Novel Machine Learning Approach Outperforms Traditional Approaches in Major Depressive Disorder Clinical Trials: Identifying Subpopulations Based on Treatment Response

Joseph Geraci[1,2,3,4], Bessi Qorri[1], Mike Tsay[1], Christian Cumbaa[1], Paul Leonchyk[1], Larry Alphs[1], and Luca Pani[5,6]

[1]NetraMark Corp, [2]Department of Pathology and Molecular Medicine, Queen's University, [3]Centre for Biotechnology and Genomic Medicine, Medical College of Georgia, Augusta University, [4]Arthur C. Clarke Center for Human Imagination, School of Physical Sciences, University of California San Diego , [5]Miller School of Medicine, University of Miami, [6]University of Modena and Reggio Emilia

INTRODUCTION

CHALLENGE OF HETEROGENEITY IN MAJOR DEPRESSION CLINICAL TRIALS

Clinical trials for central nervous system (CNS) disorders like major depressive disorder (MDD) grapple with heterogeneous patient populations, making it challenging to identify effective therapies with predictive biomarkers to optimize treatment outcomes. Traditional machine learning (ML) methods often fail to capture the combinatorial complexity of variable interactions in such diverse datasets, leading to inadequate predictive models.

AIM

Introduce a novel analytical approach to deconstruct the patient population into explainable and unexplainable subpopulations.

METHODOLOGICAL ISSUE BEING ADDRESSED

Deconstructing heterogeneous patient population data to enhance predictive modeling and patient stratification to advance personalized medicine in CNS disorders.

METHODS

DATASET

CAN-BIND trial exploratory escitalopram drug arm (n=172) with MDD patients.

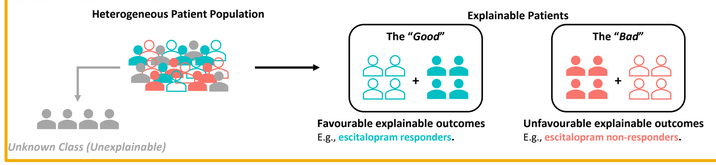
Primary Outcome: ≥50% reduction in MADRS scores from baseline over 8 weeks.

Data Types: Over 362 variables per patient including clinical scales: CGI, SEXFX, DARS, SHAPS, MINI, MADRS, BRIAN, YMRS, QIDS, QLESQ, PSQI, and SPAQ and over 20,000 genetic methylation variables.

MACHINE LEARNING APPROACH

Traditional ML Models: Several ML methods were chosen to augment with this novel mathematically augmented ML, but we report on the best performing methods which included Logistic regression with lasso feature selection, Random Forest, XGBoost, support vector machine (SVM), and neural networks. Binary classification (response vs non-response) was applied to (80%) training and testing (20%) + cross validation.

Novel ML Model: Uses **Sub-Insight Learning**, powered by dynamical systems and novel attention mechanisms to deconstruct patient populations into explainable and unexplainable subpopulations, significantly reducing computational complexity and identifying causal clusters of variables. This approach introduces a semi-supervised learning framework by adding an unknown class to account for uncertain outcomes, transforming the problem into a multiclass classification task.



Key variables identified by the algorithm guide the refinement of inclusion/exclusion criteria and reduced feature sets for the models.

RESULTS

Traditional ML Model Performance Alone vs with Novel ML Using Clinical Scale Data

Traditional ML models yielded poor to moderate performance, with accuracy scores ranging from 55.8% to 65.71% and mean area under the curve (AUC) values between 0.49 and 0.68. These models exhibited limitations in handling the heterogeneity of the dataset, leading to lower specificity and sensitivity.

Traditional ML Performance Before Using Novel ML (Scale Data Only)

- Feature Selection Methods Used:
  - LASSO
  - Filter and Wrapper Methods
  - ANOVA F-Statistic: 10, 15, and 20 features
  - SHAP

Best Traditional ML Model: XGBoost (ANOVA F-Statistic with 20 features)

Metric	Value
Accuracy (%)	65.71
Accuracy 5-Fold CV (% ± SD)	61.43 ± 6.09
Accuracy 10-Fold CV (% ± SD)	56.76 ± 10.36
Accuracy LOO CV (% ± SD)	54.39 ± 49.81

After Applying Sub-Insight Learning:

- Accuracy improved by ~28%, with some models achieving up to 100%.
- Sensitivity increased by ~31%, enhancing the detection of true positives.
- Specificity improved by ~51%, reducing false-positive rates.
- F1-score increased by ~30%, indicating balance between precision and recall.
- Medan AUC rose by ~39%, with values approaching 0.99.

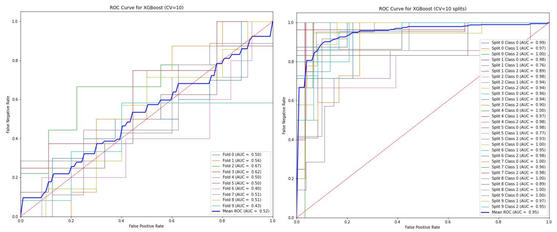
Model	Accuracy Before Novel ML (%)	Accuracy After Novel ML (%)	Improvement (%)
Logistic Regression	54.29	77.14	+22.85
XGBoost	65.71	91.43	+25.72
Random Forest	62.86	82.86	+20.00
SVM	60.00	100.00	+40.00
Neural Network	60.00	77.14	+17.14

Sub-Insight Learning enhanced the performance of all evaluated ML models, and was most pronounced in the SVM model. Adding the "unknown class" allowed models to better manage data uncertainty, and the reduced feature sets minimized overfitting. However, up to 70% of the population was poorly characterized in some instances, with the remaining 30% characterizable and robustly predicted. This is the trade-off for small populations, as found in clinical trials, that can be made to extract insights that can be used to influence subsequent trials.

Traditional ML Model Performance Alone vs with Novel ML Using Methylation Data

Comparative analysis of the XGBoost classifier's performance on methylation data before and after integrating the Sub-Insight Learning ML approach into the modeling process with key metrics: accuracy, F1 score, sensitivity, specificity, and cross-validation scores.

ROC Curves Before Sub-Insight Learning **ROC Curves After Sub-Insight Learning**



	Cross-Validation	Before Novel ML (%)	After Novel ML (%)	Improvement (%)
Accuracy Improvement	1-fold CV	62.86	100.00	+37.14
	5-fold CV	65.66	97.95	+32.29
	10-fold CV	66.32	98.00	+31.68
Sensitivity Improvement	1-fold CV	87.50	100.00	+12.50
	5-fold CV	87.50	98.63	+11.13
	10-fold CV	87.50	98.67	+11.17
Specificity Improvement	1-fold CV	9.09	100.00	+90.91
	5-fold CV	9.09	91.55	+82.46
	10-fold CV	9.09	82.02	+72.93

After integrating Sub-Insight Learning, the XGBoost classifier's accuracy increased across all cross-validation methods, sensitivity improved by ~11-13%, and specificity improved by 73-91%, suggesting a more balanced model effectively reducing false positives and false negatives. While 50% of patients were not classifiable, the insights from the resulting explainable subpopulations was able to significantly alter the p-value for the corresponding trial simulations.

CONCLUSIONS AND SIGNIFICANCE

Obtaining patients for clinical trials is expensive, frequently resulting in small datasets. The focus on AI in recent years has made it clear that massive data sets with many samples are required, but there remains a need to discover how to use these advancements to improve clinical trials. A novel mathematical innovation was discovered to improve the ability for ML to learn from smaller data sets by discovering high effect size subpopulations and the precise driving variables to train reliable models. This sub-insight learning approach demonstrates this capability and shows utility in analyzing heterogeneous datasets in CNS clinical trials. By decomposing patient populations into explainable subgroups and focusing on key predictive variables, this novel approach enhances model performance, underscoring its potential to revolutionize data analysis in MDD clinical trials, and can lead to more accurate treatment response, improved patient stratification, and optimized inclusion/exclusion criteria, advancing personalized medicine in CNS disorders.

DISCLOSURES & ACKNOWLEDGEMENTS

J.G is the founder of NetraMark and is a significant shareholder of NetraMark Holdings, which is a publicly traded company. B.Q., L.P., M.T, C.C.,P.L., and L.A. are employed by NetraMark. L.P.'s disclosures: AbbVie, USA; Acadia, USA; Alexion, Italy; BCG, Switzerland; Boehringer Ingelheim International GmbH, Germany; Compass Pathways, UK; EDRA-LSWR Publishing Company, Italy; Ferrer, Spain; Gedeon-Richter, Hungary; GLG-Institute, USA; Immunogen, USA; Inpeco SA, Switzerland; Ipsen-Abireo, France; Johnson & Johnson USA; NeuroCog Trials, USA; Novartis-Gene Therapies, Switzerland; Sanofi-Aventis-Genzyme, France and USA; NetraMark, Canada*; Otsuka, USA; Pfizer Global, USA; PharmaMar, Spain; Relmada Therapeutics, USA*; Takeda, USA; Vifor, Switzerland; WCG-VeraSci/Clinical Endpoint Solutions, USA (*options/shares).

We would like to acknowledge the individuals and organizations that have made data and the analysis used for this research available, including the Can-BIND Program, the Ontario Brain Institute, the Brain-CODE platform, and the Government of Ontario.

REFERENCES

Tsay, Mi., Geraci, J. & Agrawal, A. Next-Gen AI for Disease Definition, Patient Stratification, and Placebo Effect. doi:10.31219/OSF.IO/PC7AK.

NETRAMARK