

Appraisal of Gene Expression-Based Classifiers for Neuropsychiatric Disorders: A Meta-Regression

Authors

Ali Razavi^{1#}, Brittany Arensman^{2#}, Eric J. Barnett^{1,3}, Leo A. Lee³, Stephen V. Faraone^{1,3}, Stephen J. Glatt^{1,3,4}, and Jonathan L. Hess³

Affiliations

¹Department of Neuroscience and Physiology, Norton College of Medicine at SUNY Upstate Medical University, Syracuse, NY USA

²Emergency Medicine, Norton College of Medicine at SUNY Upstate Medical University, Syracuse, NY USA

³Department of Psychiatry and Behavioral Sciences, Norton College of Medicine at SUNY Upstate Medical University, Syracuse, NY USA

⁴Department of Public Health and Preventive Medicine, Norton College of Medicine at SUNY Upstate Medical University, Syracuse, New York, USA

#Authors made equal first-author contributions

Please address correspondence to:

Jonathan L. Hess, PhD
3732C Neuroscience Research Building
Institute for Human Performance
505 Irving Avenue
Syracuse, NY 13210
E-mail: hessjo@upstate.edu
Phone: +1 (315) 464-3269

Running title: Meta-regression of neuropsychiatric classifiers

ABSTRACT

A substantial body of research examines the potential of gene-expression-based biomarkers for diagnosing and selecting treatments for neuropsychiatric disorders, yet no clear consensus has been reached regarding the influence of controllable factors such as study design and model selection on the performance of gene-expression-based classifiers. To investigate study characteristics and methodologies that influence the accuracy of studies using transcriptomics to classify neuropsychiatric disorders, we conducted a literature review and

meta-regression of relevant studies. We extracted several characteristics from each study, including the number of samples in a training dataset, approach for model validation, and classification model. Using univariate and multi-variate mixed-effect meta-regression analyses, we estimated the association between these study characteristics and reported classification accuracies. Machine Learning (ML) models accounted for 55% of all models, Deep Learning (DL) models accounted for 20% and variations of Logistic Regression models making up the remaining 25%. Support vector machine(SVM) was the most common model type (17%).The use of withheld test samples (56%) was the most frequent approach for validating performance of classification models. We found significant associations between reported accuracies and study-rated bias risk, model type, class ratio, and validation approach. Overall, this review provides helpful insights into study characteristics that significantly influence classification accuracies and emphasizes the importance of prudent methodologies for training and evaluating classification models to mitigate biased accuracy estimates.

INTRODUCTION

Over the past two decades, researchers have explored biomarkers associated with neuropsychiatric disorders, aiming to determine their capacity for accurate and objective identification of these disorders. These biomarkers have been evaluated from many modalities including genetics, transcriptomics, proteomics, epigenetics, metabolomics, and brain structural and functional activity. Due to the complexity and overlapping of features in neuropsychiatric disorders, developing algorithms for diagnosing patients or predicting treatment outcomes has been a significant challenge (Cornblath et al., 2019). Machine learning (ML) and Deep Learning (DL) methods have emerged as powerful tools for gaining insights into complex biology and disease, demonstrating considerable potential in uncovering useful biomarker patterns.

Significant strides have been achieved in transcriptomic technology, expanding the number and diversity of RNAs that can be profiled in cells and tissues. This progress, coupled with substantial advances in genomics, has paved the way for in-depth explorations of transcript variants and splicing isoforms. These developments have prompted further investigations into the harnessing of the potential of RNA expression states as valuable diagnostic or therapeutic biomarkers. Many studies that investigated the diagnostic utility of gene expression-based biomarkers for neuropsychiatric disorders have used classification methods that range from conventional regression-based models, like logistic regression, to ML-based methods such as support vector machine or neural networks. These studies have used a wide array of study designs and procedures to evaluate model performance. This procedural variability among studies poses an opportunity to investigate how study characteristics influence model performance. Earlier reviews of classification studies have found that the size of the training sample and procedures used for assessing model performance are among the most significant predictors of classification accuracy for genomics (Barnett et al., 2024) and brain structural imaging (Zhang-James et al., 2023; Vabalas et al., 2019). To date, neuropsychiatric studies that

evaluated classification performance for ML-based methods have commonly relied upon smaller sample sizes, falling below the recommended sample size (Vabalas et al., 2019) criteria for modeling complex tasks. This is particularly pertinent for genomics datasets, where the number of features often surpasses the number of samples by several orders of magnitude.

Ascertaining tens-to-hundreds of thousands of samples for an individual classification study is generally neither practical nor feasible. Therefore, when investigators apply ML to smaller-sized datasets, it is crucial to adhere to standard procedures to ensure to prudently use the methods and evaluate them, thus mitigating the risk of bias (Vabalas et al., 2019). For example, it has been shown that k-fold cross-validation tends to strongly bias performance estimates, but nested cross-validation and designs that use dedicated training and testing split validation methods yield more accurate and less overfitted results (Vabalas et al., 2019; Quinn & Hess et al., 2024; Zhang-James et al., 2023).

Several reviews highlight the potential for using blood-based gene expression-based biomarkers to individualize treatments for neuropsychiatric disorders, including depression (Mariani et al., 2021), schizophrenia (Mamdani et al., 2013), and Alzheimer's Disease (Teunissen et al., 2022). The potential of blood-based biomarkers to reveal molecular pathways contributing to the etiology of neuropsychiatric disorders has also been reviewed, accompanied by recommendations for applying their clinical application (van de Leemput, Glatt, & Tsuang, 2016). Several studies have examined the utility of blood-based gene-expression and other 'omics profiles to serve as useful proxies for the brain. The overarching consensus from these studies is that the expression levels of genes highly expressed in both the brain and blood exhibit a moderate correlation across these tissues (Tylee et al., 2013). These findings have been expanded upon by other studies, revealing that several gene co-expression networks in the human brain are significantly and strongly preserved in the blood (Cai et al., 2010; Hess et al., 2016). This work highlights the considerable potential of blood-based gene-expression to

serve as a useful proxy of the human brain. Recent advancements in transcriptome imputation methodologies, such as *BrainGENIE*, *B-GEX*, and *TEEBot*, have further enhanced this potential by enabling gene expression profiles in the brain to be predicted based on gene expression profiles directly measured in the blood (Hess et al., 2023; Basu & Wang et al., 2021; Xu et al., 2020). The present study is a literature review and meta-regression of the gene expression-based classification studies centered on neuropsychiatric disorders aiming to investigate the study characteristics and methodologies that significantly influence the accuracy of the classification models using transcriptomics to predict neuropsychiatric disorders.

METHODS

Literature Search

We conducted a literature search in PubMed to identify original research studies that evaluated diagnostic- or treatment-response based classifiers based on traditional statistical models (such as logistic regression) or supervised ML algorithms (such as *k*-nearest neighbors, linear discriminant analysis, support vector machine, random forest, neural networks, XGBoost, and Naïve Bayes, to name a few) in the context of neuropsychiatric disorders. The scope of our review focused on studies that used gene expression exclusively or in combination with other data modalities as input into classification models. We searched PubMed and Medline using the following keywords: (Alzheimer's disease OR mild cognitive impairment OR attention-deficit/hyperactivity disorder OR schizophrenia OR depression OR bipolar disorder OR autism OR post-traumatic stress disorder OR suicide*) AND (gene expression OR transcript*) AND (classification OR machine learning OR support vector machine OR random forest OR neural network OR decision tree OR naive bayes OR XGBoost OR linear discriminant analysis OR nearest neighbor OR deep learning OR elastic net OR lasso OR gradient boosting OR autoencoder OR long-short term memory OR gated recurrent units) AND (AUC or accuracy OR

area under the curve). Our literature search yielded 260 studies and their full texts were reviewed. Studies that did not report accuracy or area under the receiver operating characteristic curve (AUC) measures were excluded. Studies that reported diagnostic accuracy of gene expression-based biomarkers solely *via* a ROC analysis without any implementation of ML or regression-based classifiers were excluded. A total of 198 studies were retained for our review.

Data Extraction

The full-text articles were reviewed and the following study features were recorded for our analysis: the number of samples in training dataset per class, number of samples in the validation set per class (if available), approach for model validation (none, *k*-fold cross-validation (CV), leave-one-out cross-validation (LOOCV), divided training/test sets), choice of classification model, and performance metrics of classification model (*i.e.*, AUC, sensitivity, specificity, or overall accuracy). If studies reported multiple accuracies for different parametrizations of the same classification model (*e.g.*, multiple SVMs trained using the top 5, 10, ... *k* differentially expressed genes), then the best-performing model was recorded for our analysis. As some studies did not report how many subjects were randomly assigned to training and validation sets, we approximated these numbers from the reported percentage of subjects included in each set (*e.g.*, 80% in training set and 20% in test set).

We evaluated studies for potential bias based on the following conditions:

- (1) Feature selection was performed exclusively on the samples designated for model training.
- (2) The training sample comprised a balanced set of classes or corrections were made to address class imbalance.
- (3) Performance of the classification model was evaluated on a set of data treated separate from the training data

Studies for which all three conditions were satisfied were assigned a “low likelihood” risk for potential bias. Studies for which one or more conditions were not satisfied or if unclear were assigned “probable/high likelihood” risk for potential bias.

A total of 92 studies were retrieved and 532 models were extracted. Ten of the models were excluded due to not reporting classification performance in terms of AUC or overall accuracy (n=2) or failing to report sample size of the training set (n=9), leaving 521 total models from 91 studies that were used in the analyses.

Statistical Analysis

We performed univariate linear mixed-effect regression analyses to estimate the association between each study characteristic and reported classification performance. In our univariate regression analyses, we specified each study characteristic as an independent variable and reported accuracy as the dependent variable. We corrected p -values obtained by our univariate models using the Benjamini-Hochberg false discovery rate (FDR) procedure. To gain additional insight into the joint effect of study characteristics on reported accuracy, we conducted a meta-regression in which we modeled all study characteristics as predictors of reported classification performance *via* a linear mixed-effect regression model. All mixed-effect regression models included a random-effect term for study to account for the non-independence among study observations. Standard errors were approximated using the Satterthwaite approach. All statistical analyses were performed in *R* (v.4.2.1) and mixed-effect regression models were fitted using the *lmerTest* package.

Secondary Analysis of Reported Accuracies in Relation to Tissue Source of Biomarkers

It is plausible that biomarkers obtained from *postmortem* brain tissue analyses exhibit greater ontological relevance to psychiatric disorders than peripherally accessible RNAs, potentially influencing classification performance. Additionally, we suspected that tissue concordance (*i.e.*, using either exclusively peripheral blood or *postmortem* brain for both model

training and validation) or discordance (*i.e.*, transition models between brain and blood samples for model training and validation) between datasets might impact classification accuracy. We conducted a secondary analysis on a subset of our entire study pool, comprising 51 unique studies and 292 classification models. These studies evaluated the performance of classification models on completely withheld samples, independent from those used for model training. Among these models, 233 were trained on gene expression profiles from *ex vivo* peripheral blood samples and tested on independent *ex vivo* peripheral blood data. Additionally, two studies trained models on gene expression profiles observed in *ex vivo* blood samples and validated them using data obtained from *postmortem* brain tissue. Nine studies trained models using gene expression profiles observed in *postmortem* brain tissue and validated them using data from *ex vivo* peripheral blood. Finally, 37 models were trained on gene expression profiles observed in *postmortem* brain and validated in independent *postmortem* brain samples. We compared reported classification accuracies between these four groups of studies using a mixed-effect type III ANOVA using a unique study identifier as a random-effect term to adjust for repeated measurements. A significance threshold of $p < 0.05$ was used for this *post-hoc* analysis.

RESULTS

Overview of Study Characteristics

Our literature search uncovered 91 eligible studies, as detailed in the PRISMA diagram provided in **Supplementary Figure 1**. A complete collection of study data is provided in **Supplementary File 1**. These studies reported classification accuracies for 521 models for 13 neuropsychiatric disorders, including: autism spectrum disorder, Alzheimer's disease, mild cognitive impairment, schizophrenia, bipolar disorder, psychosis, obsessive-compulsive disorder, attention-deficit/hyperactivity disorder, methamphetamine-associated psychosis, suicide, major depressive disorder, post-traumatic stress disorder, and Parkinson's disease.

Figure 1a summarizes the number and type of classification models reported per disorder. ML-based algorithms were used in 289 (55%) of reported classification models, DL-based models in 105 (20%), logistic regression models accounted for 85 of the models (16%) and regularized logistic regression models accounted for 42 of the models (9%). Support vector machine (SVM) was the most frequently used classification model among the studies in our review, which accounted for 17% or 86 of the 521 models. A low potential risk of bias was found for 74 out of the 91 studies; 17 studies showed probable/high potential risk of bias. Classification models for Alzheimer's disease and mild cognitive impairment accounted for 332 (63%) of all reported classifiers. The most frequent validation approach for evaluating classification model performance was *via* withheld test sample (292 models), followed by k-fold cross-validation (189 models), and LOOCV (21 models). In 19 models, classification performance was reported based solely on a training sample without the use of cross-validation or an independent test sample. An overview of the number of studies that contributed to each disorder is provided in **Figure 1b**. A complete collection of study data is provided in **Supplementary File 1**.

There was a significant difference in reported classification performance across diagnostic groups when accounting for bias level and sample size as illustrated in **Figure 2** ($p=0.029$). The mean accuracy for all models was 0.72 (SD= 0.16).

Univariate Mixed-Effect Meta-Regression of Study Characteristics and Reported Accuracy

The results obtained from univariate mixed-effect meta-regression models are summarized in **Figures 3 and 4**. We found a significant association between study-rated bias risk and classification performance ($p=0.004$). Studies rated as having high/unclear bias exhibited higher reported accuracies than studies with low observed bias. Because high risk of bias can inflate reported classification accuracy, we included study-rated bias risk as a covariate

in our downstream multivariate meta-regression when assessing the influence of other study characteristics on reported accuracy.

Our univariate mixed-effect regression analyses showed a small yet significant association between the size of the training sample and classification accuracy (**Figure 4a**). Notably, these results revealed an inverse relationship: as sample size increased, reported accuracies decreased by a rate of 0.056, 0.038, per unit increase in \log_{10} -transformed sample size for studies using k-fold CV and LOOCV respectively.

We found a significant inverse association between classification performance and the ratio of cases to controls within training samples. This association was found among studies that evaluated classification performance within independent test samples ($p=0.005$) and studies where performance was solely evaluated in the training sample ($p=0.005$) (**Figure 4b**). Type of accuracy was not significantly associated with reported accuracy and had no impact on the association between ratio of cases to controls within training samples and accuracy (**Figure 4b**).

As shown in **Supplementary Figure 2**, no significant association was found between year of study and study accuracy ($p=0.09$). However, year of study was significantly correlated with total size of training sample ($p=4.9 \times 10^{-10}$). All linear models were analyzed for heteroscedasticity through the application of Gamma family Generalized Linear Models with a log link and showed similar statistical significance.

Multivariate Mixed-Effect Meta-Regression Analysis of Study Characteristics and Reported Accuracy

Table 1 provides the summary statistics from the multivariate linear mixed-effect Regression model, which was used to estimate the conditional effects of study-level characteristics on reported accuracy of classification models. From an omnibus mixed-effect ANOVA, reported classification accuracy was found to be significantly associated with all six study characteristics. The significant association were found for the total sample size of the

training set (F-value=13.96, $p=0.0003$), type of classification model (F-value=6.10, $p=0.0004$), class ratio (F-value=7.24, $p=0.007$), approach used by studies to assess validation of model performance (F-value=7.95, $p=6 \times 10^{-5}$), the diagnostic group studied (F-value = 4.19, $p=0.0001$), and study-rated bias risk (F-value=6.59, $p=0.01$).

Post-hoc pair-wise tests revealed significant differences in classification performance among categorical study characteristics. Studies employing logistic regression-based classifiers exhibited significantly lower classification accuracies compared to those using ML models ($p=0.0002$). Logistic regression-based classifiers also exhibited significantly lower classification accuracies compared to DL models ($p=0.002$). In addition, studies using LOOCV exhibited significantly higher classification accuracies compared to studies that utilized withheld independent test sets ($p < 10^{-22}$), as well as studies that used k-fold CV ($p=0.001$). Furthermore, we found that studies that reported classification accuracies for models tasked with detecting AD/MCI were significantly higher than models tasked with detecting PTSD ($p=0.0009$). Studies that reported classification accuracies for models tasked with detecting AD/MCI were significantly higher than models tasked with detecting MDD ($p=0.002$). Studies that reported classification accuracies for ASD were significantly higher than PTSD models ($p=0.027$). Studies that reported classification accuracies for BD/SCZ were significantly higher than models focused on PTSD ($p=0.039$).

Relationship between Classification Performance and Tissue Source of Biomarkers

We found no significant difference in reported classification accuracies between studies that were trained on blood or brain data and tested on blood or brain data, suggesting that biomarker sourced from *postmortem* brain, despite their potential ontological relevance for psychiatric disorders, does not improve classification performance as compared to direct blood data (**Figure 5**). Moreover, whether the training and validation sets were acquired from the

same tissue-type (either blood or brain tissue), did not significantly impact classification accuracies.

DISCUSSION

This review summarizes the main model and study design variables impacting gene-expression classifier performance in neuropsychiatric disorders. We found that these factors significantly influence the reported performance of gene expression-based classifiers, underscoring the importance of careful interpretation. If these factors are not thoroughly considered during key stages of study design, including model training and validation, the reported performance of gene expression-based classifiers may be either over- or underestimated, which can lead to misleading interpretations of the utility of these models by investigators. We highlight recommendations from the literature on best practices for mitigating sources of bias, providing guidance for investigators to enhance the robustness and performance of their models (Vabalas, et al., 2019; Zhang-James, et al., 2023; Barnett et al., 2024).

A prevailing challenge with applying classification models to high-dimensional data is the establishment of reliable comparative benchmarks. This can be attributed to the heterogeneity of psychiatric diagnoses, novelty of modeling approaches for diagnosis, and lack of standardized reference performances. Our review provides estimates of the mean and range of reported classification accuracies of gene-expression-based classifiers for specific neuropsychiatric disorders, offering a useful reference point for benchmarking performance in future classification studies. Considering that there was an uneven representation of models in our meta-regression, it is imperative to use caution when extrapolating the estimates and interpreting how the results can be applied.

Our analyses found several trends that should be useful for designing future studies. Firstly, increasing sample size predicted decreasing accuracy with different types of validation methodologies. While this finding may appear counterintuitive, it has been documented in prior

meta-analyses of neuropsychiatric biomarker studies (Vabalas, et al., 2019; Zhang-James, et al., 2023). It is likely that studies with smaller sample sizes are more vulnerable to overfitting, resulting in overestimation of model performance. Additionally, variation in sample ascertainment between small and large studies might influence the homogeneity of study groups, potentially influencing estimates of model performance. Inflated estimates from small studies may not be replicable in larger studies that encompass more representative samples capturing the full range of heterogeneity of the disorders of interest. The technical constraints due to small sample size, along with the considerable pathophysiological and clinical heterogeneity inherent in neuropsychiatric disorders, presents challenges for generalizing classification models in future unseen data. These findings underscore the importance of high-quality datasets generated from large, diverse cohorts representative of the population.

We found a small, non-significant trend toward decreasing accuracy among reported classification models over the years, a phenomenon which may be influenced by many factors (Rajput, et al., 2022; Guo, Y. et al., 2010). This has been the case in a number of meta-analyses focused on high-dimensional data (Zhang-James, Y., et al., 2023; Xiong, S., et al., 2023). A potential source of model statistic changes over time is increased awareness on the issue of overfitting and model bias. Improved practices that limit study bias such as using standardized approaches for data pre-processing and algorithm deployment may also reduce the risk for over-fitting and over-estimation of classification performance. Choosing a model that exhibits low bias and is the right fit for the data structure is key for improving performance and generalizability of classification models. Optimizing the amount of signal captured by models while steering models away from overfitting requires careful practices to be adopted at various levels of the analysis, including feature selection and/or feature engineering, model design, training, and validation. Additionally, we cannot exclude the potential influence of bias towards better-performing models over the years—commonly known as the “file drawer effect”

(Nagarajan et al., 2017). This phenomenon, characterized by the likelihood of positive results being published, may contribute to the observed skewed outcomes over the years.

When dealing with small, cross-validation procedures such as LOOCV and k-fold CV are commonly used for assessing model performance, which allows all available data to be used for training and assessing classifiers. However, the literature suggests that cross-validation yields overly optimistic estimates of model performance compared to studies that employ divided training/test set-ups, especially if the datasets are small and exhibit high variability between the subsets of data for training and testing in each fold, or if there are imbalances between classes in the dataset (Tabe-Bordbar, S., 2018). Over-estimation of model performance can worsen when proper practices are not adhered to in order to prevent data leakage across folds as has been documented for genomic data (Barnett et al., 2024). Leakage of information between subsets of data used for training and validation can happen in numerous ways. One of the most common missteps that lead to data leakage involves applying feature selection, scaling, or transformations—such as normalization, dimensionality reduction, and oversampling—across the entire dataset prior to splitting the data into subsets for cross-validation or dividing datasets into dedicated training, validation, and/or test sets. When these techniques are applied on the complete dataset rather than solely on the training sample, information can inadvertently leak to the validation and/or test data, leading to inflated estimates of model accuracy. This phenomenon was evident in our meta-regression results, where studies exhibiting high bias through improper practices such as breaking the separation of training and validation subsets, tended to overestimate the performance of their models.

We found that ML models, particularly SVM, were the most common type of classification model employed among studies included in our meta-analysis. The popularity of SVM among studies employing gene-expression based classification methods might be attributed to the relative ease of use of such models, ability to capture higher-order interactions between features that may not be known *a priori*, and the availability of explainability methods.

Our meta-regression found that the type of classification model selected by studies had a significant influence on reported classification accuracy. Studies using conventional regression-based methods yielded lower classification accuracies than studies using more complex ML and DL models. This makes sense given that, compared with regression, the ML and DL models can model more complex relationships between features and outcomes and may better model the ground truth of psychiatric disorders which, themselves, have complex biology.

Blood, being a readily accessible source of putative biomarkers, offers practical benefits compared to the challenges associated with collecting brain tissue. Brain tissue is not readily accessible in most clinical settings. Interestingly, our results did not reveal a statistically significant difference among classification models trained and validated on these tissues. This result could be partially due to uneven sample size distribution across groups: 232 models were trained and validated on blood, two were trained on blood and validated on brain tissue, nine were trained on brain and validated on blood, and 37 were trained and validated on brain. Nevertheless, the result suggests that classification tasks are not particularly tissue-specific in training and validation, bringing into question the idea that tissue-specific or cross-tissue analyses will necessarily lead to improved classification accuracies.

There are advanced modeling methodologies holding potential promise that were not used in the pool of investigated studies. It is possible that models currently not utilized in neuropsychiatric gene expression analyses may perform better than current models (Wolpert & Macready, 1997).

Conclusions

We identified many cases of high study bias leading to either over-fitting or over-estimations of model performance, underscoring the critical importance of adhering to guidelines for the prudent use of ML- or regression-based classification models throughout the entire process, including study design, implementation, and reporting (Quinn, TP., et al., 2024).

Furthermore, the findings from our meta-regression show that studies with relatively small sample sizes produced inflated estimates of model performance, underscoring the importance of large, representative datasets. A significant portion of variability in performance of classification models observed across studies can be attributed to the type of classification model selected, the proportion of cases in the training dataset, study-rated bias risk, and approach for validating classification models. Notably, studies that employed ML- and DL-based models exhibited significantly better classification accuracies than regression-based models, even when controlling for other study factors such as sample size and bias risk. In addition, studies with larger numbers of samples in their training dataset tended to yield lower estimates of model performance, particularly for studies that employed k-fold CV or LOOCV to evaluate model performance. This may reflect that studies with relatively few samples for training and evaluating models are susceptible to over-estimating model performance. Conversely, studies employing CV as the primary method for evaluating model performance may better capture patterns capable of generalizing to future unseen data. Based on the findings from our meta-regression, greater prioritization should be given to studies employing large sample sizes and robust bias control. Leveraging larger, diverse datasets in tandem with rigorous ML techniques holds the potential to substantially enhance the generalizability of gene-expression based classifiers tailored for neuropsychiatric disorders.

FUNDING ACKNOWLEDGEMENTS

JLH is supported by grants from the U.S. National Institutes of Health (R21MH126494, R01NS128535), the Central New York Community Fund, and the NARSAD: The Brain & Behavior Research Foundation (2020 Young Investigator Award). **SJG** is supported by grants from the U.S. National Institutes of Health (R01AG064955 and R21MH126494), the U.S. National Science Foundation, the Sidney R. Baer, Jr. Foundation, and NARSAD: The Brain & Behavior Research Foundation. **SVF** is supported by the European Union's Horizon 2020

research and innovation programme under grant agreement 965381; NIH/NIMH grants U01AR076092, R01MH116037, 1R01NS128535, R01MH131685, 1R01MH130899, U01MH135970, and Supernus Pharmaceuticals. His continuing medical education programs are supported by The Upstate Foundation, Corium Pharmaceuticals, Tris Pharmaceuticals and Supernus Pharmaceuticals.

CONFLICTS OF INTEREST

Dr. Faraone in the past 36 months received income, potential income, travel expenses continuing education support and/or research support from Arbor, Aardvark, Aardwolf, AIMH, Akili, Atentiv, Axsome, Genomind, Ironshore, Johnson & Johnson/Kenvue, Kanjo, KemPharm/Corium, Medice, Noven, Ondosis, Otsuka, Rhodes, Sky Therapeutics, Sandoz, Supernus, Takeda, Tris, and Vallon. With his institution, he has US patent US20130217707 A1 for the use of sodium-hydrogen exchange inhibitors in the treatment of ADHD. He also receives royalties from books published by Guilford Press: *Straight Talk about Your Child's Mental Health*, Oxford University Press: *Schizophrenia: The Facts* and Elsevier: *ADHD: Non-Pharmacologic Interventions*. He is Program Director of www.ADHDEvidence.org and www.ADHDinAdults.com.

AR, BA, EJB, LAL, SJG, and JLH, have no conflicts of interest of report.

DATA AVAILABILITY STATEMENT

The data used in this paper were obtained from published research papers and repositories referenced throughout this manuscript. Specific details in regards to the datasets and studies investigated can be found in Supplementary File 1.

Table 1. Summary statistics from a type III mixed-effect analysis of variance (ANOVA) and mixed-effect linear regression models. The ANOVA model was used to estimate the omnibus association of each predictor variable, particularly categorical variables that had 3+ factor levels (*i.e.*, validation type and classifier type), on reported mean accuracy of the studies included in our meta-regression. Summary statistics on the right-hand side of the table (*i.e.*, not under the columns designated for the ANOVA statistics) are the regression coefficients indicating the main effects of continuous or differences in group means for 2-level categorical values on reported classification accuracies. For categorical variables with 3+ groups, we provide the pair-wise differences in estimated marginal means. In the mixed-effect ANOVA and regression models, a unique study indicator was set as the random-effect variable to adjust for non-independence between model accuracies.

Variable	Type III Mixed-effect ANOVA			Estimated marginal means (and <i>post-hoc</i> comparisons)			
	F-value	d.f.	p-value	Beta*	S.E.	t-value	p-value*
Diagnostic group*	4.1857	9, 106.42	0.0001195				
AD/MCI vs. MDD				0.16	0.039	4.10	0.002
AD/MCI vs. PTSD				0.22	0.048	4.54	0.0009
ASD vs PTSD				0.178	0.554	3.221	0.0272
BD/SCZ vs. PTSD				0.152	0.050	3.02	0.0386
Class ratio	7.3524	1, 451.03	0.0069530	0.17	0.065	2.62	0.0089
Classifier type	6.1037	3 495.92	0.0004400				
ML vs. Regression				0.07227	0.017	- 4.157	0.0002
DL vs. Regression				0.07117	0.021	3.421	0.0020
Bias risk (Probable/High vs. Low)	6.5874	1 126.63	0.0114339	0.0743	0.029	2.547	0.0122
Validation type	7.9538	3,148.35	5.928e-05				
Divided vs. k-fold CV				-0.00073	0.027	0.055	0.96
Divided vs. LOOCV				-0.1998	0.042	-4.73	<10^{-22*}
Divided vs. None				-0.0877	0.0515	-1.703	0.1170
k-fold CV vs. LOOCV				-0.1991	0.0453	-4.394	0.001
k-fold CV vs. None				-0.0869	0.0522	-1.666	0.1170
LOOCV vs. None				0.112137	0.0608	1.846	0.1170
Total sample size, log₁₀	13.958	1,139.99	0.0002713	-0.085	0.03	-2.83	0.0055

Note: Degrees of freedom for the mixed-effect regression were estimated using Satterthwaite's method. **Beta*:** indicates the regression coefficient (slope) of the main effect for continuous variables on reported accuracy, or differences in estimated marginal means of reported accuracy between categorical variables. **p-value*:** the column title denoting *p*-values indicates that *post-hoc* pairwise comparisons were adjusted for multiple comparisons with Benjamin-Hochberg false-discovery rate procedure. **:values that could not be printed with more precision. **Diagnostic group*:** As there were a large number of *post-hoc* comparisons with respect to the diagnostic group variable, we present only the statistically significant findings in the table.

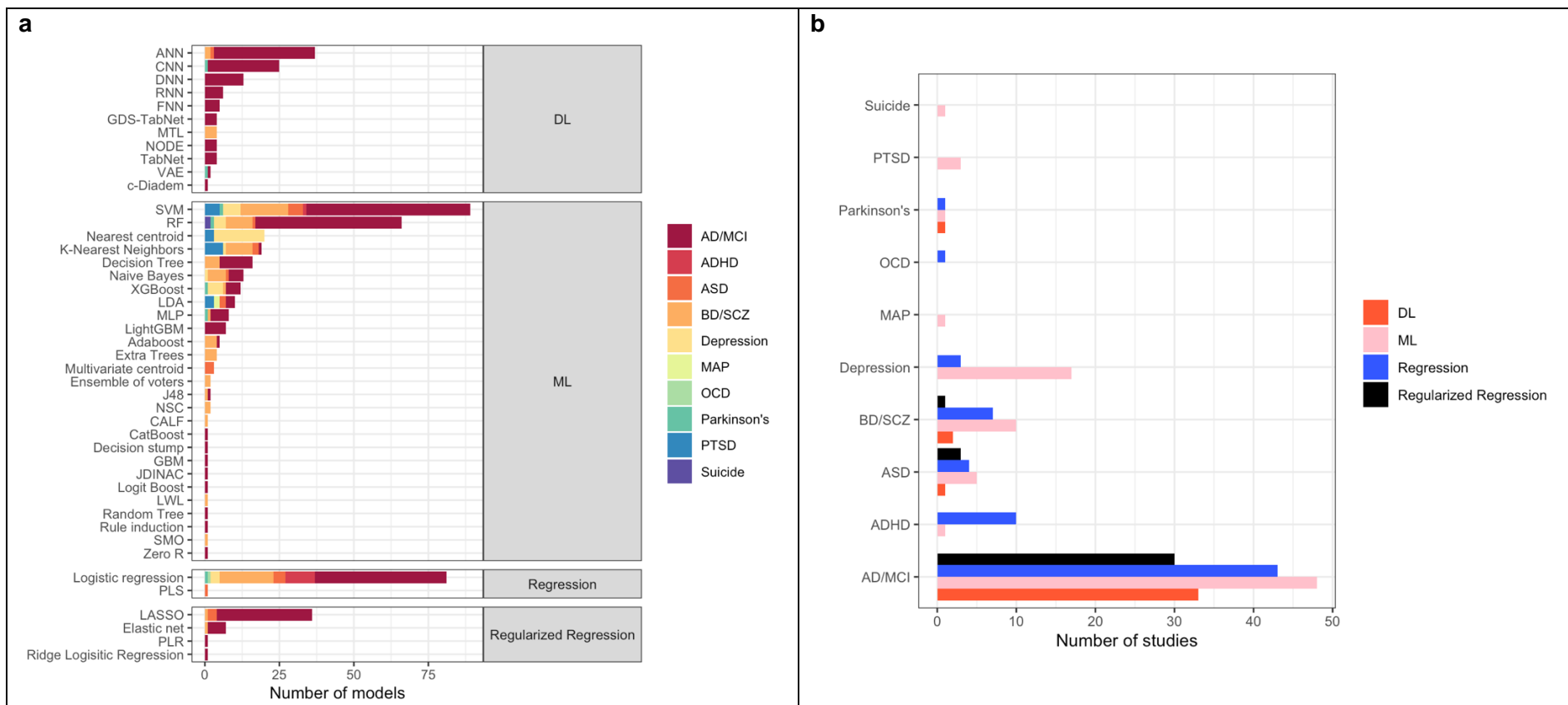


Figure 1. Graphical overview of the gene-expression based classification studies included in our meta-regression. **(a)** A bar-plot showing the number of classification models used for each of the main neuropsychiatric disorders represented among studies in our meta-regression dataset. **(b)** The number of studies featured per classification model type for each of the neuropsychiatric disorders represented in our dataset. Note: To enhance visualization clarity, we grouped Alzheimer’s disease (AD) and mild cognitive impairment (MCI)-focused classifiers into a single diagnostic label (“AD/MCI”). Similarly, we grouped schizophrenia (SCZ), bipolar disorder (BD), and psychosis into a single diagnostic label (“SCZ/Psychosis/BD”). Other abbreviations: attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), methamphetamine-associated psychosis (MAP), major depressive disorder (MDD), Parkinson’s disease (PD), post-traumatic stress disorder (PTSD).

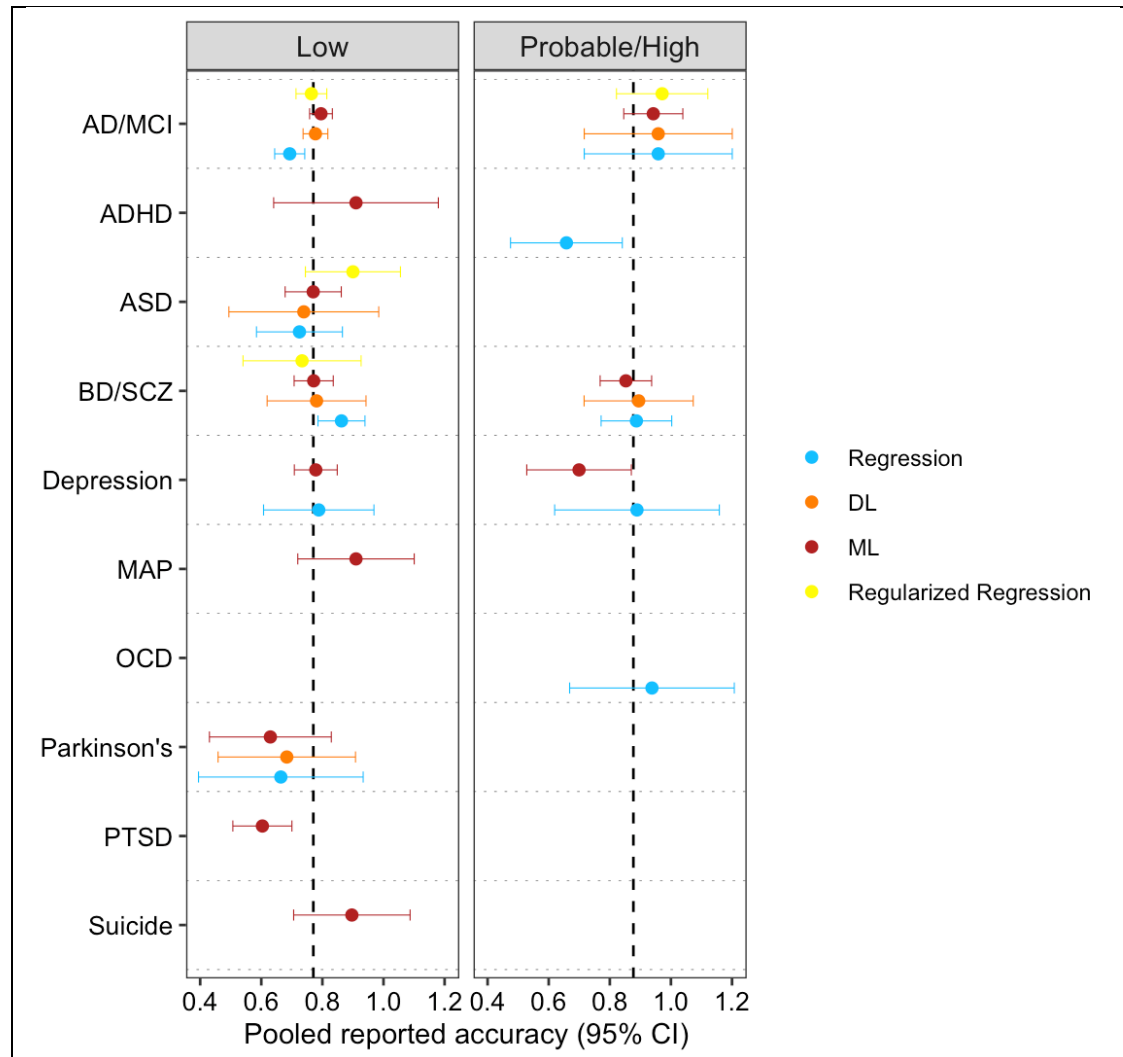


Figure 2. Forest plot showing the pooled estimated marginal means in reported accuracy for each diagnostic group. Pooled estimates of reported accuracy were obtained using a univariate linear mixed-effect model to adjust for repeated measures from studies. Estimated mean reported accuracies are color coded by type of classification model as follows: regression models appear in blue, machine learning models appear in orange, and neural networks appear in red. Vertical dashed lines designate the average pooled reported accuracy among all models according to study bias rating. There was a significant difference in reported accuracy across diagnostic groups according to mixed-effect type III ANOVA accounting for bias risk and model type ($F_{9,111.17} = 2.1663, p = 0.029$). Some categories do not include points in the figure due to limitations of estimated marginal means calculation for small sample sizes. **Abbreviations:** Alzheimer's disease (AD), mild cognitive impairment (MCI), attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), bipolar disorder (BD), methamphetamine-associated psychosis (MAP), major depressive disorder (MDD), Parkinson's disease (PD), post-traumatic stress disorder (PTSD), schizophrenia (SCZ)

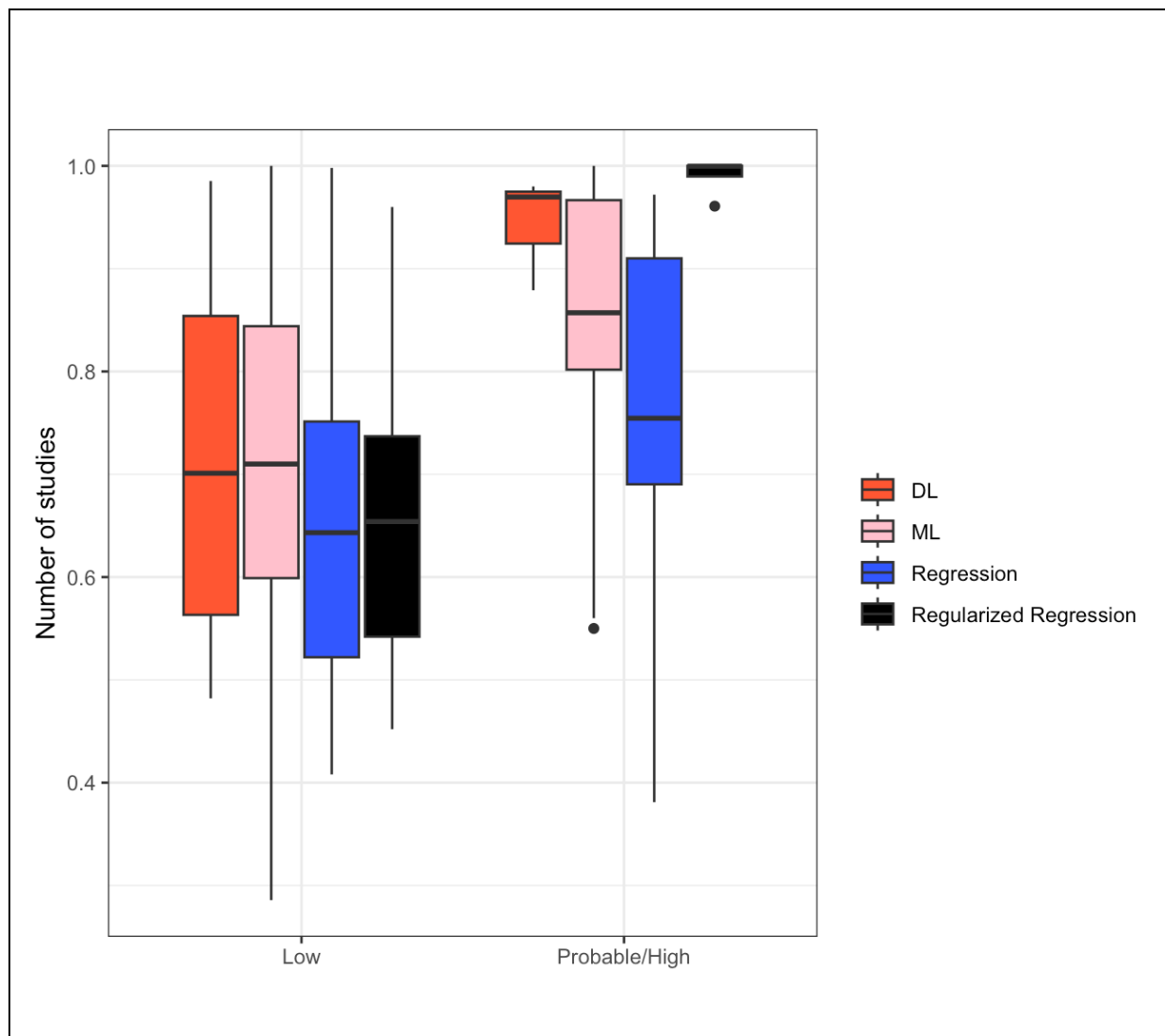


Figure 3. Accuracy of models stratified by their risk of study bias and type of classification model. **Abbreviations:** machine learning (ML), deep learning (DL). A type III mixed-effect ANOVA revealed a significant group-wise difference in reported accuracy between studies with low vs. probable/high bias ($F_{1, 291.90} = 8.3512, p = 0.004143$), but no significant difference was observed based on the interaction between bias rating and type of classification model.

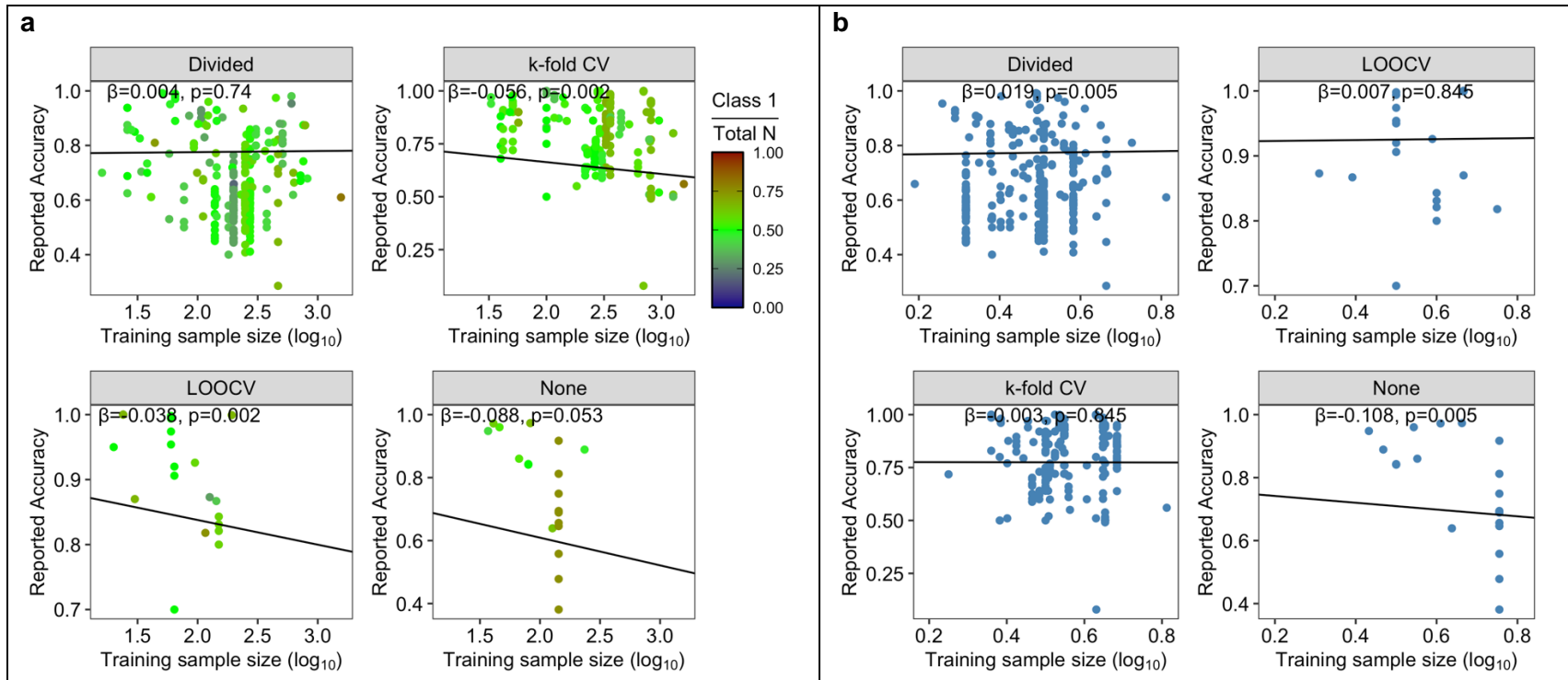
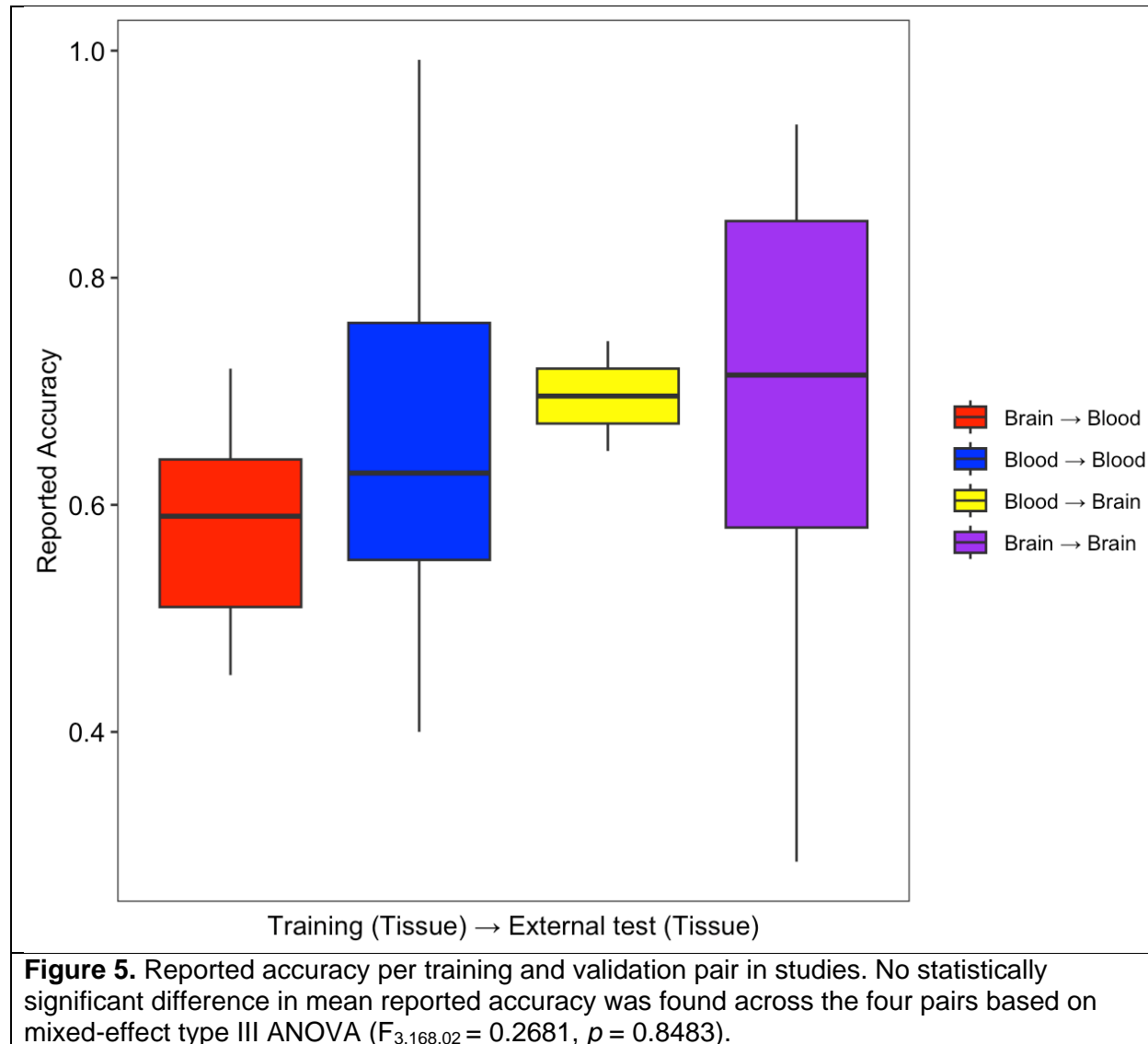


Figure 4. Scatterplots depicting the univariate linear relationship between reported classification model accuracies and **(left)** the log₁₀-transformed total sample size of the training dataset (x-axis), and **(right)** the proportion of positive classes in the training dataset. Plots were stratified by types of validation procedure used in the study. Each dot represents a single classification model. The regression coefficients and *p*-values from a univariate mixed-effect regression model are provided in each plot, which account for non-independence of the studies in our meta-regression. Black solid lines denote the best-fit relationship between variables based on mixed-effect linear regression.



REFERENCES

- Aglinskas A, Hartshorne JK, Anzellotti S. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. *Science*. 2022 Jun 3;376(6597):1070-1074. doi: 10.1126/science.abm2461. Epub 2022 Jun 2. PMID: 35653486.
- Alamro, H., Thafar, M. A., Albaradei, S., Gojobori, T., Essack, M., & Gao, X. (2023). Exploiting machine learning models to identify novel Alzheimer's disease biomarkers and potential targets. *Scientific Reports*, 13(1), 4979. <https://doi.org/10.1038/s41598-023-30904-5>
- Alsaadi T, Kassie S, Mohamed Ali O, Mozahem K, Al Fardan S, Ahmed AM. Psychiatric Comorbidity in Neurological Disorders: Towards a Multidisciplinary Approach to Illness Management in the United Arab Emirates. *Front Psychiatry*. 2019 Apr 25;10:263. doi: 10.3389/fpsy.2019.00263. PMID: 31073293; PMCID: PMC6495369.
- Alshamlan, H., Omar, S., Aljurayyad, R., & Alabduljabbar, R. (2023). Identifying effective feature selection methods for alzheimer's disease biomarker gene detection using machine learning. *Diagnostics (Basel, Switzerland)*, 13(10), 1771. <https://doi.org/10.3390/diagnostics13101771>
- Akkaya, U. M., & Kalkan, H. (2023). A New Approach for Multimodal Usage of Gene Expression and Its Image Representation for the Detection of Alzheimer's Disease. *Biomolecules*, 13(11), 1563. <https://doi.org/10.3390/biom13111563>
- Arzouni, N., Matloff, W., Zhao, L., Ning, K., & Toga, A. W. (2020). Identification of dysregulated genes for late-onset alzheimer's disease using gene expression data in brain. *Journal of Alzheimer's Disease & Parkinsonism*, 10(6), 498.
- Barnett, E. J., Onete, D. G., Salekin, A., & Faraone, S. V. (2024). Genomic machine learning meta-regression: Insights on associations of study features with reported model performance. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 21(1), 169–177. <https://doi.org/10.1109/TCBB.2023.3343808>
- Basu, M., Wang, K., Ruppin, E., & Hannenhalli, S. (2021). Predicting tissue-specific gene expression from whole blood transcriptome. *Science Advances*, 7(14), eabd6991. <https://doi.org/10.1126/sciadv.abd6991>
- Bhak, Y., Jeong, H.-O., Cho, Y. S., Jeon, S., Cho, J., Gim, J.-A., Jeon, Y., Blazyte, A., Park, S. G., Kim, H.-M., Shin, E.-S., Paik, J.-W., Lee, H.-W., Kang, W., Kim, A., Kim, Y., Kim, B. C., Ham, B.-J., Bhak, J., & Lee, S. (2019). Depression and suicide risk prediction models using blood-derived multi-omics data. *Translational Psychiatry*, 9(1), 262. <https://doi.org/10.1038/s41398-019-0595-2>
- Biomarkers consortium—The autism biomarkers consortium for clinical trials(Abc-ct)*. (n.d.). FNIH. <https://fnih.org/our-programs/biomarkers-consortium-the-autism-biomarkers-consortium-for-clinical-trials-abc-ct/>
- Breen, M. S., Tylee, D. S., Maihofer, A. X., Neylan, T. C., Mehta, D., Binder, E. B., Chandler, S. D., Hess, J. L., Kremen, W. S., Risbrough, V. B., Woelk, C. H., Baker, D. G., Nievergelt, C. M., Tsuang, M. T., Buxbaum, J. D., & Glatt, S. J. (2018). Ptsd blood transcriptome mega-analysis: Shared inflammatory pathways across biological sex and modes of trauma. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 43(3), 469–481. <https://doi.org/10.1038/npp.2017.220>
- Breen, M. S., Uhlmann, A., Nday, C. M., Glatt, S. J., Mitt, M., Metsalpu, A., Stein, D. J., & Illing, N. (2016). Candidate gene networks and blood biomarkers of methamphetamine-associated psychosis: An integrative RNA-sequencing report. *Translational Psychiatry*, 6(5), e802. <https://doi.org/10.1038/tp.2016.67>
- Cai, C., Langfelder, P., Fuller, T. F., Oldham, M. C., Luo, R., van den Berg, L. H., Ophoff, R. A., & Horvath, S. (2010). Is human blood a good surrogate for brain tissue in transcriptional studies? *BMC Genomics*, 11(1), 589. <https://doi.org/10.1186/1471-2164-11-589>

- Cao, H., Meyer-Lindenberg, A., & Schwarz, E. (2018). Comparative evaluation of machine learning strategies for analyzing big data in psychiatry. *International Journal of Molecular Sciences*, *19*(11), 3387. <https://doi.org/10.3390/ijms19113387>
- Carrillo-Roa, T., Labermaier, C., Weber, P., Herzog, D. P., Lareau, C., Santarelli, S., Wagner, K. V., Rex-Haffner, M., Harbich, D., Scharf, S. H., Nemeroff, C. B., Dunlop, B. W., Craighead, W. E., Mayberg, H. S., Schmidt, M. V., Uhr, M., Holsboer, F., Sillaber, I., Binder, E. B., & Müller, M. B. (2017). Common genes associated with antidepressant response in mouse and man identify key role of glucocorticoid receptor sensitivity. *PLoS Biology*, *15*(12), e2002690. <https://doi.org/10.1371/journal.pbio.2002690>
- Chen, D., Zhang, Y., Qiao, R., Kong, X., Zhong, H., Wang, X., Zhu, J., & Li, B. (2022). Integrated bioinformatics-based identification of diagnostic markers in Alzheimer disease. *Frontiers in Aging Neuroscience*, *14*, 988143. <https://doi.org/10.3389/fnagi.2022.988143>
- Chen, H., He, Y., Ji, J., & Shi, Y. (2019). A machine learning method for identifying critical interactions between gene pairs in alzheimer's disease prediction. *Frontiers in Neurology*, *10*, 1162. <https://doi.org/10.3389/fneur.2019.01162>
- Cheng, J., Liu, H.-P., Lin, W.-Y., & Tsai, F.-J. (2021). Machine learning compensates fold-change method and highlights oxidative phosphorylation in the brain transcriptome of Alzheimer's disease. *Scientific Reports*, *11*(1), 13704. <https://doi.org/10.1038/s41598-021-93085-z>
- Cheng, W., Zhou, S., Zhou, J., & Wang, X. (2020). Identification of a robust non-coding RNA signature in diagnosing autism spectrum disorder by cross-validation of microarray data from peripheral blood samples. *Medicine*, *99*(11), e19484. <https://doi.org/10.1097/MD.0000000000019484>
- Chiricosta, L., D'Angiolini, S., Gugliandolo, A., & Mazzon, E. (2022). Artificial intelligence predictor for alzheimer's disease trained on blood transcriptome: The role of oxidative stress. *International Journal of Molecular Sciences*, *23*(9), 5237. <https://doi.org/10.3390/ijms23095237>
- Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology*, *110*, 12–22. <https://doi.org/10.1016/j.jclinepi.2019.02.004>
- Cornblath, E. J., Lydon-Staley, D. M., & Bassett, D. S. (2019). Harnessing networks and machine learning in neuropsychiatric care. *Current Opinion in Neurobiology*, *55*, 32–39. <https://doi.org/10.1016/j.conb.2018.12.010>
- Daré, L.O., Bruand, P.E., Gérard, D. *et al.* Co-morbidities of mental disorders and chronic physical diseases in developing and emerging countries: a meta-analysis. *BMC Public Health* *19*, 304 (2019). <https://doi.org/10.1186/s12889-019-6623-6>
- Dobromylin, V. I., Megherbi, D. B., & Alzheimer's Disease Neuroimaging Initiative. (2022). Augmenting imaging biomarker performance with blood-based gene expression levels for predicting alzheimer's disease progression. *Journal of Alzheimer's Disease: JAD*, *87*(2), 583–594. <https://doi.org/10.3233/JAD-215640>
- Dufumier, B., Gori, P., Petiton, S., Louiset, R., Mangin, J. F., Grigis, A., & Duchesnay, E. (2024). Exploring the potential of representation and transfer learning for anatomical neuroimaging: application to psychiatry.
- Fan, H., Sun, X., Guo, W., Zhong, A., Niu, W., Zhao, L., Dai, Y., Guo, Z., Zhang, L., & Lu, J. (2014). Differential expression of microRNA in peripheral blood mononuclear cells as specific biomarker for major depressive disorder patients. *Journal of Psychiatric Research*, *59*, 45–52. <https://doi.org/10.1016/j.jpsychires.2014.08.007>
- Fan, H.-M., Sun, X.-Y., Niu, W., Zhao, L., Zhang, Q.-L., Li, W.-S., Zhong, A.-F., Zhang, L.-Y., & Lu, J. (2015). Altered microrna expression in peripheral blood mononuclear cells from young patients with schizophrenia. *Journal of Molecular Neuroscience: MN*, *56*(3), 562–571. <https://doi.org/10.1007/s12031-015-0503-z>

- Glatt, S. J., Tsuang, M. T., Winn, M., Chandler, S. D., Collins, M., Lopez, L., Weinfeld, M., Carter, C., Schork, N., Pierce, K., & Courchesne, E. (2012). Blood-based gene expression signatures of infants and toddlers with autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, 51(9), 934-944.e2. <https://doi.org/10.1016/j.jaac.2012.07.007>
- Glatt, S. J., Tylee, D. S., Chandler, S. D., Pazol, J., Nievergelt, C. M., Woelk, C. H., Baker, D. G., Lohr, J. B., Kremen, W. S., Litz, B. T., Tsuang, M. T., & Marine Resiliency Study Investigators. (2013). Blood-based gene-expression predictors of PTSD risk and resilience among deployed marines: A pilot study. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 162B(4), 313–326. <https://doi.org/10.1002/ajmg.b.32167>
- Gomez Rueda, H., & Bustillo, J. (2022). Brain differential gene expression and blood cross-validation of a molecular signature of patients with major depressive disorder. *Psychiatric Genetics*, 32(3), 105–115. <https://doi.org/10.1097/YPG.0000000000000309>
- Grünblatt, E., Geissler, J., Jacob, C. P., Renner, T., Müller, M., Bartl, J., Gross-Lesch, S., Riederer, P., Lesch, K.-P., Walitza, S., Gerlach, M., & Schmitt, A. (2012). Pilot study: Potential transcription markers for adult attention-deficit hyperactivity disorder in whole blood. *Attention Deficit and Hyperactivity Disorders*, 4(2), 77–84. <https://doi.org/10.1007/s12402-012-0074-6>
- Gu, X., Lai, D., Liu, S., Chen, K., Zhang, P., Chen, B., Huang, G., Cheng, X., & Lu, C. (2022). Hub genes, diagnostic model, and predicted drugs related to iron metabolism in alzheimer's disease. *Frontiers in Aging Neuroscience*, 14, 949083. <https://doi.org/10.3389/fnagi.2022.949083>
- Guo, Y., Graber, A., McBurney, R.N. *et al.* Sample size and statistical power considerations in high-dimensionality data settings: a comparative study of classification algorithms. *BMC Bioinformatics* 11, 447 (2010). <https://doi.org/10.1186/1471-2105-11-447>
- Guan, J., Yang, E., Yang, J., Zeng, Y., Ji, G., & Cai, J. J. (2016). Exploiting aberrant mRNA expression in autism for gene discovery and diagnosis. *Human Genetics*, 135(7), 797–811. <https://doi.org/10.1007/s00439-016-1673-7>
- He, K., Guo, C., Guo, M., Tong, S., Zhang, Q., Sun, H., He, L., & Shi, Y. (2019). Identification of serum microRNAs as diagnostic biomarkers for schizophrenia. *Hereditas*, 156, 23. <https://doi.org/10.1186/s41065-019-0099-3>
- Hess, J. L., Quinn, T. P., Zhang, C., Hearn, G. C., Chen, S., Kong, S. W., Cairns, M., Tsuang, M. T., Faraone, S. V., & Glatt, S. J. (2023). Braingenie: The brain gene expression and network imputation engine. *Translational Psychiatry*, 13(1), 1–10. <https://doi.org/10.1038/s41398-023-02390-w>
- Hess, J. L., Tylee, D. S., Barve, R., de Jong, S., Ophoff, R. A., Kumarasinghe, N., Tooney, P., Schall, U., Gardiner, E., Beveridge, N. J., Scott, R. J., Yasawardene, S., Perera, A., Mendis, J., Carr, V., Kelly, B., Cairns, M., Neurobehavioural Genetics Unit, Tsuang, M. T., & Glatt, S. J. (2016). Transcriptome-wide mega-analyses reveal joint dysregulation of immunologic genes and transcription regulators in brain and blood in schizophrenia. *Schizophrenia Research*, 176(2–3), 114–124. <https://doi.org/10.1016/j.schres.2016.07.006>
- Hess, J. L., Tylee, D. S., Barve, R., de Jong, S., Ophoff, R. A., Kumarasinghe, N., Tooney, P., Schall, U., Gardiner, E., Beveridge, N. J., Scott, R. J., Yasawardene, S., Perera, A., Mendis, J., Carr, V., Kelly, B., Cairns, M., Neurobehavioural Genetics Unit, Tsuang, M. T., & Glatt, S. J. (2020). Transcriptomic abnormalities in peripheral blood in bipolar disorder, and discrimination of the major psychoses. *Schizophrenia Research*, 217, 124–135. <https://doi.org/10.1016/j.schres.2019.07.036>
- Hu, V. W., & Lai, Y. (2013). Developing a predictive gene classifier for autism spectrum disorders based upon differential gene expression profiles of phenotypic subgroups. *North American Journal of Medicine & Science*, 6(3). <https://doi.org/10.7156/najms.2013.0603107>
- Jeffries, C. D., Perkins, D. O., Chandler, S. D., Stark, T., Yeo, E., Addington, J., Bearden, C. E., Cadenhead, K. S., Cannon, T. D., Cornblatt, B. A., Matalon, D. H., McGlashan, T. H., Seidman, L. J., Walker, E. F., Woods, S. W., Glatt, S. J., & Tsuang, M. (2016). Insights into

- psychosis risk from leukocyte microRNA expression. *Translational Psychiatry*, 6(12), e981. <https://doi.org/10.1038/tp.2016.148>
- Jemimah, S., AlShehhi, A., & Alzheimer's Disease Neuroimaging Initiative. (2023). C-diadem: A constrained dual-input deep learning model to identify novel biomarkers in alzheimer's disease. *BMC Medical Genomics*, 16(Suppl 2), 244. <https://doi.org/10.1186/s12920-023-01675-9>
- Ji, W., An, K., Wang, C., & Wang, S. (2022). Bioinformatics analysis of diagnostic biomarkers for Alzheimer's disease in peripheral blood based on sex differences and support vector machine algorithm. *Hereditas*, 159(1), 38. <https://doi.org/10.1186/s41065-022-00252-x>
- Jin, Y., Ren, Z., Wang, W., Zhang, Y., Zhou, L., Yao, X., & Wu, T. (2023). Classification of Alzheimer's disease using robust TabNet neural networks on genetic data. *Mathematical Biosciences and Engineering: MBE*, 20(5), 8358–8374. <https://doi.org/10.3934/mbe.2023366>
- Kalkan, H., Akkaya, U. M., Inal-Gültekin, G., & Sanchez-Perez, A. M. (2022). Prediction of alzheimer's disease by a novel image-based representation of gene expression. *Genes*, 13(8), 1406. <https://doi.org/10.3390/genes13081406>
- Kalmady, S. V., Paul, A. K., Narayanaswamy, J. C., Agrawal, R., Shivakumar, V., Greenshaw, A. J., ... & Reddy, Y. J. (2022). Prediction of obsessive-compulsive disorder: importance of neurobiology-aided feature design and cross-diagnosis transfer learning. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 7(7), 735-746.
- Karaglani, M., Gourlia, K., Tsamardinos, I., & Chatzaki, E. (2020). Accurate blood-based diagnostic biosignatures for alzheimer's disease via automated machine learning. *Journal of Clinical Medicine*, 9(9), 3016. <https://doi.org/10.3390/jcm9093016>
- Kato, T., Hayashi-Takagi, A., Toyota, T., Yoshikawa, T., & Iwamoto, K. (2011). Gene expression analysis in lymphoblastoid cells as a potential biomarker of bipolar disorder. *Journal of Human Genetics*, 56(11), 779–783. <https://doi.org/10.1038/jhg.2011.101>
- Kelly, J., Moyeed, R., Carroll, C., Luo, S., & Li, X. (2023). Blood biomarker-based classification study for neurodegenerative diseases. *Scientific Reports*, 13(1), 17191. <https://doi.org/10.1038/s41598-023-43956-4>
- Kong, S. W., Collins, C. D., Shimizu-Motohashi, Y., Holm, I. A., Campbell, M. G., Lee, I.-H., Brewster, S. J., Hanson, E., Harris, H. K., Lowe, K. R., Saada, A., Mora, A., Madison, K., Hundley, R., Egan, J., McCarthy, J., Eran, A., Galdzicki, M., Rappaport, L., ... Kohane, I. S. (2012). Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. *PLoS One*, 7(12), e49475. <https://doi.org/10.1371/journal.pone.0049475>
- Lai, C.-Y., Yu, S.-L., Hsieh, M. H., Chen, C.-H., Chen, H.-Y., Wen, C.-C., Huang, Y.-H., Hsiao, P.-C., Hsiao, C. K., Liu, C.-M., Yang, P.-C., Hwu, H.-G., & Chen, W. J. (2011). MicroRNA expression aberration as potential peripheral blood biomarkers for schizophrenia. *PLOS ONE*, 6(6), e21635. <https://doi.org/10.1371/journal.pone.0021635>
- Lai, Y., Lin, P., Lin, F., Chen, M., Lin, C., Lin, X., Wu, L., Zheng, M., & Chen, J. (2022). Identification of immune microenvironment subtypes and signature genes for Alzheimer's disease diagnosis and risk prediction based on explainable machine learning. *Frontiers in Immunology*, 13, 1046410. <https://doi.org/10.3389/fimmu.2022.1046410>
- Lai, Y., Lin, X., Lin, C., Lin, X., Chen, Z., & Zhang, L. (2022). Identification of endoplasmic reticulum stress-associated genes and subtypes for prediction of Alzheimer's disease based on interpretable machine learning. *Frontiers in Pharmacology*, 13, 975774. <https://doi.org/10.3389/fphar.2022.975774>
- Lee, T., & Lee, H. (2020). Prediction of Alzheimer's disease using blood gene expression data. *Scientific Reports*, 10(1), 3485. <https://doi.org/10.1038/s41598-020-60595-1>
- Li, J., Zhang, Y., Lu, T., Liang, R., Wu, Z., Liu, M., Qin, L., Chen, H., Yan, X., Deng, S., Zheng, J., & Liu, Q. (2022). Identification of diagnostic genes for both Alzheimer's disease and Metabolic syndrome by the machine learning algorithm. *Frontiers in Immunology*, 13, 1037318. <https://doi.org/10.3389/fimmu.2022.1037318>

- Lin, C.-H., Huang, M.-W., Lin, C.-H., Huang, C.-H., & Lane, H.-Y. (2019). Altered mRNA expressions for N-methyl-D-aspartate receptor-related genes in WBC of patients with major depressive disorder. *Journal of Affective Disorders*, *245*, 1119–1125. <https://doi.org/10.1016/j.jad.2018.12.016>
- Liu, L., Zhao, J., Chen, Y., & Feng, R. (2020). Metabolomics strategy assisted by transcriptomics analysis to identify biomarkers associated with schizophrenia. *Analytica Chimica Acta*, *1140*, 18–29. <https://doi.org/10.1016/j.aca.2020.09.054>
- Logotheti, M., Pilalis, E., Venizelos, N., Kolisis, F., & Chatziioannou, A. (2016). Studying microarray gene expression data of schizophrenic patients for derivation of a diagnostic signature through the aid of machine learning. *Biometrics & Biostatistics International Journal*, *4*(5). <https://doi.org/10.15406>
- Logotheti, M., Pilalis, E., Venizelos, N., Kolisis, F., Chatziioannou, A., Logotheti, M., Pilalis, E., Venizelos, N., Kolisis, F., & Chatziioannou, A. (2016). Development and validation of a skin fibroblast biomarker profile for schizophrenic patients. *AIMS Bioengineering*, *3*(4), 552–565. <https://doi.org/10.3934/bioeng.2016.4.552>
- Ludwig, N., Fehlmann, T., Kern, F., Gogol, M., Maetzler, W., Deutscher, S., Gurlit, S., Schulte, C., von Thaler, A.-K., Deuschle, C., Metzger, F., Berg, D., Suenkel, U., Keller, V., Backes, C., Lenhof, H.-P., Meese, E., & Keller, A. (2019). Machine learning to detect alzheimer's disease from circulating non-coding rnas. *Genomics, Proteomics & Bioinformatics*, *17*(4), 430–440. <https://doi.org/10.1016/j.gpb.2019.09.004>
- Madar, I. H., Sultan, G., Tayubi, I. A., Hasan, A. N., Pahi, B., Rai, A., Sivanandan, P. K., Loganathan, T., Begum, M., & Rai, S. (2021). Identification of marker genes in Alzheimer's disease using a machine-learning model. *Bioinformation*, *17*(2), 348–355. <https://doi.org/10.6026/97320630017348>
- Maj, C., Azevedo, T., Giansanti, V., Borisov, O., Dimitri, G. M., Spasov, S., Alzheimer's Disease Neuroimaging Initiative, Lió, P., & Merelli, I. (2019). Integration of machine learning methods to dissect genetically imputed transcriptomic profiles in alzheimer's disease. *Frontiers in Genetics*, *10*, 726. <https://doi.org/10.3389/fgene.2019.00726>
- Mamdani, F., Martin, M. V., Lencz, T., Rollins, B., Robinson, D. G., Moon, E. A., Malhotra, A. K., & Vawter, M. P. (2013). Coding and noncoding gene expression biomarkers in mood disorders and schizophrenia. *Disease Markers*, *35*(1), 11–21. <https://doi.org/10.1155/2013/748095>
- Mariani, N., Cattane, N., Pariante, C., & Cattaneo, A. (2021). Gene expression studies in Depression development and treatment: An overview of the underlying molecular mechanisms and biological processes to identify biomarkers. *Translational Psychiatry*, *11*(1), 354. <https://doi.org/10.1038/s41398-021-01469-6>
- Miller, J. B., & Kauwe, J. S. K. (2020). Predicting clinical dementia rating using blood rna levels. *Genes*, *11*(6), 706. <https://doi.org/10.3390/genes11060706>
- Moradi, E., Marttinen, M., Häkkinen, T., Hiltunen, M., & Nykter, M. (2019). Supervised pathway analysis of blood gene expression profiles in Alzheimer's disease. *Neurobiology of Aging*, *84*, 98–108. <https://doi.org/10.1016/j.neurobiolaging.2019.07.004>
- Munkholm, K., Peijs, L., Vinberg, M., & Kessing, L. V. (2015). A composite peripheral blood gene expression measure as a potential diagnostic biomarker in bipolar disorder. *Translational Psychiatry*, *5*(8), e614. <https://doi.org/10.1038/tp.2015.110>
- Munkholm, K., Vinberg, M., Pedersen, B. K., Poulsen, H. E., Ekstrøm, C. T., & Kessing, L. V. (2019). A multisystem composite biomarker as a preliminary diagnostic test in bipolar disorder. *Acta Psychiatrica Scandinavica*, *139*(3), 227–236. <https://doi.org/10.1111/acps.12983>
- Nagarajan, P. J., Garla, B. K., Taranath, M., & Nagarajan, I. (2017). The file drawer effect: A call for meticulous methodology and tolerance for non-significant results. *Indian journal of anaesthesia*, *61*(6), 516–517. https://doi.org/10.4103/ija.IJA_280_17

- Nazeen, S., Palmer, N. P., Berger, B., & Kohane, I. S. (2016). Integrative analysis of genetic data sets reveals a shared innate immune component in autism spectrum disorder and its comorbidities. *Genome Biology*, 17(1), 228. <https://doi.org/10.1186/s13059-016-1084-z>
- Oh, D. H., Kim, I. B., Kim, S. H., & Ahn, D. H. (2017). Predicting autism spectrum disorder using blood-based gene expression signatures and machine learning. *Clinical Psychopharmacology and Neuroscience: The Official Scientific Journal of the Korean College of Neuropsychopharmacology*, 15(1), 47–52. <https://doi.org/10.9758/cpn.2017.15.1.47>
- Paranjpe, M. D., Belonwu, S., Wang, J. K., Oskotsky, T., Gupta, A., Taubes, A., Zalocusky, K. A., Paranjpe, I., Glicksberg, B. S., Huang, Y., & Sirota, M. (2021). Sex-specific cross tissue meta-analysis identifies immune dysregulation in women with alzheimer's disease. *Frontiers in Aging Neuroscience*, 13, 735611. <https://doi.org/10.3389/fnagi.2021.735611>
- Pramparo, T., Pierce, K., Lombardo, M. V., Carter Barnes, C., Marinero, S., Ahrens-Barbeau, C., Murray, S. S., Lopez, L., Xu, R., & Courchesne, E. (2015). Prediction of autism by translation and immune/inflammation coexpressed genes in toddlers from pediatric community practices. *JAMA Psychiatry*, 72(4), 386–394. <https://doi.org/10.1001/jamapsychiatry.2014.3008>
- Qi, B., Ramamurthy, J., Bennani, I., & Trakadis, Y. J. (2021). Machine learning and bioinformatic analysis of brain and blood mRNA profiles in major depressive disorder: A case-control study. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 186(2), 101–112. <https://doi.org/10.1002/ajmg.b.32839>
- Quinn, T. P., Hess, J. L., Marshe, V. S., Barnett, M. M., Hauschild, A.-C., Maciukiewicz, M., Elsheikh, S. S. M., Men, X., Schwarz, E., Trakadis, Y. J., Breen, M. S., Barnett, E. J., Zhang-James, Y., Ahsen, M. E., Cao, H., Chen, J., Hou, J., Salekin, A., Lin, P.-I., ... Machine Learning in Psychiatry (MLPsych) Consortium. (2024). A primer on the use of machine learning to distil knowledge from data in biological psychiatry. *Molecular Psychiatry*. <https://doi.org/10.1038/s41380-023-02334-2>
- Rajput, D., Wang, WJ. & Chen, CC. Evaluation of a decided sample size in machine learning applications. *BMC Bioinformatics* 24, 48 (2023). <https://doi.org/10.1186/s12859-023-05156-9>
- Sainz, J., Prieto, C., Ruso-Julve, F., & Crespo-Facorro, B. (2018). Blood gene expression profile predicts response to antipsychotics. *Frontiers in Molecular Neuroscience*, 11, 73. <https://doi.org/10.3389/fnmol.2018.00073>
- Saylan, C., & Yilancioglu, K. (n.d.). Classification of schizophrenia and bipolar disorder by using machine learning algorithms. *The Journal of Neurobehavioral Sciences*. Scopemed. (n.d.). Retrieved May 1, 2024, from <https://www.scopemed.org/?mno=206492>
- Sekaran, K., Alsamman, A. M., George Priya Doss, C., & Zayed, H. (2023). Bioinformatics investigation on blood-based gene expressions of Alzheimer's disease revealed ORAI2 gene biomarker susceptibility: An explainable artificial intelligence-based approach. *Metabolic Brain Disease*, 38(4), 1297–1310. <https://doi.org/10.1007/s11011-023-01171-0>
- Sh, Y., Liu, B., Zhang, J., Zhou, Y., Hu, Z., & Zhang, X. (2021). Application of artificial intelligence modeling technology based on fluid biopsy to diagnose alzheimer's disease. *Frontiers in Aging Neuroscience*, 13, 768229. <https://doi.org/10.3389/fnagi.2021.768229>
- Sharma, A., & Dey, P. (2021). A machine learning approach to unmask novel gene signatures and prediction of Alzheimer's disease within different brain regions. *Genomics*, 113(4), 1778–1789. <https://doi.org/10.1016/j.ygeno.2021.04.028>
- Shigemizu, D., Akiyama, S., Asanomi, Y., Boroevich, K. A., Sharma, A., Tsunoda, T., Matsukuma, K., Ichikawa, M., Sudo, H., Takizawa, S., Sakurai, T., Ozaki, K., Ochiya, T., & Niida, S. (2019). Risk prediction models for dementia constructed by supervised principal component analysis using miRNA expression data. *Communications Biology*, 2, 77. <https://doi.org/10.1038/s42003-019-0324-7>

- Smith, D. L., & Held, P. (2023). Moving toward precision PTSD treatment: Predicting veterans' intensive PTSD treatment response using continuously updating machine learning models. *Psychological Medicine*, 53(12), 5500–5509. <https://doi.org/10.1017/S0033291722002689>
- Song, X., Liu, X., Liu, F., & Wang, C. (2021). Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis. *International Journal of Medical Informatics*, 151, 104484. <https://doi.org/10.1016/j.ijmedinf.2021.104484>
- Struyf, J., Dobrin, S., & Page, D. (2008). Combining gene expression, demographic and clinical data in modeling disease: A case study of bipolar disorder and schizophrenia. *BMC Genomics*, 9, 531. <https://doi.org/10.1186/1471-2164-9-531>
- Sun, D., Peng, H., & Wu, Z. (2022). Establishment and analysis of a combined diagnostic model of alzheimer's disease with random forest and artificial neural network. *Frontiers in Aging Neuroscience*, 14, 921906. <https://doi.org/10.3389/fnagi.2022.921906>
- Tabatabaie S, Emad A, Zhao SD, Sinha S. A closer look at cross-validation for assessing the accuracy of gene regulatory networks and models. *Sci Rep*. 2018 Apr 26;8(1):6620. doi: 10.1038/s41598-018-24937-4. PMID: 29700343; PMCID: PMC5920056.
- Takahashi, M., Hayashi, H., Watanabe, Y., Sawamura, K., Fukui, N., Watanabe, J., Kitajima, T., Yamanouchi, Y., Iwata, N., Mizukami, K., Hori, T., Shimoda, K., Ujike, H., Ozaki, N., Iijima, K., Takemura, K., Aoshima, H., & Someya, T. (2010). Diagnostic classification of schizophrenia by neural network analysis of blood-based gene expression signatures. *Schizophrenia Research*, 119(1–3), 210–218. <https://doi.org/10.1016/j.schres.2009.12.024>
- Tang, H., Liang, J., Chai, K., Gu, H., Ye, W., Cao, P., Chen, S., & Shen, D. (2023). Artificial intelligence and bioinformatics analyze markers of children's transcriptional genome to predict autism spectrum disorder. *Frontiers in Neurology*, 14, 1203375. <https://doi.org/10.3389/fneur.2023.1203375>
- Tatara, Y., Yamazaki, H., Katsuoka, F., Chiba, M., Saigusa, D., Kasai, S., Nakamura, T., Inoue, J., Aoki, Y., Shoji, M., Motoike, I. N., Tamada, Y., Hashizume, K., Shoji, M., Kinoshita, K., Murashita, K., Nakaji, S., Yamamoto, M., & Itoh, K. (2023). Multiomics and artificial intelligence enabled peripheral blood-based prediction of amnesic mild cognitive impairment. *Current Research in Translational Medicine*, 71(1), 103367. <https://doi.org/10.1016/j.retram.2022.103367>
- Teunissen, C. E., Verberk, I. M. W., Thijssen, E. H., Vermunt, L., Hansson, O., Zetterberg, H., van der Flier, W. M., Mielke, M. M., & Del Campo, M. (2022). Blood-based biomarkers for Alzheimer's disease: Towards clinical implementation. *The Lancet. Neurology*, 21(1), 66–77. [https://doi.org/10.1016/S1474-4422\(21\)00361-6](https://doi.org/10.1016/S1474-4422(21)00361-6)
- Tian, Y., Lu, Y., Cao, Y., Dang, C., Wang, N., Tian, K., Luo, Q., Guo, E., Luo, S., Wang, L., & Li, Q. (2022). Identification of diagnostic signatures associated with immune infiltration in Alzheimer's disease by integrating bioinformatic analysis and machine-learning strategies. *Frontiers in Aging Neuroscience*, 14, 919614. <https://doi.org/10.3389/fnagi.2022.919614>
- Tsuang, M. T., Nossova, N., Yager, T., Tsuang, M.-M., Guo, S.-C., Shyu, K. G., Glatt, S. J., & Liew, C. C. (2005). Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: A preliminary report. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 133B(1), 1–5. <https://doi.org/10.1002/ajmg.b.30161>
- Tylee, D. S., Chandler, S. D., Nievergelt, C. M., Liu, X., Pazol, J., Woelk, C. H., Lohr, J. B., Kremen, W. S., Baker, D. G., Glatt, S. J., Tsuang, M. T., & Marine Resiliency Study Investigators. (2015). Blood-based gene-expression biomarkers of post-traumatic stress disorder among deployed marines: A pilot study. *Psychoneuroendocrinology*, 51, 472–494. <https://doi.org/10.1016/j.psyneuen.2014.09.024>
- Tylee, D. S., Hess, J. L., Quinn, T. P., Barve, R., Huang, H., Zhang-James, Y., Chang, J., Stamova, B. S., Sharp, F. R., Hertz-Picciotto, I., Faraone, S. V., Kong, S. W., & Glatt, S. J. (2017). Blood

- transcriptomic comparison of individuals with and without autism spectrum disorder: A combined-samples mega-analysis. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 174(3), 181–201. <https://doi.org/10.1002/ajmg.b.32511>
- Tylee, D. S., Kawaguchi, D. M., & Glatt, S. J. (2013). On the outside, looking in: A review and evaluation of the comparability of blood and brain “-omes.” *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 162B(7), 595–603. <https://doi.org/10.1002/ajmg.b.32150>
- Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLOS ONE*, 14(11), e0224365. <https://doi.org/10.1371/journal.pone.0224365>
- van de Leemput, J., Glatt, S. J., & Tsuang, M. T. (2016). The potential of genetic and gene expression analysis in the diagnosis of neuropsychiatric disorders. *Expert Review of Molecular Diagnostics*, 16(6), 677–695. <https://doi.org/10.1586/14737159.2016.1171714>
- Varatharajah, Y., Ramanan, V. K., Iyer, R., Vemuri, P., & Alzheimer’s Disease Neuroimaging Initiative. (2019). Predicting short-term mci-to-ad progression using imaging, csf, genetic factors, cognitive resilience, and demographics. *Scientific Reports*, 9(1), 2235. <https://doi.org/10.1038/s41598-019-38793-3>
- Vawter, M. P., Philibert, R., Rollins, B., Ruppel, P. L., & Osborn, T. W. (2018). Exon array biomarkers for the differential diagnosis of schizophrenia and bipolar disorder. *Molecular Neuropsychiatry*, 3(4), 197–213. <https://doi.org/10.1159/000485800>
- Verma, P., & Shakya, M. (2022). Machine learning model for predicting Major Depressive Disorder using RNA-Seq data: Optimization of classification approach. *Cognitive Neurodynamics*, 16(2), 443–453. <https://doi.org/10.1007/s11571-021-09724-8>
- Viñas, R., Joshi, C.K., Georgiev, D. *et al.* Hypergraph factorization for multi-tissue gene expression imputation. *Nat Mach Intell* 5, 739–753 (2023). <https://doi.org/10.1038/s42256-023-00684-8>
- Wang, L.-J., Li, S.-C., Lee, M.-J., Chou, M.-C., Chou, W.-J., Lee, S.-Y., Hsu, C.-W., Huang, L.-H., & Kuo, H.-C. (2018). Blood-borne microrna biomarker evaluation in attention-deficit/hyperactivity disorder of han chinese individuals: An exploratory study. *Frontiers in Psychiatry*, 9, 227. <https://doi.org/10.3389/fpsy.2018.00227>
- Wang, T., Chen, K., Li, H., Dong, S., Su, N., Liu, Y., Cheng, Y., Dai, J., Yang, C., & Xiao, S. (2015). The feasibility of utilizing plasma MiRNA107 and BACE1 messenger RNA gene expression for clinical diagnosis of amnesic mild cognitive impairment. *The Journal of Clinical Psychiatry*, 76(2), 135–141. <https://doi.org/10.4088/JCP.13m08812>
- Wang, Y., Cheng, C., Zhang, Z., Wang, J., Wang, Y., Li, X., Gao, R., Wang, Z., Fang, Y., Wang, J., Wang, M., Fan, Q., Periya, S., Zhang, H., Tsuang, M. T., & Liew, C.-C. (2018). Blood-based dynamic genomic signature for obsessive-compulsive disorder. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 177(8), 709–716. <https://doi.org/10.1002/ajmg.b.32675>
- Watanabe, S.-Y., Numata, S., Iga, J.-I., Kinoshita, M., Umehara, H., Ishii, K., & Ohmori, T. (2017). Gene expression-based biological test for major depressive disorder: An advanced study. *Neuropsychiatric Disease and Treatment*, 13, 535–541. <https://doi.org/10.2147/NDT.S120038>
- Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1), 67–82.
- Wu, X., Li, R., Hong, Q., & Chi, X. (2022). Development and validation of a novel diagnostic model for childhood autism spectrum disorder based on ferroptosis-related genes. *Frontiers in Psychiatry*, 13, 886055. <https://doi.org/10.3389/fpsy.2022.886055>
- Wu, X., Zhu, L., Zhao, Z., Xu, B., Yang, J., Long, J., & Su, L. (2022). Application of machine learning in diagnostic value of mRNAs for bipolar disorder. *Nordic Journal of Psychiatry*, 76(2), 81–88. <https://doi.org/10.1080/08039488.2021.1937311>

- Xie, Y., Xiao, L., Chen, L., Zheng, Y., Zhang, C., & Wang, G. (2021). Integrated analysis of methylomic and transcriptomic data to identify potential diagnostic biomarkers for major depressive disorder. *Genes*, 12(2), 178. <https://doi.org/10.3390/genes12020178>
- Xiong, S., Chen, W., Jia, X. *et al.* Machine learning for prediction of asthma exacerbations among asthmatic patients: a systematic review and meta-analysis. *BMC Pulm Med* 23, 278 (2023). <https://doi.org/10.1186/s12890-023-02570-w>
- Xu, A., Kouznetsova, V. L., & Tsigelny, I. F. (2022). Alzheimer's disease diagnostics using mirna biomarkers and machine learning. *Journal of Alzheimer's Disease: JAD*, 86(2), 841–859. <https://doi.org/10.3233/JAD-215502>
- Xu, W., Liu, X., Leng, F., & Li, W. (2020). Blood-based multi-tissue gene expression inference with Bayesian ridge regression. *Bioinformatics*, 36(12), 3788–3794. <https://academic.oup.com/bioinformatics/article/36/12/3788/5819142>
- Yamagata, H., Tsunedomi, R., Kamishikiryo, T., Kobayashi, A., Seki, T., Kobayashi, M., Hagiwara, K., Yamada, N., Chen, C., Uchida, S., Ogihara, H., Hamamoto, Y., Okada, G., Fuchikami, M., Iga, J., Numata, S., Kinoshita, M., Kato, T. A., Hashimoto, R., ... Nakagawa, S. (2023). Interferon signaling and hypercytokinemia-related gene expression in the blood of antidepressant non-responders. *Heliyon*, 9(1), e13059. <https://doi.org/10.1016/j.heliyon.2023.e13059>
- Yi, Z., Li, Z., Yu, S., Yuan, C., Hong, W., Wang, Z., Cui, J., Shi, T., & Fang, Y. (2012). Blood-based gene expression profiles models for classification of subsyndromal symptomatic depression and major depressive disorder. *PLOS ONE*, 7(2), e31283. <https://doi.org/10.1371/journal.pone.0031283>
- Yu, J. S., Xue, A. Y., Redei, E. E., & Bagheri, N. (2016). A support vector machine model provides an accurate transcript-level-based diagnostic for major depressive disorder. *Translational Psychiatry*, 6(10), e931. <https://doi.org/10.1038/tp.2016.198>
- Zhang, H., Xie, Z., Yang, Y., Zhao, Y., Zhang, B., & Fang, J. (2017). The correlation-base-selection algorithm for diagnostic schizophrenia based on blood-based gene expression signatures. *BioMed Research International*, 2017, 7860506. <https://doi.org/10.1155/2017/7860506>
- Zhang-James, Y., Razavi, A. S., Hoogman, M., Franke, B., & Faraone, S. V. (2023). Machine learning and mri-based diagnostic models for adhd: Are we there yet? *Journal of Attention Disorders*, 27(4), 335–353. <https://doi.org/10.1177/10870547221146256>
- Zhao, S., Bao, Z., Zhao, X., Xu, M., Li, M. D., & Yang, Z. (2021). Identification of diagnostic markers for major depressive disorder using machine learning methods. *Frontiers in Neuroscience*, 15, 645998. <https://doi.org/10.3389/fnins.2021.645998>
- Zhao, X., Kang, J., Svetnik, V., Warden, D., Wilcock, G., David Smith, A., Savage, M. J., & Laterza, O. F. (2020). A machine learning approach to identify a circulating microRNA signature for alzheimer disease. *The Journal of Applied Laboratory Medicine*, 5(1), 15–28. <https://doi.org/10.1373/jalm.2019.029595>
- Zhou, J., Hu, Q., Wang, X., Cheng, W., Pan, C., & Xing, X. (2019). Development and validation of a novel and robust blood small nuclear RNA signature in diagnosing autism spectrum disorder. *Medicine*, 98(45), e17858. <https://doi.org/10.1097/MD.0000000000017858>
- Zhu, L., Wu, X., Xu, B., Zhao, Z., Yang, J., Long, J., & Su, L. (2021). The machine learning algorithm for the diagnosis of schizophrenia on the basis of gene expression in peripheral blood. *Neuroscience Letters*, 745, 135596. <https://doi.org/10.1016/j.neulet.2020.135596>