Comparing Machine Learning to Evidence-Based Decision-Making Techniques for Identification of Mood, Trauma, and Behavior Disorders

Zhuoyu Shi, Kalil Manara, Alberto Stefana, & Eric Youngstrom
Department of Psychology & Neuroscience, University of North Carolina at Chapel Hill

COLLEGE OF ARTS AND SCIENCES
Psychology and Neuroscience

INTRODUCTION

Introducing Computational Models to Clinical Diagnosis

- Diagnosis lays the foundation for other clinical activities
- Current clinical practice is imperfect & may have biased results
- Algorithms perform better
  - 13% increase in accuracy using statistical predictions vs clinical methods
  - Mechanical prediction substantially outperformed clinical prediction in 33%–47% of studies examined
  - Superiority for mechanical-prediction techniques was consistent, regardless of the task, clinicians' amounts of experience, or the types of data being examined

Machine Learning

Algorithms that enable computers to learn from and make predictions or decisions based on data

	ML Models	Traditional Models
Design Logic	Designed to learn patterns from data; self-adapted to changes & different data	Based on established statistical theories; require human help in specificing model structure and selecting variables
Complexity	Capable of modeling non-linear relationships (multi-dimentional)	Simpler and focus on linear relationships and interactions
Interpretability	Less interpretable as a black box	More interpretable with clear math formulation

Variables

- PGBI-Depression & Hypo/Biphasic: The full PGBI has 73 items, with scores ranging from 0 to 3; 46 items focusing on depression & 28 items focusing on hypomanic/biphasic scales.
- PGBI10M: PGBI Mania scale; focusing on the items best discriminating bipolar from nonbipolar diagnoses
- PGBI-Sleep: Sleep disturbance
- PGBI 7Up & 7Down: Seven hypomanic/biphasic and seven depressive items selected for optimal psychometrics in a self-report format

Variable	Nomogram	Logistic Regression	LASSO	SVM	RF
PGBI10M	X	X	X	X	X
Family Bipolar History	X	X	X	X	X
Sex		X	X	X	X
Youth Age		X	X	X	X
Race		X	X	X	X
PGBI-depression			X	X	X
PGBI-hypo/depression			X	X	X
PGBI-sleep			X	X	X
PGBI 7 Up			X	X	X
PGBI 7 Down			X	X	X
Diagnosis Count			X	X	X
Other Diagnoses			X	X	X
2-Way Interaction			X	X	X

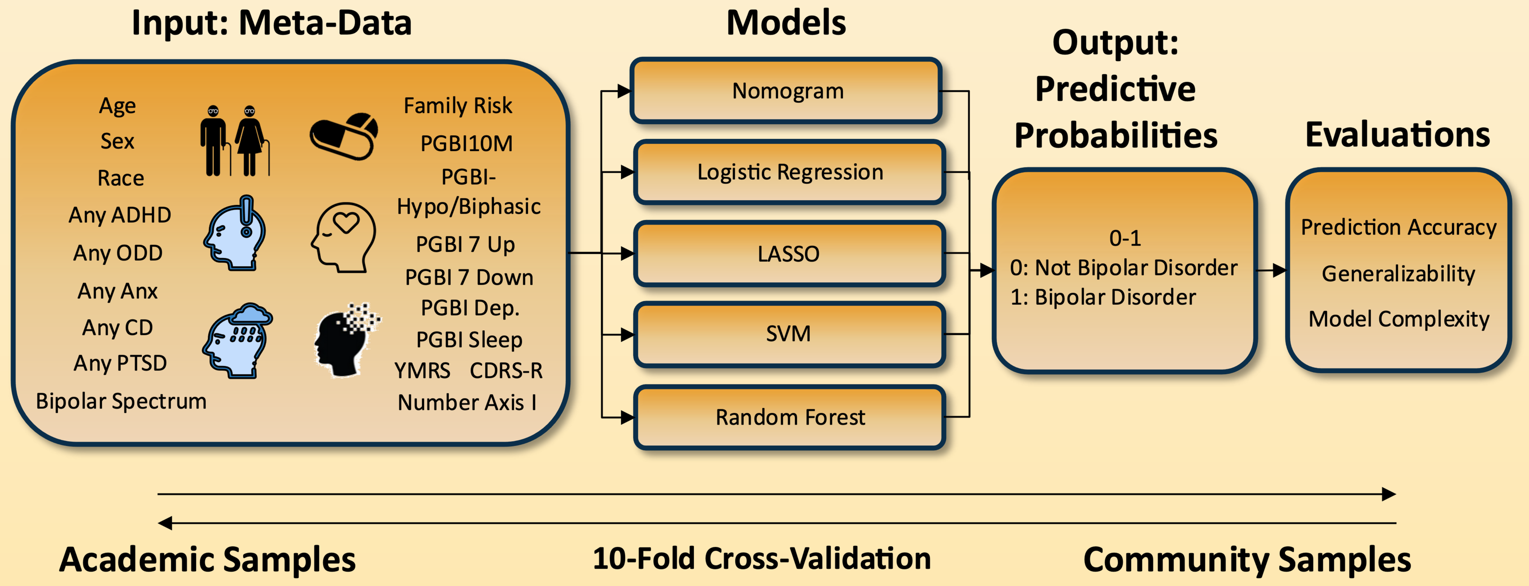
METHODS

Data

- Age Range: Youths between 5-18 years old
- The academic dataset (*N* = 550) was collected at a clinic within a university's psychiatry department
- The community dataset (*N* = 511) was a randomly selected group that sought mental health and behavioral services for their children

	Community Dataset (N = 511)	Academic Dataset (N = 550)	Effect Size
Demographics			
Age, Years (SD)	10.53 (3.41)	11.40 (3.23)	.26
Male, % (n)	60% (205)	60% (217)	.01
White, % (n)	6% (31)	79% (433)	.74
Clinical Scales			
PGBI10M	7.47 (6.35)	10.13 (7.88)	.37
PGBI-hypo/biphasic	19.70 (14.22)	24.66 (16.84)	.32
PGBI-depression	24.48 (21.49)	36.19 (25.67)	.49
7 Up	4.11 (3.83)	5.16 (4.61)	.25
7 Down	3.21 (4.04)	6.24 (5.28)	.64
PGBI-sleep scale	4.06 (4.18)	5.87 (4.74)	.41
Family History of Bipolar	32% (165)	35% (194)	.03
Any Attention-Deficit/Hyperactivity	66% (338)	54% (295)	-.13
Any Oppositional Defiant Disorder	38% (196)	30% (167)	-.08
Any Conduct Disorder	12% (61)	8% (44)	-.07
Any Anxiety Disorder	27% (138)	8% (45)	-.25
Any Posttraumatic Stress Disorder	11% (54)	2% (11)	-.18

Analysis Pipeline



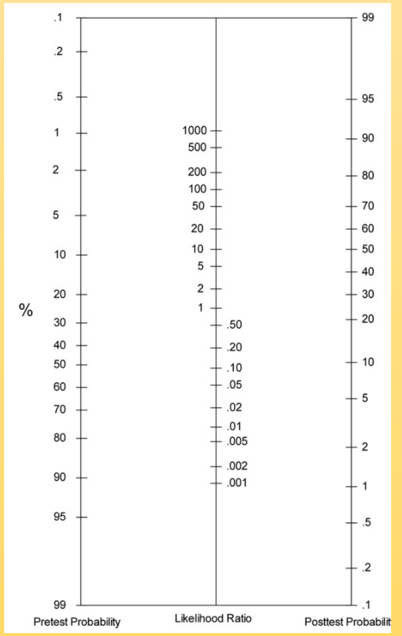
MODELS

Nomogram

- A graphical calculating device
- An efficient way to compute results with pencil & paper
- High interpretability

Logistic Regression

- A statistical method for binary classification
- Binary classification problem
- High interpretability



LASSO

- Regression model
- Classification problem with many features
- Medium interpretability

Support Vector Machine

- Supervised learning model used for classification and regression
- Widely used in different classification problems
- Low interpretability, less intuitive with N-dimensional

Random Forest

- Linear Ensemble Learning method
- Handling a large dataset with higher dimensionality while generating high accuracy predictions with less tuning of parameters
- Low interpretability, as it is an ensemble method

RESULTS

AUC Comparison

* Benchmark models from Youngstrom et al. 2018 study

Model	Academic Dataset (N = 550)	Community Dataset (N = 511)	Shrinkage upon External Cross-Validation
Multipredictor Nomogram*	.882	.775	.107
Logistic Regression (5df)*	.890	.775	.115
LASSO*	.902	.801	.101
Reversed LASSO*	.864	.830	.034
SVM	.926	.737	.189
Reversed SVM	.713	.843	.130
Random Forest	.999	.791	.208
Reversed Random Forest	.824	**.999**	.175
Diagnosis Upper Limit	.925	.925	0

Factors Ranking

```
> print(academic_importance_df_sorted)
                            Feature  Importance
bh10:comorbid         bh10:comorbid   9.1704430
bh10                           bh10   8.5284341
bh10:lbhpgb9           bh10:lbhpgb9   7.9391436
bh10:pgbi7up           bh10:pgbi7up   7.6009621
lbhpgb9:comorbid   lbhpgb9:comorbid   7.2148100
bfamrisk:lbhpgb9   bfamrisk:lbhpgb9   6.2774417
pgbi7up:lbhpgb9     pgbi7up:lbhpgb9   6.2660001
agechild:bh10         agechild:bh10   6.0912290
lbhpgb9                     lbhpgb9   5.4869276
bh10:bfamrisk         bh10:bfamrisk   5.1746328
```

```
> print(community_importance_df_sorted)
                              Feature  Importance
lbhpgb9:comorbid     lbhpgb9:comorbid   2.0208863
bh10:comorbid           bh10:comorbid   1.8256973
pgbsleep:lbhpgb9     pgbsleep:lbhpgb9   1.7698587
agechild:lbhpgb9     agechild:lbhpgb9   1.7362529
pgbi7up:comorbid     pgbi7up:comorbid   1.5570592
ldeppgb9:comorbid   ldeppgb9:comorbid   1.5568081
pgbi7down:lbhpgb9   pgbi7down:lbhpgb9   1.5511001
bh10:lbhpgb9             bh10:lbhpgb9   1.5182587
pgbsleep:ldeppgb9   pgbsleep:ldeppgb9   1.5015342
pgbsleep:comorbid   pgbsleep:comorbid   1.4865265
```


```
> print(academic_internal_importance_df_sorted)
                            Feature  Importance
bh10                           bh10   6.4847853
bh10:whiteyn1         bh10:whiteyn1   6.3715946
bh10:lbhpgb9           bh10:lbhpgb9   5.9517822
bfamrisk:lbhpgb9   bfamrisk:lbhpgb9   5.8307835
bh10:comorbid         bh10:comorbid   5.7483715
bh10:bfamrisk         bh10:bfamrisk   5.7340437
bh10:ldeppgb9         bh10:ldeppgb9   5.5829160
bh10:pgbi7up           bh10:pgbi7up   5.4779405
bfamrisk:anyadhd   bfamrisk:anyadhd   5.2055279
whiteyn1:lbhpgb9   whiteyn1:lbhpgb9   5.1277685
```

```
> print(community_internal_importance_df_sorted)
                              Feature  Importance
bh10:comorbid           bh10:comorbid   5.4619903
pgbsleep:lbhpgb9     pgbsleep:lbhpgb9   5.4387119
agechild:bh10           agechild:bh10   5.3658218
agechild:pgbi7up     agechild:pgbi7up   5.0389239
bh10:pgbi7up             bh10:pgbi7up   4.9526968
pgbi7up                       pgbi7up   4.5682268
bh10:bfamrisk           bh10:bfamrisk   4.5658583
ldeppgb9:comorbid   ldeppgb9:comorbid   4.5339432
pgbi7down:comorbid pgbi7down:comorbid   4.5245807
pgbsleep                     pgbsleep   4.4297456
```

ACKNOWLEDGMENTS

Special thanks to Dr. Eric Youngstrom, Dr. Alberto Stefana, and Kalil Manara for making this idea possible. In addition, thanks to Dr. Aysenil Belger and Dr. Oscar Gonzalez for their guidance as a part of my honors thesis committee.