Psychiatry Research 278 (2019) 27–34

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Relative importance of symptoms, cognition, and other multilevel variables for psychiatric disease classifications by machine learning

Julie Walsh-Messinger[a,b,1,*], Haoran Jiang[c,1], Hyejoo Lee[d], Karen Rothman[e], Hongshik Ahn[c,1], Dolores Malaspina[f,1]

[a] Department of Psychology, University of Dayton, Dayton, OH, United States
[b] Department of Psychiatry, Wright State University Boonshoft School of Medicine, Dayton, OH, United States
[c] Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY, United States
[d] Korea Institute of Science and Technology, Seoul, Republic of Korea
[e] Department of Psychology, University of Miami, Coral Gables, FL, United States
[f] Icahn Medical School at Mount Sinai, New York, NY, United States

ARTICLE INFO

Keywords:
Machine learning
Nosology
Schizophrenia
Schizoaffective disorder
Major depressive disorder, Bipolar disorder

ABSTRACT

This study used machine-learning algorithms to make unbiased estimates of the relative importance of various multilevel data for classifying cases with schizophrenia (n = 60), schizoaffective disorder (n = 19), bipolar disorder (n = 20), unipolar depression (n = 14), and healthy controls (n = 51) into psychiatric diagnostic categories. The Random Forest machine learning algorithm, which showed best efficacy (92.9% SD: 0.06), was used to generate variable importance ranking of positive, negative, and general psychopathology symptoms, cognitive indexes, global assessment of function (GAF), and parental ages at birth for sorting participants into diagnostic categories. Symptoms were ranked most influential for separating cases from healthy controls, followed by cognition and maternal age. To separate schizophrenia/schizoaffective disorder from bipolar/unipolar depression, GAF was most influential, followed by cognition and paternal age. For classifying schizophrenia from all other psychiatric disorders, low GAF and paternal age were similarly important, followed by cognition, psychopathology and maternal age. Controls misclassified as schizophrenia cases showed lower nonverbal abilities, mild negative and general psychopathology symptoms, and younger maternal or older paternal age. The importance of symptoms for classification of cases and lower GAF for diagnosing schizophrenia, notably more important and distinct from cognition and symptoms, concurs with current practices. The high importance of parental ages is noteworthy and merits further study.

1. Introduction

The DSM-5 (American Psychiatric Association, 2013) and the ICD-11 (World Health Organization, 1992) list specific observable criteria that must be fulfilled to make categorical diagnoses. In contrast, the evolving NIMH Research Domain Criteria (RDoC) aims to identify cross-cutting domains of dysfunction that transcend different diagnostic categories and can even occur at sub-threshold levels in some persons without a psychiatric diagnosis (Bzdok and Meyer-Lindenberg, 2018). Both approaches are meaningful attempts to classify psychiatric phenomena that have enigmatic origins, but each is limited by human subjectivity, including diagnostic traditions and prevailing concepts on the neural determinants of mental illness.

Machine learning classification methods can provide objective, unbiased information to inform or complement these categorical and continuous systems. Machine learning classification procedures are hypothesis-independent approaches in which algorithms are first "learned" from a training set of data from individuals with known conditions and then applied to multimodal data to determine the relative importance of different variables for the classification. Such methods are already employed for medical decision-making that involves complex multilevel data, such as health assessments, treatment decisions and evaluation of risk factors (Buzaev et al., 2016; Senders et al., 2018; Sinha et al., 2018). With respect to schizophrenia, machine learning approaches have classified schizophrenia from bipolar disorder cases with 88% accuracy and separated schizophrenia cases from healthy participants with 90% accuracy (Schnack et al., 2014). Machine learning algorithms have also predicted good versus poor treatment

* Corresponding author at: Department of Psychology, University of Dayton, 300 College Park, Dayton, OH 45469- 1430, United States.
E-mail address: jmessinger1@udayton.edu (J. Walsh-Messinger).
[1] Shared first (Walsh-Messinger and Jiang) and senior (Ahn and Malaspina) authorship.

https://doi.org/10.1016/j.psychres.2019.03.048
Received 24 November 2018; Received in revised form 28 March 2019; Accepted 29 March 2019
Available online 29 March 2019
0165-1781/

outcome in schizophrenia (Koutsouleris et al., 2016), predicted conversion to psychosis in an ultra-high risk sample (Mechelli et al., 2017) and also differentiated first-episode psychosis from ultra-high risk cases (Pettersson-Yeo et al., 2013). It should be noted, however, that findings across machine learning studies for psychiatric diagnosis and classification are variable, and the replicability and generalizability of these models to other samples may be limited (Bzdock and Meyer-Lindenberg, 2018).

To our knowledge, no study has employed machine learning to determine the relative importance of different types of multilevel data for separating schizophrenia cases from those with other psychiatric disorders and healthy controls. The present study applied machine-learning algorithms to a large sample of rigorously assessed research participants to determine the relative importance of symptoms, cognition, past and current function and parental ages for classifying schizophrenia and other psychiatric diagnoses. We hypothesized that machine learning could be used to generate unbiased determinations of the most influential variables for classifying psychiatric disorders for research and to inform classification strategies. A secondary goal of the study was to examine the individual features of any healthy control participants that were classified as having schizophrenia by machine-learning algorithms. This information might be especially pertinent for defining which multilevel data is most relevant for schizophrenia. As there are several machine learning tools available, a preliminary analysis was first conducted to choose the method with the highest efficacy for the data.

2. Methods

2.1. Participants

Clinical research data was uniformly collected from 113 psychiatric cases and 51 healthy comparison participants between 1995 and 2010 as part of several NIMH funded research studies conducted at two public hospitals in New York City. Healthy controls did not meet criteria for a DSM-IV Axis I disorder within the past two years, were not taking any psychiatric medications, and had no family history of psychosis in any first or second-degree relative. Institutional Review Board approval was obtained for all studies and all participants provided written informed consent.

2.2. Measures

All participants underwent the same comprehensive assessments. The *Diagnostic Interview for Genetic Studies* (Nurnberger et al., 1994) was conducted by clinicians with a master's degree or higher to assess psychiatric symptoms and disorders. The machine learning algorithm was trained using final "best estimate" DSM-IV diagnoses (American Psychiatric Association, 2000), which were determined upon agreement by two senior diagnosticians at a consensus conference attended by all clinical raters. Other assessment data included parental ages at birth, current and past Global Assessment of Function (GAF), the *Positive and Negative Syndrome Scale* (PANSS; Kay et al., 1987), which was used to generate separate positive, negative and general psychopathology scale scores, and the Verbal Comprehension (VCI), Working Memory (WMI), Processing Speed (PSI) and Perceptual Organization (POI) index scores from the *Wechsler Adult Intelligence Scale, Third Edition* (WAIS-III; Wechsler, 1997).

2.3. Models and statistical analysis

The efficacy of four widely used machine-learning methods, Random Forest (RF), linear discriminant analysis (LDA), AdaBoost, and Support Vector Machine (SVM), was first examined for efficacy in classifying psychiatric cases versus healthy controls. Accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were compared across the models, and RF was determined to be the superior method overall and was therefore employed for the study. The RF method (Breiman, 2001) uses an ensemble of multiple decision trees as tools to map the variables into possible classes (see Ahn et al., 2007) for detailed methodology). The RF model consisted of 500 trees analyzed with the R statistical software package randomForest (Liaw and Wiener, 2002) using default settings. Any missing observations were imputed by the mean for continuous variables and mode for categorical variables.

The response sets were evaluated using three different groupings: 1) healthy control = 0, all psychiatric cases = 1 (schizophrenia, schizoaffective, bipolar, and major depressive disorder cases); 2) schizophrenia and schizoaffective cases = 0, affective disorder cases = 1 (bipolar and major depressive disorders); 3) schizophrenia cases = 0, other psychiatric disorder cases = 1 (schizoaffective, bipolar and major depressive disorders). For each analysis we examined the mean decrease in accuracy (MDA), which provides a measure of the decrease in accuracy of the model if the variable is not included (Breiman, 2001). A higher MDA indicates greater importance of the variable for accurate classification. We also tested the utility of the algorithms to conduct three-way and five-way classifications across all the psychiatric cases and healthy control participants.

Next, we examined which healthy controls had been misclassified as psychiatric cases by each of the four machine-learning models by randomly dividing the control group into three groups of roughly equal size: G1, G2 and G3. First, G1 and G2 plus the 113 cases formed the learning set and G3 served as the test set. Second, G1 and G3 plus the 113 cases formed the learning set and G2 served as the test set. Finally, G2 and G3 plus the 113 cases were in the learning set and G1 served as the test set. For SVM, linear kernel was used because it performed better than radial basis kernel for these data. The individual variables for each control participant classified as a case by any of the algorithms were inspected to identify which variables were consistent with schizophrenia case categorization.

3. Results

Diagnoses of the study participants included schizophrenia (n = 60), schizoaffective disorder (n = 19), bipolar disorder (n = 20), major depressive disorder (n = 14) and no psychiatric disorder (n = 51). Mean values for all variables are displayed in Table 1 for each diagnostic group. There were no significant differences in any of the clinical measures across the two data collection sites (all p's > .300); however, there were significant differences in sex (X^2 = 6.03, p = .01) and in diagnostic distribution across the two sites (see Table 2). Results presented in Table 3 demonstrate the superior efficiency of RF in comparison to the other algorithms for all models except model III, for which Adaboost was minimally stronger. Performance metrics of RF for classifying the participants into diagnostic categories is shown in Fig. 1a–c.

3.1. . Diagnostic classification

3.1.1. . All psychiatric cases versus healthy control participants

For classifying psychiatric cases from healthy control participants, the RF algorithms demonstrated an average accuracy of 92.9% (SD, 0.6) (see Table 3a and b). As depicted in Fig. 2a, the most influential measures were general psychopathology symptoms (MDA = 0.107) and positive symptoms (MDA = 0.082). There were weaker effects for negative symptoms (MDA = 0.04), and even smaller effects for cognition (WMI MDA = 0.01; PSI MDA = 0.009; POI MDA = 0.008; VCI MDA = 0.007) and younger maternal age (MDA = 0.007).

3.1.2. . Schizophrenia and schizoaffective disorder versus affective disorder cases

The RF algorithms differentiated schizophrenia and schizoaffective

Table 1
Descriptive statistics of the data for all 164 participants and grouped by best estimate research diagnoses.

	All		Control		Schizoaffective		Schizophrenia		Bipolar disorder		Major depression	
	N	*M* (SD)	*N*	*M* (SD)	*N*	*M* (SD)	*N*	*M* (SD)	*N*	*M* (SD)	*N*	*M* (SD)
Age	164	35.0 (11.0)	51	33.1 (11.6)	19	36.5 (9.3)	60	35.1 (10.7)	20	34.5 (8.9)	14	40.5 (14.0)
Age of onset	103	22.5 (7.3)	–	–	17	22.1 (7.7)	58	23.2 (6.1)	16	21.2 (8.0)	12	21.6 (10.8)
Mother's age at birth	122	27.8 (6.4)	40	29.2 (6.3)	11	24.3 (5.1)	46	27.9 (6.6)	16	27.1 (7.7)	9	26.7 (4.8)
Father's age at birth	115	31.4 (6.9)	37	32.2 (6.6)	10	27.1 (7.1)	46	32.7 (6.3)	15	30.6 (8.1)	7	27.1 (6.0)
WAIS-III Full Scale IQ	132	100.6 (17.6)	37	106.0 (13.0)	17	95.6 (19.7)	50	96.7 (20.1)	18	100.4 (13.9)	10	109.0 (15.7)
WAIS-III Working Memory	133	85.8 (31.8)	37	94.0 (31.4)	17	73.5 (36.1)	51	76.2 (33.0)	18	99.0 (16.4)	10	102.3 (17.2)
WAIS-III Verbal Comprehension	133	92.3 (34.0)	37	98.1 (31.6)	17	83.1 (42.4)	51	82.3 (36.6)	18	108.4 (16.6)	10	108.6 (11.2)
WAIS-III Perceptual Organization	131	84.6 (32.3)	37	86.5 (29.7)	17	77.0 (39.3)	50	76.6 (35.6)	18	97.0 (11.7)	10	107.3 (19.4)
WAIS-III Processing Speed	131	81.2 (34.5)	37	90.3 (32.7)	17	67.8 (37.9)	50	68.4 (35.9)	18	97.7 (14.2)	9	106.4 (14.4)
PANSS Positive Symptoms	136	10.4 (5.6)	40	7.0 (0.0)	15	12.4 (4.4)	57	12.9 (7.0)	12	8.4 (1.7)	12	9.8 (4.6)
PANSS Negative Symptoms	138	11.7 (5.5)	41	7.6 (1.6)	15	13.7 (4.9)	57	13.6 (5.8)	13	11.9 (5.8)	12	14.1 (5.9)
PANSS General Psychopathology	138	23.9 (8.8)	41	16.8 (1.7)	15	27.4 (8.1)	57	26.6 (10.0)	13	25.8 (4.2)	12	29.1 (9.0)
Current GAF	74	32.4 (19.6)	–	–	14	29.0 (12.2)	42	25.7 (14.1)	12	43.7 (24.2)	6	54.5 (17.9)
Past GAF	74	40.2 (19.1)	–	–	14	35.4 (13.8)	42	37.1 (16.7)	12	44.5 (24.0)	6	55.8 (16.9)

Table 2
Number of participants for each diagnosis across the two data collection sites.

	Site I	Site II
	N	*N*
Healthy control	44	7
Schizophrenia	44	16
Schizoaffective	14	5
Bipolar	20	0
MDD	14	0

disorder cases from unipolar and bipolar affective disorder cases with a mean accuracy of 90.3% (SD, 0.8) (Table 3a and b). Function, particularly current GAF (MDA = 0.087) but also past GAF (MDA = 0.059), was most important for dichotomizing the schizophrenias from the affective disorders (see Fig. 2b). Cognition was much less important (PSI MDA = 0.039; VCI MDA = 0.024; WMI MDA = 0.024; POI MDA = 0.012), and older paternal age (MDA = 0.014) was more important than positive symptoms (MDA = 0.010).

3.1.3. . *Schizophrenia cases versus other psychiatric disorder cases*

The mean accuracy of the RF algorithms for classifying schizophrenia cases from all other psychiatric disorder cases was 78.5% (SD, 2.2) (Table 3a and b). Current GAF (MDA = 0.057) and older paternal age (MDA = 0.055) were comparably the most influential measures for classifying schizophrenia (Fig. 2c). These effects were substantially more important than past GAF (MDA = 0.026) and several cognitive indices (PSI MDA = 0.018; VCI MDA = 0.016; POI MDA = 0.0014). Finally, general psychopathology symptoms (MDA = 0.007), younger mother's age at birth (MDA = 0.006) and working memory (MDA = 0.006) also influenced the classification of schizophrenia, although these were much less important than current GAF and older paternal age.

As GAF was highly important for classifying schizophrenia, multiple regression with missing value mean imputation was employed to evaluate the degree to which parental ages, psychiatric symptoms and cognitive indices uniquely contributed to current GAF scores in cases with schizophrenia and schizoaffective disorder ($N = 68$). Together the variables explained only 23% of the variance in current GAF. General psychopathology was the only variable to be significantly associated ($t = 2.261$, $p = .028$), with VCI marginally associated ($t = -1.981$, $p = .053$).

3.2. *More complex models*

The RF algorithms also performed well in a three-way classification partitioning healthy controls, the combined group of schizophrenias, and the affective disorder cases. Twenty repetitions of 5-inner fold cross validations (CV) had an accuracy of 90.3% (SD, 0.8). A five-way classification of the participants to partition healthy controls, schizophrenia, schizoaffective, bipolar and major depression cases using RF yielded an accuracy of 70.9% (SD, 1.0). SVM also performed well in these multiple classification schemes, although there were poor results from LDA and AdaBoost approaches.

3.2.1. *Qualitative descriptions of misclassified healthy control participants*

Some healthy control participants were classified as psychiatric cases by the algorithms, including 10 controls by RF, 8 controls by SVM, 24 controls by LDA and 16 controls by AdaBoost with some overlap (Table 4). Five of the 51 controls were classified as psychiatric cases by all four models and four controls were classified as cases by three of the models. Of these nine controls, four females and five males, seven who were classified as cases by RF were then further evaluated using a three-group model (schizophrenia vs. all other psychiatric disorder vs. controls) and a five-group model (schizophrenia vs. schizoaffective vs. depression vs. bipolar vs. control), as shown in the center and right columns in Table 4.

The RF approach classified three of the seven controls (A, B and E) as schizophrenia cases in both the 3-group and 5-group analyses. The first, control A (6), was also classified as schizophrenia by RF and SVM, and as schizoaffective disorder by LDA. Control A had a young mother (16 years), was divorced and unemployed, had earned an associate degree, was of average intelligence, and displayed slight (minimal) elevations in negative symptoms and general psychopathology. The second (control B), classified as schizophrenia by three models (RF, SVM and LDA) in the three-way comparison and all four models in the four-way comparison, was married, employed part-time, had average intelligence but had only earned a high school degree, and demonstrated slight (minimal) elevations in negative symptoms and general psychopathology. Control E, classified as schizophrenia by RF, SVM, and LDA in the three-way and four-way comparisons, had an older father (52 years), was single, unemployed, had completed some college courses but not earned a degree, displayed a mild elevation in negative symptoms, and had low average PSI and POI scores, which were notably weaker than his average VCI and WMI scores. No control classified as an “other psychiatric disorder” by RF in the 3-way analysis was subsequently classified as schizophrenia in the 5-way analysis.

Regarding the other algorithms, SVM classified two other healthy controls (C, D) as schizophrenia cases. Control C was single, earned a bachelor's degree, was unemployed, had moderately high negative symptoms and Average VCI, WMI, and PSI, but Low Average POI. In contrast, Control D had a younger mother (22 years), was single, employed full-time, had completed some college courses but had not earned a degree, and showed substantial scatter in her performance on the WAIS-III indices: her VCI was in the superior, POI was high average,

Table 3
The efficacy of the four machine-learning algorithms and performance of Random Forest (RF).

3a. Accuracy of the four machine-learning algorithms for diagnostic classification across the three classification groups models

	RF mean (SD)	SVM mean (SD)	LDA mean (SD)	Adaboost mean (SD)
Model I: All psychiatric cases vs healthy controls	0.919 (0.010)	0.913 (0.014)	0.837 (0.022)	0.892 (0.013)
Model II: Schizophrenia and schizoaffective disorder cases vs affective disorder cases	0.902 (0.010)	0.879 (0.019)	0.873 (0.020)	0.885 (0.017)
Model III: Schizophrenia cases vs all other psychiatric disorder cases	0.778 (0.022)	0.753 (0.031)	0.754 (0.024)	0.808 (0.022)

3b. Performance of RF algorithms for diagnostic classification across the three classification groups models

	Sensitivity mean (SD)	Specificity mean (SD)	Balanced accuracy mean (SD)	PPV mean (SD)	NPV mean (SD)
Model I: All psychiatric cases vs healthy controls	0.939 (0.013)	0.875 (0.018)	0.907 (0.012)	0.943 (0.007)	0.868 (0.023)
Model II: Schizophrenia and schizoaffective disorder cases vs affective disorder cases	0.900 (0.007)	0.916 (0.028)	0.908 (0.015)	0.968 (0.011)	0.759 (0.026)
Model III: Schizophrenia cases vs all other psychiatric disorder cases	0.781 (0.019)	0.793 (0.037)	0.787 (0.021)	0.823 (0.034)	0.732 (0.032)

Fig. 1. (a) ROC curve: classification algorithms into cases and control, (b) ROC curve: classification algorithms into schizophrenias (Schizophrenia & Schizoaffective) and affective disorders (Bipolar & MDD), (c) ROC curve: classification algorithms into schizophrenia and other psychiatric disorders.

PSI was average, and WMI was low average. Adaboost classified two other healthy controls (F, H) as schizophrenia cases, who were educated with a Bachelor's (F) or graduate (H) degree but evidenced a mild elevation in general psychopathology combined with Low Average PSI and POI but Average VCI and WMI. Control F was single and employed part-time while control H was divorced and retired. Adaboost classified another healthy control as a bipolar case (I).

No single profile accounted for these participants' inclusions as cases. None had any positive symptoms whereas most had mild negative symptoms and general psychopathology. Almost all had average or better VCI and WMI, but four had low average POI and three had low average PSI. One had a 16 year-old mother and another's father was 52 years at birth. More than half of the controls classified as cases did not

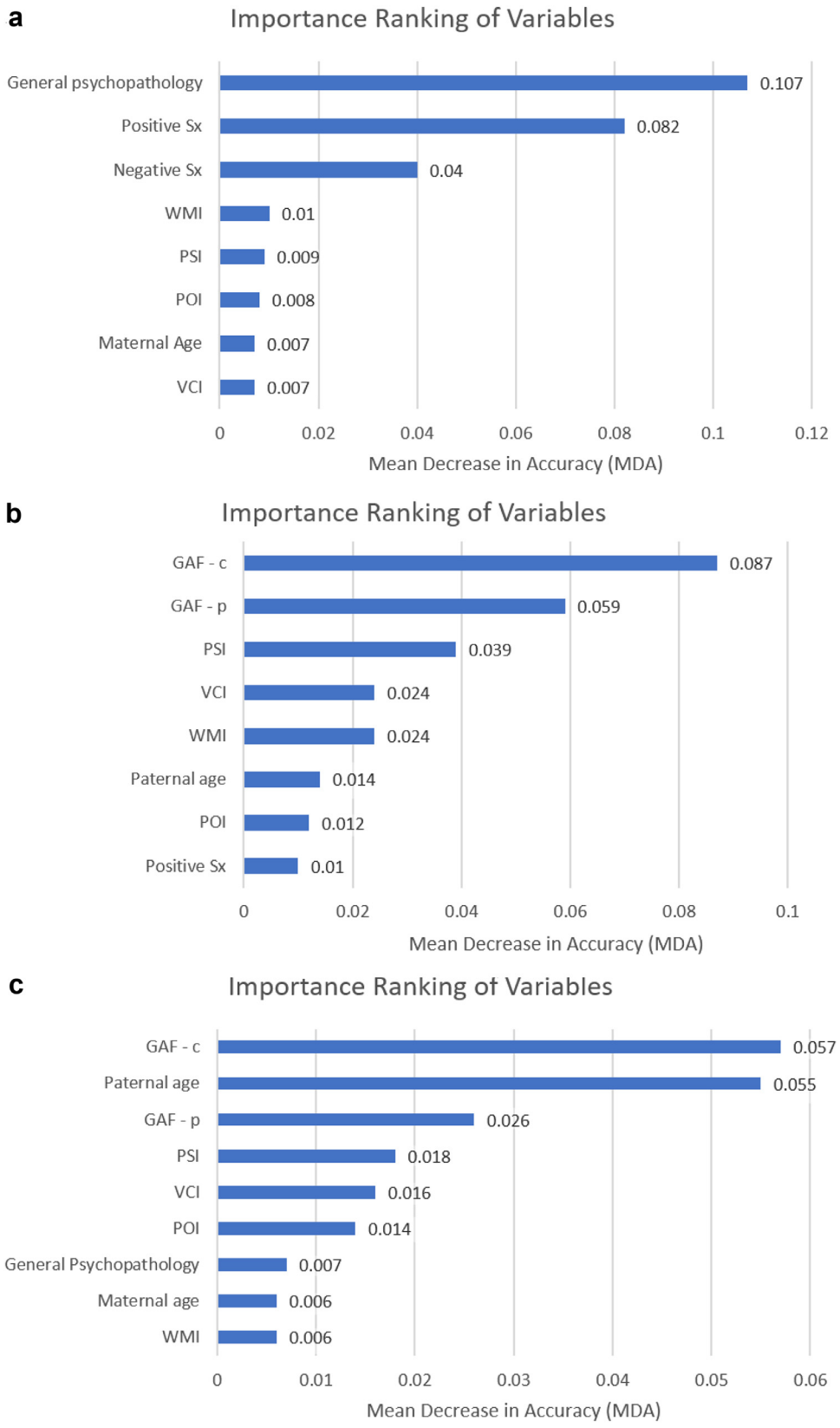

Fig. 2. Variable importance ranking and mean decrease in accuracy (MDA) for the Random Forest (RF) model, (a) classifying psychiatric cases from healthy controls, (b) classifying cases with schizophrenia and schizoaffective disorder from affective disorder cases, (c) classifying schizophrenia cases from cases with other psychiatric disorders

GAF-c = Current Global Assessment of Function; GAF-p = Past Global Assessment of Function; WAIS III Indexes: VCI = Verbal Comprehension Index; WMI = Working Memory Index; POI = Perceptual Orientation Index; PSI = Processing Speed Index.

Table 4
Participants in the healthy control group predicted to be cases by the respective classification methods: Random Forest (RF), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and AdaBoost.

ID	Two classes Classified as a psychiatric case by 3 or 4 of the algorithms	Three classes Classified as schizophrenia vs. another psychiatric disorder vs. control subject	Five classes Classified as schizophrenia vs. schizoaffective vs. depression vs. bipolar vs. control subject
A	4 methods (RF, SVM, LDA, AdaBoost)	Schizophrenia (RF, SVM, LDA)	Schizophrenia (RF, SVM) Schizoaffective (LDA)
B	4 methods (RF, SVM, LDA, AdaBoost)	Schizophrenia (RF, SVM, LDA) Other Psychiatric Disorder (AdaBoost)	Schizophrenia (RF, SVM, LDA, AdaBoost)
C	3 methods (SVM, LDA, AdaBoost)	Schizophrenia (SVM)	Schizophrenia (SVM)
D	4 methods (RF, SVM, LDA, AdaBoost)	Other Psychiatric Disorder (RF, SVM, LDA)	Schizophrenia (SVM) Schizoaffective (LDA)
E	3 methods (SVM, LDA, AdaBoost)	Schizophrenia (RF, SVM, LDA)	Schizophrenia (RF, SVM, LDA)
F	3 methods (RF, LDA, AdaBoost)	Other Psychiatric Disorder (RF, LDA, AdaBoost)	Schizophrenia (AdaBoost)
G	3 methods (RF, LDA, AdaBoost)	Control	Control
H	4 methods (RF, SVM, LDA, AdaBoost)	Other Psychiatric Disorder (RF, SVM, AdaBoost)	Schizophrenia (AdaBoost)
I	4 methods (RF, SVM, LDA, AdaBoost)	Other Psychiatric Disorder (RF, SVM, AdaBoost)	Bipolar (LDA, AdaBoost)

complete a four-year college degree, only one was employed full-time, and another was presumably employed prior to early retirement.

4. Discussion

Hypothesis-free machine-learning algorithms identified symptoms as the most influential variables for classifying psychiatric cases from controls, ranking cognition as much less important and of similar influence as young maternal age. While this supports current diagnostic practices, we had considered the possibility that cognitive measures might be of similar or greater importance. In contrast, it was current and past global assessment of function scores (GAF) that best dichotomized the group of schizophrenias from affective disorders. Cognitive measures and paternal age were less important, but still more influential than psychotic symptoms. This underlines the importance of not misdiagnosis a condition as schizophrenia based on psychosis, which can be a feature of many disorders. Low GAF scores remained valuable for the specific classification of schizophrenia, apart from schizoaffective disorder and other conditions, but paternal age was similarly influential. Less highly ranked, although still influential, were verbal comprehension, perceptual organization, and processing speed followed by general psychopathology, maternal age and working memory.

Inspection of the characteristics of the seven healthy controls who were misclassified as schizophrenia cases illuminated sub-threshold negative symptoms and a cognitive pattern of intact (average or superior) verbal comprehension with low average scores in perceptual organization, processing speed or working memory. One had a younger mother (16 years) and another had an older father (52 years). Negative symptoms and nonverbal deficits indices may represent the most salient sub-threshold symptoms in healthy subjects for schizophrenia vulnerability. The relative sparing of verbal comprehension in most controls who were classified as cases suggests that verbal comprehension deficits may be a key vulnerability for the expression of schizophrenia. Indeed, the relative importance of verbal comprehension and processing speed over other cognitive indices in classifying schizophrenia is consistent with this contention. Such importance of impaired verbal comprehension for classifying schizophrenia parallels the findings from another machine learning study that found language usage, assessed by automated algorithms, was highly accurate in predicting the conversion to psychosis among high risk youth (Corcoran et al., 2018). Global assessments of function scores were not obtained for the controls, so function did not contribute to the misclassification of controls as cases. Nonetheless, while education and employment were not included as variables, the controls who were misclassified as cases tended to have less education and low employment compared to the other controls, also consistent with schizophrenia vulnerability.

In part, these results confirm the efficacy of the current symptom-based clinical practices that employ observable criteria, such as the DSM-5 (American Psychiatric Association, 2013), for psychiatric classification. The preeminent importance of lower current and past global functioning for separating schizophrenia-related disorders from affective disorders also confirms the necessity of the DSM-5 requirement of impaired function (B Criteria) for a schizophrenia diagnosis. Importantly, while general psychopathology was correlated with current GAF scores, more than three-quarters of the variance in current GAF was not explained by the other study variables, suggesting that functional incapacity is not simply a consequence of symptoms and/or cognitive deficits. Rather, reduced function may be related to a core pathology that could also give rise to symptoms and cognitive deficits.

These findings suggest that current and past function may best capture the “fundamental features” of the disease, hypothesized by Bleuler to include autism, changes in the structure of the person, ambivalence, avolition, thought disorder and the affective-emotional and affect-expressive changes that were distinct from organic dementias (Bleuler, 1950/1908; Urfer-Parnas et al., 2010). The central importance of function for schizophrenia pathology is further supported by a recent clinical high-risk study that found a significant drop in social function occurred in the clinical high-risk cases who converted to psychosis compared to cases that did not convert within two years, along with a similar trend-level decline for their role function (Carrión et al., 2018).

Cognitive deficits are common in schizophrenia and have been considered as possible explanations for the disease (Jirsaraie et al., 2018) although many persons with far greater cognitive limitations have spared social functioning. If indeed cognitive impairments are secondary to a latent fundamental lesion, then interventions designed to improve cognition may not target the central deterministic pathology that integrates and activates social and affective drives to function in society. Importantly, psychotic symptoms, which guide most clinicians to the diagnosis, were least important for the classification.

Of the cognitive measures, working memory was the most important one for differentiating all psychiatric cases from controls, whereas processing speed and verbal comprehension were most important for differentiating schizophrenia from other disorders. This finding supports the transdiagnostic importance of working memory as a risk factor for psychopathology in general (Huang-Pollock et al., 2017), perhaps tapping a brain activation abnormality that is relevant to multiple disorders as a final pathway (McTeague et al., 2017). Supporting our results, processing speed, followed by social cognition, also best distinguished persons with schizophrenia from community residents using the MATRICS Consensus Cognitive Battery (MCCB;

Kern et al., 2011), which was not included in the present study.

The unbiased algorithm in this study showed older paternal age was similar in importance to low function for classifying schizophrenia. Given the inclusion of symptoms, cognition and function as variables for the algorithms, the separate importance of parental ages is intriguing. Advancing paternal age is an established risk factor for schizophrenia, proposed to be associated with *de novo* genetic changes in the paternal germ line that advantage the affected clone of spermatogonia to divide more rapidly, thus expanding in proportion to other spermatogonia as paternal age advances (Goriely et al., 2013). Younger maternal age is also consistently linked to the risk for psychiatric conditions in offspring (McGrath et al., 2014). Earlier papers proposed that offspring risks for men who reproduced later and women who did so early could reflect the inheritance of the behavioral deficits of impulsive mothers or social deficiencies of men with delayed reproduction (Sandin et al., 2016). However, these results show parental age effects are separate from cognition and symptoms, implicating other biological pathways. The ages of parents might, for example, epigenetically influence offspring behavior to introduce useful variation among siblings for group survival (Eisenberg and Kuzawa, 2018), which could be prone to disruption at the extremes of parental ages.

The main output for this study was the variable importance rankings, which reflect the proportion of observations that would be incorrectly classified were not included. By comparing several different machine-learning classification models, Random Forest (RF) was shown to be superior in terms of overall accuracy for classifying cases from controls (93%). The superiority of RF is attributable to its use of an ensemble of multiple decision trees as tools that map the variables into possible classes, which is more accurate than the other methods, which are based on individual classifier models (Kohavi,1995). While RF also demonstrated a 90% accuracy in dichotomizing the schizophrenias from affective disorders, it was less accurate for classifying schizophrenia *per se* (79%). This lesser accuracy of the RF algorithm for classifying schizophrenia mirrors the real-world findings of low reliability for schizophrenia and schizoaffective disorder diagnoses based on the clinical instrument utilized for the present study, the Diagnostic Interview for Genetic Studies, particularly when the conditions are separately considered (Faraone et al., 1996).

A final strength of the present study is the misclassification of seven healthy participants as cases by three or four of the different machine-learning algorithms. These results provide a qualitative picture of subthreshold illness features in some healthy controls and suggest the possibility that spared verbal capacity is protective, which are goals of the RDoC approach (Kozak and Cuthbert, 2016). In addition, as half of these controls were less than 30 years of age, they are potentially still at risk for the disorder, further supporting strengths of machine learning algorithms.

There are some limitations to the present study. While participants across the sites did not differ in age or clinical measures, there were significant sex and diagnostic distribution differences that could have influenced our results. Follow up studies are needed to verify the present results, and the addition of genomic information and life exposures would provide further information about key features that differentiate schizophrenia. Additionally, due to the small sample size of misclassified controls, supporting the efficacy of the machine learning algorithms, we relied on qualitative descriptions to understand why the algorithms may have misclassified the controls. Our qualitative findings should be considered preliminary and followed up with quantitative analyses in larger samples. Still, the results of the present study demonstrate that machine-learning methods can provide a novel window on salient domains and etiologies for psychiatric conditions and support the utility of current criteria based diagnostic practices.

Funding

Hongshik Ahn: Ministry of Science, ICT and Future Planning, Korea, under the “ICT Consilience Creative Program” (IITP-2017-R0346-15-1007) supervised by the Institute for Information & Communications Technology Promotion. Dolores Malaspina: NIMH R01MH066428; NIMH RC1MH088843; NIMH K24MH001699.

Conflict of interest

None to report.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:https://doi.org/10.1016/j.psychres.2019.03.048.

References

Ahn, H., Moon, H., Fazzari, M.J., Lim, N., Chen, J.J., Kodell, R.L., 2007. Classification by ensembles from random partitions of high-dimensional data. Comput. Stat. Data Anal. 51 (12), 6166–6179.

American Psychiatric Association, 2000. Diagnostic and Statistical Manual of Mental Disorders, fifthth edition. American Psychiatric Association, Washington, DC Text Revision.

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, fifth ed. American Psychiatric Association, Washington, DC.

Bleuler, E., 1950/1908. Dementia Praecox or the Group of Schizophrenias. International Universities Press, New York, NY J. Zinkin, TransOriginal work published 1908.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Buzaev, I.V., Plechev, V.V., Nikolaeva, I.E., Galimova, R.M., 2016. Artificial intelligence: neural network model as the multidisciplinary team member in clinical decision support to avoid medical mistakes. Chronic Dis. Transl. Med. 2 (3), 166–172.

Bzdok, D., Meyer-Lindenberg, A., 2018. Machine learning for precision psychiatry: opportunities and challenges. Biol. Psychiatry Cognit. Neurosci. Neuroimaging 3 (3), 223–230.

Carrión, R.E., Auther, A.M., McLaughlin, D., Olsen, R., Addington, J., Bearden, C.E., et al., 2018. The global functioning: social and role scales—further validation in a large sample of adolescents and young adults at clinical high risk for psychosis. Schizophr. Bull. https://doi.org/10.1093/schbul/sby126. [Epub ahead of print].

Corcoran, C.M., Carrillo, F., Fernández-Slezak, D., Bedi, G., Klim, C., Javitt, D.C., Bearden, C.E., Cecchi, G.A., 2018. Prediction of psychosis across protocols and risk cohorts using automated language analysis. World Psychiatry 17 (1), 67–75.

Eisenberg, D.T.A., Kuzawa, C.W., 2018. The paternal age at conception effect on offspring telomere length: mechanistic, comparative and adaptive perspectives. Philos. Trans. R. Soc. Lond. B Biol. Sci. 373 (1741), 20160442.

Faraone, S.V., Blehar, M., Pepple, J., Moldin, S.O., Norton, J., Nurnberger, J.I., Malaspina, D., Kaufmann, C.A., Reich, T., Cloninger, C.R., DePaulo, J.R., Berg, K., Gershon, E.S., Kirch, D.G., Tsuang, M.T., 1996. Diagnostic accuracy and confusability analyses: an application to the Diagnostic Interview for Genetic Studies. Psychol. Med. 26 (2), 401–410.

Goriely, A., McGrath, J., Hultman, C., Wilkie, A., Malaspina, D., 2013. 'Selfish spermatogonial selection': a novel mechanism for the association between advanced paternal age and neurodevelopmental disorders. Am. J. Psychiatry 170 (6), 599–608.

Huang-Pollock, C., Shapiro, Z., Galloway-Long, H., Weigard, A., 2017. Is Poor Working Memory a Transdiagnostic Risk Factor for Psychopathology? J. Abnorm. Child Psychol. 45 (8), 1477–1490.

Jirsaraie, R.J., Sheffield, J.M., Barch, D.M., 2018. Neural correlates of global and specific cognitive deficits in schizophrenia. Schizophr. Res. 201, 237–242.

Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale, (PANSS) for schizophrenia. Schizophr. Bull. 13 (2), 261–276.

Kern, R.S., Gold, J.M., Dickinson, D., Green, M.F., Nuechterlein, K.H., Baade, L.E., Keefe, R.S., Mesholam-Gately, R.I., Seidman, L.J., Lee, C., Sugar, C.A., Marder, S.R., 2011. The MCCB impairment profile for schizophrenia outpatients: results from the MATRICS psychometric and standardization study. Schizophr. Res. 126 (1–3), 124–131.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Mellish, C.S. (Ed.), Proceedings of the 14th International Joint Conference on Artificial Intelligence. 6. pp. 1137–1143.

Koutsouleris, N., Kahn, R.S., Chekroud, A.M., Leucht, S., Falkai, P., Wobrock, T., Derks, E.M., Fleischhacker, W.W., Hasan, A., 2016. Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. Lancet Psychiatry 3 (10), 935–946.

Kozak, M.J., Cuthbert, B.N., 2016. The NIMH research domain criteria initiative: background, issues, and pragmatics. Psychophysiology 53 (3), 286–297.

Liaw, A., Wiener, M., 2002. Classification and regression by randomforest. R News 2 (3), 18–22.

McGrath, J.J., Petersen, L., Agerbo, E., Mors, O., Mortensen, P.B., Pedersen, C.B., 2014. A comprehensive assessment of parental age and psychiatric disorders. JAMA Psychiatry 71 (3), 301–309.

McTeague, L.M., Huemer, J., Carreon, D.M., Jiang, Y., Eickhoff, S.B., Etkin, A., 2017. Identification of common neural circuit disruptions in cognitive control across

psychiatric disorders. Am. J. Psychiatry 174 (7), 676–685.
Mechelli, A., Lin, A., Wood, S., McGorry, P., Amminger, P., Tognin, S., McGuire, P., Young, J., Nelson, B., Yung, A., 2017. Using clinical information to make individualized prognostic predictions in people at ultra high risk for psychosis. Schizophr. Res. 184, 32–38.
Nurnberger, J.I., Blehar, M.C., Kaufmann, C.A., York-Cooler, C., Simpson, S.G., Harkavy-Friedman, J., Severe, J.B., Malaspina, D., Reich, T., 1994. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genet. Initiative Arch. Gen. Psychiatry 51 (11), 849–859.
Pettersson-Yeo, W., Benetti, S., Marquand, A.F., Dell'acqua, F., Williams, S.C., Allen, P., Prata, D., McGuire, P., Mechelli, A., 2013. Using genetic, cognitive and multi-modal neuroimaging data to identify ultra-high-risk and first-episode psychosis at the individual level. Psychol. Med. 43 (12), 2547–2562.
Sandin, S., Schendel, D., Magnusson, P., Hultman, C., Surén, P., Susser, E., Grønborg, T., Gissler, M., Gunnes, N., Gross, R., Henning, M., Bresnahan, M., Sourander, A., Hornig, M., Carter, K., Francis, R., Parner, E., Leonard, H., Rosanoff, M., Stoltenberg, C., Reichenberg, A., 2016. Autism risk associated with parental age and with increasing difference in age between the parents. Mol. Psychiatry 21 (5), 693–700.
Schnack, H.G., Nieuwenhuis, M., van Haren, N.E., Abramovic, L., Scheewe, T.W., Brouwer, R.M., Hulshoff Pol, H.E., Kahn, R.S., 2014. Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. Neuroimage 84, 299–306.
Senders, J.T., Staples, P.C., Karhade, A.V., Zaki, M.M., Gormley, W.B., Broekman, M.L.D., Smith, T.R., Arnaout, O., 2018. Machine learning and neurosurgical outcome prediction: a systematic review. World Neurosurg 109, 476–486.
Sinha, M., Jupe, J., Mack, H., Coleman, T.P., Lawrence, S.M., Fraley, S.I., 2018. Emerging technologies for molecular diagnosis of sepsis. Clin. Microbiol. Rev. 31 (2) e00089-17.
Urfer-Parnas, A., Mortensen, E.L., Parnas, J., 2010. Core of schizophrenia: estrangement, dementia or neurocognitive disorder? Psychopathology 43 (5), 300–311.
Wechsler, D., 1997. Wechsler Adult Intelligence Scale, Third Edition, WAIS-III): Administration and Scoring Manual. The Psychological Corporation, San Antonio, TX.
World Health Organization, 1992. The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. World Health Organization, Geneva.