Predicting anxiety treatment outcomes with machine learning

Marija Stanojevic Lesley A. Norris Philip C. Kendall Zoran Obradovic
Temple University
Philadelphia, PA, USA
{marija.stanojevic, lesley.norris, philip.kendall, zoran.obradovic}@temple.edu

***Abstract*—Youth anxiety disorders are highly prevalent and associated with considerable concurrent functional impairments. According to the State of the World's Children report, 13% of youth between 10 and 19 years old have a diagnosed mental health disorder, 40% of which are anxious and depressive disorders. In a typical longitudinal anxiety clinical study, many explanatory variables are observed in a few patients. As patients drop or miss appointments, collected data has a high missing rate in explanatory and predicted variables. We suggest using machine learning methods to improve understanding of treatments and prediction of outcomes in such studies. We propose machine learning-based imputation for understanding youth anxiety data containing features with high missing rates. In the dataset used, the missing rate of features is up to 80%, making them impossible to use in traditional analysis. Our results show that the proposed iterative imputation with a bag of elastic net regressions imputes missing data better than traditional imputation methods and allow for the best prediction result. We investigate imputation and prediction performance change when using jointly data from multiple studies, where each study has a different bias and missing rate. Leveraging joint dataset allows for predicting the therapy outcome in longitudinal studies with few patients. Additionally, we can now impute or predict features and diagnoses not reported by the clinical study. In conducted experiments, pooling data from nine different studies resulted in 9.3% smaller imputation and 33% lower prediction errors, respectively. Results have higher confidence than when studies are considered separately. We also explored the performance of imputation and prediction in the domain adaptation case of withdrawn patients, in which 50% improvement is obtained when data from all studies are used to impute and train the model.**

***Index Terms*—anxiety, machine learning, imputation, outcome prediction**

I. Introduction

A recent Nature editorial [1] suggests that youth mental health diagnoses have reached alarming levels, with one longitudinal study [2] suggesting that up to 36.7% of youth ages 9-13 (N=1420) will meet criteria for at least one psychiatric disorder diagnosis by the age of 16. Within the range of youth psychiatric disorders, youth anxiety disorders are the most prevalent. For example, The National Comorbidity Survey of US Adolescents [3] (N = 10,123) found that anxiety disorders were the most common disorders observed among adolescents ages 13-18, with an earlier age of onset (6 years) than behavioral (11 years), mood (13 years) and substance use (15 years) disorders. Converging evidence was given in a separate study conducted by the World Health Organization [4], which estimated that 3.6% of children aged 10-14 and 4.6% of adolescents aged 15-19 experience an anxiety disorder. A recent study [5] found that lifetime diagnoses of anxiety or depression among children ages 6-17 increased from 5.4% in 2003 to 8.4% in 2011-12. Beyond high frequency, youth anxiety disorders are associated with considerable short- and long-term functional impairments across multiple domains if left untreated [6], [7].

Efficacious intervention is needed, and machine learning can be a valuable tool to leverage in making evidence-based treatment allocation decisions. However, before using machine learning for implementing personalized treatments, there are problems of data quality that must be addressed. Data from youth anxiety treatment studies have multiple challenges making them hard to analyze by clinical psychologists and off-the-shelf statistical tools are deemed ineffective. To the best of our knowledge, only one youth anxiety treatment study [8] uses a machine learning model to predict treatment outcomes within a small sample (N=124) and there is no analysis on quality of used methods for this application. Therefore, this paper examines how machine learning can be effectively utilized to understand youth anxiety data and help in predicting prognosis and assigning treatments.

We consider different challenges in data that emerge due to the collection of multi-informant, randomized clinical trial data, particularly in any effort to pool data across trials to create harmonized datasets and due to longitudinal nature of treatment. Firstly, single studies have more features than patients, even after domain-based clinical feature selection. For example, the most extensive anxiety treatment study to date [9] observed 488 patients, but there are more than 1000 features collected from each study. Secondly, as large questionnaire batteries are collected from youth and their caregivers over treatment and typically through a follow-up period, the amount of patients who withdraw from treatment or do not fill in all the information is significant. Therefore, patient data can be missing a large fraction of features and predicted variables. Thirdly, the protocol varies slightly across studies based on study objectives. Fourthly, youth diagnostic comorbidity is the norm, leading to multiple predictive variables. Finally, data is influenced by recorded diagnoses, physicians who conducted studies, and geographical regions. For example, ethnicity is not recorded in many parts of the world, but it is commonly

Funded by NIH F31 MH123038-01

documented in the US.

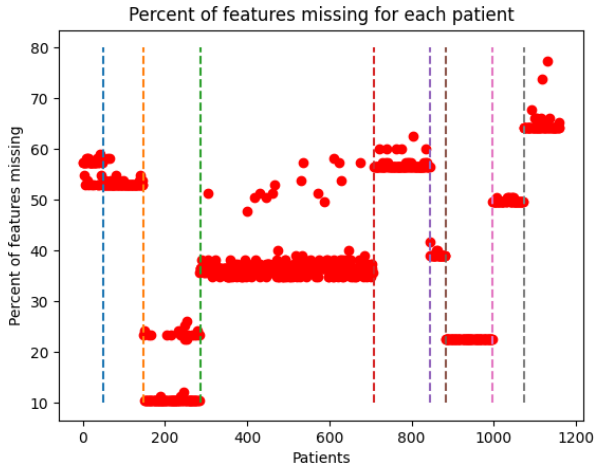

Fig. 1. Percent of missing features per patient. Total features=108. Vertical lines separate patients from different studies.

To address these challenges, we use data from nine youth anxiety studies [10], [11], [12], [9], [13], [14], [15], [16] and [17]. Figure 1 shows missing rate for each patient and it is clearly influenced by study the data came from. Missing rate is between 10% and 80%. There is no patient with fully observed data.

The proposed methods and results of our analysis can be applied to other longitudinal studies with high missing rate, such as clinical trials, observational studies, and in future, for personalized medical diagnostic and treatment using self-observational, device-based, and other longitudinal data which collection depends on patient's promptness.

Our contributions are as follows:

1) Proposed and evaluated more complex machine learning-based imputation methods than those used in clinical psychology intervention studies.
2) Evaluated if data from multiple studies with different biases and missing rates can help better imputation and data understanding.
3) Examined the tool's performance on domain transfer in which target domain are patients who did not finish the treatment.

II. Methodology

Methodology section describes proposed and leveraged machine learning approaches and Experiments section gives details on experimental settings and data. Please, see steps of the implemented system in the Appendix [1].

Imputation methods. Modern imputation methods (MICE, KI, FCKI, missForest) utilize an iterative process to achieve better results. We propose the integration of iterative imputation with machine learning models that are good for the representation of small linear data, such as regression with Ridge, Elastic Net, or Lasso regularization. We believe those techniques will take advantage of correlation among features that DT and RF-based models (missForest) cannot capture. In addition, we expect linear models to learn better representation given that data is mainly numerical. On the other side, the proposed model will not require fully observed samples like MICE, KI, and FCKI do.

As regularized regression models work with fully imputed data, we use initial mean imputation of unknown values. Then, we propose two-level iterative imputation. The outer level iteration step finishes after all features have been imputed within that step. We consider an application where only a few features (number of sessions, age, sex, study, and starting condition) are fully observed. In contrast, some of the relevant features have a missing rate of 80%. Therefore, the inner iteration step always handles the feature with the lowest missing rate among the features not handled in the current outer iteration step.

In addition, as missForest is leveraging multiple decision trees, we propose and evaluate using AdaBoost and Bagging for feature imputation instead of simple regularized regression. Therefore, we combine multiple elastic net (ET) models using those boosting and bagging techniques to predict missing values of the feature with the lowest missing rate that was not already imputed in the current outer iteration step.

Youth anxiety data has missing values in explanatory (X) and predicted (y) variables. Moreover, predicted variables (diagnoses) are partially correlated as youth anxiety often has associated comorbidities. Since the proposed algorithm uses correlations among observed data, it, like all imputers, underestimates the prediction error. To avoid leak between X and y variables and reduce the error underestimation, we propose training separate imputing models on X and y variables.

Prediction methods. Once datasets are imputed, we trained and tested diverse regression models as we hypothesized that random forest used in [8] is not the best model for regression task where most of the variables are numerical.

Joining multiple studies. Given many features, a small number of samples, and a high percentage of missing values in data from each study, we propose combining studies into a single dataset to achieve better prediction confidence. Joined dataset also allows us to predict diagnoses that were not recorded or reported in a single study and to use and evaluate methods for studies with too small a sample size. Since all the studies have different biases and missing rates, we examine when joining data has the most benefits.

Domain adaptation. Youth anxiety studies work with youth over an extended period, leading to some participants dropping out during the study, while a few come to the final evaluation. It would be helpful to predict outcomes of dropout patients as well. However, those data have even higher missing rate and an additional bias. We argue that the proposed imputation approach using joined data from multiple studies can improve transfer learning prediction for the case of withdrawn patients.

[1]https://github.com/marija-stanojevic/anxiety_ml

III. EXPERIMENTS

The proposed and applicable state-of-the-art methods are evaluated on data from youth anxiety studies.

A. Data

Data is integrated from studies done in the last 30 years by different investigator teams in different countries (primarily in the United States) but using a similar randomized controlled trial approach and within samples of youth with primary anxiety disorders. The main difference between studies is the treatments they utilized (i.e., treatment protocol, session number). Variables contain information answered in a survey. Independent Evaluators (IEs) trained to reliability assessed youth and caregiver diagnoses using a semi-structured diagnostic interview. Diagnoses were given a severity level between 0-8, with higher numbers indicating greater severity. Youth and their caregivers also filled out questionnaires to assess basic demographic information. As original data from studies contained too many variables, features for this study are selected on the basis of domain knowledge from the clinical psychology literature on treatment outcome predictor and moderators.[2]

The severity of diagnoses before the treatment is a feature, while the severity of diagnoses after treatment is a predicted variable. Severity scores are considered to be numerical variables in this study. All categorical features (sex, race, ethnicity, study, and starting condition) are fully observed. All imputed features are numerical.

Table I shows the statistics of each study and the joined dataset. It is noticeable that the number of features is often similar to the number of samples and all except the most extensive study have less than 140 patients. The missing rate is above 40%, except in the biggest study, and seven out of nine studies do not have data for at least one of the predicted variables. These statistics do not include patients who did not finish the treatment. Youth who dropped early (n=202) were extracted and used to understand the properties of transfer learning to withdrawn patients. Study 9 has an immense missing rate (70%) and does not include any data for predicted variables.

B. Baseline methods

Commonly used imputation methods are based on: 1) statistics (mean, median, mode, random, MCMC); 2) nearest-neighbors (KNN, FKM, KI, FCKI, MICE), 3) decision trees and random forests (decision tree with a surrogate, missForest) and 4) matrix decomposition techniques (soft imputation, singular value decomposition (SVD)). We compare all of those algorithms if they are applicable to the youth anxiety dataset.

Unfortunately, FKM, KI, FCKI, and MICE cannot be used as they require fully observed samples. We use EM instead of MCMC as literature considers those of the same performance. Our missForest is named II with Extra Trees, and DT with a surrogate is one iteration of II with DT.

[2]The list of features and software implementation can be found here: https://github.com/marija-stanojevic/anxiety_ml.

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%

TABLE I
STATISTICAL OVERVIEW OF YOUTH ANXIETY DATA. DATA COMES FROM STUDIES S1-S9 IN THE ORDER REFERENCED IN INTRODUCTION.

While our method proposes using iterative imputation with regularized regression (ridge, elastic net, or lasso) and boosting (AdaBoost) and bagging algorithms ensembled with those regressions, we also test other regression options in integration with iterative imputation: 1) linear regression; 2) Bayesian ridge (BR) regression; 3) orthogonal matching pursuit (OMP); 4) Bayesian automatic relevance determination regression (ARD); 5) KNN; 6) random forest regression (RF) and 7) gradient boosting regression ensembled with decision trees (GB).

To evaluate the influence of imputation on the prediction task and to understand domain adaptation performance, we use the same regression algorithms we use as part of iterative imputation.

C. Experimental settings

Cleaning. Values 888 (if the measure was not included in the trial) and 999 (other missing values) are replaced with unknown values. Study code and treatment used in the study are replaced with nine binary features for each study and eight for each treatment type. Race variable is substituted with four features representing "White", "Black", "Asian", and "Other" races. Patients who did not finish the treatment are separated into another dataset for the domain adaptation task.

Imputation evaluation. To evaluate the proposed and state-of-the-art imputation methods, 10% of non-missing values in the dataset are randomly masked, and then imputation algorithms are trained and evaluated on that dataset. Root mean square error (RMSE) is calculated on masked data by comparing original and imputed values (Full RMSE). To understand method robustness, we repeated each imputation on ten randomly masked dataset versions and reported mean and standard deviation (STD). When imputation is learned and performed on each study separately, the datasets are joined before RMSE is calculated on all data at once (Avg study RMSE). To understand imputation performance on each study separately, we also calculate imputation RMSE for data from each study separately. Imputation on study level is performed on five randomly masked versions of data from the study, and average and standard deviation of RMSE are reported.

Comparison of the imputated datasets on prediction task. While some off-the-shelf imputing methods can only

apply imputation on the same dataset on which imputation was trained, the proposed imputing methods can be trained on the existing data and used for imputing unseen data. For the prediction task, data of patients who finished treatment is split into training, validation, and test with the ratio of 70:15:15. Then, the features in each of them are normalized. The imputation model is trained on the training part of the dataset, but the validation and test parts are just imputed. An imputed training dataset is used to learn prediction models. Prediction is evaluated using RMSE on non-imputed predicted values of validation and test data to avoid underestimation due to imputation of predicted variables. The RMSE is also calculated on all predicted values in validation and test data and compared to previously calculated RMSE to understand the underestimation caused by imputation. In both cases, RMSE is calculated on validation and test dataset and then averaged to find RMSE confidence interval.

To evaluate prediction performance using each study separately, we would ideally split them in training, validation, and testing and then use the same imputation and prediction process as in the previous paragraph. However, many of the datasets are too small and have a high missing rate. Therefore, many validation and test datasets have more features than samples and cannot be processed. Consequently, studies are imputed only using joined dataset as described in the previous paragraph. Then, imputed data is split into studies, and each of the studies' data is further split into training, validation, and test. Finally, the prediction process described in the previous paragraph is applied.

Domain adaptation task. Features of withdrawn patients are normalized. Then, imputation RMSE mean and standard deviation on withdrawn data is calculated using five randomly masked then imputed versions of the dataset. Next, the imputer is trained and evaluated on the same dataset. Additionally, the imputation model trained on the original dataset is evaluated on five randomly masked then imputed versions of the withdrawn patients' dataset. As in the original imputation step, two independent imputing models are used for features and predicted variables to avoid data leaks. Models trained on the joined dataset and each study separately are used to predict the withdrawn patients' dataset values. Finally, the RMSE of predicted values is calculated on non-imputed predicted values and all predicted variables.

IV. Results

RMSE values for imputation are calculated using the difference between actual and 10% of randomly masked values in that dataset. RMSE values for prediction show how well the seriousness of diagnoses is predicted. The intensity scale is 0-8, where values 0-3 are milder, and 4-8 are severe. Prediction RMSE values are reported only on the difference between predicted and observed data, excluding imputations of predicted variables.

	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	$\mathbf{3.809 \pm 0.181}$	$\mathbf{4.195 \pm 2.869}$

TABLE II
IMPUTATION EVALUATED ON JOINED DATASET (FULL RMSE) AND AVERAGE RMSE OF IMPUTATION ON SEPARATE STUDIES (AVG STUDY RMSE)

A. Imputation results

Imputation is evaluated with RMSE on normalized dataset when 1) joined dataset is imputed (Full RMSE) and 2) each study is imputed separately, imputed studies are combined, and RMSE is calculated (Avg study RMSE). Table II shows results for both cases. The first part of the table gives baseline methods, and the second part contains the proposed methods.

The best performing baseline method is II with extra trees, which is a substitute to missForest, and that results align with the literature. However, six proposed methods have better Full RMSE than II with extra trees, and a bag of elastic nets performs the best among all. II combined with boosting and bagging of regularized regression models gives the two best results. However, all six models outperforming II with extra trees integrate II and regularized regression. Three proposed methods have better average study RMSE, and a bag of elastic nets also has the best result in this scenario.

B. Performance comparison between joined dataset and single studies

Table II shows that using joined dataset increases RMSE performance by 9.3% for the best imputation model (II with a bag of elastic net regressions). However, more importantly, it decreases the variability of imputation performance given the randomly masked data. The best imputation method has almost 16 times smaller STD when a joined dataset is used compared to single studies' imputation.

Iterative imputation with the linear model produces infinite predictions due to a small percentage of known values within a single study.

Table III shows RMSE of imputation per each study. Study 9 has high imputation RMSE, which is a consequence of the high missing rate (70%) and no prediction variables in that dataset. Because of this bias in the joined dataset, only three

Study	Imp. RMSE	Pred. RMSE	Coverage
S1	2.885 ± 0.842	1.029 ± 0.311	25%|1.7%
S2	3.983 ± 0.201	0.9034 ± 0.041	28%|4.1%
S3	3.229 ± 0.164	1.335 ± 0.192	60%|16%
S4	3.018 ± 0.087	1.359 ± 0.062	51%|39%
S5	2.805 ± 0.188	1.529 ± 0.030	24%|5.3%
S6	4.942 ± 1.207	1.439 ± 0.087	30%|2.1%
S7	1.594 ± 0.120	1.618 ± 0.358	35%|8.0%
S8	3.661 ± 0.549	1.814 ± 0.475	19%|2.1%
S9	11.639 ± 1.284	Impossible	23%|3.0%

TABLE III
PERFORMANCE OF IMPUTATION (BAGGING WITH ELASTICNET) AND PREDICTION (BAYESIAN REGRESSION) ON IMPUTED DATA FOR EACH STUDY.

studies' datasets have worse imputation RMSE than the joined dataset.

The coverage column shows the percent of columns that the study observes compared to the joined dataset (left) and the percent of missing values used to evaluate this study's imputation compared to the joined dataset (right). Coverage shows high variability in known values (1.7% - 39%) and covered features (19% - 60%) among the studies.

We can see that studies with high imputation RMSE (S2, S6, S9) also have low confidence in that result which signalizes that correlation among available data is not enough to impute well. That is not surprising given that those three datasets have a missing rate of 61%-70%.

In Figure 1 studies 1 and 2 look similar. However, statistics in the Table I can explain the significantly higher RMSE on S2 as a factor of its higher overall missing rate.

Since S4 is the most significant dataset from a single study, imputation is very confident and with lower RMSE mean and std. Therefore, if 55 features and 6 predicted variables that dataset has were enough, training imputer and predictor using only S4 instead of joined dataset would be a good choice.

When statistics from Tables I and III are combined, it is clear that imputation confidence and RMSE are the most influenced by the missing rate of the dataset and then by its size and structure of missing values. Also, imputation RMSE is not a great predictor of prediction RMSE, but imputation confidence is highly correlated with prediction confidence. While proposed imputation methods perform better than methods commonly used in literature, that advantage does not guarantee better prediction on a particular small study with a high missing percent.

Figure 2 shows that using joined dataset for training stabilizes and improves imputation RMSE. Better prediction results depend more on predictors than imputation methods when imputations are chosen between our proposed predictors as they are all performing well. Despite the predictor choice, AdaBoost and bagging of elastic net regressors are performing slightly better than the other imputation methods. Notice that imputation with random forest and gradient boosting of decision trees gives the worst prediction results.

While the best prediction models give acceptable RMSE, prediction with the worst model trained on each study separately is not meaningful given the scale of the predicted value.

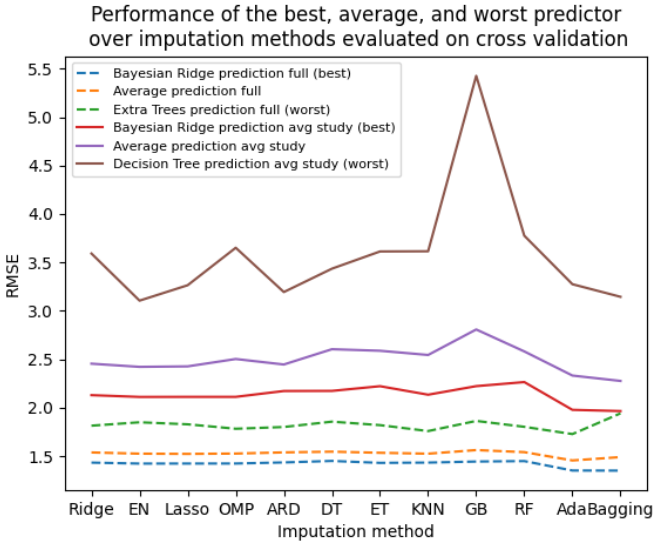

Fig. 2. RMSE of the best, an average of all predictors, and worst over different imputation methods on test data. **Prediction full** lines describe prediction made on a joined dataset, and **prediction avg study** lines show average prediction RMSE when a different model is trained on each study.

The importance of the joined dataset is also evident as even the worst prediction model on the joined dataset is better than the best prediction model trained on single datasets.

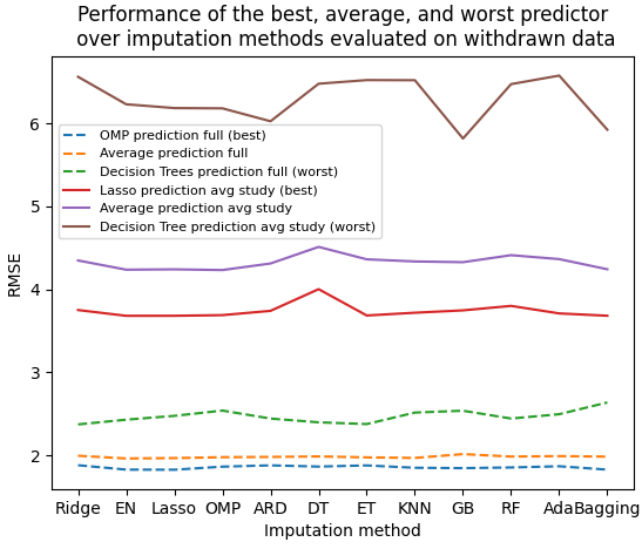

Fig. 3. RMSE of the best, an average of all predictors, and the worst predictor over different imputation methods on data from withdrawn patients. Lines have the same meaning as in Figure 2.

C. Domain adaptation - withdrawn patients

To evaluate the domain adaptation properties of learned predictors, we substitute the test dataset with the dataset of withdrawn patients and record RMSE for each combination of imputation and prediction methods used. Figure 3 shows RMSE of prediction given different imputation methods. Performance patterns on withdrawn patients are similar to the

test dataset, but prediction RMSE is worse for each type of predictor. Although the best predictors are still linear, they are not the same as in Figure 2. As hypothesized, the importance of joined datasets is even more visible as all predictors trained on single datasets give meaningless predictions with RMSE higher than 4 points. Predictors achieve a 50% RMSE decrease when using the joined dataset. While all the imputation models perform similarly on the best predictors, bagging of elastic nets has a small advantage in this case.

V. Conclusion

Youth anxiety is a rising problem for which treatment testing studies are small, and data have significant challenges. We start the process of building and evaluating helpful machine learning tools for this task.

While the dataset's missing rate, size, and structure of missing values influence the ability to predict youth anxiety well, we have shown that choosing the proposed iterative imputation with the bagging of elastic net regressions gives a slight advantage despite the chosen predictor. While chosen predictors are integral, joining datasets as proposed is more significant in successful prediction, especially when the target variable is from a different domain, such as among withdrawn patients.

There is a vast gap in data and method understanding and software availability. Improving ML methods for analyzing this kind of data would help many other psychological studies and other longitudinal medical research studies with small samples, many features, and a high missing rate.

While our paper gives the first directions on handling youth anxiety data with machine learning, additional work is required to achieve better results on this data. Also, more can be done to understand how prediction and imputation relate when the percentage of missing data is this high. Finally, further collaboration between clinical psychologists and machine learning scientists is required to collect more data and create meaningful, better-performing, and widely used tools. Discussion of potential ethical challenges around imputing is a requisite for applying this machine learning system into clinical settings. It is important to evaluate bias that imputation induces on an individual level.

Acknowledgments

We thank D. H. Bodden, S. M. Bogels, M. H. Nauta, E. De Haan, J. Ringrose, C. Appelboom, A. G. Brinkman, K. C. M. M. J. Appelboom-Geerts, E. Flannery-Schroeder, S. M. Panichelli-Mindel, M. Southam-Gerow, A. Henin, M. Warman, J. L. Hudson, E. Gosch, C. Suveg, A. Scholing, P. M. Emmelkamp, R. B. Minderaa, J. T. Walkup, A. M. Albano, J. Piacentini, B. Birmaher, S. N. Compton, J. T. Sherrill, G. S. Ginsburg, M. A. Rynn, J. McCracken, B. Waslick, S. Iyengar, J. S. March, J. J. Wood, M. Southam-Gerow, B. C. Chu, M. Sigman, M. A. Villabø, M. Narayanan, S. N. Compton, S. P. Neumer, K. Aafjes-van Doorn, C. Kamsteeg, J. Bate and M. Aafjes for providing the anonymized data from their studies cited in the paper.

This research was supported in part by the NIH F31 MH123038-01 grant.

References

[1] "Young people's mental health is finally getting the attention it needs," Nature, Oct 2021, vol. 598, no. 7880, pp. 235–236.

[2] E. J. Costello, S. Mustillo, A. Erkanli, G. Keeler, and A. Angold, "Prevalence and development of psychiatric disorders in childhood and adolescence," Archives of General Psychiatry, 2003, vol. 60, no. 8, pp. 837-844, doi:10.1001/archpsyc.60.8.837.

[3] K. R. Merikangas, J. P. He, M. Burstein, S. A. Swanson, S. Avenevoli, L. Cui, J. Swendsen, "Lifetime prevalence of mental disorders in U.S. adolescents: Results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A)," Journal of the American Academy of Child and Adolescent Psychiatry, 2010, vol. 49, pp. 980-989. doi:10.1016/j.jaac.2010.05.017.

[4] "Adolescent mental health," World Health Organization, 2021, https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health, accessed on 8th Sep 2022.

[5] B. H. Bitsko, J. R. Holbrook, R. M. Ghandour, S. J. Blumberg, S. N. Visser, R. Perou, J. Walkup, "Epidemiology and impact of healthcare provider diagnosed anxiety and depression among US children," Journal of Developmental and Behavioral Pediatrics, Jun 2018, vol. 39, no. 5, pp. 395-403, doi: 10.1097/DBP.0000000000000571.

[6] A. J. Swan, and P. C. Kendall, "Fear and missing out: Youth anxiety and functional outcomes," Clinical Psychology: Science and Practice, 2016, vol. 23, no. 4, p. 417.

[7] C. A. Essau, P. M. Lewinsohn, B. Olaya, and J. R. Seeley, "Anxiety disorders in adolescents and psychosocial outcomes at age 30," Journal of affective disorders, 2014, 163, pp. 125-132.

[8] E. R. Lebowitz, S. Zilcha-Mano, M. Orbach, Y. Shimshoni, and W. K. Silverman, "Moderators of response to child-based and parent-based child anxiety treatment: a machine learning-based analysis," Journal of Child Psychology and Psychiatry, 2021.

[9] J. T. Walkup, A. M. Albano, J. Piacentini, B. Birmaher, S. N. Compton, J. T. Sherrill, G. S. Ginsburg, M. A. Rynn, J. McCracken, B. Waslick, S. Iyengar, J. S. March, and P. C. Kendall, "Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety," New England Journal of Medicine, 2008, vol. 359, no. 26, pp. 2753–2766.

[10] P. C. Kendall, "Treating anxiety disorders in children: results of a randomized clinical trial," Journal of consulting and clinical psychology, 1994, vol. 62, no. 1, 100.

[11] P. C. Kendall, E. Flannery-Schroeder, S. M. Panichelli-Mindel, M. Southam-Gerow, A. Henin, and M. Warman, "Therapy for youths with anxiety disorders: A second randomized clincal trial," Journal of consulting and clinical psychology, 1997, vol. 65, no. 3, 366.

[12] P. C. Kendall, J. L. Hudson, E. Gosch, E. Flannery-Schroeder, and C. Suveg, "Cognitive-behavioral therapy for anxiety disordered youth: a randomized clinical trial evaluating child and family modalities," Journal of consulting and clinical psychology, 2008, vol. 76, no. 2, 282.

[13] M. A. Villabø, M. Narayanan, S. N. Compton, P. C. Kendall, and S. P. Neumer, "Cognitive–behavioral therapy for youth anxiety: An effectiveness evaluation in community practice," Journal of consulting and clinical psychology, 2018, vol. 86, no. 9, 751.

[14] J. J. Wood, J. C. Piacentini, M. Southam-Gerow, B. C. Chu, and M. Sigman. "Family cognitive behavioral therapy for child anxiety disorders," Journal of the American Academy of Child and Adolescent Psychiatry, 2006, vol. 45, no. 3, pp. 314–321.

[15] D. H. M. Bodden, S. M. Bögels, M. H. Nauta, E. De Haan, J. Ringrose, C. Appelboom, A. G. Brinkman, and K. C. M. M. J. Appelboom-Geerts, "Child versus family cognitive-behavioral therapy in clinically anxious youth: An efficacy and partial effectiveness study," Journal of the American Academy of Child and Adolescent Psychiatry, 2008, vol. 47, no. 12, pp. 1384–1394.

[16] M. H. Nauta, A. Scholing, P. M. G. Emmelkamp, and R. B. Minderaa, "Cognitive-behavioral therapy for children with anxiety disorders in a clinical setting: No additional effect of a cognitive parent training," Journal of the American Academy of Child and Adolescent Psychiatry, 2003, vol. 42, no. 11, pp. 1270–1278.

[17] W. K. Silverman, W. M. Kurtines, J. Jaccard, and A. A. Pina. "Directionality of change in youth anxiety treatment involving parents: an initial examination," Journal of consulting and clinical psychology, 2009, vol. 77, no. 3, 474.