

Generating Better Macroecological Data From Literature OR **Gold** in the **Garbage Dump**

Noam Ross ([@noamross](#)) , Allison M. White, Cale Basaraba, Brooke Watson, Erica Johnson, Karissa Whiting, Melanie Kirshenbaum, Jacob Kotcher, Ayomide Sokale, Mushtaq Dualeh, Zach Matson, Nchedo Ezekoli, Carlos Zambrana-Torrel, Peter Daszak



Origins at #ESA2017

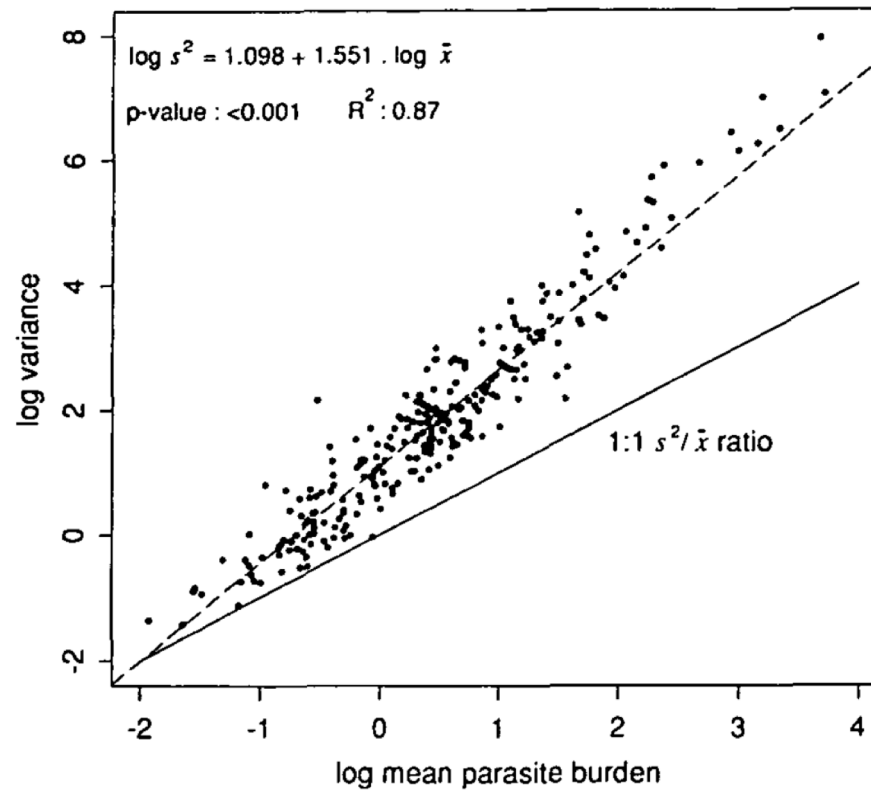
Noam: "We're kind of scraping the barrel with these host-pathogen-interactions-from-the-literature papers."

John Drake: "No, we keep finding new things."

Noam: "We're kind of scraping the barrel with these host-pathogen-interactions-from-the-literature papers"

John Drake: "No, we keep finding new things"

Carl Boettiger: "The problem is that we're still combing through these papers by hand."



Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review

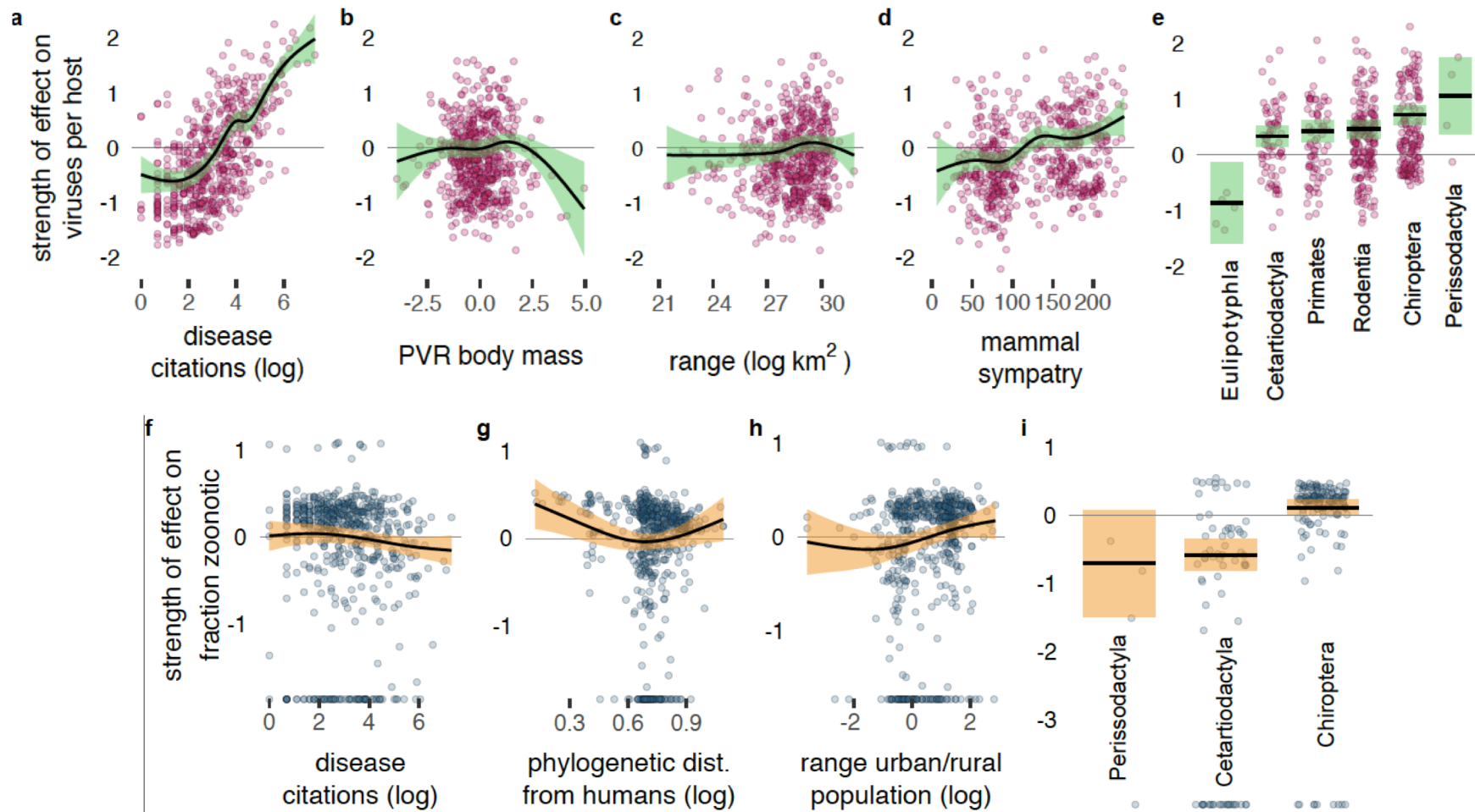
D. J. SHAW¹ and A. P. DOBSON²

¹Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK

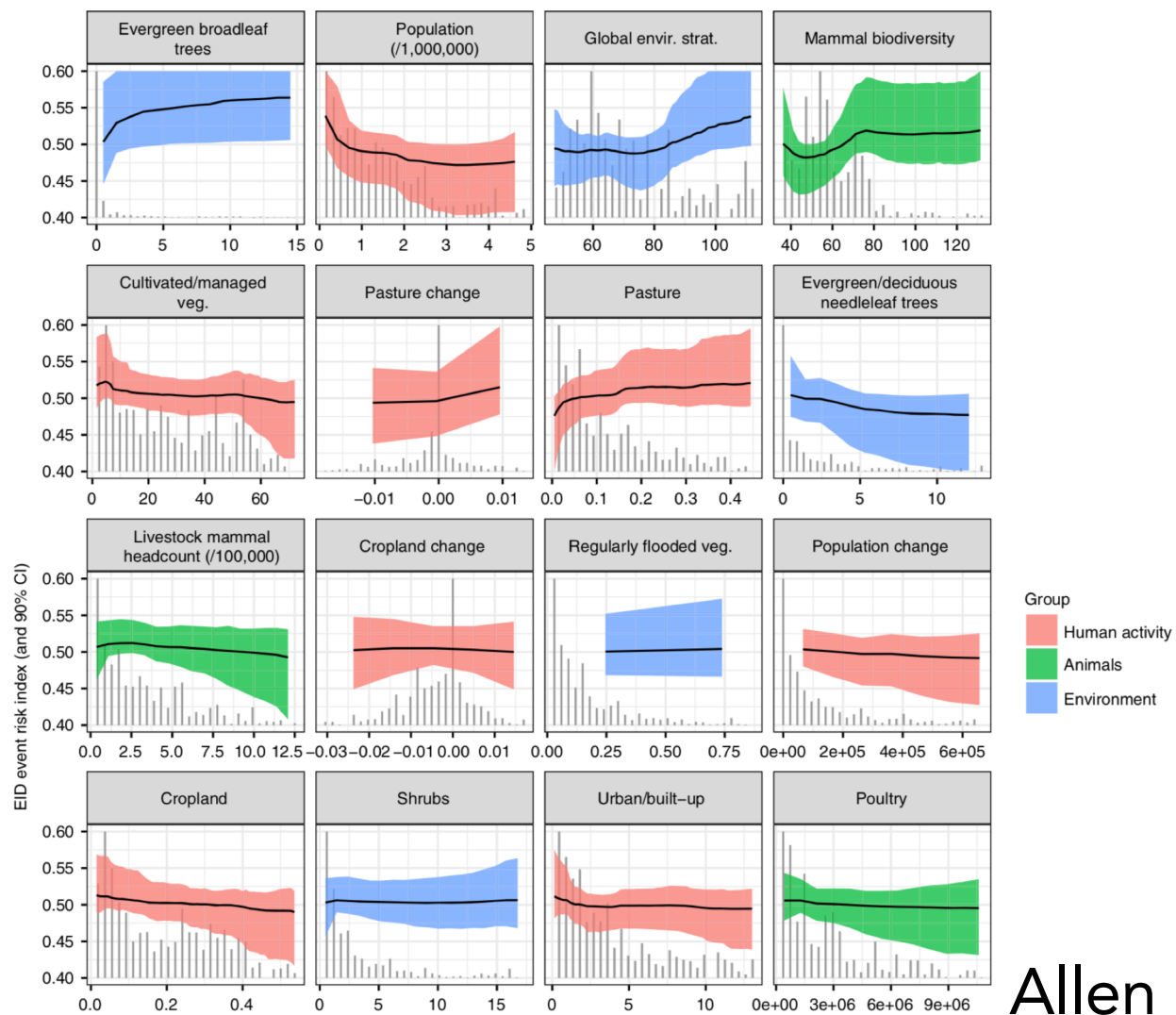
²Department of Ecology and Evolutionary Biology, Eno Hall, Princeton University, Princeton, New Jersey, USA

SUMMARY

Dobson and Shaw (1995)



Olival et al. (2017)



Allen et al. (2017)

Dobson and
Shaw (1995)

263 Host-Parasite
Associations

Allen et al. (2017)

747 literature sources

Olival et al. (2017)

1317 literature sources



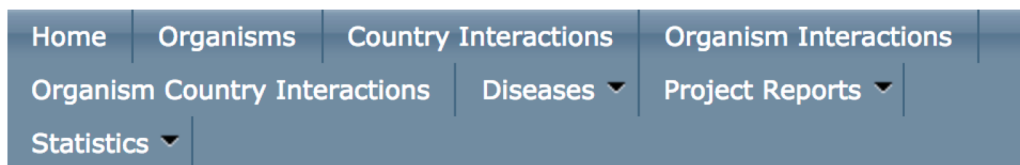
Welcome to the Interaction Web Database. This is a nonprofit, cooperative database containing published data on species interaction networks. It is hosted by the [National Center for Ecological Analysis and Synthesis](#), at the University of California, Santa Barbara, U.S.A.

You must read our [Privacy Policy](#) before using the Interaction WebDatabase website

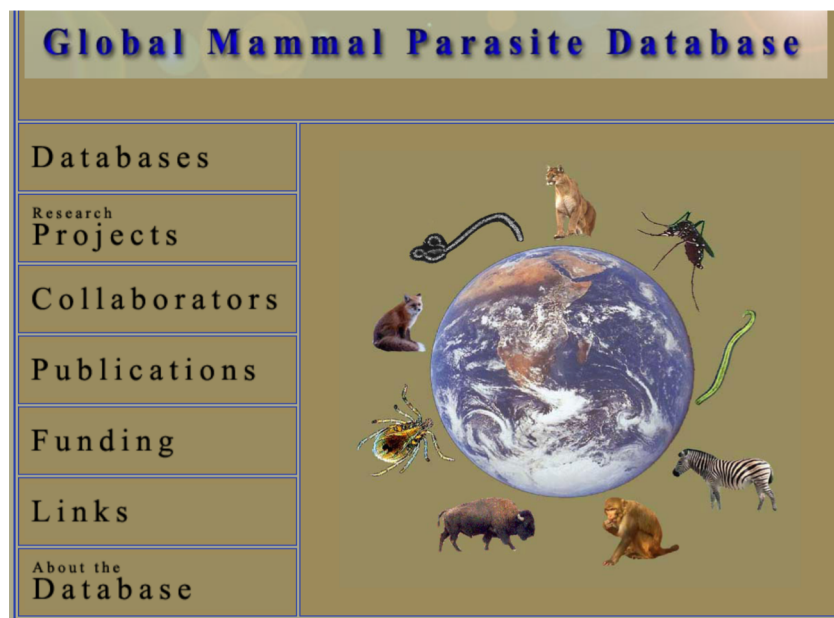


What is GloBI?

Global Biotic Interactions (GloBI) provides [open access](#) to species interaction data (e.g., predator-prey, pollinator-plant, pathogen-host, parasite-host) by combining existing [open datasets](#) using [open source](#) software.



Welcome to the [ENHanCED](#) Infectious Diseases (EID2) database. This evidence-based database annotates and integrates existing data on vectors, hosts and their pathogens. The database enables users to obtain:







Landfill Mining

**A good idea defeated
by high soil processing
costs?**

**US and European
experiences.**

(A real slide from someone's very different conference)

A new landfill mining project: Macroecology of Antibiotic Resistance Emergence

Identify global correlates of emergence of new
AMR mutations

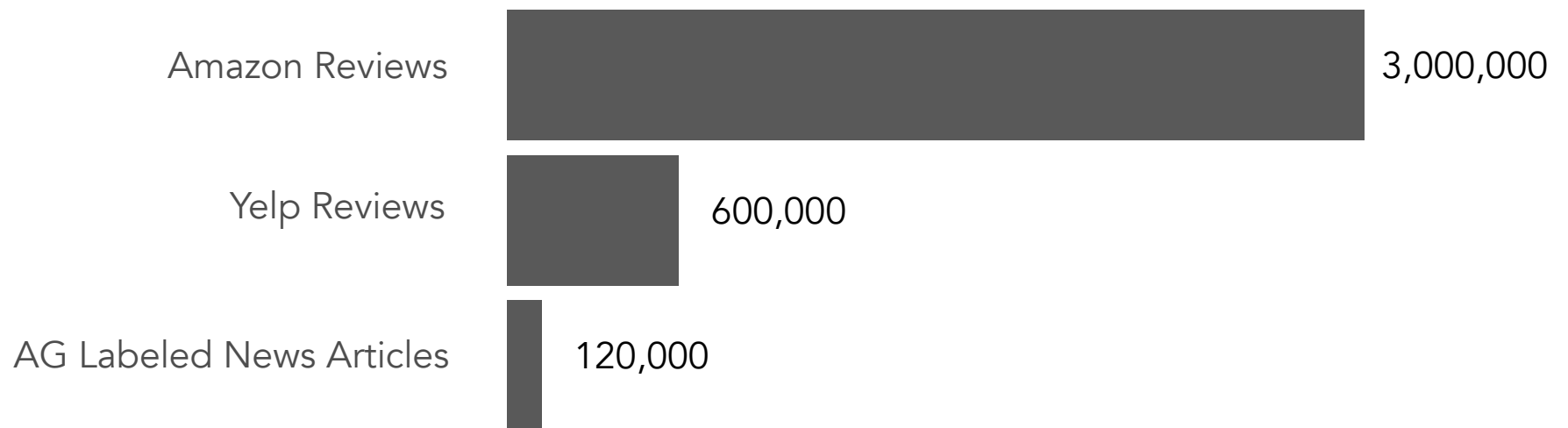
Use machine learning and text classification to
identify relevant events from the literature



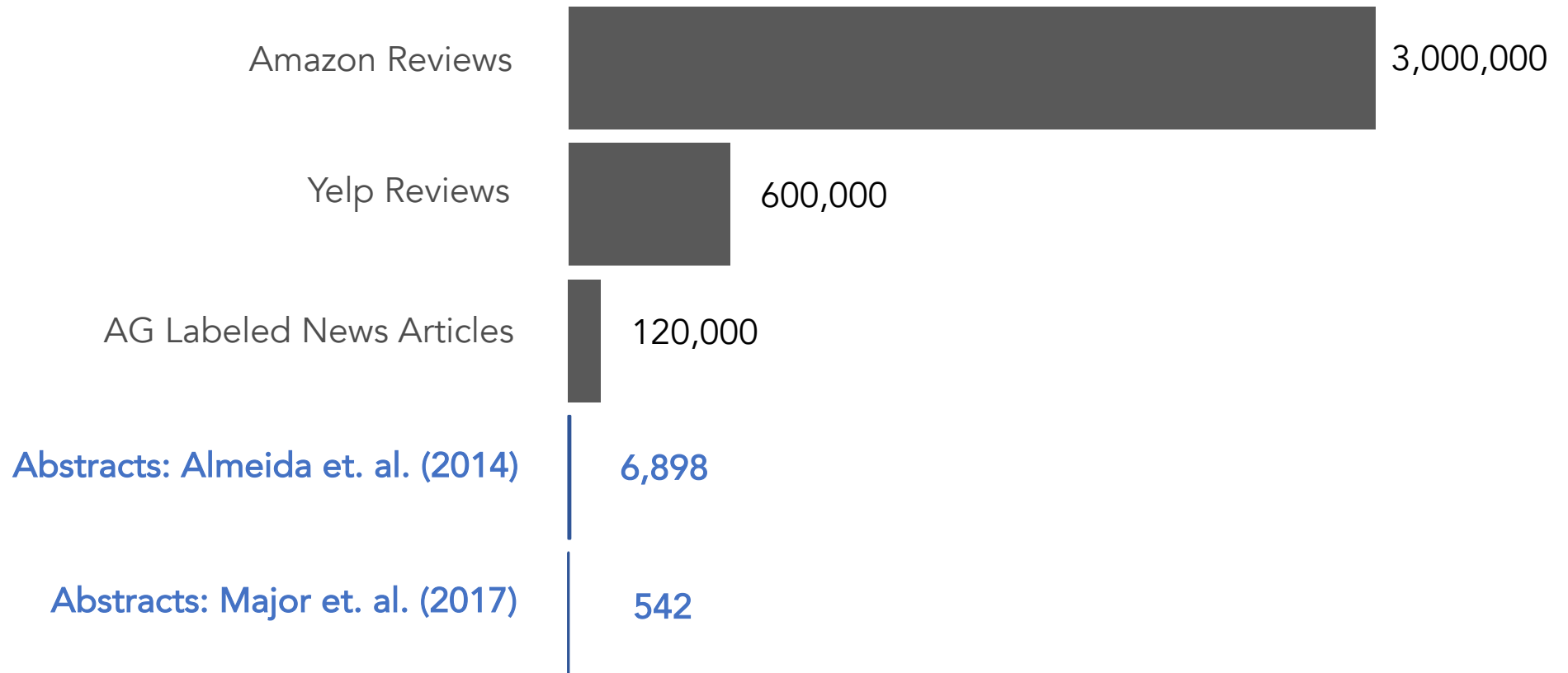
A **Wicked** Text-Classification Task

Does an article or report describe at least one **clinical case** of infectious disease in a **human patient** caused by a **novel resistance** to a particular **antimicrobial combination** in **bacteria**?

Data Set Sizes in Neural-Net Text Classification Research



Data Set Sizes in Neural-Net Text Classification Research



24,000 articles found via search for
AMR keywords

Each abstract classified by **2** readers

1000+ hours of screening effort

~**25+** person-weeks

Very Deep Convolutional Networks for Text Classification

Alexis Conneau
Facebook AI Research
aconneau@fb.com

Holger Schwenk
Facebook AI Research
schwenk@fb.com

Yann Le Cun
Facebook AI Research
yann@fb.com

Loïc Barrault
LIUM, University of Le Mans, France
loic.barrault@univ-lemans.fr

GloVe: Global Vectors for Word Representation

Jeffrey Pennington, Richard Socher, Christopher D. Manning
Computer Science Department, Stanford University, Stanford, CA 94305
jpennin@stanford.edu, richard@socher.org, manning@stanford.edu

Character-level Convolutional Networks for Text Classification*

Xiang Zhang Junbo Zhao Yann LeCun
Courant Institute of Mathematical Sciences, New York University
719 Broadway, 12th Floor, New York, NY 10003
{xiang, junbo.zhao, yann}@cs.nyu.edu

Enriching Word Vectors with Subword Information

Piotr Bojanowski* and Edouard Grave* and Armand Joulin and Tomas Mikolov
Facebook AI Research
{bojanowski, egrave, ajoulin, tmikolov}@fb.com

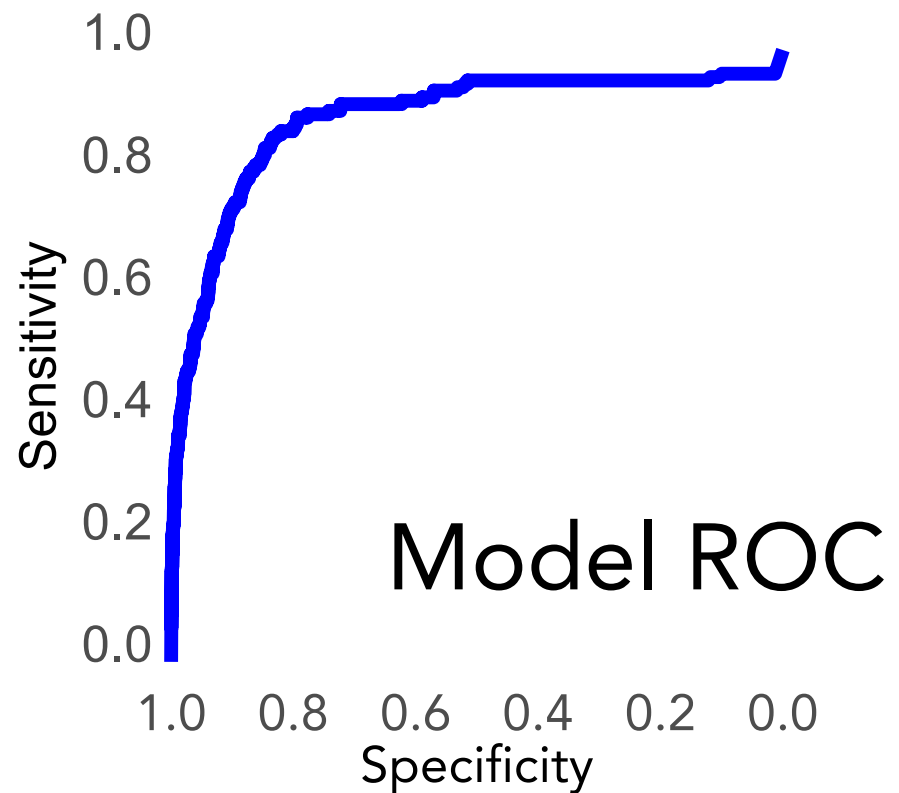
*fast*Text



Keras

7% of articles
were true cases

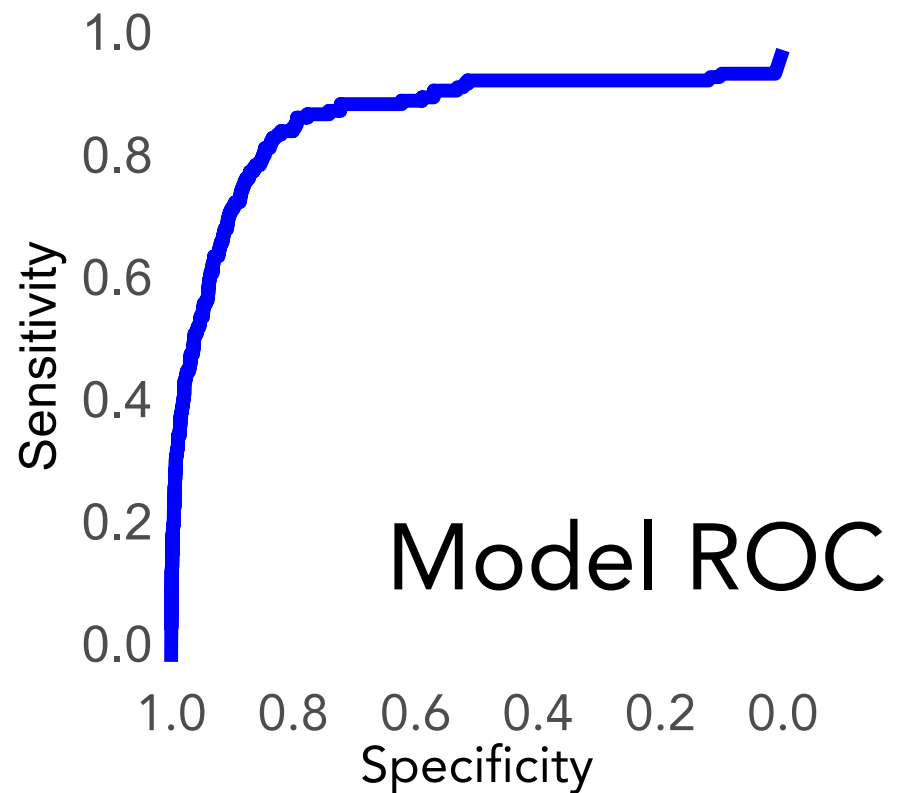
85% balanced
accuracy



7% of articles
were true cases

85% balanced
accuracy

We can get **95%** of
events automating **65%** of screening



Next: Automated Semantic Extraction

We present the first known case of osteomyelitis associated to an CTX-M-15-producing *A. hydrophila* isolate carrying also the *bla*_{TEM-1} gene, as well as genes conferring resistance to aminoglycosides, *aac(3)-IIa*, and to aminoglycosides and ciprofloxacin, *aac(6')-Ib-cr*.

A 37-year-old woman was hospitalised in May of 2009 for an open fracture grade III of the right tibial pilon. External fixators were placed and empiric treatment with ciprofloxacin plus vancomycin was established. Seven

The patient was treated with meropenem (1g tid) plus amikacin (750 mg bid) for a course of 30 days, and treatment with meropenem alone was continued for 15 days more. The patient's progress was monitored by acute phase reactants, serial imaging studies and clinical evolution.

Machine learning may not help us infer mechanisms but it can help find data!

Publish **negative** data from your literature review!

Noam Ross (@noamross) , Allison M. White, Cale Basaraba, Brooke Watson, Erica Johnson, Karissa Whiting, Melanie Kirshenbaum, Jacob Kotcher, Ayomide Sokale, Mushtaq Dualeh, Zach Matson, Nchedo Ezekoli, Carlos Zambrana-Torrel, Peter Daszak



USAID
FROM THE AMERICAN PEOPLE

PREDICT



NVIDIA ACCELERATED COMPUTING



Microsoft

| AI for Earth



EcoHealth Alliance