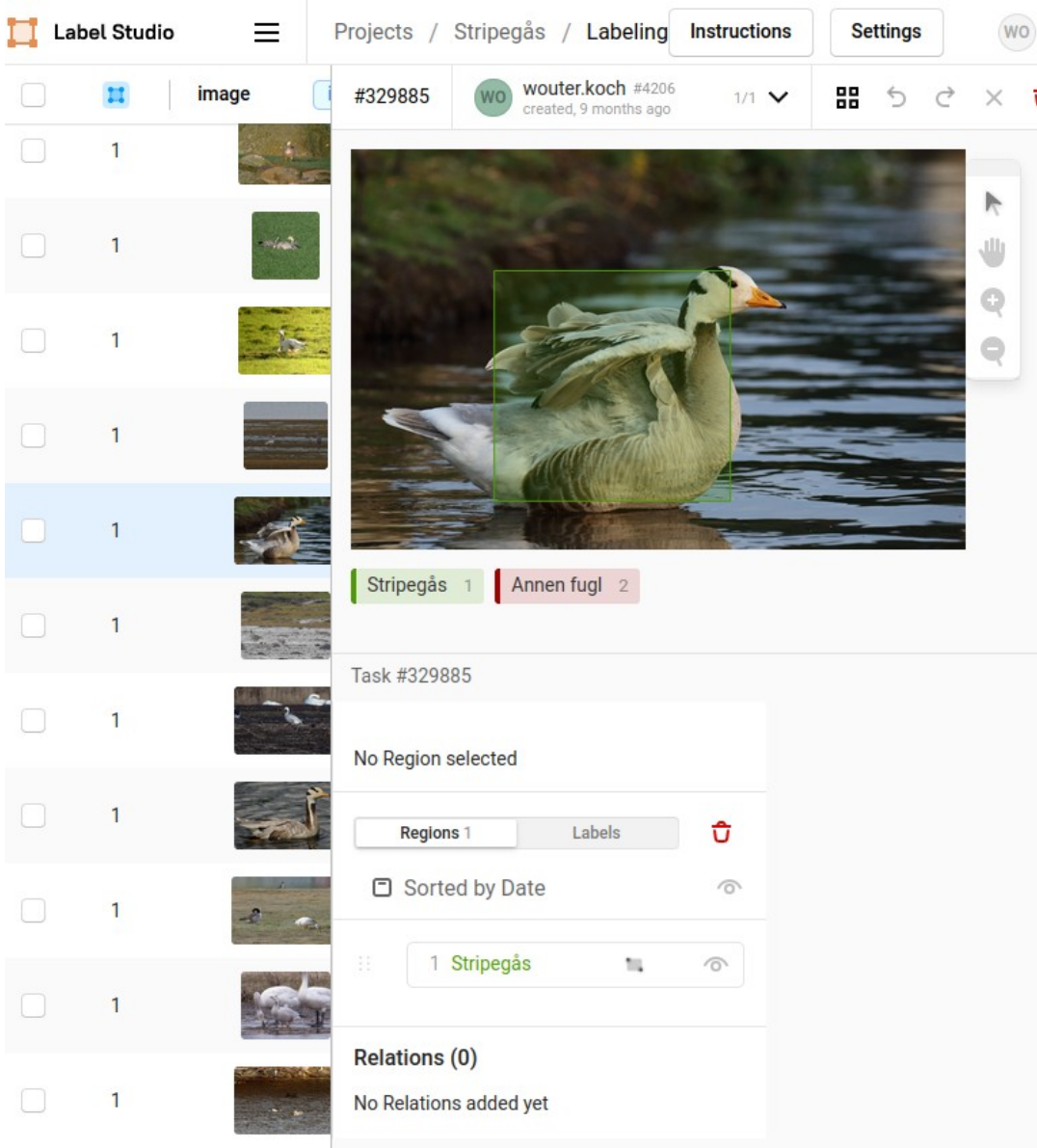
Fewer other species?

Are the pictures better?

Close by or better zoom?

Fewer other species?

Pictures annotated with help from volunteers

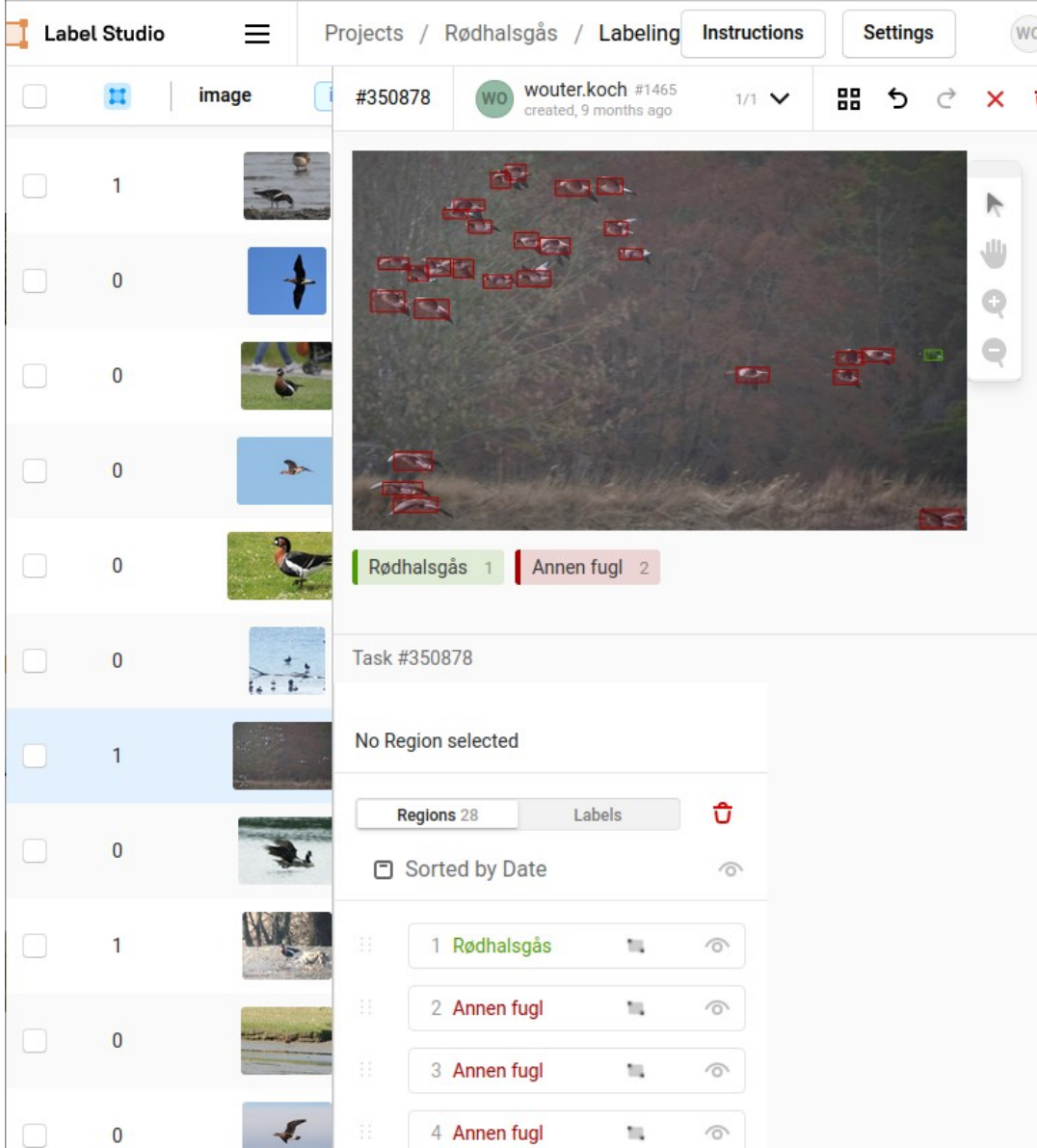


Are the pictures better?

Close by or better zoom?

Fewer other species?

Pictures annotated with help from volunteers



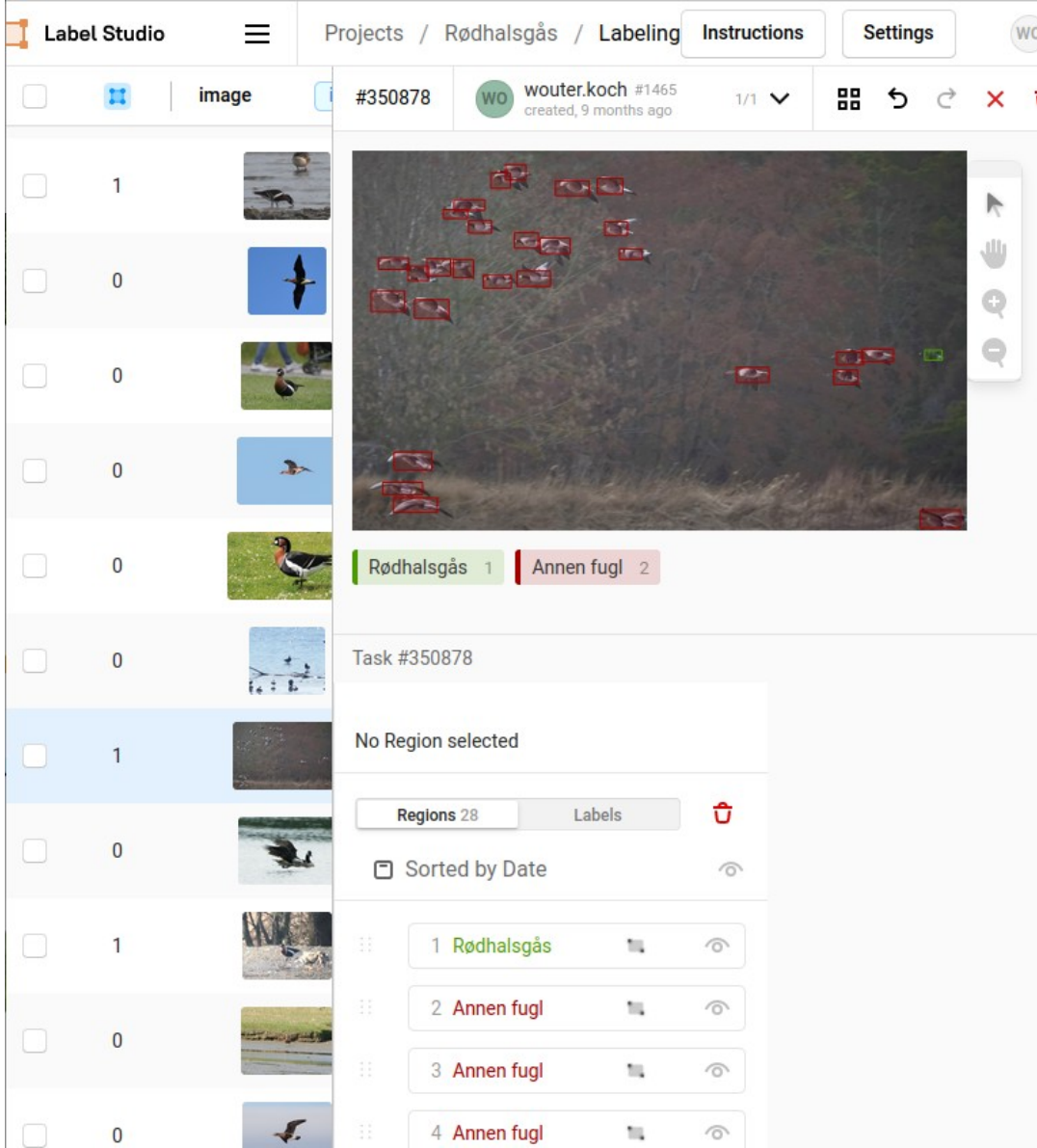
Are the pictures better?

Close by or better zoom?

Fewer other species?

Pictures annotated with help from volunteers

Provides image quality measures



Are the pictures better?

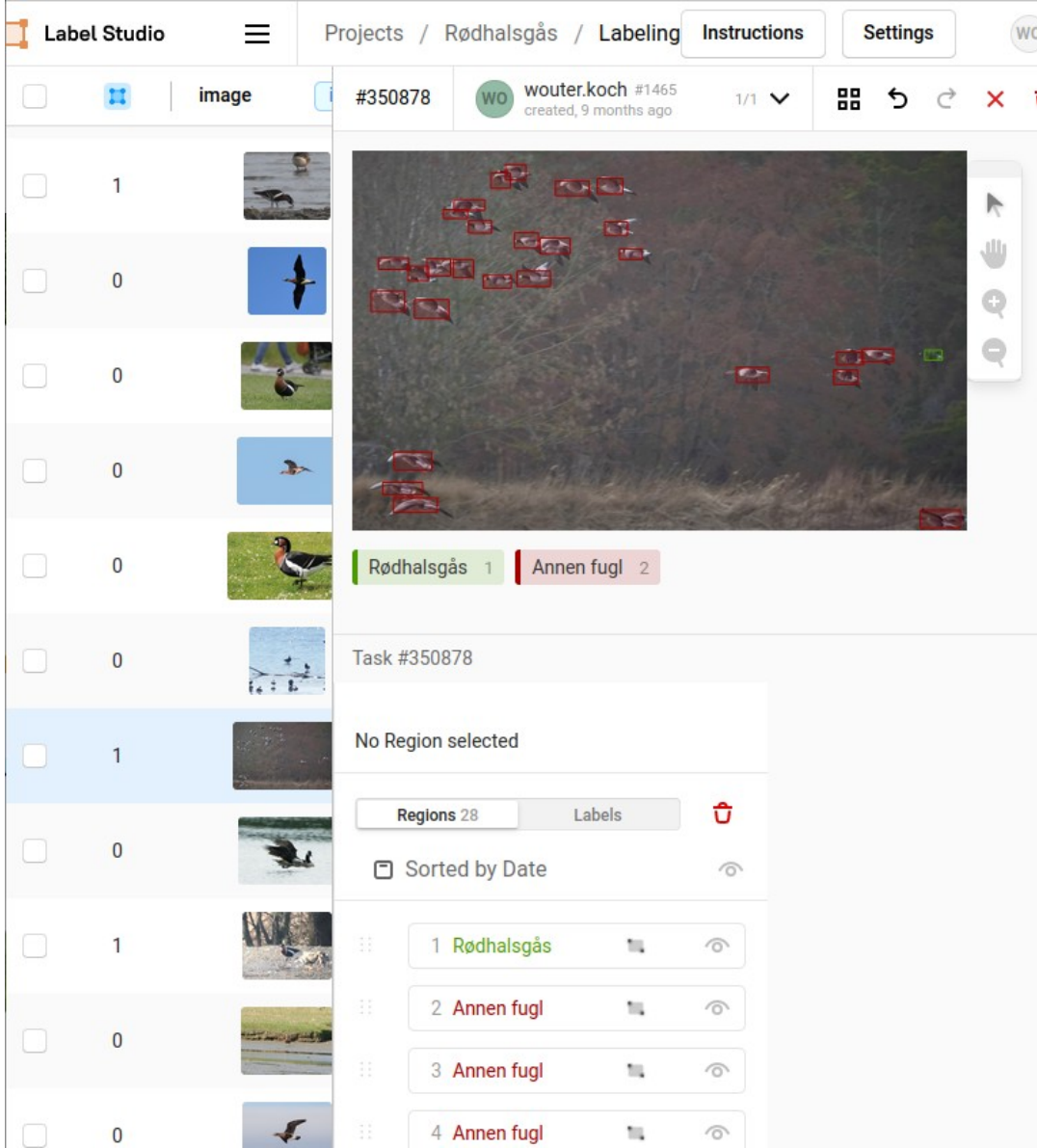
Close by or better zoom?

Fewer other species?

Pictures annotated with help from volunteers

Provides image quality measures

Does not correlate with performance



Is it the kind of bird?

Larger birds, different habitat, migratory, other behavior?

Is it the kind of bird?

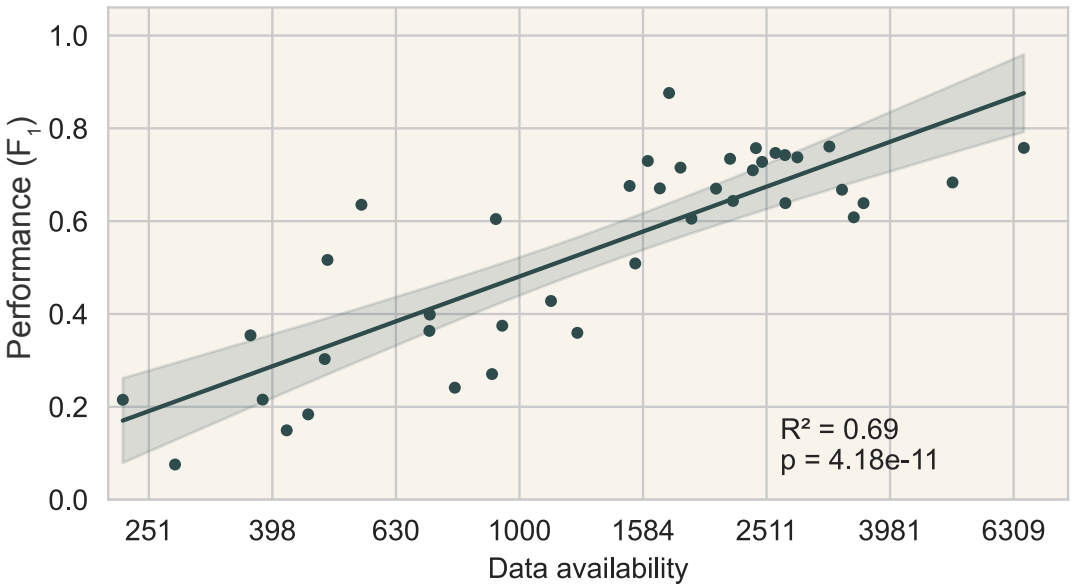
Larger birds, different habitat, migratory, other behavior?

None of these correlate with the AI model performance

Recognizability

The AI model finds some species easier to recognize

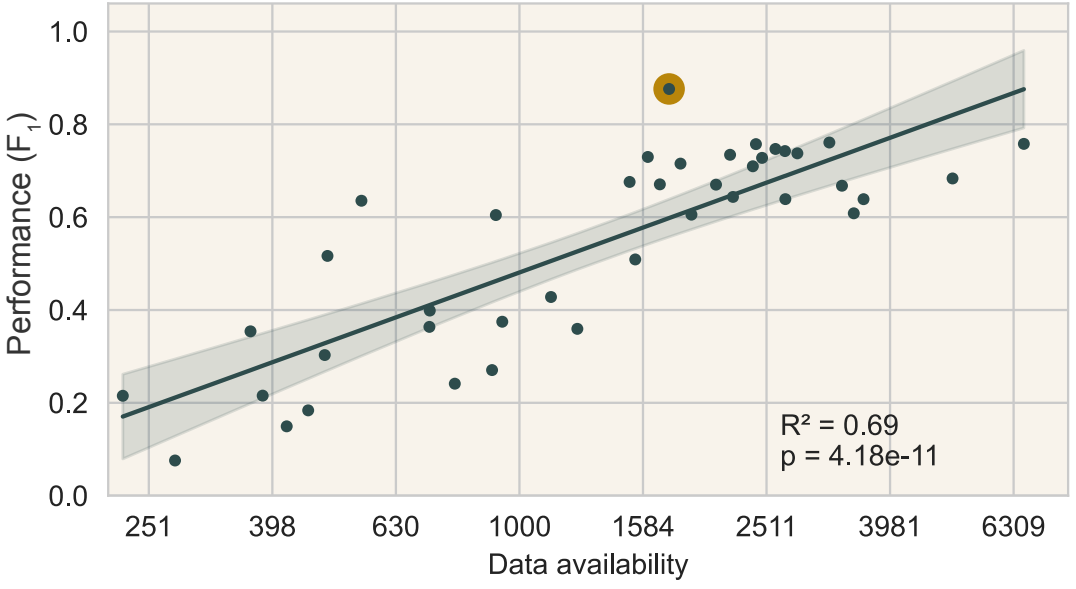
People do too, and report more



Recognizability

The AI model finds some species easier to recognize

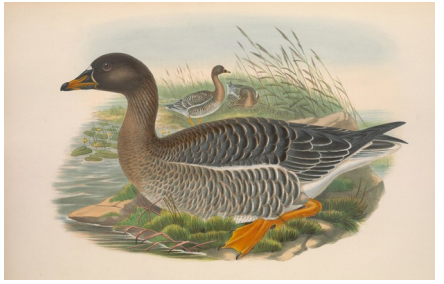
People do too, and report more

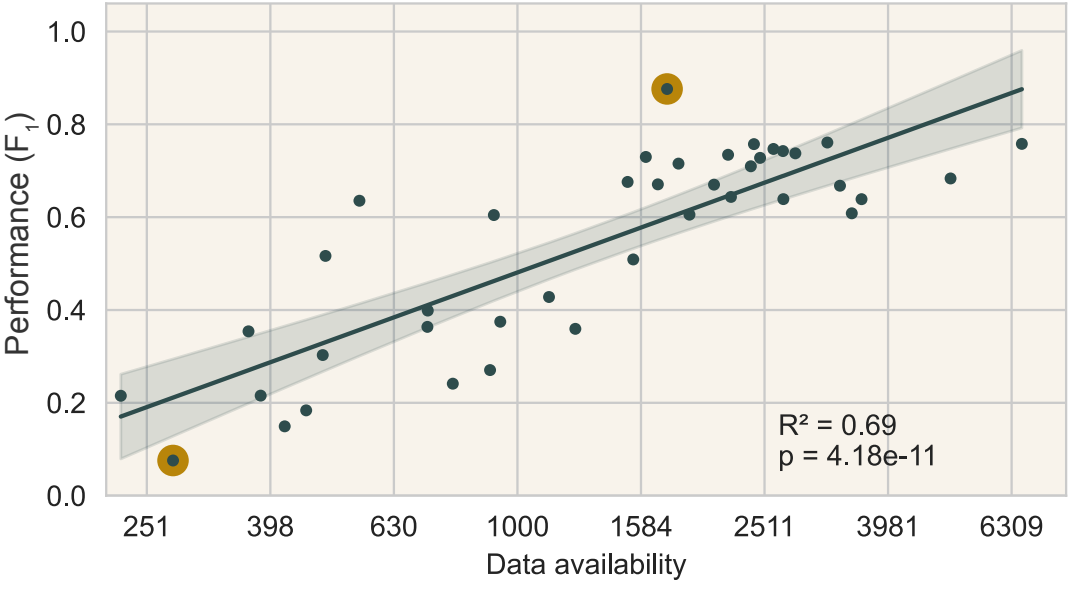


Recognizability

The AI model finds some species easier to recognize

People do too, and report more



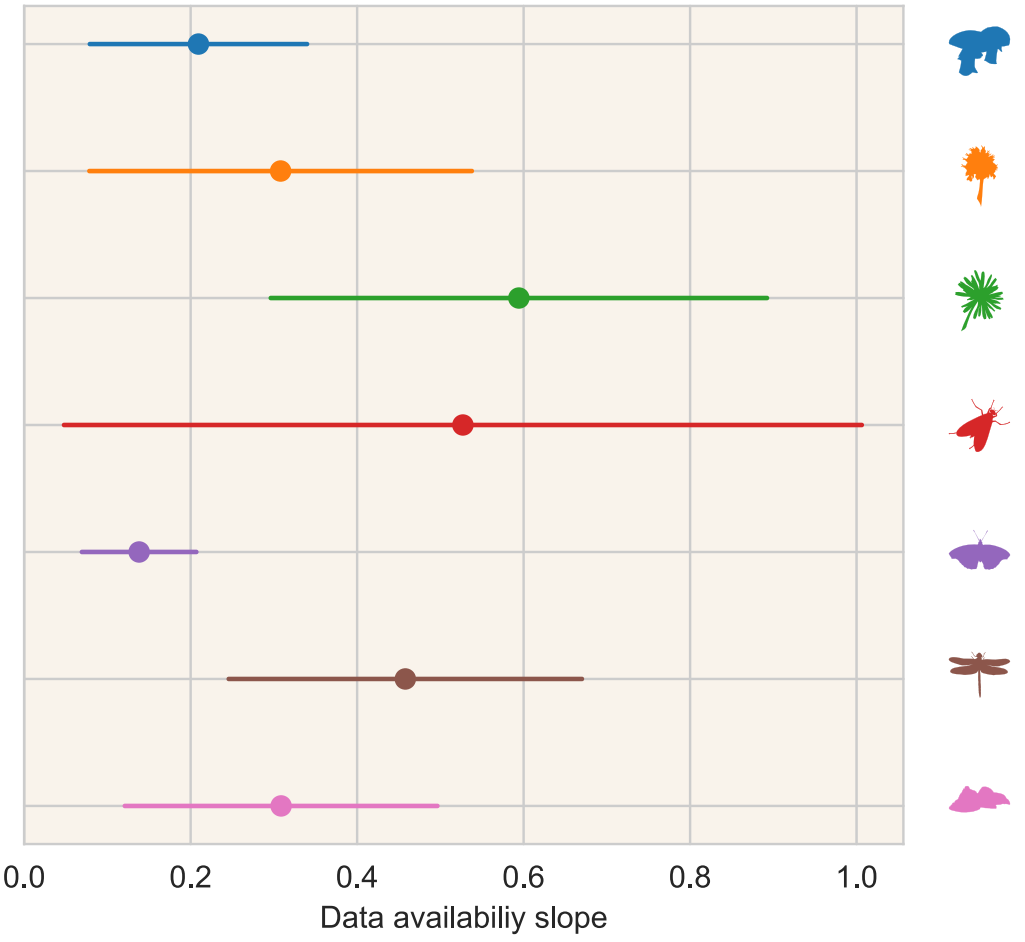


Recognizability

The AI model finds some species easier to recognize

People do too, and report more

Same pattern for other orders



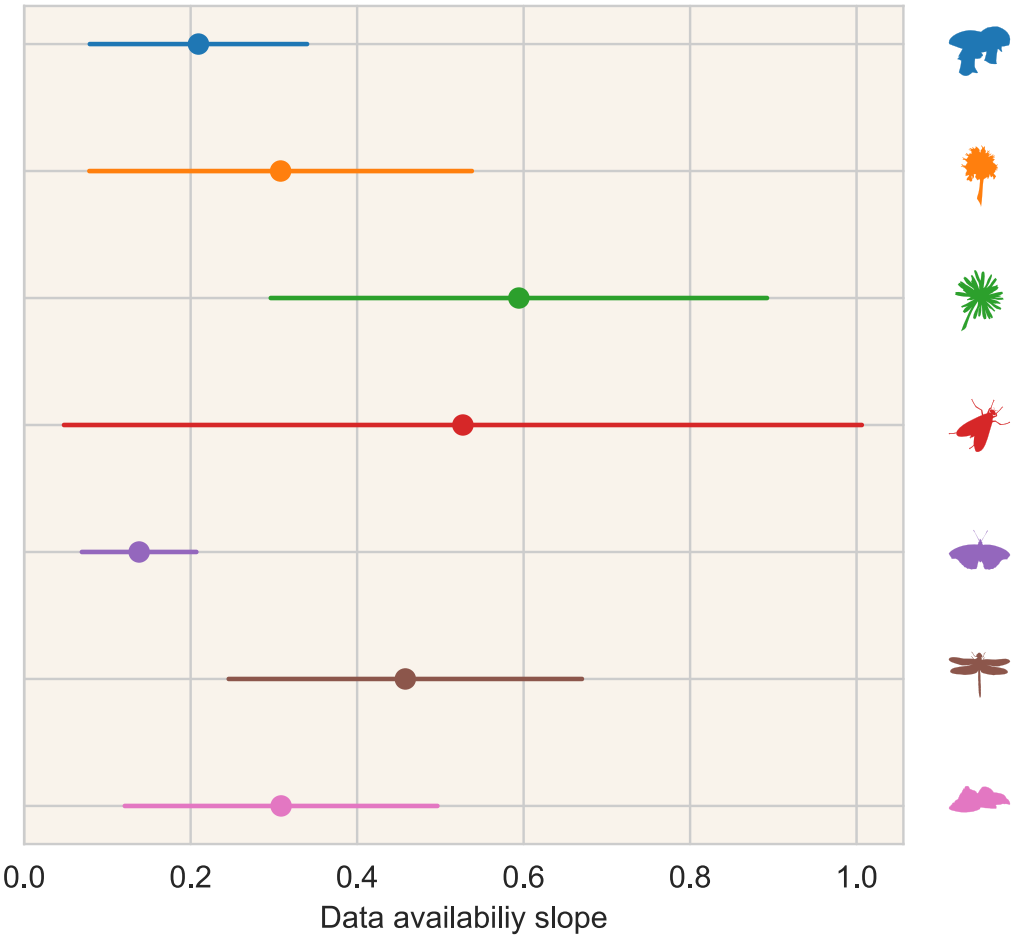
Recognizability

The AI model finds some species easier to recognize

People do too, and report more

Same pattern for other orders

Increased data likely of easier species



Final thoughts

Citizen science = valuable data + engagement

We need to cope with bias and other issues in citizen science

But there is more we can do than simply accepting the data “as is”