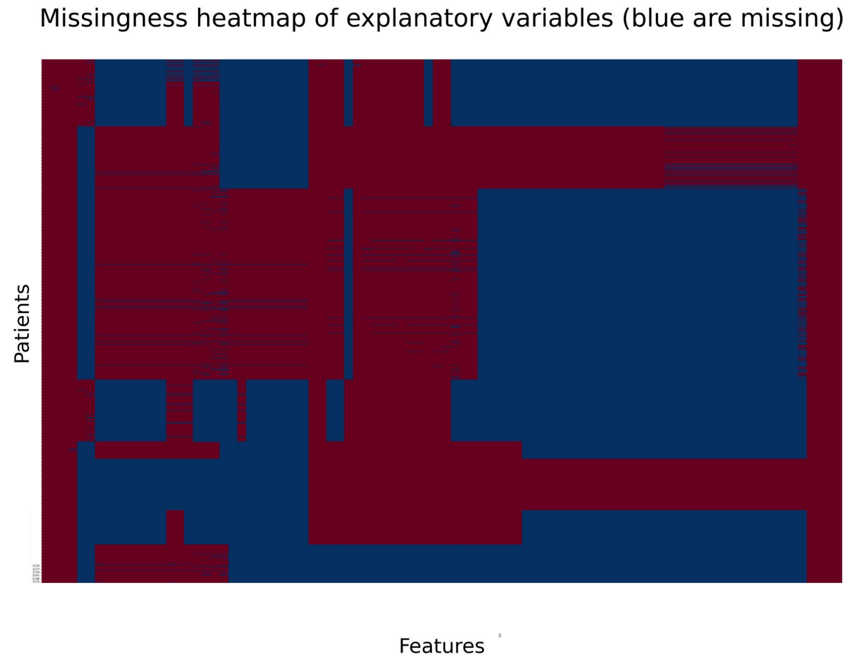
Predicting anxiety treatment outcomes with machine learning

Marija Stanojevic, Lesley A. Norris, Philip C. Kendall, Zoran Obradovic
Temple University, Philadelphia, USA

Presented by Zoran Obradovic at ICMLA Conference, 12-14th December 2022

Introduction

- 26% (2009) -> 44% (2021) of US adolescents with anxiety or depression
- Self-harm increased by 88% from 2001 to 2019
- Survey data of children and parents: before and after treatment
- Data from 9 studies, 108 features, 7 outcomes
- Features: demographic, scale 0-8 (severity) and 0-100 (t-distribution)
- Outcomes severity 0-3 (mild - no diagnosis), 4-8 (disorder)

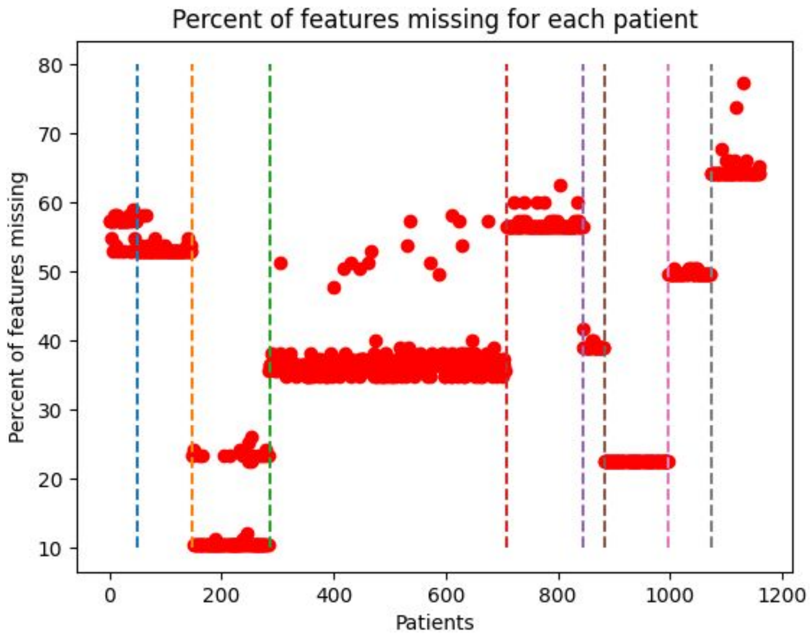


Heatmap of missing features per patient (108 features, 1161 patients)

Heatmap of missing labels per patient (108 features, 1161 patients)

Objectives

- Most of the published research contains up to 30% of missing data
- Goals:
  - Examine which imputation methods are the best
  - Determine best prediction models
  - Can using different datasets jointly lead to better generalization?
  - Can we predict outcomes for patients who didn’t finish?



Study	Samples	Features	Predicted variables	% missing
S1	49	27	6	43%
S2	99	30	6	61%
S3	138	66	7	57%
S4	422	55	6	14%
S5	137	26	4	40%
S6	38	33	6	61%
S7	114	37	7	42%
S8	76	21	6	53%
S9	88	25	0	70%
Joined	1161	108	7	40%

Percent of missing features per patient. Total features=108. Vertical lines separate data from different studies.

Statistical overview of youth anxiety data. Data comes from studies S1-S9 in the order referenced in introduction.

Methods (1)

- **Imputation:**
  - **baselines**: mean, median, KNN, Soft Impute, SVD, EM, II DT, II ET
  - **proposed**: iterative imputation (two-level iteration) + ensambles of advanced regression algorithms (random forest, bag of elastic nets (EN) and decision trees (DT), ada boost (DT, EN) and gradient boosting (DT))
    - inner level iterates over features from the least missing one
    - outer level iterates over dataset until convergence
- **Prediction:**
  - baseline: random forest (used in only ML anxiety study)
  - proposed: advanced regression model and their ensembles

Methods (2)

- **Joining 9 studies:**
  - allowing prediction of outcomes not available from data
- **Domain adaptation**:
  - applying models learned on patients who finished the treatment to withdrawn patients which final medical status is known
- **RMSE evaluation** of imputation and prediction
- Imputation evaluation: **mask 10% of known data randomly**
- Prediction train, cv, test split: **70:15:15**
- Reported prediction evaluation only on known outcomes
  - RMSE is better when evaluating on imputed outcomes

Results (1)

Imputation evaluated on joined dataset (Full RMSE) and average RMSE of imputation on separate studies (Avg study RMSE)

Iterative imputation with bag of ElasticNet models is the best

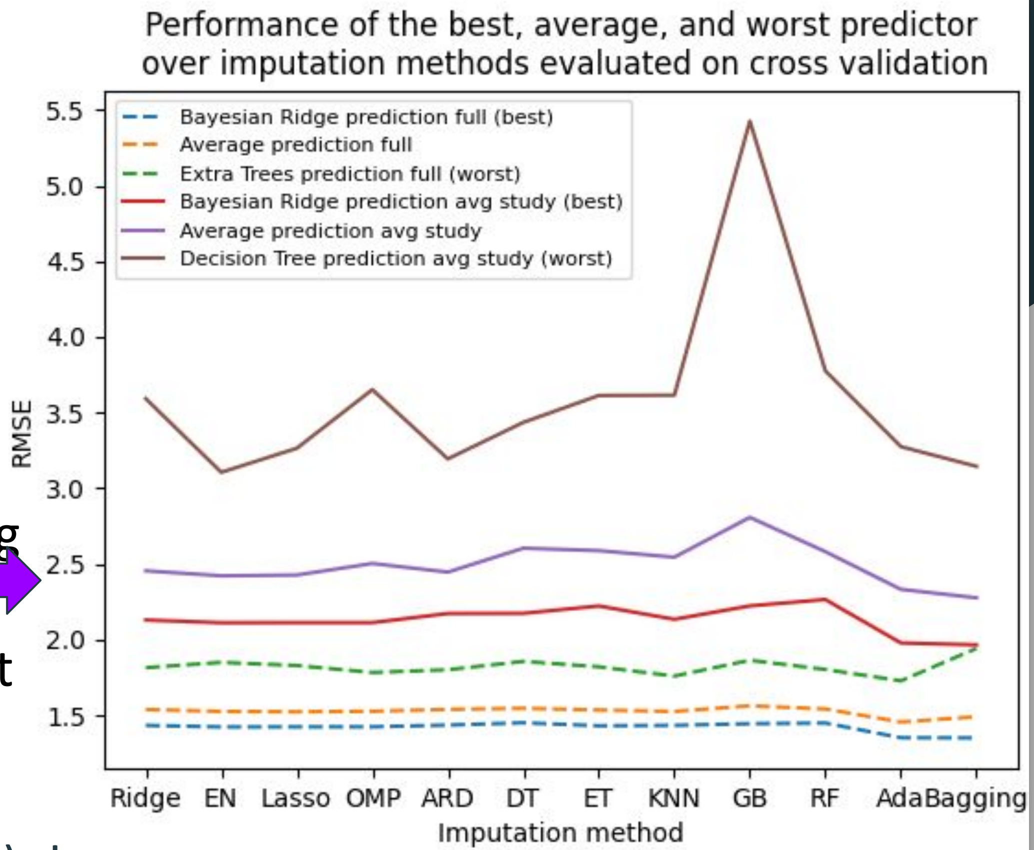
- 4.4% Full RMSE improvement over baseline

Joining datasets is valuable for datasets missing >60% of data

	Regression alg.	Full RMSE	Avg study RMSE
Baseline methods	Mean	5.758 ± 0.188	5.683 ± 3.916
	Median	5.832 ± 0.210	5.965 ± 4.275
	KNN	4.665 ± 0.202	4.837 ± 3.183
	Soft Impute	4.990 ± 0.188	5.298 ± 2.834
	SVD	6.527 ± 0.377	9.104 ± 5.293
	EM	7.960 ± 0.265	7.857 ± 5.626
	II DT	5.633 ± 0.085	6.131 ± 3.489
	II Extra Trees	3.984 ± 0.177	4.300 ± 2.038
Proposed methods	II Linear	8.619 ± 1.947	Not valid
	II BR	4.056 ± 0.115	4.451 ± 2.830
	II Ridge	4.284 ± 0.398	5.882 ± 3.251
	II Elastic Net	3.886 ± 0.145	4.430 ± 2.992
	II Lasso	3.935 ± 0.129	4.410 ± 2.909
	II OMP	3.982 ± 0.229	4.291 ± 2.669
	II ARD	7.225 ± 3.549	8.863 ± 9.864
	II KNN	4.206 ± 0.188	4.716 ± 2.829
	II RF	4.059 ± 0.204	4.213 ± 2.183
	II GB	3.913 ± 0.148	4.457 ± 2.379
	II Ada Boost	3.903 ± 0.180	4.476 ± 3.322
	II Bagging	$\mathbf{3.809 \pm 0.181}$	$\mathbf{4.195 \pm 2.869}$

Results (2)

RMSE of the best, an average of all predictors, and worst over different imputation methods on test data. Prediction full lines describe prediction made on a joined dataset, and prediction avg study lines show average prediction RMSE when a different model is trained on each study.

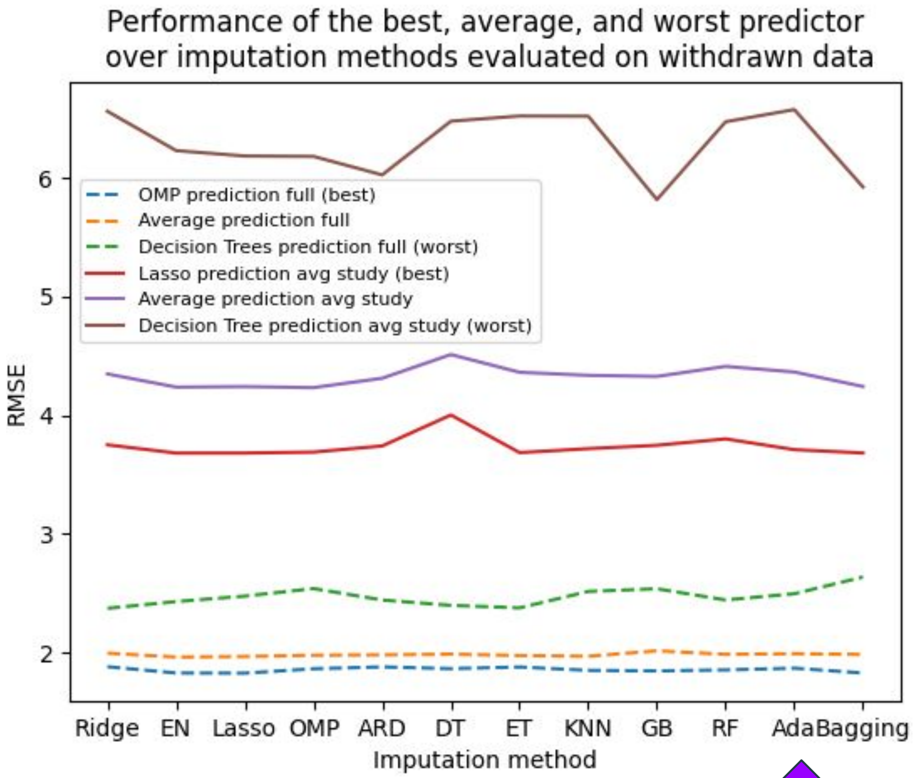


Predicting over all data (9 studies) decreases imputation error by 9.3% and prediction error by 33%

Results (3)

Joining datasets prediction error is 50% smaller on withdrawn patients

- Better generalization of both imputation and prediction
- Joining more important than imputation/prediction methods



RMSE of the best, an average of all predictors, and the worst predictor over different imputation methods on data from withdrawn patients. Lines have the same meaning as in Figure 5.

Conclusions

- The most predictive features:
  - Extracted by lasso regression over the best imputation method
  - Diagnosis before the treatment
  - Treatment type 5 (placebo pill)
  - Parent's OCD (negative)
- Bayesian Ridge and Linear Regressions don't converge due to high percent of missing data
- MICE, MissForest, Bayesian Gauss Multiple Imputation, KI, FCKI imputations don't work as all examples have some missing features

Questions