

ARTICLE

Open Access

Machine learning classification of ADHD and HC by multimodal serotonergic data

A. Kautzky¹, T. Vanicek¹, C. Philippe², G. S. Kranz^{1,3}, W. Wadsak^{1,3,4}, M. Mitterhauser^{2,5}, A. Hartmann⁶, A. Hahn¹, M. Hacker², D. Rujescu⁶, S. Kasper¹ and R. Lanzenberger¹

Abstract

Serotonin neurotransmission may impact the etiology and pathology of attention-deficit and hyperactivity disorder (ADHD), partly mediated through single nucleotide polymorphisms (SNPs). We propose a multivariate, genetic and positron emission tomography (PET) imaging classification model for ADHD and healthy controls (HC). Sixteen patients with ADHD and 22 HC were scanned by PET to measure serotonin transporter (SERT) binding potential with [¹¹C] DASB. All subjects were genotyped for thirty SNPs within the *HTR1A*, *HTR1B*, *HTR2A* and *TPH2* genes. Cortical and subcortical regions of interest (ROI) were defined and random forest (RF) machine learning was used for feature selection and classification in a five-fold cross-validation model with ten repeats. Variable selection highlighted the ROI posterior cingulate gyrus, cuneus, precuneus, pre-, para- and postcentral gyri as well as the SNPs *HTR2A* rs1328684 and rs6311 and *HTR1B* rs130058 as most discriminative between ADHD and HC status. The mean accuracy for the validation sets across repeats was 0.82 (± 0.09) with balanced sensitivity and specificity of 0.75 and 0.86, respectively. With a prediction accuracy above 0.8, the findings underlying the proposed model advocate the relevance of the SERT as well as the *HTR1B* and *HTR2A* genes in ADHD and hint towards disease-specific effects. Regarding the high rates of comorbidities and difficult differential diagnosis especially for ADHD, a reliable computer-aided diagnostic tool for disorders anchored in the serotonergic system will support clinical decisions.

Introduction

The most common neurodevelopmental disorder, attention-deficit and hyperactivity disorder (ADHD), affects up to 10% of children with symptoms often persisting throughout the whole lifespan and predisposes to comorbidities like major depressive disorder (MDD)¹. However, substantial fluctuation of prevalence was reported between and across nations, likely owed to disputed diagnostic criteria that are mostly based on behavioral symptoms rather than objective biomarkers. While pathognomonic for many psychiatric disorders, the lack of biomarkers for ADHD is particularly baneful due to the overlap of core symptoms with other frequent psychiatric

disorders as mood, anxiety and personality disorders. Diagnosis of adult ADHD is further hindered by retrospective assessment of symptoms in the childhood. Disputes among opinion leaders on ADHD and debates over misuse of ADHD treatment like methylphenidate (MPH) have encouraged research exploring objective over subjective ADHD predictors, so far with modest success^{2,3}.

Genetics were expected to resolve disparate findings and explain heterogeneity, especially in ADHD with heritability estimated to exceed 70%^{4,5}. Some candidate gene studies associated variants implicated in the monoaminergic neurotransmission with ADHD^{6,7}, while the GWAS mostly highlighted genes that previously received less attention and are trickier to mesh with established etiologic theories⁸. However, genetic studies did not impact ADHD diagnosis or treatment yet⁹. Consequently, the translation to the clinic is lacking so far.

Neuroimaging and data-driven diagnostics that naturally come along with it were considered a corrective to

Correspondence: R. Lanzenberger (rupert.lanzenberger@meduniwien.ac.at)

¹Department of Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria

²Department of Biomedical Imaging and Image-guided Therapy, Division of Nuclear Medicine, Medical University of Vienna, Vienna, Austria

Full list of author information is available at the end of the article

© The Author(s) 2020



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

the issues of subjective symptoms and heterogeneity. Basic tools like electro-encephalography (EEG) as well as more advanced techniques as magnetic resonance imaging (MRI) have now been fully established in ADHD research¹⁰. While misbalance of dopaminergic and noradrenergic neurotransmission is putatively the main biological substrate of ADHD, data on in vivo neuroreceptor binding are scarce due to the resource intensive nature of positron emission tomography (PET). While serotonergic neurotransmission is considered a pivotal substrate of affective disorders, the role of serotonin is not sufficiently understood in ADHD^{11,12}. Data from animal models as well as pharmacological and genetic studies point toward involvement of serotonin in ADHD and atomoxetine, a well-established drug in ADHD treatment, has been demonstrated to block the serotonin transporter in addition to its noradrenergic properties¹³. Emotional dysregulation with mood swings and irritability, closely linked to serotonergic pathways, has lately been discussed as an additional core symptom of ADHD^{14,15}. Additionally, comorbid mood disorders are frequent in ADHD. Nevertheless, only few PET studies have targeted the serotonin system in ADHD so far. Earlier studies on the serotone transporter (SERT) binding did not support differences while recently altered interregional connectivity of SERT binding in the hippocampus and precuneus of ADHD patients compared to control subjects was demonstrated¹⁶.

With the influx of advanced statistics into neuropsychiatric research, a copious amount of machine learning studies targeted ADHD classification. Algorithms based on EEG and MRI features reported accuracies ranging from hardly above chance level to beyond 90%¹⁰. However, no studies combined imaging and genetic predictors up to this point. A reliable diagnostic tool for ADHD may be especially relevant to precision medicine in psychiatry. Since serotonergic transmission has to some extent been demonstrated to pilot the disorders, the focus of this study was classification of ADHD and healthy individuals based on multimodal serotonergic data.

Methods

Subjects

ADHD subjects derive from a previously reported study on SERT binding measured with [¹¹C]DASB¹⁶.

In short, 16 patients with adult ADHD (aged 31.9 ± 10.9 standard deviation (SD), seven females) were recruited through the outpatient clinic for ADHD and affective disorders at the Department of Psychiatry and Psychotherapy, Medical University of Vienna. Twenty-two healthy control subjects (aged 33.19 ± 10.3 SD, nine females) were recruited through advertisement at the Department of Psychiatry and Psychotherapy. ADHD patients were required to be free of neuropsychiatric

medication for at least three months. None of the HC were previously exposed to any psychopharmacologic treatment. All study related procedures were approved by the Ethics Committee of the Medical University of Vienna. All participants consented in written form to partake in the study after extensive explanation of the study protocol.

Subjects were screened for any somatic or neurological disorder by assessment of physical and neurological status, laboratory tests including urine drug and pregnancy tests and electrocardiography. Comorbid psychiatric disorders were assessed with the structured clinical interview for DSM-IV (SCID-I, SCID-II). Subjects with severe comorbidities or any substance abuse or addiction other than nicotine were excluded. ADHD symptomatology was evaluated by Conners' Adult ADHD Diagnostic Interview (CAADID, Conners 1999).

Genotyping procedures

Genotyping protocols were published previously, please see for details ref.¹⁷. In summary, EDTA blood tubes of 9 ml were collected and the QiaAmp DNA blood maxi kit was applied for DNA isolation (Qiagen, Hilden, Germany). The iPLEX assay was used for genotyping on a mass spectrometer (MassARRAY MALDI-TOF). Typer 3.4 (Sequenom, San Diego, CA, USA) was utilized for genotype assignment after selection of the allele-specific extension products. Quality control required to surpass a threshold of 80% individual and 99% SNP call rate identity of genotyped CEU trios (Coriell Institute for Medical research, Camden, NJ).

Thirty SNPs of four genetic key players of the serotonergic system, the *HTR1A*, *HTR1B*, *HTR2A* and *TPH2* genes, were selected for this analysis based on the literature. All SNPs were coded numerically for the number of minor alleles, therefore ranging from 0 to 2. SNPs were determined based on the literature. For an overview of baseline characteristics, including genotypes (Table 1).

PET data acquisition

All PET and MRI scans were carried out at the Department of Biomedical Imaging and Image-guided Therapy, Division of Nuclear Medicine, Medical University of Vienna. A full-ring scanner (General Electric Medical Systems, Milwaukee, WI, USA) in 3D acquisition mode was used. For all subjects, the state-of-the-art radiotracer [¹¹C]DASB was used to quantify SERT binding as protocolled previously¹⁸. In summary, for tissue attenuation correction a transmission scan was obtained for five minutes with retractable 68 Ge rod sources. Data acquisition of the actual scan started with a bolus i.v.-injection of [¹¹C]DASB. A series of 50 consecutive time frames (12 × 5 s, 6 × 10 s, 3 × 20 s, 6 × 30 s, 4 × 1 min, 5 × 2 min, 14 × 5 min) was carried out, resulting

Table 1 Baseline characteristics with sex, age and genotypes for the total sample, HC and ADHD.

<i>n</i>	Total, 38	HC, 22	ADHD, 16
Age (mean ± SD)	32.51 ± 10.5	33.19 ± 10.3	31.9 ± 10.9
Sex (female/male)	16/22	9/13	7/9
Mean Global BP _{ND}	0.37 ± 0.36	0.37 ± 0.36	0.36 ± 0.35
<i>HTR1A</i> rs6295	13/16/9	6/10/6	3/6/7
<i>HTR1A</i> rs878567	13/16/9	6/10/6	7/6/3
<i>HTR1A</i> rs1423691	13/16/9	6/10/6	7/6/3
<i>HTR1A</i> rs10042486	13/16/9	6/10/6	3/6/7
<i>HTR1B</i> rs6296	4/18/16	2/10/10	2/8/6
<i>HTR1B</i> rs6298	16/18/4	10/10/2	6/8/2
<i>HTR1B</i> rs130058	21/15/2	16/4/2	5/11/0
<i>HTR1B</i> rs13212041	25/12/1	13/8/1	12/4/0
<i>HTR2A</i> rs6311	11/18/9	7/13/2	4/5/7
<i>HTR2A</i> rs6313	11/18/9	7/13/2	4/5/7
<i>HTR2A</i> rs1328684	16/20/2	4/16/2	12/4/0
<i>HTR2A</i> rs1923886	10/19/9	5/10/7	5/9/2
<i>HTR2A</i> rs2224721	17/16/5	8/10/4	9/6/1
<i>HTR2A</i> rs2770296	23/13/2	12/9/1	11/4/1
<i>HTR2A</i> rs6561332	9/20/9	4/11/7	5/9/2
<i>HTR2A</i> rs6561333	13/18/2	10/6/1	3/12/1
<i>HTR2A</i> rs7322347	11/22/5	7/12/3	4/10/2
<i>HTR2A</i> rs7984966	21/17/0	13/9/0	8/8/0
<i>HTR2A</i> rs7997012	13/19/6	7/11/4	6/8/2
<i>HTR2A</i> rs9316233	18/13/7	8/9/5	10/4/2
<i>HTR2A</i> rs9534495	9/18/11	5/10/7	4/8/4
<i>TPH2</i> rs1386493	28/8/2	16/5/1	12/3/1
<i>TPH2</i> rs1386497	29/7/2	16/5/1	13/2/1
<i>TPH2</i> rs1487275	22/12/4	11/8/3	11/4/1
<i>TPH2</i> rs1487278	26/10/2	16/4/2	10/6/0
<i>TPH2</i> rs1843809	29/7/2	15/6/1	14/1/1
<i>TPH2</i> rs4570625	25/12/1	17/5/0	8/7/1
<i>TPH2</i> rs7305115	13/19/6	9/9/4	4/10/2
<i>TPH2</i> rs10879352	17/14/7	10/7/5	7/7/2

For each single nucleotide polymorphism counts for major allele homozygotes, heterozygotes and minor allele homozygotes are provided (in this order). *ADHD* attention-deficit and hyperactivity disorder, *MDD* major depressive disorder, *HC* healthy control, *SD* standard deviation, *BP_{ND}* non-displaceable binding potential.

in a measurement time of 90 min in total. FORE-ITER, an iterative filtered back-projection algorithm, was used for reconstructing the measured data in volumes of 35 transaxial sections (128 × 128 matrix). For this step, the

spatial resolution was at 4.36 mm full-width at half maximum 1 cm next to the center of the field of view.

SERT quantification

The protocol for data quantification was reported previously, including preprocessing carried out in SPM12 (Wellcome Trust Centre for Neuroimaging, London, UK; <http://www.fil.ion.ucl.ac.uk/spm/>)¹⁶. In summary, the means of all time frames without visually observable head motion was used for realignment of each time frame of the dynamic PET scans. All subjects also underwent MRI scans on a 3 Tesla Philips scanner (Achieva, 3D T1 FFE weighted sequence, 0.88 mm slice thickness, 0.8 × 0.8 mm in-plane resolution). Summed PET images (integral) from realigned data was co-registered to T1-weighted images. Next, spatial normalization of the T1-weighted images was performed. Transformation of the co-registered PET images into MNI standard space was achieved by application of the obtained transformation matrices to the dynamic PET data. Finally, computation of voxel-wise images of BP_{ND} values was carried out with PMOD image analysis software, version 3.509 (PMOD Technologies Ltd., Zurich, Switzerland; <http://www.pmod.com>) and the multilinear reference tissue model with two parameters (MRTM2)¹⁹. The cerebellar grey matter without vermis and venous sinus was assigned as the reference region due to negligible availability of SERT in this region^{20,21}.

Non-displaceable binding potential (BP_{ND}) values were extracted for regions defined according to the automated anatomical atlas (AAL). Mean values were calculated from BP_{ND} for the left and right hemispheres. Thus, a total of 49 cortical and subcortical ROI was included in the analyses.

Statistics

A classification model for ADHD and HC was computed with genetic predictors, imaging predictors, all predictors as well as the top performing predictors, for each fold, respectively.

Computations were performed with the statistical software “R” (<https://www.R-project.org/>). The package “randomForest” was used for application of the eponymous algorithm (RF)^{22,23}. In short, RF is an ensemble tree classification tool that randomly selects subsamples of observations and builds a decision tree for optimal splitting of these observations according to an outcome variable by a combination of predictors. For each split, the best performing predictor out of a random selection is applied. Generally, a higher number of predictors allowed for selection leads to optimal splits but also low diversity of the individual trees. Therefore, restricting the number of features can generate models that perform worse in the training set but are more flexible when exposed to new data. Here, 3000 trees were grown (*ntree* = 3000) for each

model to enable multiple predictions for all patients. Classification was performed with a five-fold cross-validation (CV) design to allow optimal validation in absence of an independent test set²⁴. If hyperparameters must be tuned, nested CV is the gold-standard technique to prevent data leakage from training to the validation phase. For RF only the number of features randomly selected at each split (mtry) can be optimized; however, there is a standard of using the square root of the number of predictors. To prevent overfitting, no optimization of mtry was performed for this analysis.

For variable selection, a combination of established algorithms “Boruta” and “varSelRF” for “R” were used^{23,25}. Comparable to permutation-based importance evaluations, “Boruta” doubles the predictors included in the model by generating “shadow predictors” that show randomly interchanged values for each observation. Then 500 iterations of RF are run and only those predictors performing better than the best “shadow predictor” by a p -value threshold of 0.01 are preserved. These relevant predictors were then included in a backwards variable elimination algorithm, “varSelRF”. The best performing combination of predictors was then applied to the test set corresponding to each fold of the CV.

The whole CV procedure was repeated ten times and average accuracy is reported. See also Fig. 1 for a synopsis of the CV design.

There is no established method of power calculation for RF. Research indicated stable predictive capabilities of RF and comparable machine learning algorithms when enough observations are available, even in high dimensional data with the number of variables surpassing that of observations^{26,27}. For this dataset, a ratio of 79 predictors to 38 subjects was observed.

In addition to the results produced by RF, a mixed model was computed with the “lme” package for “R”. Linear mixed regression models for BP_{ND} with diagnosis, ROI and the most informative genetic predictors included as fixed effects and subject as random effect were built. Main and interaction effects (up to three-way) were computed. Mixed model results were corrected for the number of tests and models with a corrected threshold of $p \leq 0.001$. Based on these results, logistic regression models for each ROI and SNP with diagnosis as outcome variable and the respective ROI/SNP interaction term were computed. Logistic regression results were not corrected.

Results

Classification

Using all available predictors in a five-fold CV model yielded an accuracy of 0.80 (± 0.13). Restricting the predictors to either ROI or SNPs only, accuracies dropped to 0.58 (± 0.15) and 0.62 (± 0.15), respectively. Using only the

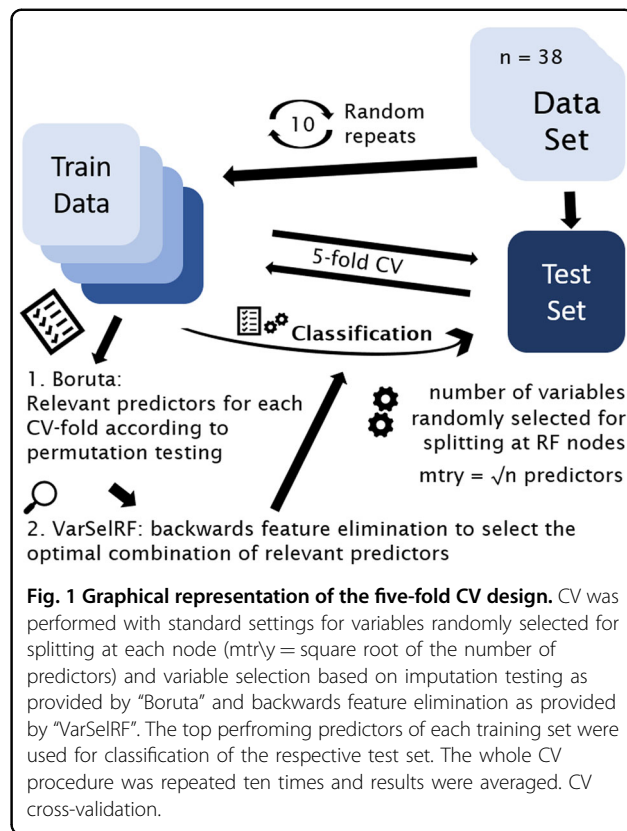


Table 2 Performance evaluators of the classification model for ADHD and HC.

Classification HC vs ADHD	Sens	Spec	PPV	NPV	Accuracy
SNPs ($n = 30$)	0.56	0.66	0.55	0.67	0.62 (± 0.15)
ROI ($n = 49$)	0.50	0.73	0.50	0.61	0.58 (± 0.15)
All ($n = 79$)	0.82	0.80	0.75	0.86	0.80 (± 0.13)
FS SNPs	0.71	0.80	0.72	0.79	0.76 (± 0.13)
FS ROI	0.65	0.76	0.66	0.75	0.71 (± 0.16)
FS SNP + ROI	0.75	0.86	0.80	0.83	0.82 (± 0.09)

Sensitivity, specificity, PPV, NPV and accuracy are provided for models based on only ROI or SNP predictors, all predictors, and predictors highlighted by feature selection. All results are averaged over 10 repeats of five-fold cross-validation. ADHD attention-deficit and hyperactivity disorder, HC healthy control, PPV positive predictive value, NPV negative predictive value, FS feature selection.

top performing predictors boosted the accuracy to 0.76 (± 0.13) and 0.71 (± 0.16) for SNPs and ROI, respectively. Optimal results were achieved combining the selected SNPs and ROI, with an accuracy of 0.82 (± 0.09) and positive and negative predictive values (PPV and NPV) of 0.80 and 0.83, indicating the probability of correct classification for and ADHD and HC, respectively

For a detailed overview of the classification outcome with all evaluation parameters (Table 2).

Variable selection

Concerning variable selection, most predictors showed a high agreement between training sets. The posterior cingulate gyrus was selected by 100% of models, followed by cuneus (48%), precuneus (22%), middle temporal gyrus (22%), precentral gyrus (20%), paracentral lobule (18%), postcentral gyrus (18%) and temporal pole (18%). Among the SNPs, *HTR2A* rs1328684 was selected by all models, followed by *HTR1B* rs130058 (82%). Other than that, only rs6311 (12%) and rs6313 (8%) were selected, which showed identical allelic expression. For a graphical depiction of importance measurements and probability rates of the top scoring predictors being selected in the training sets of each fold (Fig. 2).

Mixed model and logistic regression statistics

Mixed model revealed three-way associations for *HTR1B* SNP rs130058 ($F = 1.67, p < 0.0001$), *HTR2A* SNP rs1328684 ($F = 1.61, p = 0.001$), *HTR2A* rs6313 ($F = 2.38, p < 0.0001$) and *TPH2* rs1843809 ($F = 2.08, p < 0.0001$). Significant results of the mixed models are presented in Table 3, section A.

Logistic regression revealed interaction effects for the temporal pole ($z = -2.29, p = 0.022$) as well as posterior cingulate gyrus ($z = -2.14, p = 0.032$) with *HTR1B* rs130058, none of which remained significant after correction. Significant results of the logistic regression models are presented in Table 3, section B.

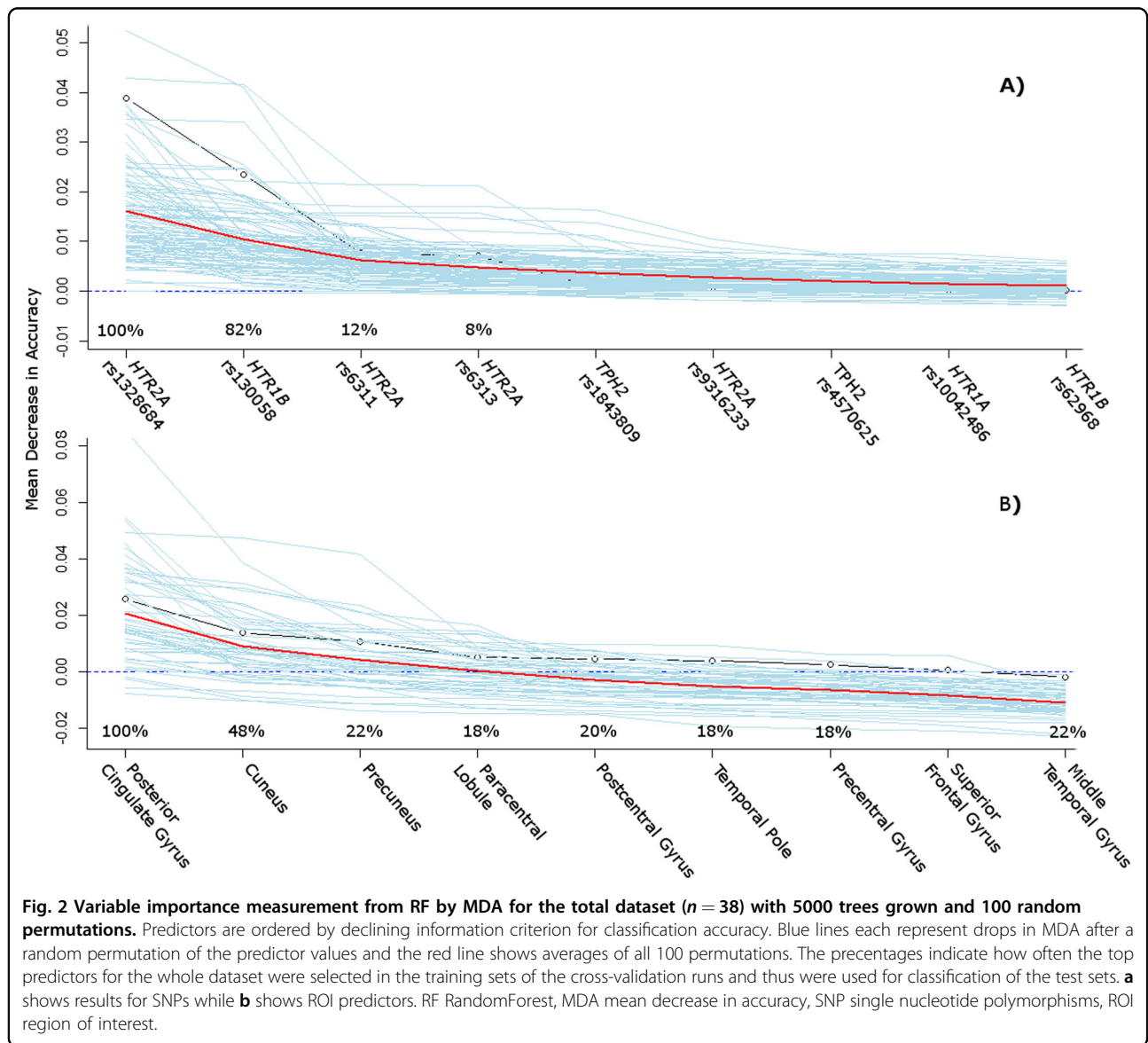


Fig. 2 Variable importance measurement from RF by MDA for the total dataset ($n = 38$) with 5000 trees grown and 100 random permutations. Predictors are ordered by declining information criterion for classification accuracy. Blue lines each represent drops in MDA after a random permutation of the predictor values and the red line shows averages of all 100 permutations. The percentages indicate how often the top predictors for the whole dataset were selected in the training sets of the cross-validation runs and thus were used for classification of the test sets. **a** shows results for SNPs while **b** shows ROI predictors. RF RandomForest, MDA mean decrease in accuracy, SNP single nucleotide polymorphisms, ROI region of interest.

Table 3 A) Linear mixed model results for effects of diagnosis, ROI and the top performing SNPs on BP_{ND}. B) Binomial logistic regression results for classification of diagnosis.

A) Mixed model	DF numerator	DF denominator	F value	p-value
Group * ROI * <i>HTR2A</i> rs1328684	49	1666	1.61	0.001
Group * ROI * <i>HTR1B</i> rs130058	49	1666	9.49	<0.0001
Group * ROI * <i>HTR2A</i> rs6313	49	1666	2.38	<0.0001
Group * ROI * <i>TPH2</i> rs1843809	49	1666	2.08	<0.0001
B) Logistic regression model	Estimator	Std. Error	Z value	p-value
Temporal Pole * <i>HTR1B</i> rs130058	-21.76	9.52	-2.29	0.022
PCG * <i>HTR1B</i> rs130058	-17.93	8.38	-2.14	0.032

Models were built for each of the top performing ROI and, respectively, for each interaction with performing SNPs, according to the machine learning classification model. Interaction effects are marked with *. Only significant results are shown (uncorrected). Mixed model results with $p \leq 0.001$ stayed significant after correction for multiple testing. None of the logistic regression results withstood correction.

ROI region of interest, SNP single nucleotide polymorphisms, BP_{ND} non-displaceable binding potential, DF Degrees of Freedom, PCG Posterior Cingulate Gyrus.

Discussion

Evaluating PET imaging and genetic predictors anchored within the serotonergic system, a moderate accuracy of 0.82 could be achieved for classification of ADHD and HC. Beyond the utility as a diagnostic tool, these results advocate different and recognizable serotonergic properties for ADHD and HC.

Predictor evaluation based on importance can adumbrate interaction effects that would not reach statistical significance in conventional association as thousands of predictor combinations are assessed for model building. Concerning the most prominent features in this analysis, SERT BP_{ND} within the three ROI posterior cingulate gyrus, cuneus and precuneus as well as SNPs rs130058 of *HTR1B* and rs1328684 of *HTR2A* were selected consistently by variable importance measures. All the anatomical structures labelled by these ROIs have previously been implicated in ADHD pathology as part of the default mode network (DMN)²⁸. Altered DMN activation during sustained attention paradigms in ADHD could thereby be partly redeemed by methylphenidate^{29–31}. Also both SNPs rs130058 and rs1328684 were implicated to mediate DMN abnormalities measured by MRI in a recent candidate gene study in PTSD³². *HTR1B* rs130058 was previously associated with ADHD as well as frequent comorbidities as substance dependence disorders^{33,34}. Although *HTR1B* was highlighted as a top gene by candidate gene reviews in ADHD³⁵, findings for rs130058 were overall inconsistent^{36–38}. In this sample, ADHD patients showed an increased frequency of rs1328684 minor alleles (1.69 vs 1.22) and decreased frequency of rs130058 minor alleles (1.25 vs 1.75). Concerning other predictors, the prominent SNPs rs6311 and rs6313, which showed complete linkage in our sample, did hardly impact classification results. Among other positive

reports, rs6311 was associated with ADHD in a rather large candidate gene study, but overall no consistent associations were reported so far^{35,39}.

However, the literature on in vivo SERT binding is limited. A previous study using single-photon emission computed tomography and the radioligand [¹²³I]FP-CIT, binding to dopamine and serotonin transporters, did not demonstrate serotonergic binding differences in 17 ADHD patients compared to HC⁴⁰. These findings were supported by a PET study using the tracer [¹¹C]MADAM that reported similar mean binding between eight ADHD patients and HC in several ROI, including prefrontal cortex, thalamus and putamen⁴¹. However, these results must be interpreted in the light of different radioligands and small sample sizes. The most extensive analysis of SERT binding in ADHD reported so far showed overall decreased BP_{ND} in 25 ADHD patients compared to age and sex matched controls applying the current gold-standard radioligand [¹¹C]DASB. Strongest effects were observed in the striatum, insula and anterior cingulate cortex; however, these results did not withstand correction for multiple testing in the post hoc analyses. Interestingly, only interregional molecular correlations of SERT binding between the hippocampus and the precuneus withstood correction thresholds, indicating that an interplay of brain regions may better portray abnormalities of serotonergic transmission in ADHD¹⁶. These thoughts are in line with moderate to high accuracy despite the lack of single predictor association results in our sample.

There is a particularly abundant literature for ADHD classification models to put these results into perspective with. While advanced statistics and especially machine learning methods have been established in all of neuropsychiatric research in recent years, this particular

boom in ADHD is probably owed to the heterogeneous nature and lack of objective biomarkers contrasted by overlapping clinical phenotypes with highly subjective symptoms and frequent comorbidities. Usually, diagnostic models were based on EEG or MRI data and aimed at automated classification of ADHD or MDD and HC.

Thereby, no algorithm proved to be superior to the other frequently applied techniques such as RF, support vector machines (SVM), neural networks or gradient boosting machines. This observation has been shared throughout different application areas of machine learning and culminated in the “no-free-lunch”-theorem, meaning that comparative algorithm performance cannot be generalized and is dependent on structure and context of the data as well as the model⁴². Nevertheless, RF and SVM may be the most commonly applied and best-established algorithms in imaging-based research^{10,43}. While SVM was demonstrated to often outperform RF by means of accuracy⁴⁴, RF may be more resilient to overfitting in small datasets as no hyper-parameter tuning is necessary and the generalization error does not increase with trees grown²². Furthermore, in contrast to SVM, both feature selection and classification can be performed with RF. Consequently, for this investigation all analyses were performed with RF.

A contemporary review suggested accuracy between 0.6 and 0.8 for published MRI based algorithms that conform with methodological standards concerning validation and feature selection in ADHD research¹⁰. Surveying the reports for MRI based models for ADHD diagnosis, accuracies above 0.9 attract attention⁴³. However, higher accuracies reported by some imaging studies may be owed to circular analysis or other intrusions of information between training and test samples. Although EEG based machine learning algorithms have been supported by the Food and Drug Association (FDA), preliminary results are hindered by the same issues⁴⁵.

The focus of this study on genetic imaging applying PET instead of in comparison easily obtainable MRI data brings about a considerably extenuated sample size of 34 subjects compared to some reported MRI based classification algorithms. While there is no other PET imaging and genetic machine learning study for comparison, study populations from MRI studies ranged from few dozens to several hundred subjects. Interestingly, recent meta analyses and reviews have emphasized a curious finding of decline of accuracy with increased sample size across studies despite oppositional effects within studies^{46,47}. While the majority of published studies featured below 100 observations, a decline on accuracy with sample size was observed. This may partly be explained by the contrast of narrow study settings to the heterogeneity of phenotypes in the clinical routine, which are better reproduced by larger, more natural samples. Along these

lines, larger samples are usually collected in multi-center approach and slight differences in implementation of study protocols or data acquisition and interpretation among contributing centers may explain a reduction in accuracy as machine learning algorithms can easily be disrupted by data disparity. On the other hand, however, accuracies may be inflated in small samples despite optimal validation protocols. Furthermore, smaller studies may be more prone to publication bias as low accuracy samples are probably underreported.

Our results must be interpreted cautiously due to the lack of external validation, constituting the most important limitation. The latter was not possible as to our knowledge there was no comparative sample on SERT binding in ADHD measured with [11 C]DASB that could have been used for validation. Thus, we cannot rule out overoptimistic PPV and NPV in our model. Although 38 subjects can be considered a large sample for a PET neuroimaging analysis, the observation count is marginal for machine learning classification. The CV design with fold-specific feature selection, refrainment from further parameter tuning and averaging across repeats can be regarded as state of the art but cannot substitute low sample size and lack of external validation²⁴. Along these lines, the moderate standard deviation across the repeats of the CV procedure must be noted and indicates that results may still be dependent on the data context. In line with these considerations, prediction accuracy based on MRI data from a recently published large sample of MDD patients thoroughly analyzed throughout a machine learning competition did not surpass 0.65⁴³. These lower accuracies but may be closer to the actual clinical value of currently available models. While an accuracy higher than 0.8 generally indicates good discrimination⁴⁸, the cut-offs necessary for clinical application are primarily dependent on already available screening and diagnostic tests and the expected ratio of observed cases. For example, an easily applicable screening test designed for detecting the few cases among a predominant number of controls must show good sensitivity while diagnostic tests also need high specificity to prevent false-positive outcomes. Considering the cost-intensive nature of PET and, to a lesser degree, also MRI, an imaging-based classification model can currently only fulfill a role as specialized diagnostic tool in clinically challenging cases as proposed for classification of psychosis⁴⁹. Consequently, current data do not support the viability of solely imaging-based algorithms for clinical applications, neither regarding ADHD nor MDD. Along these lines, clinical predictors such as scores of the Conners' Adult ADHD Rating Scale can most likely increase accuracies but were kept out of this analysis as two clearly distinct samples, healthy controls and ADHD patients, were compared. Keeping in mind the symptom overlap between ADHD and frequent comorbidities, a

clinical and bio-marker based transdiagnostic classification model may be clinically meaningful even with moderate accuracy.

To summarize, we propose a diagnostic prediction model for ADHD and HC based on multimodal serotonergic data. Thereby, we present the first PET based classification model for ADHD and expand on previous designs based solely on a single data type. We cannot yet advocate clinical applicability of this diagnostic model but present a step towards the goal of precision medicine in psychiatry. More importantly, our findings support different serotonergic profiles in ADHD and HC, reflected by distinct SERT and *HTR1B* as well as *HTR2A* activity, and especially put emphasis on the rs130058 and rs1328684 polymorphisms.

Acknowledgements

We are grateful to the technical and medical teams of the PET Centre, Medical University of Vienna, especially to K. Kletter, R. Dudczak, L.-K. Mien, J. Ungersboeck, C. Rami-Mark, D. Haeusler, T. Traub-Weidinger and L. Nics. Furthermore, we would like to thank A. Höfllich, C. Kraus, E. Akimova, M. Hienert, M. Spies, G. Gryglewski, C. Spindelegger, U. Moser, M. Fink, D. Winkler, R. Frey, A. Kutzelnigg, M. Stamenkovic, C. Klier, B. Hackenberg, A. Konstantinidis, D. Meshkat, J. Losak, P. Huf and P. Baldinger-Melich for medical support, and M. Savli for technical support. Further, G.M. James and H. Sigurdardottir deserve our gratitude for supporting data analyses and we thank K. Papageorgiou for supervision of clinical work. This work was supported by the Austrian Science Fund (FWF) grant numbers P27141 and KLI 516 to R.L. This research was conducted by pooling data from studies supported by grants of the Oesterreichische Nationalbank (Anniversary Fund, project numbers: 13675, 13214) and an intramural grant of the Department of Psychiatry and Psychotherapy (Forschungskostenstelle). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author details

¹Department of Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria. ²Department of Biomedical Imaging and Image-guided Therapy, Division of Nuclear Medicine, Medical University of Vienna, Vienna, Austria. ³Department of Rehabilitation Sciences, Hong Kong Polytechnic University, Hung Hom, Hong Kong. ⁴Center for Biomarker Research in Medicine CBmed, Graz, Austria. ⁵Ludwig Boltzmann Institute Applied Diagnostics, Vienna, Austria. ⁶Department of Psychiatry, University of Halle, Halle, Germany

Author contributions

A.K. and T.V. were preparing the manuscript and were responsible for data preparation and statistics. G.K. was assisting with statistical procedures and data preparation. A.H. supervised PET data analyses. C.P. was responsible for radiochemical preparations. W.W. and M.M. were responsible for all PET-related procedures and respective methodological contribution. M.H. was supervising the PET imaging procedures and revising on the manuscript. D.R. and A.H. were responsible for genetic analyses. S.K. advised on preparation of the manuscript, was involved in planning the study design and supervised clinical procedures. R.L. was responsible for planning and supervision of all study related procedures and manuscript preparation.

Conflict of interest

S.K. received grants/research support, consulting fees and/or honoraria within the last three years from Angelini, AOP Orphan Pharmaceuticals AG, AstraZeneca, Eli Lilly, Janssen, KRKA-Pharma, Lundbeck, Neuraxpharm, Pfizer, Pierre Fabre, Schwabe and Servier. R.L. received conference speaker honorarium within the last three years from Shire and support from Siemens Healthcare regarding clinical research using PET/MR, and is shareholder of BM Health GmbH since 2019. W.W. received speaker honoraria from GE Healthcare, research grants from DSD, BSM Diagnostica and ABX and is a part time employee of CBmed GmbH (Graz, Austria). M. Mitterhauser received speaker

honoraria from GE Healthcare, research grants from DSD, BSM Diagnostica, BAYER and ABX and is a part time employee of CBmed GmbH (Graz, Austria). G. S.K. received travel grants from Roche and Pfizer. The other authors report no potential conflict of interest with relevance to this work.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 23 October 2019 Revised: 20 February 2020 Accepted: 28 February 2020

Published online: 07 April 2020

References

- Faraone, S. V. & Biederman, J. What is the prevalence of adult ADHD? Results of a population screen of 966 adults. *J. Atten. Disord.* **9**, 384–391 (2005).
- Thome, J. et al. Biomarkers for attention-deficit/hyperactivity disorder (ADHD). A consensus report of the WFSBP task force on biological markers and the World Federation of ADHD. *World J. Biol. Psychiatry* **13**, 379–400 (2012).
- Uddin, L. Q., Dajani, D. R., Voorhies, W., Bednarz, H. & Kana, R. K. Progress and roadblocks in the search for brain-based biomarkers of autism and attention-deficit/hyperactivity disorder. *Transl. Psychiatry* **7**, e1218 (2017).
- Faraone, S. V. et al. Molecular genetics of attention-deficit/hyperactivity disorder. *Biol. Psychiatry* **57**, 1313–1323 (2005).
- Fliers, E. A. et al. Genome-wide association study of motor coordination problems in ADHD identifies genes for brain and muscle function. *World J. Biol. Psychiatry* **13**, 211–222 (2012).
- Faraone, S. V. & Khan, S. A. Candidate gene studies of attention-deficit/hyperactivity disorder. *J. Clin. Psychiatry* **67**(Suppl 8), 13–20 (2006).
- van der Meer, D. et al. The serotonin transporter gene polymorphism 5-HTTLPR moderates the effects of stress on attention-deficit/hyperactivity disorder. *J. Child Psychol. Psychiatry* **55**, 1363–1371 (2014).
- Franke, B., Neale, B. M. & Faraone, S. V. Genome-wide association studies in ADHD. *Hum. Genet.* **126**, 13–50 (2009).
- Thapar, A. Discoveries on the Genetics of ADHD in the 21st Century: new findings and their implications. *Am. J. Psychiatry*. <https://www.ncbi.nlm.nih.gov/pubmed/30111187> (2018).
- Pulini, A. A., Kerr, W. T., Loo, S. K., Lenartowicz, A. Classification accuracy of neuroimaging biomarkers in attention-deficit/hyperactivity disorder: effects of sample size and circular analysis. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6310118/> (2018).
- Albert, P. R. & Benkelfat, C. The neurobiology of depression-revisiting the serotonin hypothesis. II. Genetic, epigenetic and clinical studies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120535 (2013).
- Albert, P. R., Benkelfat, C. & Descarries, L. The neurobiology of depression-revisiting the serotonin hypothesis. I. Cellular and molecular mechanisms. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 2378–2381 (2012).
- Ding, Y. S. et al. Clinical doses of atomoxetine significantly occupy both norepinephrine and serotonin transports: Implications on treatment of depression and ADHD. *Neuroimage* **86**, 164–171 (2014).
- van Stralen, J. Emotional dysregulation in children with attention-deficit/hyperactivity disorder. *Atten. Defic. Hyperact. Disord.* **8**, 175–187 (2016).
- Kutlu, A., Akyol Ardıc, U. & Ercan, E. S. Effect of methylphenidate on emotional dysregulation in children with attention-deficit/hyperactivity disorder + oppositional defiant disorder/conduct disorder. *J. Clin. Psychopharmacol.* **37**, 220–225 (2017).
- Vanicek, T. et al. Altered interregional molecular associations of the serotonin transporter in attention deficit/hyperactivity disorder assessed with PET. *Hum. Brain Mapp.* **38**, 792–802 (2017).
- Sigurdardottir, H. L. et al. Effects of norepinephrine transporter gene variants on NET binding in ADHD and healthy controls investigated by PET. *Hum. Brain Mapp.* **37**, 884–895 (2016).
- Langzenberger, R. et al. Prediction of SSRI treatment response in major depression based on serotonin transporter interplay between median raphe nucleus and projection areas. *Neuroimage* **63**, 874–881 (2012).
- Innis, R. B. et al. Consensus nomenclature for in vivo imaging of reversibly binding radioligands. *J. Cereb. Blood Flow. Metab.* **27**, 1533–1539 (2007).

20. Parsey, R. V. et al. Altered serotonin 1A binding in major depression: a [¹¹C]-WAY100635 positron emission tomography study. *Biol. Psychiatry* **59**, 106–113 (2006).
21. Gryglewski, G. et al. Simple and rapid quantification of serotonin transporter binding using [(11)C]DASB bolus plus constant infusion. *Neuroimage* **149**, 23–32 (2017).
22. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
23. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R. N.* **2**, 18–22 (2002).
24. Varoquaux, G. et al. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *Neuroimage* **145**(Pt B), 166–179 (2017).
25. Kursa, M. B. & Rudnicki, W. R. Feature selection with the boruta package. *J. Stat. Softw.* **36**, 1–13 (2010).
26. Chen, C. C. et al. Methods for identifying SNP interactions: a review on variations of Logic Regression, Random Forest and Bayesian logistic regression. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **8**, 1580–1591 (2011).
27. Roetker, N. S. et al. Assessment of genetic and nongenetic interactions for the prediction of depressive symptomatology: an analysis of the Wisconsin Longitudinal Study using machine learning algorithms. *Am. J. Public Health* **103**(Suppl 1), S136–S144 (2013).
28. Bush, G. Attention-deficit/hyperactivity disorder and attention networks. *Neuropsychopharmacology* **35**, 278–300 (2010).
29. Kowalczyk, O. S., et al. Methylphenidate and atomoxetine normalise frontoparietal underactivation during sustained attention in ADHD adolescents. *Eur. Neuropsychopharmacol.* <https://www.ncbi.nlm.nih.gov/pubmed/31358436> (2019).
30. Christakou, A. et al. Disorder-specific functional abnormalities during sustained attention in youth with Attention Deficit Hyperactivity Disorder (ADHD) and with autism. *Mol. Psychiatry* **18**, 236–244 (2013).
31. Rubia, K. Cognitive neuroscience of attention deficit hyperactivity disorder (ADHD) and its clinical translation. *Front Hum. Neurosci.* **12**, 100 (2018).
32. Miller, M. W. et al. 5-HT2A gene variants moderate the association between PTSD and reduced default mode network connectivity. *Front. Neurosci.* **10**, 299 (2016).
33. Cao, J., LaRocque, E. & Li, D. Associations of the 5-hydroxytryptamine (serotonin) receptor 1B gene (HTR1B) with alcohol, cocaine, and heroin abuse. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **162B**, 169–176 (2013).
34. Muller, D. et al. Evidence of sexual dimorphism of HTR1B gene on major adult ADHD comorbidities. *J. Psychiatr. Res.* **95**, 269–275 (2017).
35. Gizer, I. R., Ficks, C. & Waldman, I. D. Candidate gene studies of ADHD: a meta-analytic review. *Hum. Genet.* **126**, 51–90 (2009).
36. Ickowicz, A. et al. The serotonin receptor HTR1B: gene polymorphisms in attention deficit hyperactivity disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **144B**, 121–125 (2007).
37. Smoller, J. W. et al. Association between the 5HT1B receptor gene (HTR1B) and the inattentive subtype of ADHD. *Biol. Psychiatry* **59**, 460–467 (2006).
38. Ribases, M. et al. Exploration of 19 serotonergic candidate genes in adults and children with attention-deficit/hyperactivity disorder identifies association for 5HT2A, DDC and MAOB. *Mol. Psychiatry* **14**, 71–85 (2009).
39. Pazvantoglu, O. et al. The relationship between the presence of ADHD and certain candidate gene polymorphisms in a Turkish sample. *Gene* **528**, 320–327 (2013).
40. Hesse, S., Ballaschke, O., Barthel, H. & Sabri, O. Dopamine transporter imaging in adult patients with attention-deficit/hyperactivity disorder. *Psychiatry Res.* **171**, 120–128 (2009).
41. Karlsson, L. et al. Serotonin transporter in attention-deficit hyperactivity disorder—preliminary results from a positron emission tomography study. *Psychiatry Res.* **212**, 164–165 (2013).
42. Gomez, D. & Rojas, A. An empirical overview of the no free lunch theorem and its effect on real-world machine learning classification. *Neural Comput.* **28**, 216–228 (2016).
43. Gao, S., Calhoun, V. D., Sui, J. Machine learning in major depression: from classification to treatment outcome prediction. *CNS Neurosci. Ther.* <https://www.ncbi.nlm.nih.gov/pubmed/30136381> (2018).
44. Statnikov, A., Wang, L. & Aliferis, C. F. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinforma.* **9**, 319 (2008).
45. Snyder, S. M., Rugino, T. A., Hornig, M. & Stein, M. A. Integration of an EEG biomarker with a clinician's ADHD evaluation. *Brain Behav.* **5**, e00330 (2015).
46. Schnack, H. G. & Kahn, R. S. Detecting neuroimaging biomarkers for psychiatric disorders: sample size matters. *Front. Psychiatry* **7**, 50 (2016).
47. Chu, C. et al. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage* **60**, 59–70 (2012).
48. Hosmer, D. W., Lemeshow, S., Sturdivant, R. X. *Applied Logistic Regression*. 3rd edn. (Wiley, 2013).
49. Koutsouleris, N. et al. Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. *JAMA Psychiatry* **75**, 1156–1172 (2018).