Classifying environmental impacts with a coupled manual-automated literature review

Gretchen L. Stokes[1], John Flores[1], Jesse P. Wong[2], Connor Morang[1], Abigail J. Lynch[3], & Samuel J. Smidt[4]

[1]School of Natural Resources & Environment, University of Florida; [2]Department of Environmental Science and Policy, George Mason University; [3]National Climate Adaptation Science Center, US Geological Survey; [4]Soil and Water Sciences Department, University of Florida

THE GEOLOGICAL SOCIETY OF AMERICA® CONNECTS 2021 10–13 October Portland, Oregon, USA

Poster 84-12

1 INTRODUCTION

Addressing **rapid environmental changes** has quickly increased the body of literature on global ecological changes. **Systematic literature reviews** are important for synthesizing accumulated evidence and informing areas of publishing needs. Synthesis reviews can support **evidence-based management** and decision-making and highlight potential research gaps or publishing bias.

Traditional literature synthesis for extracting information from high volumes of text data are **expensive**, **inefficient**, and **not easily replicable**. There is a growing need for improved methods that rapidly and **efficiently synthesize** the ever-growing body of scientific literature.

Automated machine learning techniques can help increase efficiency in the manual review process by objectively learning as they process increasing amounts of relevant data. **Natural language processing** (NLP) uses **automated text classification**, article categorization, and topical extraction method to expedite literature review with text labels from a training dataset. This study provides an example of NLP for classifying literature with anthropogenic drivers and documented threats from the field of fisheries science.

RESEARCH QUESTION: How can machine learning improve efficiency of environmental literature reviews?

2 METHODS

The goal of this research was to evaluate the efficacy of coupled human and machine learning methods for reviewing large volumes of literature.

1. Screen articles for eligibility using inclusion criteria
2. Review and classify articles manually and extract data
3. Train machine learning models using manually classified dataset
4. Classify articles using machine learning models and evaluate performance metrics

We conducted a **systematic literature review** using the Web of Science search engine to identify studies that documented one or more **direct threats to inland fisheries** at a basin scale. We then performed an initial inclusion eligibility screening to identify abstracts describing one or more documented, direct threats.

Manually reviewed abstracts (n=4,336) served as the training and testing data for automated classification of the remaining abstracts (n=4,092) using **machine learning by NLP**. We used four types of text classification algorithms: binary logistic regression, naïve Bayesian classification, linear single vector machine, and k-nearest neighbor and evaluated model performance using **recall**, **precision**, and **F1 score**.

Once screened, the same **independent reviewers** extracted data from all abstracts classified as direct, including those classified using machine learning.

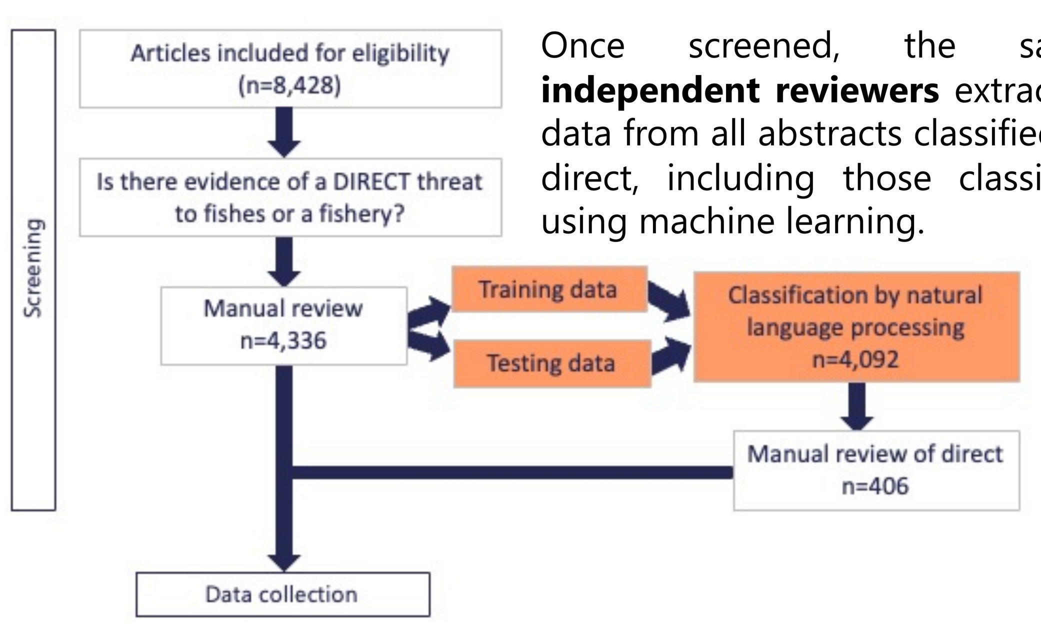

Fig. 1. Process for screening articles and reviewing them using manual and automated techniques

3 RESULTS

The results of this study synthesize **documented anthropogenic effects** on fishes, highlight research gaps from the absence of studies linking certain drivers to **direct impacts**, and demonstrate opportunities for improving **efficiency in literature** reviews through integrated machine learning and natural language processing approaches. Results suggest machine learning may be most useful for **eliminating extraneous literature** during preliminary review steps and point to the need for improved **refinement of machine learning** processes and noise reduction in data for fields (e.g., earth sciences) where syntax may be less standardized or structured.

LITERATURE SELECTION

We screened 9,361 abstracts from **45 major river basins** published in **1,008** distinct **peer-reviewed journals** from 1990 to 2020.

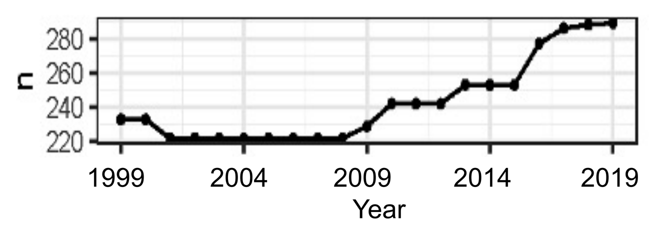

Fig. 2 Abstract counts per year in ecology

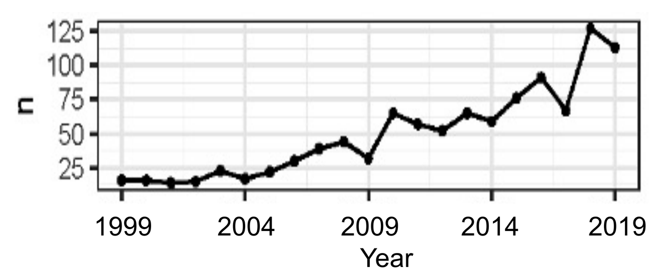

Fig. 3 Abstract counts per year in inland fisheries

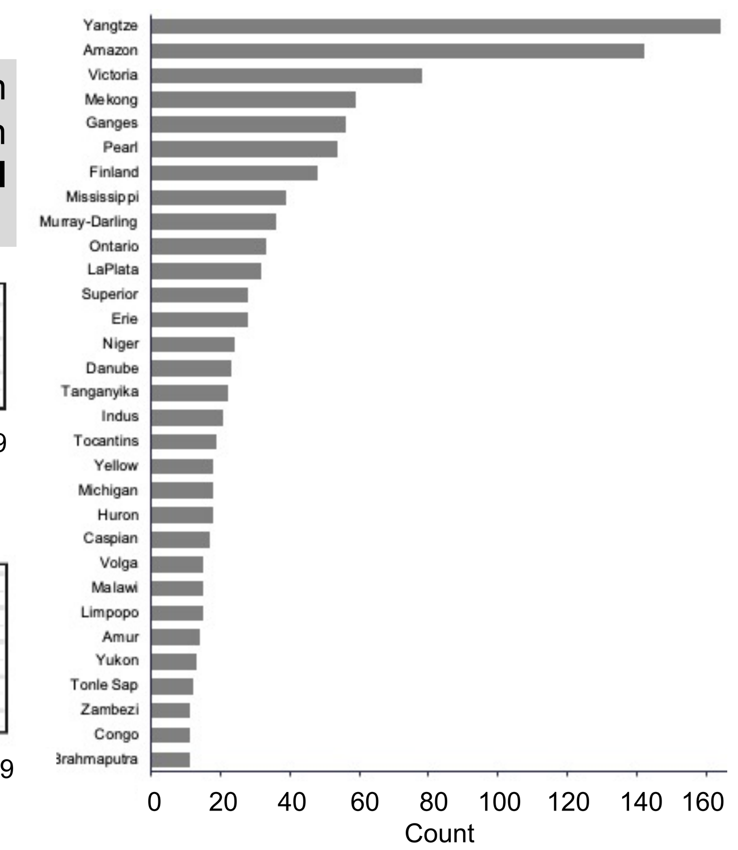

Fig. 5. Natural language processing performance metrics

Direct threat ratios were **40% greater** for manually reviewed abstracts than machine learning derived abstracts. Of the abstracts screened manually apart from automated classification (n=5,269), **16.7%** (n=881) were **direct threats** included for review and **82.3%** (n=4,388) **were excluded**.

THREATS TO FISHERIES

The most common documented drivers of threats to fisheries were **pollution** (33%, n=379), **dams**, and **fishing pressure** (each 17%, n=196, 195 respectively). All other drivers contributed less than 5% each to all articles with direct threats. While a small contributor of articles, **climate change** was documented in a relatively high proportion of basins (n=21).

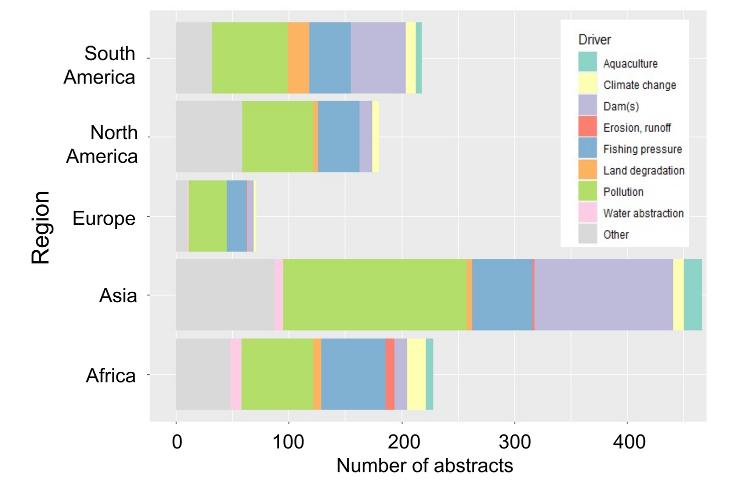

Fig. 6. Total counts of publications by region and driver type

Impact distribution strongly corresponded with top drivers: pollution was linked to 87% of **bioaccumulation** impacts, dams to 56% of **fragmentation** impacts, and fishing pressure to 93% of **overfishing** impacts. **Biodiversity loss** and **disease** were observed in 8% and 4% of articles. 70% of articles included documented effects on **multiple species** and 28% on single species. The **data supported logical links**: aquaculture linked to disease, climate change to biodiversity loss, erosion to sedimentation, land use change to fragmentation and biodiversity loss, and water abstraction to changes in flow.

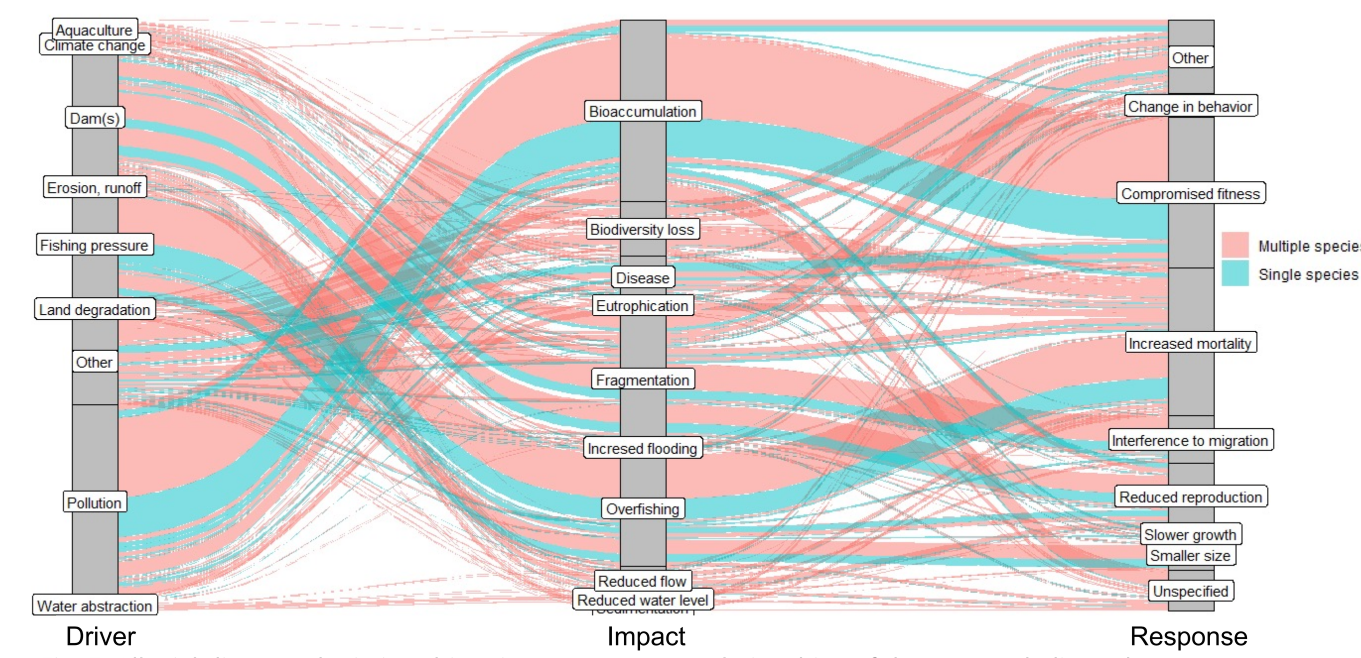

Fig. 7. Alluvial diagram depicting driver-impact-response relationships of documented, direct threats to fisheries in 45 hydrological basins; colors represent impacts on multiple (pink) and individual (blue) species.

MACHINE LEARNING CLASSIFICATION

Of the abstracts screened by NLP (n=4,092) using the **best performing model** (linear regression), 9.9% (n=406) were classified as direct threats. These were classified with **~65% precision** (i.e., positive predictive value). Manual review revealed 67% (n=251) were indeed correctly classified by NLP classification and were suitable for inclusion for data review.

Table 1. Natural language processing performance metrics

	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Direct	0.56	0.45	0.50	0.64	0.40	0.50	1.00	0.03	0.06	0.49	0.36	0.41
Indirect	0.90	0.93	0.92	0.89	0.96	0.92	0.84	1.00	0.92	0.88	0.93	0.91
Average Accuracy			0.86			0.87			0.84			0.84

Of the models used, **linear single vector machine** (SVM), and **k-nearest neighbor** (kNN) classification consistently underperformed and produced especially poor recall and F1 outputs for the study targets (direct threats). **Naïve Bayesian classification** and logistic regression performed similarly, with 11.1% higher recall in the Bayes model and 12.5% higher precision in the regression model for direct classification. After **iterative testing** and **comparative review** of model performance using the training data set, we selected the **logistic regression** for its **consistently highest accuracy** in classifying unclassified articles. The model selected approximately 10% of unclassified articles as having 'direct' threats, with 65% precision that those classified as 'direct' were correct. **Post-hoc NLP algorithms** applied to the manually reviewed set of machine-classified articles revealed **similar model performance**.

Despite the difference in basins and geographic locations represented in the two sets of articles (i.e., those manually classified contained articles for different basins than those automatically classified), **post-hoc model performance** shows only **nominal improvement** in the classification of direct threats. As expected from the addition of correctly classified direct articles (reintegrated after automated classification and manual review), **model improvements** occurred only with direct classification metrics, while decreased performance occurred for indirect classification. There was a 2% reduction in overall accuracy for both naïve Bayes and regression models. As such, we found **no evidence to suggest bias** from the selection of articles for manual and machine classification.

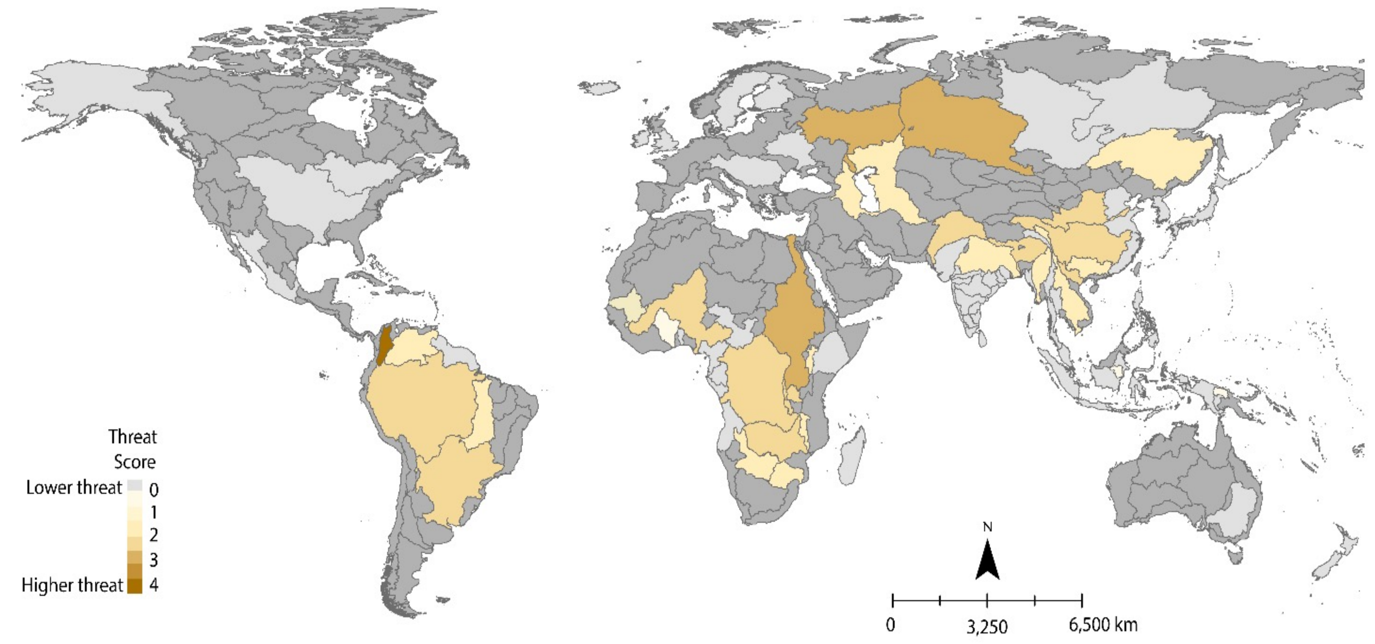

Fig. 8. Threat scores derived from direct threats to fisheries, as documented by systematic abstract review of 45 major basins important to fisheries, where darker colors represent basins with higher mean threat scores and lighter colors represent basins with lower mean threat scores.

4 DISCUSSION

This study advances the understanding of trends with initial literature synthesis using NLP for inland fisheries. Motivated by the need for understanding driver-impact associations of fish for the development of indicators and proxy measures of stressors to inland fisheries, this study compiled existing driver-response-impact links in the literature and applied a paired method for literature classification.

Results **align with expected relationships** of known causal relationships and highlight a need for documenting responses of some **known, but underrepresented, relationships** in the literature of drivers of change and their impact on species responses.

NLP performance points both to the usefulness of machine learning for literature synthesis, and the necessity of human contributions. Low to intermediate NLP text classification performance may signal noisy text from **unstructured or unstandardized data** reporting in fisheries literature, or ecology literature more broadly.

CONCLUSIONS

1. Documented links suggest all direct drivers of anthropogenic change exist in all major fisheries; some relationships are strongly documented.
2. Documented links reflect publishing and research bias by basin locations and types of studies; these biases create noticeable gaps in the literature.
3. Despite relatively standardized abstract structure and length, there is still inherent human objectiveness in the literature and widely variable use of words and interpretability.
4. NLP may be most useful for improving efficiency in preliminary steps of fisheries literature synthesis (i.e., classifying extraneous or redundant articles not pertinent to the study.
5. NLP and automated machine learning performance for fisheries and ecology literature may benefit from integrated noise reduction techniques.
6. Both humans and machine learning are necessary for effective and efficient literature review. With exclusively humans, reviews are costly and not replicable; with only computers, there is no training data and no "right" answer.

KEY REFERENCES

- Adger, W.N. (2006). Vulnerability. *Glob. Environ. Change*, 16, 268-281.
- Dudgeon, D. *et al.* (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biol. Rev.*, 81(2), 163-182.
- Lynch, A.J. *et al.* (2016). The social, economic, and ecological importance of inland fishes and fisheries. *Environ. Reviews*, 24, 1-8.
- Polanin, J. R. *et al.* (2019). Best practice guidelines for abstract screening large-evidence systematic reviews and meta-analyses. *Res. Synth. Methods*, 10(3), 330-342.

Acknowledgments

We thank Chelsie Romulo for conceptualization of machine learning application and use, Siddhesh Ambokar for technical assistance with machine learning algorithms and constructive feedback. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1842473. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Contact Information:
gstokes@ufl.edu
ssmidt@ufl.edu

Land and Water Lab
University of Florida
landandwaterlab.org

UF UNIVERSITY of FLORIDA