IECF 2024 Conference
The 4th International Conference on Forests
23–25 September 2024 | Online
MDPI

Machine Learning to Predict Selective Browsing Behavior

Zsofia Katona[1], Krisztian Katona[2,3]

[1]Vrije Universiteit Amsterdam [2] Hungarian University of Agriculture and Life Sciences

[3]National Laboratory for Health Security, Hungarian University of Agriculture and Life Sciences

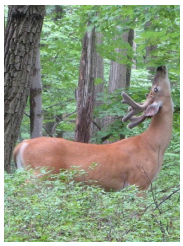
INTRODUCTION & AIM

Ungulate animals have a central impact in shaping forest ecosystems. This impact is partially exercised by selective browsing, which helps regulate the competition between different plant species and might even shape the effect of climate change on forest habitats[1]. However, this beneficial effect is often neglected in comparison to economically damaging ones[2].

Therefore, accurately predicting selective browsing behavior could help develop forest management policies which advocate biodiversity as well as economic interests.

The **aim of the project** was to:

- Train three machine learning (ML) models to predict the extent of selective browsing on woody plants.
- Integrate the highest performing model into an easy-to-use pipeline which can be applied to new forest data.



METHOD

RAW DATASET

Data originally collected in 7 forested areas in Hungary at a total of 2100 sampling points[3].

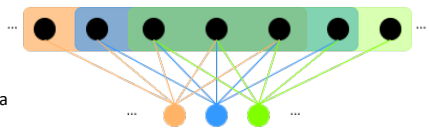
At each sampling point, per-species **woody plant supply** (total of 29 species) and ungulate **browsing activity** is recorded.

PRE-PROCESSING

Sliding-window approach applied to decrease extreme data sparseness

ORIGINAL INSTANCES:
Sampling points

NEW INSTANCES:
Forest sections aggregating data on 5 adjoining sampling points



MODEL TRAINING

INPUT

- Composition of plant supply at the sampling point (shoots per species)
- Composition of plant supply across the whole forest (spread and proportion of each species)

PREDICTION

Proportion of shoots (%) to be browsed by ungulates for each plant species

MACHINE LEARNING MODELS TESTED

- Random Forest
- Gradient Boosted Regressor
- Zero-Inflated Beta Regressor

Trained after **hyperparameter grid search** in cross-validation.

EVALUATION

THREE ASPECTS OF PERFORMANCE EVALUATED

ACCURACY
Mean absolute error

ROBUSTNESS TO SMALL SAMPLE SIZES
Change in accuracy after training on 100%, 75% and 50% of the sample

ROBUSTNESS AGAINST ZERO-COLLAPSING ESTIMATES
Ability to correctly predict larger browsing extents despite the rareness of such instances

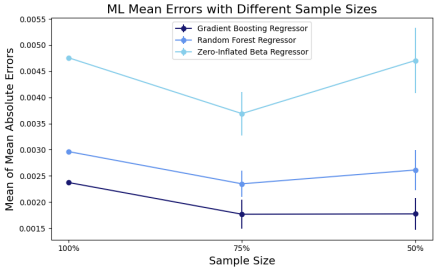
APPLICATION

Best-performing model integrated into a Jupyter Notebook pipeline which is **easy to apply to new forest data**.

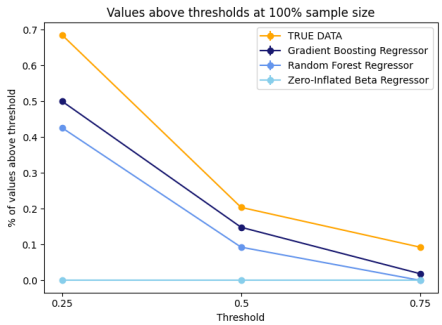
RESULTS & DISCUSSION

The **Gradient Boosting Regressor performs best** at all sample sizes, both in terms of accuracy and ability to correctly predict larger browsing extents. In any scenario, its **mean error is below 0.25%pt**.

This result was confirmed in a two-way ANOVA test (between model choice and sample size) at $p << 0.001$.



Mean absolute errors in case of the whole sample, and in case of reduced samples simulating less data collection effort. Several runs on randomly removed observations create variation.



Proportion of browsing extent values correctly predicted to be above certain thresholds using the full sample. Orange line indicates the proportion of such values in the true data. Results are close to identical in case of all sample sizes.

The most important limitations include:

- **Constraints introduced by the sliding window approach**: an animal must be able to perceive the complete plant supply in a 40-meter radius to assume selective behavior.
- Prediction pipeline may **only be used in forests which comprise woody plants present in the training data**, and not other species.
- **Performance is difficult to assess over complete forests** (instead of individual sample sections) due to **lack of data**.

Therefore, **future research** could help:

- Collect browsing data for larger and more varied forested areas.
- Construct and test ML approaches on such large datasets.
- Integrate ML solutions into GUI-based applications to further promote accessibility to forest managers and researchers.

CONCLUSION

- Machine learning can be an effective tool to predict selective browsing behavior for mindful forest management.
- Data availability is crucial for ML in ecological research, making efforts to publish ecological datasets that much more valuable.

REFERENCES

[1] K. Katona, et al. "Ungulate browsing shapes climate change impacts on forest biodiversity in Hungary." *Biodiversity and Conservation* 22 (2013): 1167-1180.
[2] Á. Fehér, K. Katona, and L. Szemethy. "Okozhatnak-e a csülkös vadfajok térségi szintű problémákat a Mátra erdőiben." *Vadbiológia* 18 (2016): 17-26.
[3] Á. Fehér, et al. "Monitoring of ungulate impact in Hungarian forested Natura 2000 sites." *Review on Agriculture and Rural Development* 3.1 (2014): 126-130.

This research was supported by the National Research, Development and Innovation Office in Hungary within the framework of the National Laboratory for Health Security programme (RRF-2.3.1-21-2022-00006).