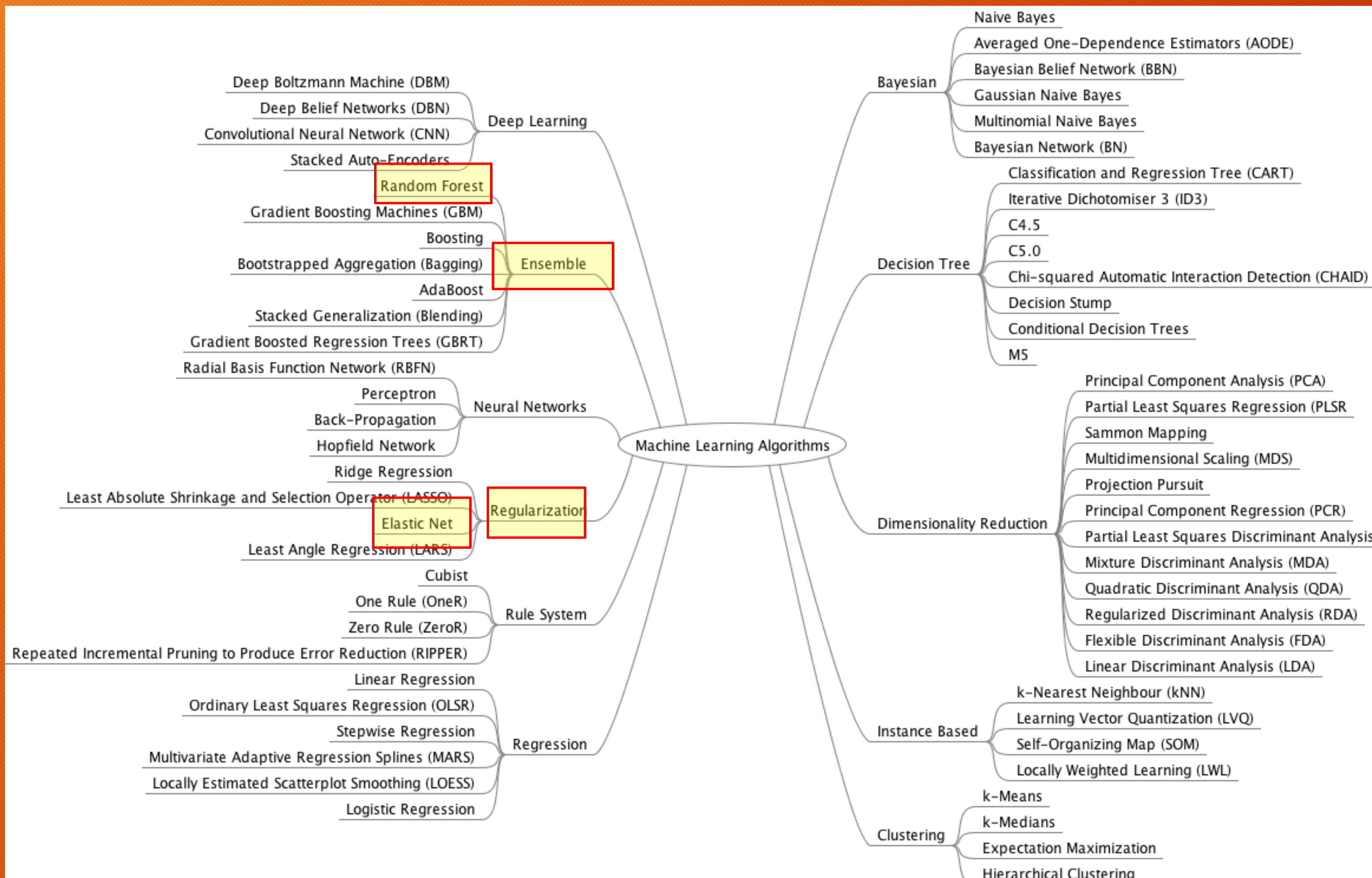
Machine Learning and Internet-Based Treatments: Opportunities and Challenges

Jason Shumake, Rahel Pearson, Derek Pisner, and Christopher Beevers
Institute for Mental Health Research
The University of Texas at Austin



Provided by Jason Brownlee

Typical data sets for treatment studies

- Clinic- or lab-based studies
  - Bountiful measures, limited observations
- Large healthcare databases
  - Bountiful observations, limited measures

Where ensembles and deep learning shine

- ✓Bountiful measures + bountiful observations (AKA "Big Data")
  - Complex patterns can be both discovered and validated
- Clinic- or lab-based studies?
  - Bountiful measures, limited observations
  - Deep phenotypes, but complexity cannot be validated
  - Expect little or no improvement over simpler models
- Large healthcare databases?
  - Bountiful observations, limited measures
  - Shallow phenotypes, so complexity cannot be discovered
  - Expect little or no improvement over simpler models

How do we get the very large, feature-rich data sets we need?

Psychological Medicine

cambridge.org/psm

Original Article

Cite this article: Pearson R, Pisner D, Meyer B, Shumake J, Beevers CG (2018). A machine learning ensemble to predict treatment outcomes following an Internet intervention for depression. *Psychological Medicine* 1–12. https://doi.org/10.1017/S003329171800315X

A machine learning ensemble to predict treatment outcomes following an Internet intervention for depression

Rahel Pearson[1], Derek Pisner[1], Björn Meyer[2,3], Jason Shumake[1] and Christopher G. Beevers[1]

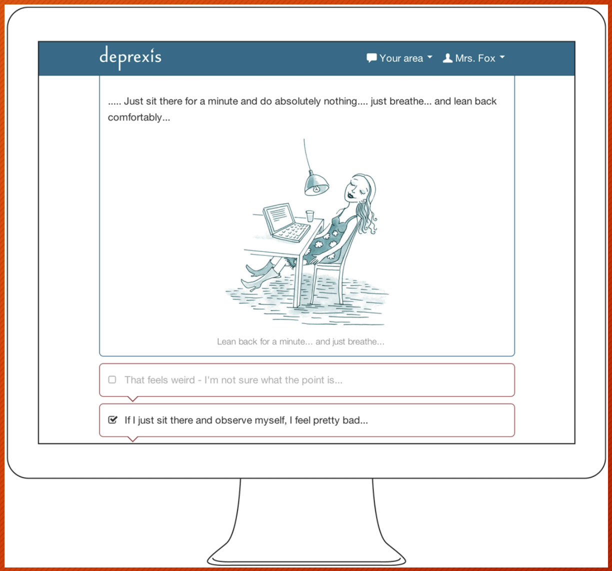
[1]Department of Psychology, Institute for Mental Health Research, University of Texas at Austin, Austin, TX, USA; [2]Gaia AG, Hamburg, Germany and [3]University of London, London, England, UK

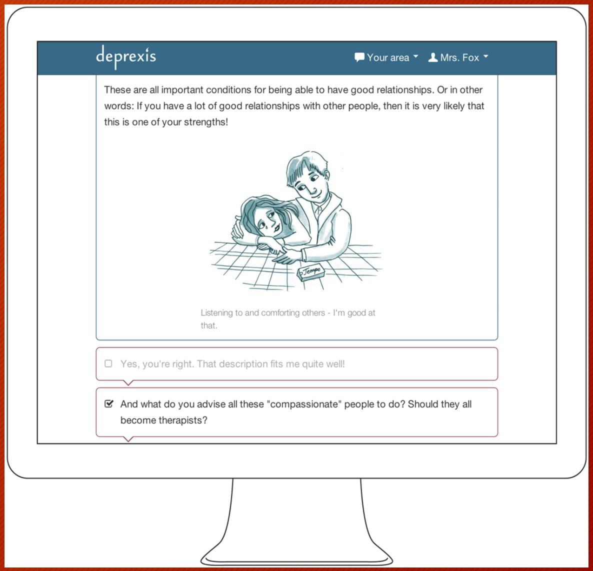
Abstract

Data from clinical trial of Deprexis

Includes:

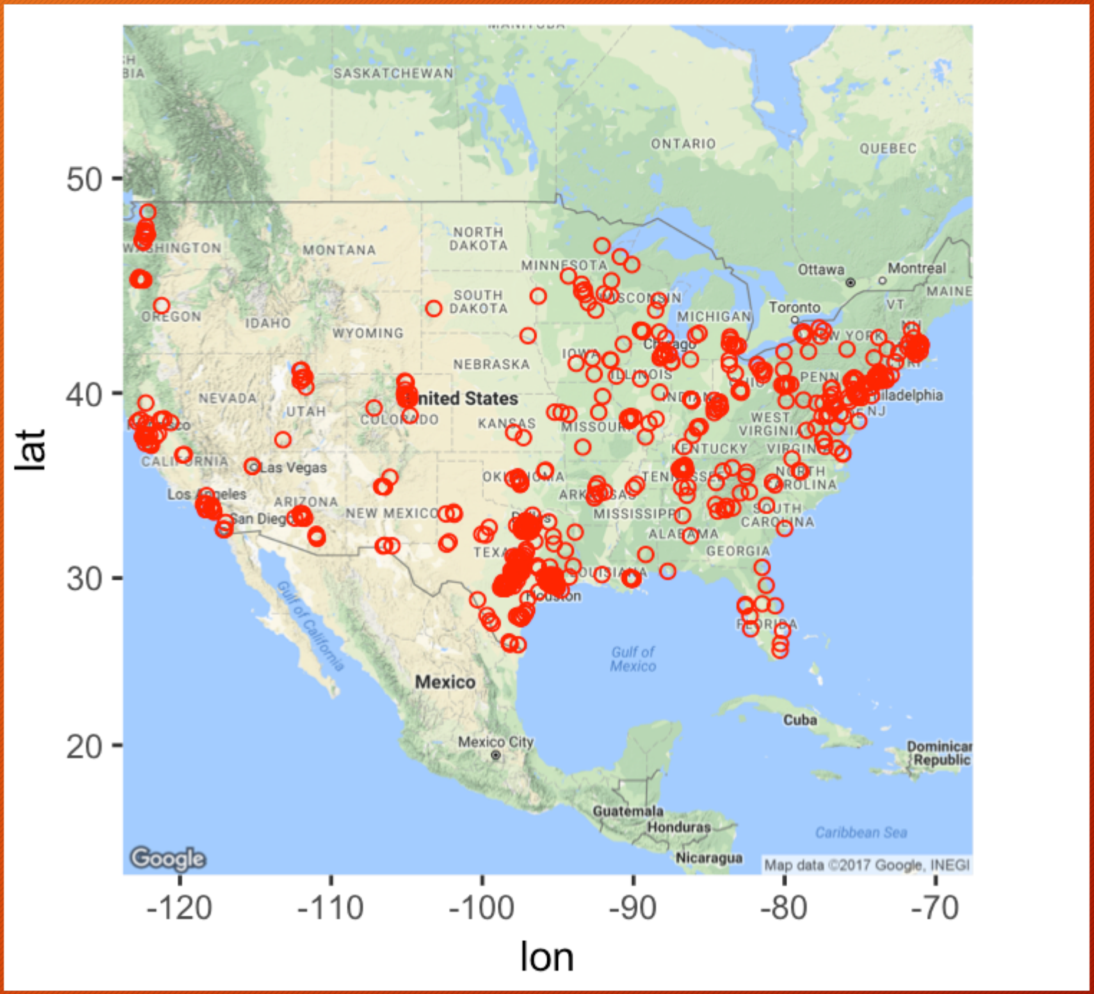
- Psycho-education
- Behavioral Activation
- Cognitive Modification
- Relaxation
- Acceptance and Mindfulness
- Problem-Solving
- Interpersonal Skills
- Positive Psychology
- But also: Dream work and Childhood experiences





Data from clinical trial of Deprexis

- 8-week course of Internet intervention (*N* = 283)
- Candidate predictors (*P* = 120)
  - Baseline depression (QIDS and HRSD items)
  - Demographics
  - Other Psychopathology (PDSQ scales)
  - Treatment expectations
  - Sheehan Disability Scale
  - History of early life stress (risky families)
  - Family history of mental illness
  - Antidepressant usage
  - “ZNA”

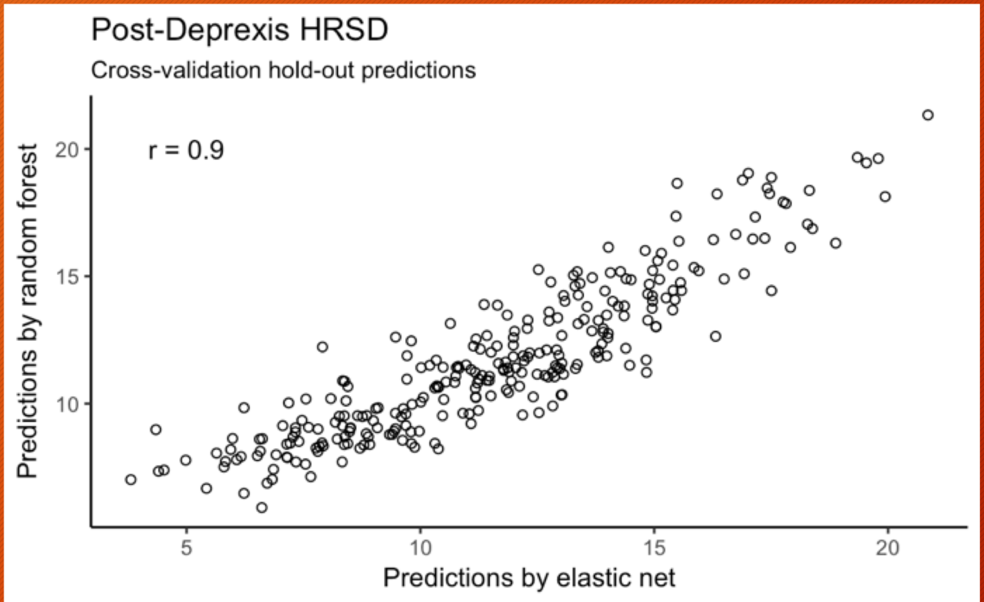

lat
50
40
30
20
-120
-110
-100
-90
-80
-70
lon
SASKATCHEWAN
ONTARIO
QUEBEC
WASHINGTON
MONTANA
NORTH DAKOTA
MINNESOTA
Ottawa
Montreal
MAINE
OREGON
IDAHO
WYOMING
SOUTH DAKOTA
WISCONSIN
MICHIGAN
Toronto
VT
NEW YORK
NEBRASKA
IOWA
Chicago
NEVADA
UTAH
United States
ILLINOIS
INDIANA
PENN
Philadelphia
NJ
COLORADO
KANSAS
MISSOURI
WEST VIRGINIA
VIRGINIA
CALIFORNIA
Las Vegas
KENTUCKY
OKLAHOMA
TENNESSEE
NORTH CAROLINA
Los Angeles
ARIZONA
NEW MEXICO
ARKANSAS
MISSISSIPPI
SOUTH CAROLINA
San Diego
ALABAMA
TEXAS
GEORGIA
LOUISIANA
Houston
FLORIDA
Gulf of California
Gulf of Mexico
Mexico
Cuba
Dominican Republic
Mexico City
Guatemala
Honduras
Caribbean Sea
Nicaragua
Google
Map data ©2017 Google, INEGI

“ZNA”: ZIP code predictors

- Some examples:
  - median household income
  - ethnic/racial diversity
  - population density
  - crime rate
  - access to mental healthcare providers

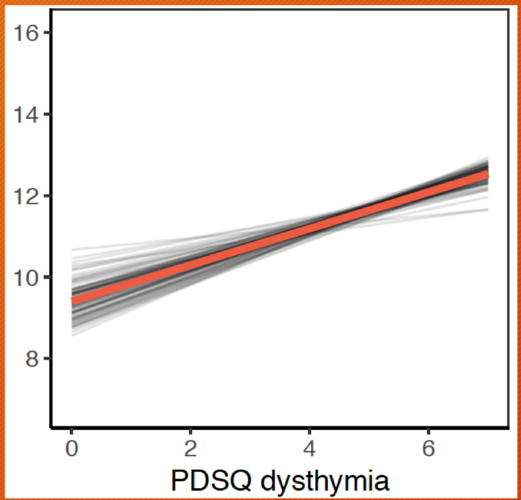
	Prediction R^2	95% CI
HRSD		
Benchmark	0.17	0.07–0.26
Random forest	0.23	0.14–0.31
Elastic net	0.24	0.14–0.33
Random forest/elastic net ensemble	0.25	0.16–0.33
Gain for ensemble model	**+0.08**	**+0.008 to +0.15**
Disability		
Benchmark	0.20	0.10–0.31
Random forest	0.24	0.13–0.34
Elastic net	0.24	0.15–0.33
Random forest/elastic net ensemble	0.25	0.16–0.35
Gain for ensemble model	**+0.05**	**−0.003 to +0.10**
IDAS-Well Being		
Benchmark	0.18	0.08–0.27
Random forest	0.26	0.19–0.34
Elastic net	0.29	0.19–0.40
Random forest/elastic net ensemble	0.29	0.21–0.38
Gain for ensemble model	**+0.12**	**+0.05 to +0.19**

Elastic Net and Random Forest make similar predictions

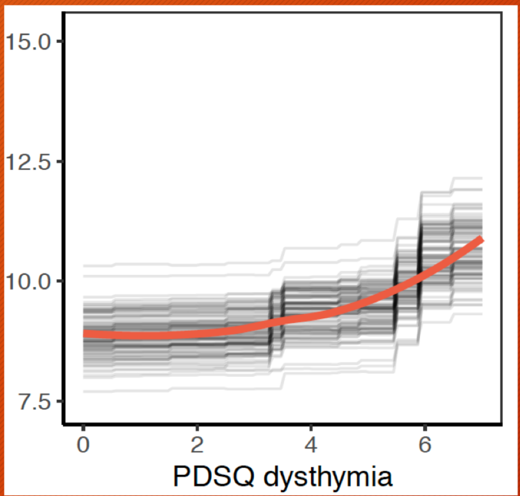


Partial Dependence of HRSD on Dysthymia

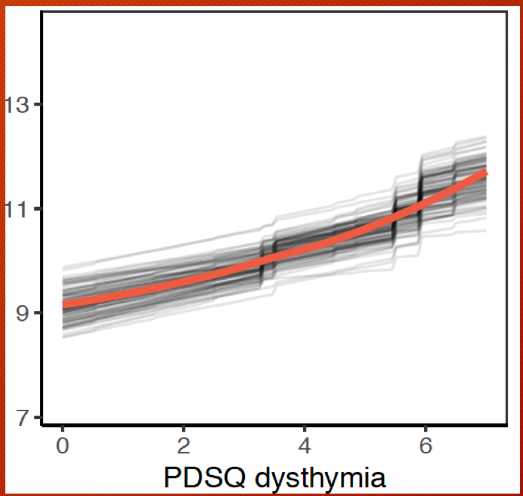
Elastic Net



Random Forest

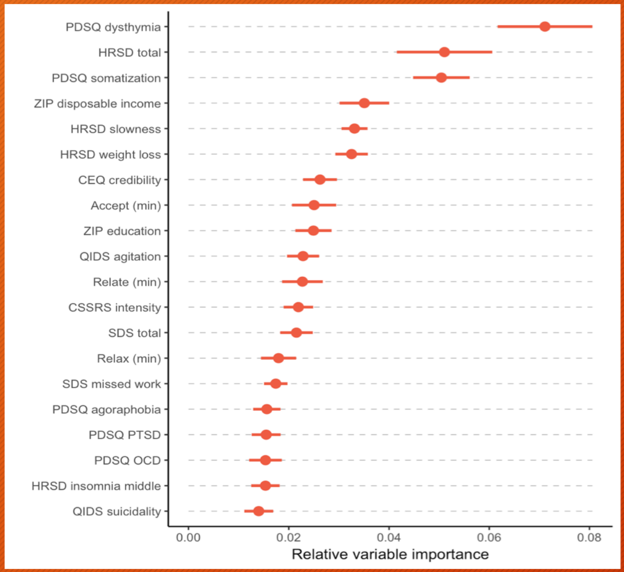


Elnet-RF Ensemble

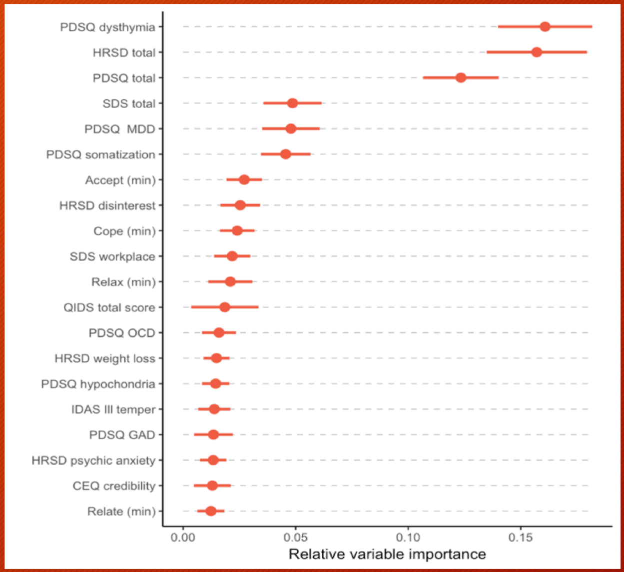


Relative importance of predictors of HRSD outcome

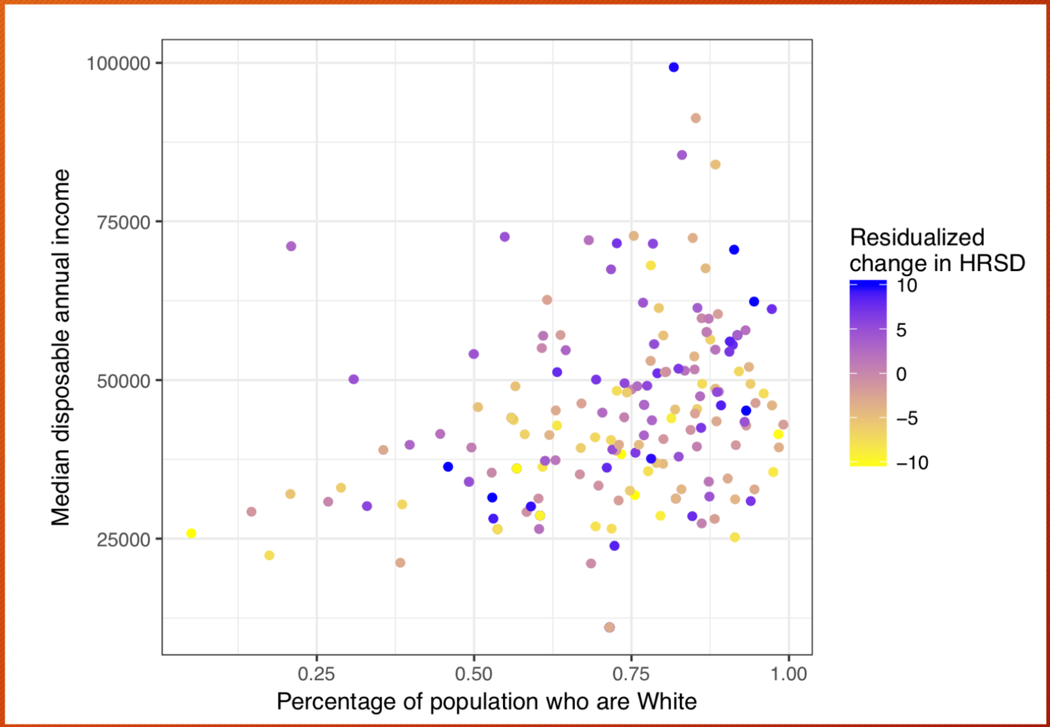
Elastic Net

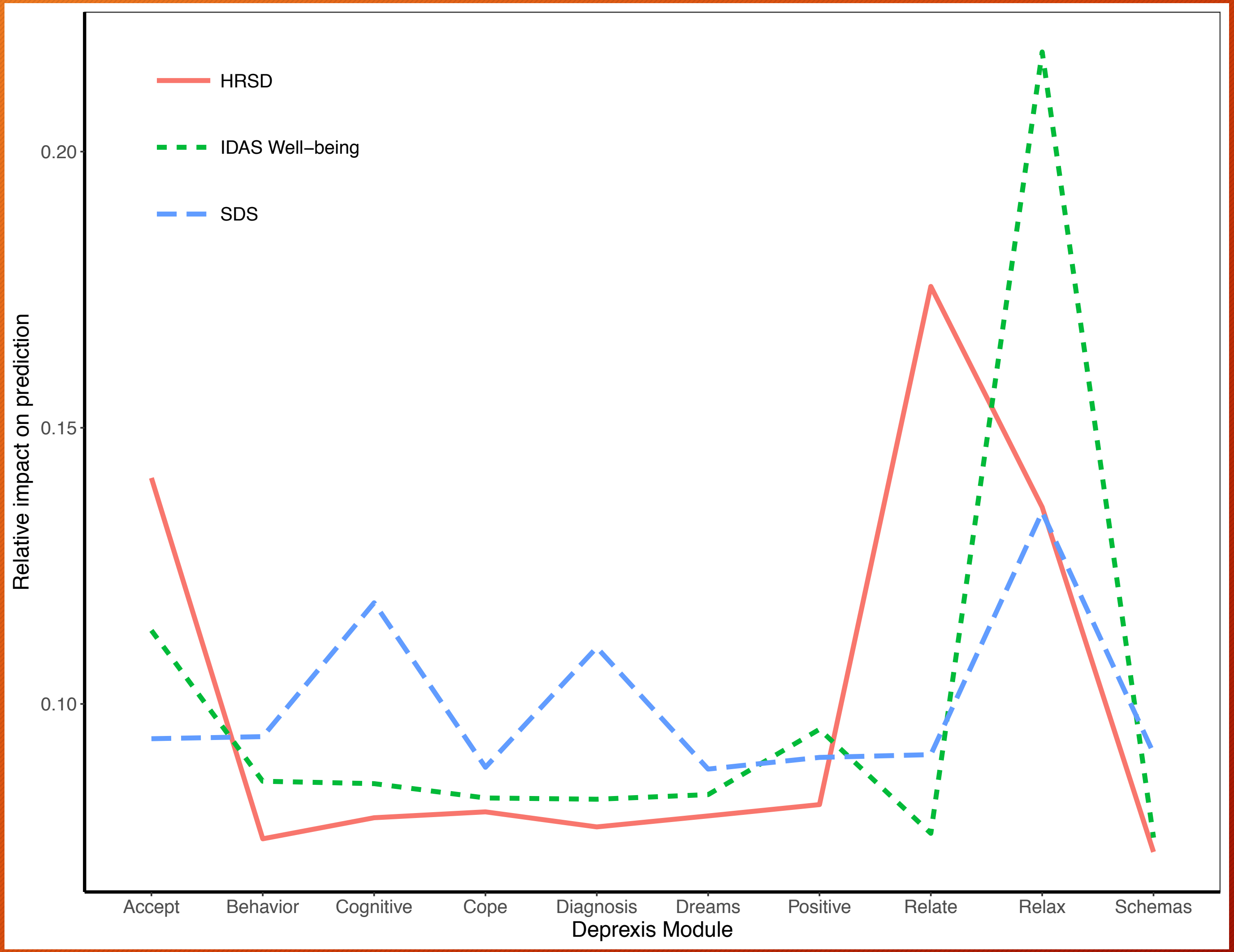


Random Forest

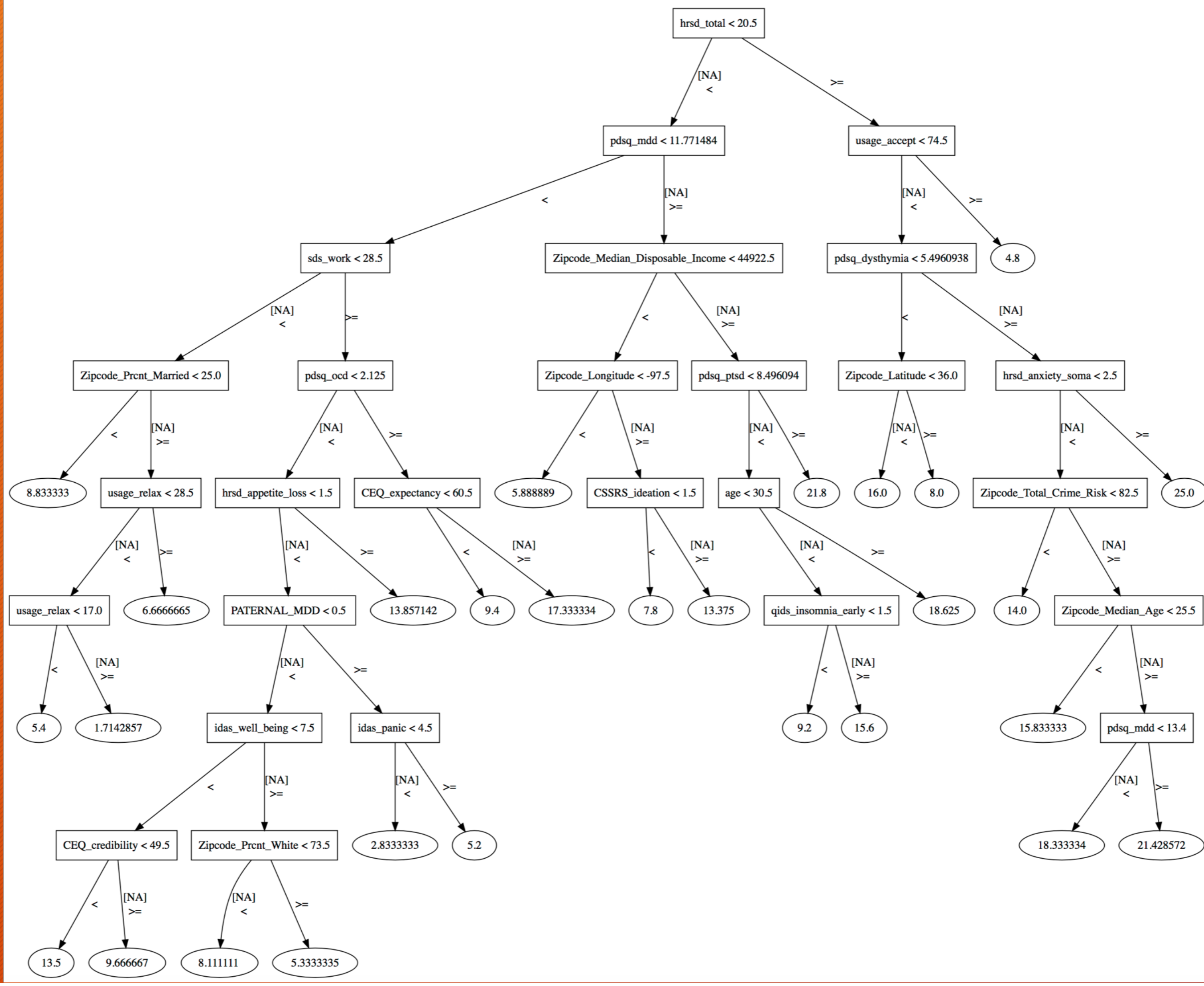


Better outcomes for those living in less affluent ZIP codes



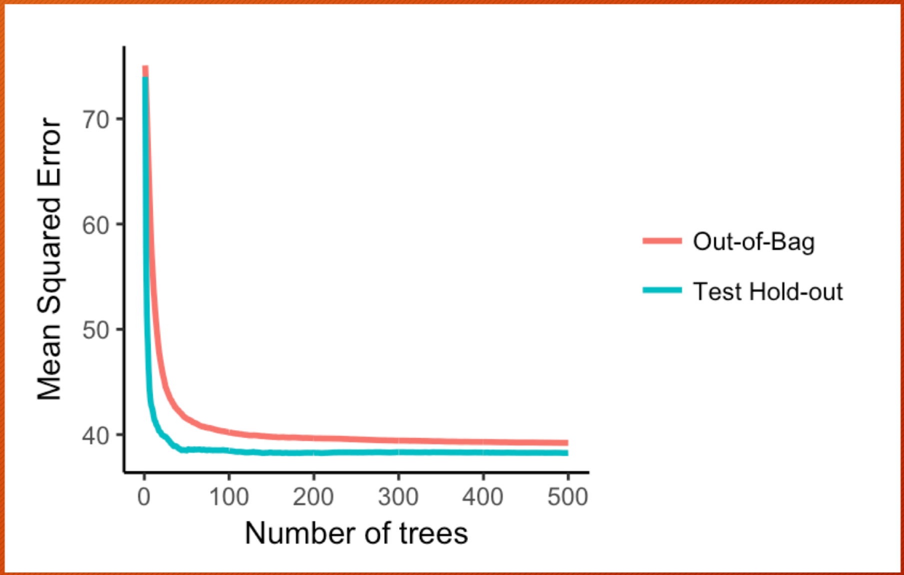

HRSD
IDAS Well–being
SDS
Relative impact on prediction
0.20
0.15
0.10
Accept
Behavior
Cognitive
Cope
Diagnosis
Dreams
Positive
Relate
Relax
Schemas
Deprexis Module

Tree 0

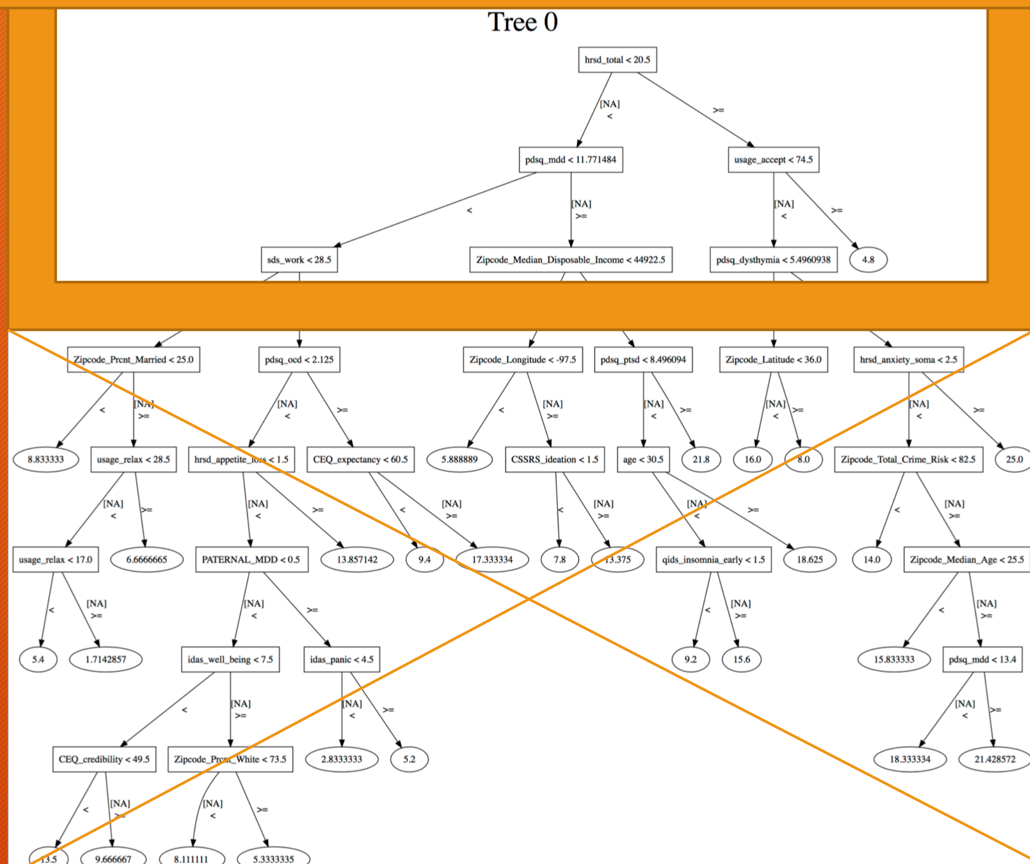


hrsd_total < 20.5
[NA]
<
>=
pdsq_mdd < 11.771484
usage_accept < 74.5
sds_work < 28.5
Zipcode_Median_Disposable_Income < 44922.5
pdsq_dysthymia < 5.4960938
4.8
Zipcode_Prcnt_Married < 25.0
pdsq_ocd < 2.125
Zipcode_Longitude < -97.5
pdsq_ptsd < 8.496094
Zipcode_Latitude < 36.0
hrsd_anxiety_soma < 2.5
8.833333
usage_relax < 28.5
hrsd_appetite_loss < 1.5
CEQ_expectancy < 60.5
5.888889
CSSRS_ideation < 1.5
age < 30.5
21.8
16.0
8.0
Zipcode_Total_Crime_Risk < 82.5
25.0
usage_relax < 17.0
6.6666665
PATERNAL_MDD < 0.5
13.857142
9.4
17.333334
7.8
13.375
qids_insomnia_early < 1.5
18.625
14.0
Zipcode_Median_Age < 25.5
5.4
1.7142857
idas_well_being < 7.5
idas_panic < 4.5
9.2
15.6
15.833333
pdsq_mdd < 13.4
CEQ_credibility < 49.5
Zipcode_Prcnt_White < 73.5
2.8333333
5.2
18.333334
21.428572
13.5
9.666667
8.111111
5.3333335

Averaging dumb trees = smart forest



N = 200


Tree 0
hrsd_total < 20.5
pdsq_mdd < 11.771484
usage_accept < 74.5
sds_work < 28.5
Zipcode_Median_Disposable_Income < 44922.5
pdsq_dysthymia < 5.4960938
4.8
Zipcode_Prcnt_Married < 25.0
pdsq_ocd < 2.125
Zipcode_Longitude < -97.5
pdsq_ptsd < 8.496094
Zipcode_Latitude < 36.0
hrsd_anxiety_soma < 2.5
8.833333
usage_relax < 28.5
CEQ_expectancy < 60.5
5.888889
CSSRS_ideation < 1.5
age < 30.5
21.8
16.0
8.0
Zipcode_Total_Crime_Risk < 82.5
25.0
usage_relax < 17.0
6.6666665
PATERNAL_MDD < 0.5
13.857142
9.4
17.333334
7.8
13.375
qids_insomnia_early < 1.5
18.625
14.0
Zipcode_Median_Age < 25.5
5.4
1.7142857
idas_well_being < 7.5
idas_panic < 4.5
9.2
15.6
15.833333
pdsq_mdd < 13.4
CEQ_credibility < 49.5
2.8333333
5.2
18.333334
21.428572
13.5
9.666667
8.111111
5.3333335

N = 2000

Opportunities and Challenges

- Depression prognosis either has huge amount of irreducible error or depends on complex phenotypes.
- Ensemble machine learning methods—like random forest—have potential to identify these phenotypes, but will require sufficiently large samples that can provide multiple examples of these phenotypes.
- We need scalable data collection and therapeutic interventions to realize full potential of these methods.

Acknowledgements

Mood Disorders Lab: https://labs.la.utexas.edu/beevers/

Funding	
R21MH110758	MACHINE LEARNING AND PERSONALIZED PROGNOSIS FOR DEPRESSION TREATMENT
R01DA032457	GENETIC INFLUENCES ON DUAL PROCESSING MODES OF REWARD AND PUNISHMENT LEARNING
R56MH108650	CONTRIBUTION OF GENOME-WIDE VARIATION TO COGNITIVE VULNERABILITY TO DEPRESSION

NIH

NIMH National Institute of Mental Health

NIDA NATIONAL INSTITUTE ON DRUG ABUSE