



OPEN Exploring voice as a digital phenotype in adults with ADHD

Georg G. von Polier^{1,2,3,9}✉, Eike Ahlers^{4,5,9}, Julia Volkening^{1,6}, Jörg Langner⁶, Kaustubh R. Patil^{1,7}, Simon B. Eickhoff^{1,7}, Florian Helmhold^{6,8}, Agnieszka Ewa Krautz⁶ & Daina Langner⁴

Current diagnostic procedures for attention deficit hyperactivity disorder (ADHD) are mainly subjective and prone to bias. While research on potential biomarkers, including EEG, brain imaging, and genetics is promising, it has yet to demonstrate clinical utility. Dopaminergic signaling alternations and executive functioning, crucial to ADHD pathology, are closely related to voice production. Consistently, previous studies point to alterations in voice and speech production in ADHD. However, studies investigating voice in large clinical samples allowing for individual-level prediction of ADHD are lacking. Here, 387 ADHD patients, 204 healthy controls, and 100 psychiatric controls underwent standardized diagnostic assessment. Subjects provided multiple 3-minutes speech samples, yielding 920 samples. Based on prosodic voice features, random forest-based classifications were performed, and cross-validated out-of-sample accuracy was calculated. The classification of ADHD showed the best performance in young female participants (AUC = 0.87) with lower performance in older participants and males. Psychiatric comorbidity did not alter the classification performance. Voice features were associated with ADHD-symptom severity as indicated by random forest regressions. In summary, prosodic features seem to be promising candidates for further research into voice-based digital phenotypes of ADHD.

Keywords ADHD, Voice as a biomarker, AI diagnostics, Digital biomarker, Machine learning, Precision psychiatry

Attention Deficit Hyperactivity Disorder (ADHD) is a neurodevelopmental condition defined by symptoms of inattention, hyperactivity, and impulsivity, which impair quality of life, social, and academic outcome¹. The condition is highly prevalent worldwide, with estimates of 5% prevalence in childhood and about 2.5% in adults². Current ADHD diagnostic procedures, according to international guidelines, are built around rater-dependent assessments, including a diagnostic interview, as well as self- and third-party-reports and rating scales³. These are, however, subjective procedures and therefore prone to biases⁴. The limitations of current diagnostic criterion standards are widely acknowledged, and the clinical practice is criticized for mis-, over- or under-diagnosing ADHD⁵. Thus, developing biomarkers is an important field in psychiatric research to help improve the diagnostic accuracy and provide treatment guidance in the context of precision psychiatry.

Extensive research has been dedicated to exploring biomarkers in ADHD, including neuropsychological tests, electroencephalography (EEG), structural and functional brain imaging, and genetics^{6,7}. However, to date, research does not support immediate clinical applications of these biomarkers for ADHD, mostly because they lack convincing accuracy and/or have limited practicality. The importance of confirming the reliability of these biomarkers through larger cohort studies that also account for sex differences has been emphasized⁸. One possible avenue to develop clinically feasible biomarkers is the development of high-dimensional digital biomarkers based on multiple clinically feasible objective measures and appropriate statistical models, including machine learning (ML)⁶.

¹Institute of Neuroscience and Medicine Brain and Behaviour, Forschungszentrum Jülich, Wilhelm-Johnen-Str, 52528 Jülich, Germany. ²Clinic for Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, University Leipzig, Leipzig, Germany. ³Clinic for Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, RWTH Aachen University, Aachen, Germany. ⁴Institute of Psychiatry Campus Benjamin Franklin, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin Institute of Health, Berlin, Germany. ⁵Clinic for Psychiatry, Psychotherapy and Psychosomatic medicine, Johanniter-Krankenhaus Treuenbrietzen, Treuenbrietzen, Germany. ⁶PeakProfiling GmbH, Berlin, Germany. ⁷Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. ⁸Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen, Tübingen, Germany. ⁹Georg G. von Polier and Eike Ahlers contributed equally to this work. ✉email: g.von.polier@fz-juelich.de

One component of future, high-dimensional biomarkers could be based on voice assessments. The study of voice in ADHD is motivated for several reasons. First, from a neurobiological perspective, altered dopamine signaling is thought to play an important role in the pathophysiology of ADHD, amongst others related to deficits in executive functioning (EF) and motor control⁹. Moreover, dopamine is intimately involved in motor behaviors and plays a central role in vocal production¹⁰. Speech production is rated as one of the most complex motor behaviors, based on the coordination of more than 100 muscles, including laryngeal, supralaryngeal, and respiratory muscles. Thus, based on the hypothesis of altered dopamine signaling in ADHD, changes in speech production will possibly occur in individuals with ADHD. Secondly, EF deficits are core symptoms of ADHD and are linked to speech production. Stronger EF is associated with more accurate articulation in children and better speech performance (articulatory control and fluency of language output) in adults^{11–14}. Consistent with this, speech production deficits in ADHD have been linked to impairments in working memory and EF^{15,16}. Children with ADHD, and sometimes their parents, poorly modulate voice volume, often speak louder and for longer periods and show signs of hyperfunctional voice disorder^{17,18}. They have higher subglottal pressure and lower transglottal airflow, likely due to increased muscle tone in the glottis^{17,19}. Adults with ADHD show lower articulatory accuracy and slower speech rates compared to controls, with articulatory accuracy negatively correlated with symptom severity²⁰. Articulation requires complex motor control dependent on self-regulation and inhibition. Lastly, recent work by Li et al. suggests that prosodic and deep-learned linguistic features can distinguish ADHD from healthy controls, achieving a classification accuracy of 0.78²². In summary, changes in voice and speech production likely occur in ADHD and thus could serve as part of a high-dimensional biomarker. The aim of this study is to determine whether prosodic voice features recorded in brief speech tasks may allow for a differentiation of ADHD from healthy controls and subjects with other psychiatric diseases using a machine learning -based approach.

Materials and methods
Participants

A large heterogeneous group of 563 adults were recruited at our specialized adult ADHD outpatient clinic, 387 of whom were subsequently diagnosed with ADHD (Table 1), and 100 of whom were diagnosed with other mental disorders (psychiatric controls PC) (Table S3) but did not fulfill the diagnostic criteria for ADHD. Further, 76 patients were excluded due to positive drug screenings ($n=55$), technical deficits of the recordings ($n=7$), or because they showed subclinical ADHD symptoms but did not fulfill the criteria for ADHD or other mental disorders after the diagnostic evaluation ($n=14$). All participants were asked to provide voice samples before undergoing a standardized diagnostic procedure. In addition, 204 non-patient adults were recruited as healthy controls (HC) through public announcement. Healthy controls had no history of past or present neuropsychiatric conditions. All participants were aged between 18 and 59 years, and gave written informed consent (Box S1 for full inclusion/ exclusion criteria). A detailed sample description is provided in Table 1 and Table S2.

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human subjects/patients were approved by the ethics committee of the Charité Universitätsmedizin, Berlin, Germany; the approval number is EA4/014/10. The study was registered as a clinical trial (ClinicalTrials.gov Identifier: NCT01104623).

Diagnostic procedure

All patients were assessed by trained and licensed psychologists and psychiatrists affiliated with the specialized ADHD adult outpatient clinic of the Charité University Hospital, Berlin, Germany. A full diagnostic workup, including ADHD diagnoses, was obtained through a multi-informant, multi-method approach with psychological and medical assessments. The diagnostic procedure was structured in accordance with recommended practice of national and international guidelines²³ and represents criterion standards for the diagnosis of ADHD in adults. It included a diagnostic interview of mental diseases according to DSM criteria, a review of client history (including developmental, medical, academic, and social background) and relevant documentation, behavioral observations, completion of self- and third party-report, and standardized ADHD-specific rating scales both

	male	female	total
All participants	388	379	767
Age in years / mean (SD)	34.3 (10.5)	33.4 (9.8)	34.4 (10.6)
ADHD (all)	221	166	387
ATT	117	81	198
COM	104	85	189
ADHD with comorbidity	117	84	201
HC	75	129	204
PC	36	64	100
Excluded	56	20	76

Table 1. Sample description. *ATT* inattentive ADHD subtype, *COM* combined inattentive and hyperactive / impulsive ADHD subtype, *HC* healthy controls, *PC* psychiatric controls

for childhood and adulthood. These evaluations were completed over two sessions with up to 8 h of assessment (Table 2). To evaluate ADHD-related symptom severity, a clinical interview was performed by a senior clinician according to DSM-related criteria²³, using the ADHD-DC scale²⁴. In addition to 18 items covering inattention, hyperactivity, and impulsivity symptoms, the age of onset, symptoms related burden, general burden, and reduced social contacts were rated on 0–3-point Likert scales: 22 items in total with sum scores ranging from 0 to 66.

Voice recordings

High-quality audio recordings were obtained by a recording technician in a quiet, dedicated room free of external noise sources before each clinical evaluation, ensuring that participants and recording technicians were unaware of the diagnostic results. Prior to the actual recording, instructions for the participants were recorded and played back through an interface (“Fast Track” Interface (MediTECH Electronic GmbH) to minimize the effects of interactions with the recording technicians and ensure a standardized recording procedure. The recording instructions included a brief rehearsal phase, during which no data was recorded. This was done to ensure that participants were familiar with the instructions and the recording process. Participants spoke into a headset microphone (C555L AKG), and signals were recorded in 16-bit/22.5 kHz sampling rate on a PC, followed by cutting and labeling procedures as preparation for the sound analysis via the “Fast Track” Interface (MediTECH Electronic GmbH).

All participants were asked to speak the following utterances: spoken single vowels and consonants, reading out single words, counting from one to ten (two trials), and free speech of around 2 min duration on a free topic, e.g. the last weekend or holidays (Table S4). Previous research indicates that voice production varies depending on the cognitive demands of the specific speech task^{25,26}.

Each clinical appointment was preceded by the standardized recording task described above. Since most participants had multiple clinical visits, multiple recordings were obtained from the same individuals using identical procedures. All recordings followed this consistent protocol, including repeated voice recordings before each follow-up appointment. For the final analysis, a maximum of three recordings per participant were included as separate data points.

From an initial 1029 voice recordings from 767 participants (including participants that were later excluded), 24 recordings were excluded due to technical problems of the recordings (e.g., low volume) or multiple recordings, i.e., more than three recordings of the same person. From the remaining 1005 recordings, 85 recordings were excluded: (a) due to positive drug screenings of the subjects ($n=60$), because participants showed subclinical ADHD-symptoms but did not qualify for ADHD or other mental disorders ($n=24$) or due to severe cold ($n=1$). Thus, a total of 920 recordings were included in the classification steps, using 1 to 3 recordings per participant. A total of 71 recordings were made while patients were on medication (Methylphenidate or Atomoxetine), either prescribed prior to their initial presentation at our clinic or initiated following our diagnostic assessment.

Feature generation and machine learning

Paralinguistic features were calculated, focusing solely on prosodic information without analyzing semantic content. Only utterances based on free speech and counting were included in the analyses. Following a *Fourier* transformation of the voice recordings²⁷, paralinguistic features were calculated in two feature groups pertaining to (1) the contour of loudness and (2) the contour of speech melody (i.e. the course of F_0)²⁸⁻³¹. In brief, loudness was transformed to loudness as perceived by humans (i.e. Sone) based on the model of subjective loudness by Zwicker³². The transformed loudness was then averaged over 24 different time spans with durations, ranging

Psychiatric assessment		
	Patients	Healthy controls
Psychiatric/ psychological examination, psychiatric history	+	+
SCID I - screening	+	+
SCID I / II – diagnostic interview	+	
BDI II	+	
Medical examination	+	
Medical and psychiatric history	+	+
Blood tests, including screening for thyroid conditions	+	
Drug tests	+	+
ADHD specific assessment		
	Patients	Healthy controls
ADHD-checklist for DSM-IV ¹	+	+
Wender Utah rating scale, short version ²	+	+
Semi-structured clinical interview following DSM-IV-TR	+	
Second party reports, school reports	+	

Table 2. Diagnostic procedure. 1: Diagnostic procedure according to the guidelines of the German Society for Psychiatry, Psychology and Neurology, Psychosomatics (DGPPN)²⁴, 2: citation ⁶⁵, German version ⁶⁶; SCID structured clinical interview part I/ II; BDI II – Beck Depression Inventory-II

from 0.020 s (s) to 4.032 s (Box S2). Thus, we obtained 24 contours of loudness, each representing temporal characteristics of speech sound across two orders of magnitude of time scales: from very short timespans (20 ms) informative of ‘micro prosodic’ structures up to longer time spans (4 s) capturing voice relations between words. The first derivative of all timeframes was calculated and subsequently means, standard deviations, peaks and quantiles were calculated.

For a second set of features pertaining to changes pitch, F_0 was extracted and transformed on a logarithmic scale. The resulting values were again averaged over the same 24 different time spans (Box S2), and again after calculating the first derivative means, standard deviations, peaks and quantiles were calculated³³. To account for statistical effects related to the dimension time of the recordings, the curves of the respective features were fitted to multiple high-dimensional vectors using proprietary code of “PeakProfiling”. In total 6145 raw features were calculated. Differences of this approach with frequently used voice analyses are given in Table 3.

Prior to the classification, initial filtering was performed to eliminate features with very low ($<2.2 \times 10^{-16}$) variance ($n=0$) and variables with less than 100 unique values ($n=4$), resulting in 6141 features that were included in further analyses. Random forest (RF) classifications were conducted using MATLAB Version 2023a, employing the TreeBagger algorithm. The model comprised 500 decision trees, and performance evaluation was achieved through 10-fold cross-validation repeated 100 times to ensure robustness and mitigate overfitting. Prior to classification, potential confounding variables, namely sex, age, and education, were accounted for. Each feature underwent linear regression to regress out these confounds in a cross-validation consistent manner. This involved estimating the regression model on the training fold and then applying the model to both the training and validation folds, ensuring that the residuals, which are the confound-free features, were retained for further analysis. Classification performance was quantified using the area under the receiver operating characteristic curve (AUC-ROC), referred to as AUC.

To assess the prediction capacity of the voice features regarding continuous variables such as ADHD symptom severity according to the ADHD-DC scale, but also potential confounder such as age and education, we calculated RF based regression models with 100 trees again within the cross-validation folds described above with 100 repetitions. Mean absolute error of the regression and Pearson’s correlation coefficients between actual and predicted variables were calculated as performance measures. To compare Pearson’s coefficients, we first calculated Fisher r -to- z transformations and then performed a conservative version of t -tests as proposed by Nadeau and Bengio³⁴ to account for the possibility of bias related to cross-validation analyses.

To assess prediction accuracy across different subgroups (e.g., sex and age groups), we repeated all analyses in the subgroups male, female and in two age groups of similar sample size, split at the median age of 32 years. To explore differences between these subgroups, group comparisons were performed using t -tests in a similar manner as described above. To evaluate the relevance of specific features and speech tasks for the classification of ADHD, the feature importance was assessed as defined by the permutation of out-of-bag predictions as implemented in MATLAB³⁵. The top ten features were extracted for further evaluation.

Results
Descriptive results

ADHD subjects exhibited higher symptom severity than HC and PC subjects (Table 4). Comparisons with excluded subjects remained non-significant. Within the ADHD group, symptoms in the subscale ‘inattention’ were more pronounced than in the subscale ‘hyperactivity/impulsivity’; however, the latter showed a higher variance, as indicated by a significant Levene’s test. Moreover, female ADHD participants showed more pronounced ‘hyperactivity’ than male ADHD participants. Symptom severity was similar among age groups of comparable sample sizes after a median split of age (i.e., 18–31.9 years and 32–59 years).

Classification

RF classification was calculated, controlling for effects of age, sex, and education based on voice features. Cross-validated out-of-sample accuracies are provided in Table 5; Fig. 1. The classification over the complete sample (ADHD patients versus HC based on 801 recordings) resulted in an AUC of 0.77, which is invariant to classification threshold. Repeating the classification with speaker-stratified cross-validation produced largely consistent results (AUC differences ranging from 0.00 to 0.03 in subanalyses, detailed in Supplementary Table S6), with no change in AUC for the complete sample.

general approach	current approach
Spectrogram employing short-time Fourier transform	Spectrogram employing short-time Fourier transform
Calculating the root-mean-square value for each time frame of the spectrogram Generating a single contour of energy over time	Calculating the loudness value (Zwicker) for each time frame of the spectrogram Generating a contour of loudness over time
	Smoothing the contour by 24 time spans ¹ , so obtaining 24 different contours
Calculating statistics of the single contour – e.g. max, min, mean, standard deviation	Calculating statistics of the 24 contours - max, min, mean, standard deviation, and higher level features derived from the slopes including the curvature, the number of peaks and valleys and the relations between them

Table 3. Comparison of general and current approach to loudness analysis. 1 time spans range between 20 ms and 4032 ms, details in Box S2 of the supplement

	Clinical Interview M ± SD	Inattention M ± SD	Hyperactivity/ Impulsivity M ± SD	WURS-k M ± SD	BDI - II M ± SD
ADHD (all)	40.0 ± 8.3	7.0 ± 1.7 ²	4.9 ± 2.5 ²	38.9 ± 12.9	13.3 ± 8.9
ADHD - male	40.5 ± 8.3	7.0 ± 1.7	4.7 ± 2.5*	39.4 ± 12.9	12.7 ± 8.7
ADHD - female	39.3 ± 8.3	6.9 ± 1.6	5.2 ± 2.6*	38.3 ± 12.9	14.0 ± 9.2
ADHD - ATT	35.1 ± 7.3**	7.0 ± 1.6	2.8 ± 1.6**	36.6 ± 11.6**	13.3 ± 9.0
ADHD - COM	44.7 ± 6.6**	6.9 ± 1.8	7.0 ± 1.1**	41.5 ± 13.7**	13.0 ± 8.9
ADHD < 32 y ¹	40.2 ± 8.0	7.0 ± 1.5	4.7 ± 2.6	38.0 ± 11.8	13.7 ± 9.0
ADHD ≥ 32 y ¹	39.5 ± 8.9	6.8 ± 1.8	5.0 ± 2.5	39.8 ± 13.7	12.8 ± 8.8
HC	5.2 ± 4.4***	0.5 ± 0.8***	0.5 ± 0.8***	6.6 ± 6.5***	3.2 ± 2.8***
PC	23.3 ± 12.2***	3.3 ± 2.5***	2.7 ± 2.3***	23.7 ± 13.8***	14.9 ± 8.8***
Excl. Particip.	35.9 ± 11.6	5.5 ± 2.4	4.5 ± 2.6	39.1 ± 15.4	16.5 ± 9.3

Table 4. Test psychology. *ATT* inattentive ADHD subtype, *COM* combined inattentive and hyperactive/impulsive ADHD subtype, *HC* healthy controls, *PC* psychiatric controls, *WURS-k* Wender-Utah-rating-scale short form, *BDI* Beck's depression inventory; * $p < 0.05$ in comparisons of sex; ** $p < 0.001$ in comparisons ATT vs. COM; *** $p < 0.001$ in comparisons ADHD vs. HC and ADHD vs. PC; 1: Comparisons between age-groups of ADHD-participants remained non-significant; 2: Comparing the variance of inattention and hyperactivity with Levene's test indicates higher variance in hyperactivity than in inattention

Group		AUC	Precision	Recall	F1
All		0.77	0.60	0.75	0.66
Male		0.70	0.51	0.90	0.66
Female		0.83	0.71	0.78	0.74
one recording/ participant		0.76	0.63	0.73	0.67
excl. stimulants		0.78	0.64	0.77	0.70
excl. comorbid ADHD		0.78	0.70	0.73	0.72
ADHD, subtype ATT		0.79	0.73	0.75	0.74
ADHD, subtype COM		0.80	0.72	0.75	0.73
Age 18–31	All	0.80	0.68	0.79	0.73
(n = 385 recordings)	Male (n = 202)	0.75	0.53	0.72	0.62
	Female (n = 183)	0.87	0.82	0.81	0.81
Age 32–59	all	0.72	0.55	0.75	0.63
(n = 416 recordings)	Male (n = 193)	0.59 ¹	0.50	NaN	0.50
	Female (n = 223)	0.74	0.57	0.73	0.64

Table 5. Classification ADHD vs. HC. *ATT* inattentive ADHD subtype, *COM* combined inattentive and hyperactive/impulsive ADHD subtype, *HC* healthy controls, 1: This calculation should be interpreted with caution, as it suffers from class imbalance with only 30 HC

ADHD-subtype, sex, age

The classification showed a largely similar performance in subjects with combined symptoms (i.e., including hyperactivity and impulsivity) compared to predominantly inattentive subjects. Analyzing both sexes separately pointed toward a higher AUC in females than in male participants ($t = 3.3$, $p = 0.001$). A higher AUC in female subjects was also evident when analyzing separate age groups (age < 32: $t_{\text{male/female}} = 2.2$, $p = 0.03$; age > 32: $t_{\text{male/female}} = 2.3$, $p = 0.03$, Table 5) and in the non-comorbid ADHD sample (male: AUC = 0.68; female AUC = 0.79; $t = 3.3$, $p = 0.001$). Comparisons of age groups indicated superior classification performance in the younger sample (AUC = 0.80 vs. AUC = 0.71; $t = 2.0$, $p = 0.04$), with the highest performance in young female participants (AUC = 0.86).

Comorbidity, psychiatric controls, number of speech samples, medication

The classification of a subsample of ADHD subjects with no comorbidity resulted in a comparable AUC of 0.78. Calculating the classification of ADHD participants against a psychiatric control (PC) group, which presented for the evaluation of ADHD, but did not meet the clinical criteria of ADHD, showed a lower prediction accuracy (AUC = 0.60, Precision = 0.91, Recall = 0.61, F1 = 0.71). Notably, the PC group still exhibited elevated symptoms of ADHD as indicated by elevated symptom scores in the clinical interview of ADHD compared to HC (mean sum score psychiatric controls (PC): 23 vs. 5 in HC; mean sum score 40 in ADHD, Table 4). Subgroup details (age, sex) are provided in Supplementary Table S5.

Furthermore, utilizing a single recording per subject yielded a classification performance that was nearly equivalent to that achieved with multiple recordings (0.76 vs. 0.77). Excluding subjects who took ADHD

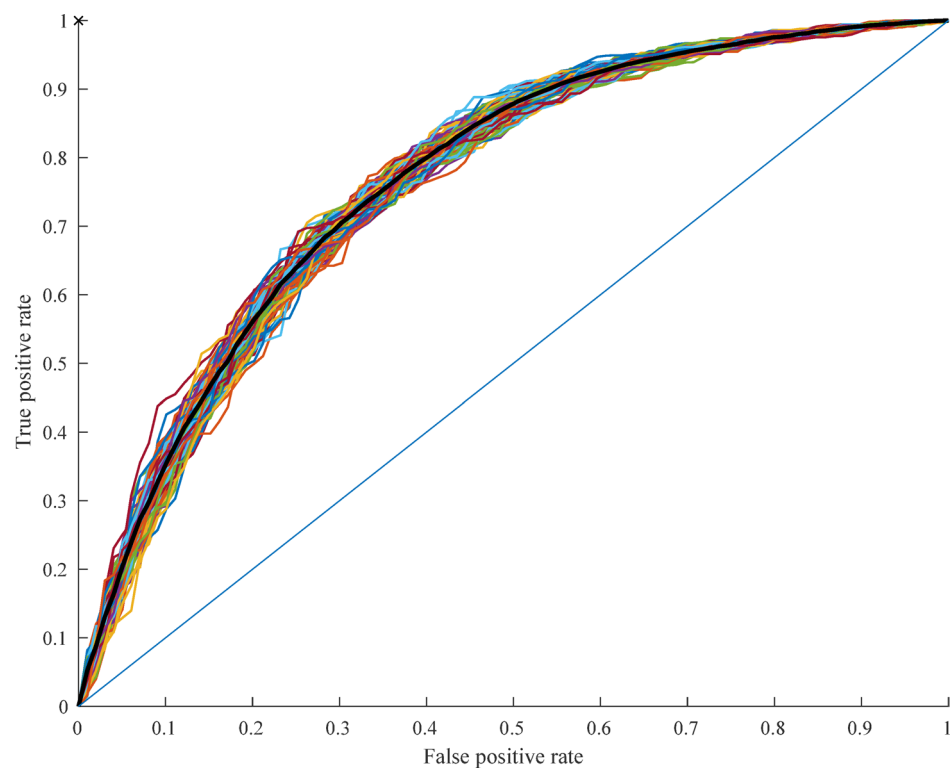


Fig. 1. Classification ROCs for (all) ADHD patients versus healthy controls (801 recordings). Random-forest classification, 10-fold cross-validation; AUC = 0.77.

	Group	<i>n</i>	<i>r</i>	<i>t</i>	MAE
ADHD clinical interview ¹	All	876	0.45	14.7	12.5
ADHD clinical interview ¹	Female	457	0.49	12.1	12.9
ADHD clinical interview ¹	Male	419	0.29	6.3	12.0
Hyperactivity	All	876	0.36	11.6	2.3
Inattention	All	876	0.39	12.9	2.6
Education	All	920	0.33	10.5	0.7
Age (years)	All	920	0.52	18.5	7.2
Beck's depression inventory ¹	All	701	0.24	6.6	7.1
ADHD clinical interview ¹	adhd	554	0.30	6.7	6.3
Hyperactivity	adhd	554	0.39	9.2	2.0
Inattention	adhd	554	0.19	3.9	1.3

Table 6. Correlation coefficients of true and predicted clinical variables. Pearson's correlation coefficients are given based on RF regression; MAE mean absolute error, BDI Beck's depression inventory; *n* = number of voice samples; 1: total symptoms score *p*-value for all analyses < 0.001

medication (methylphenidate, atomoxetine) before the voice recording did not alter the classification performance, however the number of affected recordings was modest (*n* = 71; 8.9% of the analyzed sample).

Correlations of symptom severity using random forest regression

Correlations of true and predicted symptoms based on voice features were calculated using RF regressions. The results are given in Table 6; Fig. 2. Analyses performed in the ADHD group only indicated a higher association of true and predicted hyperactive/ impulsive symptoms than inattention symptoms ($r_{\text{hyper}} = 0.39$; $r_{\text{inattention}} = 0.19$; comparison: $t = 3.6$, $p < 0.001$). Extending the analysis of ADHD symptom severity and voice features to all participants (ADHD, PC, HC) yielded a higher association of true and predicted symptoms ($r = 0.45$, $t = 14.7$, $p < 0.001$), in particular in female subjects ($r_{\text{female}} = 0.49$; $r_{\text{male}} = 0.29$; comparison $t = 3.4$, $p < 0.001$).

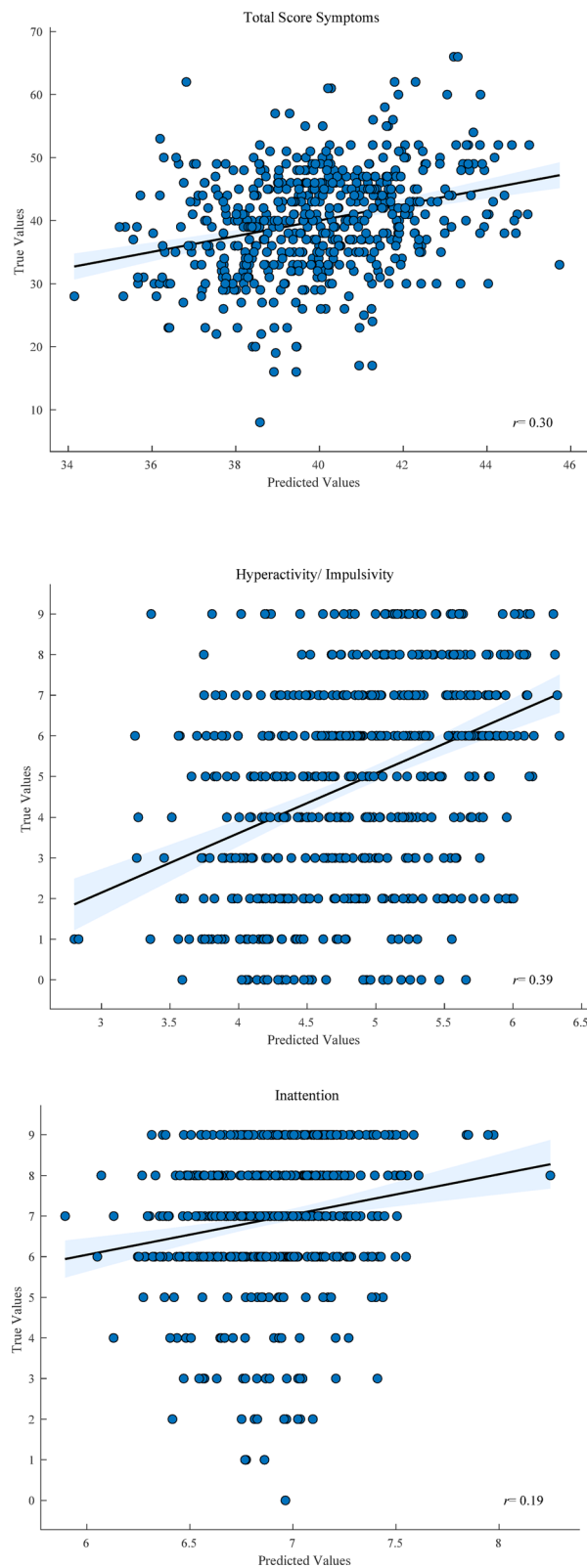


Fig. 2. Correlation of true and predicted symptom severity scores (from voice features) in ADHD subjects across scales of the ADHD Clinical Interview: total score symptoms, hyperactivity, inattention. Mean slope and 95% confidence intervals depicted; r = Pearson's correlation coefficient; all $p < 0.001$.

Confounding variables

To analyze confounders likely influencing voice analysis, we performed predictions of sex, age, and education (highest educational level) using RF classification and RF regression as appropriate in the entire sample (ADHD, PC, HC). Sex was classified with AUC=0.98. The correlations of true and predicted age ($r=0.52$, $t=18.5$, $p<0.001$) and education ($r=0.33$, $t=10.5$, $p<0.001$) likewise pointed to relevant associations with voice features and were thus included in all analyses as confounders of no interest. Moreover, sex ($t=5.4$, $p<0.001$), age ($t=-1.7$, $p=0.09$) and education ($t=5.9$, $p<0.001$) were associated with ADHD status and therefore included as covariates of no interest, as described above. Given concerns about changes in results due to confound leakage³⁶, we re-ran all analysis without confound removal. The results were largely identical, with changes of AUC between 0.00 and 0.05.

Since depression has been linked to changes of loudness in past research³⁷, we calculated associations of depression scores with voice features ($r=0.24$, $t=6.6$, $p<0.001$). However, depression scores were not related to ADHD symptom severity; thus, we did not include it as covariate in the main analysis. Including depression severity as covariate in a secondary analysis did not change the results.

Analysis of feature importance for the prediction of ADHD

The ten most important features were equally (50%) distributed to free speech (2 min), based on loudness changes over time frames between 0.02 and 0.252 s, and changes of F_0 in the counting task in timeframes of 0.02 to 0.1 s.

Discussion

In the context of a growing interest in analyzing voice as a potential biomarker of mental health³⁸, to the best of our knowledge, we have conducted the first study to investigate prosodic features in a large, heterogeneous sample of adults with ADHD. The results support and extend previous findings of paralinguistic abnormalities in ADHD and point toward the possibility of predicting ADHD in unseen individuals, particularly in early adulthood. The classification showed good prediction accuracy in differentiating ADHD from HC, and this was confirmed in associations of true and predicted symptom severity.

Previous studies in the emerging field of automatic assessments of mental disorders using speech were restricted to mood and psychotic disorders³⁸, and they showed comparable prediction accuracies to those reported in this study. For instance, depression can be predicted from (paralinguistic) voice features with an accuracy of about 80%³⁹. The prediction of schizophrenia has been successful using linguistic features, such as semantic relatedness, differentiating between individuals with schizophrenia or those at a high risk to develop acute symptoms and healthy controls⁴⁰. Studies that have investigated the prediction of ADHD from biological signals other than voice point towards a similar performance of approaches based on neuropsychological assessment⁴¹, EEG-measures⁴², questionnaires⁴³ or resting state fMRI⁴⁴. Thus, the findings in this study are comparable to previous research using voice to predict mental disorders and research relying on other biological signals to predict ADHD.

Moreover, correlations between true and predicted symptom scores, both in the whole sample and within the ADHD group, indicate that voice features were able to predict ADHD symptom severity. The continuous association of voice-features and ADHD severity, in turn, may help to interpret the lower prediction accuracy of ADHD from a PC group. The PC group presented for a diagnostic workup to rule out ADHD and showed markedly elevated ADHD symptoms as compared to HC. After an extensive diagnostic procedure, the participants in this group were not diagnosed with ADHD but with various other mental disorders, mostly affective disorders. Thus, since the classifier seems to be sensitive to lower levels of ADHD symptoms as indicated by the RF-regression analyses of symptom severity above, the classification is likely less suited to differentiate between ADHD-suspect subjects with lower symptoms and ADHD subjects who fulfill the clinical criteria. Moreover, the PC group was heterogeneous with regards to mental disorders, and this may contribute to the limited differentiation from ADHD. Future studies with larger sample size may be able to predict ADHD with higher precision and better differentiate ADHD participants from individuals with low or moderate ADHD symptoms. Likewise, other studies investigating potential biomarkers of ADHD point to a restricted differentiation from PC, such as studies of neuropsychological tests in combination with actigraphy⁴⁵ or studies using MRI⁴⁶.

A strength of the current approach is related to a robust classification performance of ADHD in the presence of comorbid mental disorders. The classifier showed a similar performance to differentiate ADHD from HC with and without comorbidity. Regarding the role of comorbidity, we additionally controlled the analyses for depression symptom severity with no changes in the AUC. Of note, participants with severe clinical comorbidity, such as schizophrenia or severe affective disorders were excluded from the analysis. Adult ADHD patients frequently present with latent or lifetime comorbid conditions⁴⁷, thus compromising the diagnostic accuracy of ADHD in clinical practice⁴⁸. Overall, the limited influence of comorbidity in this study is encouraging in terms of potential clinical applications; replication of our results pending. Since our data point to a relatively robust accuracy particularly in younger and female participants - possibly independent of comorbidity - automated voice analysis could evolve as a valuable addition to support the diagnostic process, considering marked diagnostic challenges amongst others due to comorbidity in this patient group⁴⁹.

Evaluating the prediction accuracy of inattentive and hyperactivity/impulsivity symptom scores, the RF-regression analyses pointed to a higher correlation of true and predicted symptoms of hyperactivity than of inattention in the ADHD group. Importantly, we noted a higher variance of the hyperactivity/impulsivity symptoms than of inattention in the ADHD group, as indicated by a significant Levene's test. Thus, the higher correlation of true and predicted hyperactivity/impulsivity scores may be explained by a better ability for the RF regression to learn from a sample with a larger variance. Using the whole sample (i.e., ADHD, PC, HC), inattention and hyperactivity subscales showed similar prediction scores of true and predicted symptoms (with

similar variance of both variables in this group). In line with this, the higher correlation of true and predicted hyperactivity/impulsivity symptoms did not translate into a significantly better prediction of the subgroup with combined (inattention and hyperactive/impulsive) symptoms.

With regard to emotion regulation or personality traits as a second important symptom cluster of ADHD, prior research has pointed to associations of personality traits and prosody of speech^{37,50}, including associations of impulsivity and F_0 / jitter⁵¹, that are related to the features in this study. Nilsen et al.¹⁸ reported higher speech volumes and pitch to be related to decreased inhibitory control, a key component of impulsivity. Moreover, a higher motor activity in ADHD, may result into higher subglottal pressures that have been reported in children with ADHD¹⁷. In summary, a relation of prosodic speech abnormalities and ADHD traits such as deficits in EF or hyperactivity/impulsivity seems plausible and should be evaluated in more detail in future research.

Discussing secondary findings, sex was classified with a high accuracy (AUC=0.98) using voice features. The high accuracy is in line with recent studies on predicting sex from prosodic voice features⁵². Furthermore, our data indicate a better classification performance of ADHD in female over male subjects, and this pertained to subsamples with different age-groups and non-comorbid ADHD-subjects. To some extent the sex related differences may be explained by slightly higher hyperactivity scores in female ADHD, but other measures of symptom severity or comorbidity did not differ between sexes. Thus, we assume that the voice features vary between male and female subjects and are possibly more pronounced in female ADHD subjects, though this must be investigated in future studies including the underlying biological mechanisms. One possible explanation may be related to sex specific dopamine receptor expression and functioning⁵³; given the important role of dopamine for voice production¹⁰, sex specific dopamine function may pertain to voice. Previous studies on voice in psychiatric disorders have mostly not differentiated between sex due to limited sample size, but this has been requested³⁸.

With respect to age, the voice features in this study were able to predict age with a comparable if slightly lower performance to other approaches predicting age from prosodic features⁵⁴. Research on changes of prosody in aging indicates higher scores of hoarseness, instability and breathiness in higher age⁵⁵, reflected by a diminished harmony to noise ratio⁵⁶. Moreover, with increasing age, the subglottal pressure decreases in line with a decrease in overall muscle mass, and subjects compensate for this with increased expiratory airflow⁵⁷. In study with children, increased hoarseness has been reported in ADHD^{17,58} as well as an increased subglottal pressure and decreased airflow¹⁷. Thus, if healthy subjects in older age may show increased hoarseness, which could also occur in ADHD, a possible difference in prosodic features due to hoarseness may decrease with older age and thus explain the decreased classification performance in older subjects. Since the finding on hoarseness in ADHD have been reported in children, our hypotheses related to age should be tested in adult populations in future research.

Stimulants have been reported to impact voice and prosody in a few small studies, including reports of lower F_0 , increased jitter⁵⁹, and increased hoarseness⁶⁰. In our study, we did not note differences with regard to classification accuracy and intake of stimulants, however the subsample with intake of stimulants was relatively small.

To understand which features are relevant to differentiate ADHD from HC, the analyses of feature importance firstly point to changes in loudness as an important differentiating feature. This finding supports previous research that identified voice anomalies in ADHD, frequently pertaining to loudness¹⁵. Breznitz et al. reported differences in temporal speech patterns and physical features of vocalization in boys with ADHD⁶¹ compared to PC (reading disabilities) and HC. Children with ADHD showed increased loudness, and children with combined ADHD type were louder, and showed lower F_0 ^{16,62}. We assume that subtle changes in voice based on higher subglottal pressure in ADHD¹⁷ may result in relevant changes of loudness to differentiate ADHD from healthy subjects.

Secondly, feature importance data indicate that speech test selection plays an important role in the classification of ADHD. Five of the ten features with the highest feature importance were related to counting (from one to ten) and five to free speech. It has been shown in previous research, that speech task (complexity) is related to vocal features, e.g. vocal variability differs between spontaneous speech and reading aloud⁶³ and a larger vocal variability was observed in a picture description task than in recalling autobiographical memories⁶⁴. Since EF mirrored by task complexity is related to voice and EF is known to be compromised in ADHD, a combination of speech tasks with varying complexity may be relevant for the differentiation of ADHD from HC and should be included in future trials.

In summary, we were able to differentiate ADHD patients from HC using voice features with good results in early adulthood and in female subjects. Strengths of the present study include the large, heterogeneous sample, the assessment of HC and PC and the application of criterion diagnostic procedures. The study's limitations include an imbalance in subgroup sizes, which may have particularly affected performance in male participants. Additionally, the exact time of day for each recording was not documented, which may have influenced the results.

We detected ADHD associated vocal patterns that likely reflect a disorder related vocal hyperfunction possibly related to altered dopamine signaling. These vocal patterns may be related to personality traits such as impulsivity, impaired EF, and changes in motor functioning pertaining to speech in ADHD. Neurobiologically, these traits are related to dopamine, which is thought to play an important role in the pathogenesis of ADHD⁹. Replication pending, a strength of voice-based features to predict ADHD might be the lack of impact from psychiatric comorbidity, that frequently occurs in ADHD and complicates the diagnostic process.

Given the feasibility and low cost to record and analyze voice, we see a potential value for future clinical application as a digital biomarker and encourage further investigation. A larger study sample including a larger PC group will be necessary to achieve a better differentiation of ADHD from PC with low or moderate ADHD symptoms. However, the clinical value of a voice-based screening might be in supporting the clinician to include

potential differential diagnoses rather than excluding differential diagnoses based on one test. In future research, the classification performance may increase with the addition of distinct speech features, such as speech pauses and the utilization of linguistic measures such as verbal fluency. Taken together, we consider voice analysis a promising avenue to support the diagnostic process in adult ADHD.

Data availability

The data that support the findings of this study are available from PeakProfiling GmbH with certain restrictions. The data were used under license for this study. Please contact co-author JL with requests. The code of the analyses was written in MATLAB and is available from the corresponding author (GP) upon request.

Received: 8 June 2024; Accepted: 9 May 2025

Published online: 24 May 2025

References

- Geissler, J. & Lesch, K. P. A lifetime of attention-deficit/hyperactivity disorder: diagnostic challenges, treatment and Neurobiological mechanisms. *Expert Rev. Neurother.* **11**, 1467–1484. <https://doi.org/10.1586/ern.11.136> (2011).
- Faraone, S. V. et al. Attention-deficit/hyperactivity disorder. *Nat. Rev. Dis. Primers.* **1**, 15020. <https://doi.org/10.1038/nrdp.2015.20> (2015).
- National Institute for care and excellence. *Attention deficit hyperactivity disorder: diagnosis and management*, (2018). <https://www.nice.org.uk/guidance/ng87/chapter/Recommendations#diagnosis>
- Du Rietz, E. et al. Self-report of ADHD shows limited agreement with objective markers of persistence and remittance. *J. Psychiatr. Res.* **82**, 91–99. <https://doi.org/10.1016/j.jpsychires.2016.07.020> (2016).
- 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. <https://doi.org/10.1038/tp.2017.164> (2017).
- Buitelaar, J. et al. Toward precision medicine in ADHD. *Front. Behav. Neurosci.* **16**, 900981. <https://doi.org/10.3389/fnbeh.2022.900981> (2022).
- Michellini, G., Norman, L. J., Shaw, P. & Loo, S. K. Treatment biomarkers for ADHD: taking stock and moving forward. *Translational Psychiatry.* **12**, 444. <https://doi.org/10.1038/s41398-022-02207-2> (2022).
- Chen, H. et al. Can biomarkers be used to diagnose attention deficit hyperactivity disorder? *Front. Psychiatry.* **14**, 1026616. <https://doi.org/10.3389/fpsy.2023.1026616> (2023).
- Dum, R., Ghahramani, A., Baweja, R. & Bellon, A. Dopamine receptor expression and the pathogenesis of Attention-Deficit hyperactivity disorder: a scoping review of the literature. *Curr. Dev. Disorders Rep.* <https://doi.org/10.1007/s40474-022-00253-5> (2022).
- Turk, A. Z., Marchoubeh, L., Fritsch, M., Maguire, I., SheikhBahaei, S. & G. A. & Dopamine, vocalization, and astrocytes. *Brain Lang.* **219**, 104970. <https://doi.org/10.1016/j.bandl.2021.104970> (2021).
- Faraone, S. V. et al. The world federation of ADHD international consensus statement: 208 Evidence-based conclusions about the disorder. *Neurosci. Biobehav. Rev.* **128**, 789–818. <https://doi.org/10.1016/j.neubiorev.2021.01.022> (2021).
- Netelenbos, N., Gibb, R. L., Li, F. & Gonzalez, C. L. R. Articulation speaks to executive function: an investigation in 4- to 6-year-olds. *Front. Psychol.* **9** <https://doi.org/10.3389/fpsyg.2018.00172> (2018).
- Shen, C. & Janse, E. Maximum speech performance and executive control in young adult speakers. *J. Speech Lang. Hear. Res.* **63**, 3611–3627. https://doi.org/10.1044/2020_jslhr-19-00257 (2020).
- Engelhardt, P. E., Nigg, J. T. & Ferreira, F. Is the fluency of Language outputs related to individual differences in intelligence and executive function? *Acta. Psychol.* **144**, 424–432 (2013).
- Machado-Nascimento, N., Melo, E. K. A. & Lemos, S. M. Speech-language pathology findings in attention deficit hyperactivity disorder: a systematic literature review. *Codas* **28**, 833–842. <https://doi.org/10.1590/2317-1782/20162015270> (2016).
- Hamdan, A. L. et al. Vocal characteristics in children with attention deficit hyperactivity disorder. *J. Voice.* **23**, 190–194. <https://doi.org/10.1016/j.jvoice.2007.09.004> (2009).
- Barona-Lleo, L. & Fernandez, S. Hyperfunctional voice disorder in children with attention deficit hyperactivity disorder (ADHD). A phenotypic characteristic?? *J. Voice.* **30**, 114–119. <https://doi.org/10.1016/j.jvoice.2015.03.002> (2016).
- Nilsen, E. S. et al. The influence of ADHD symptomatology and executive functioning on paralinguistic style. *Front. Psychol.* **7**, 1203. <https://doi.org/10.3389/fpsyg.2016.01203> (2016).
- Teresa, G. R. & Díaz-Román, T. M. Vocal hyperfunction in parents of children with attention deficit hyperactivity disorder. *J. Voice.* **30**, 315–321. <https://doi.org/10.1016/j.jvoice.2015.04.013> (2016).
- Etter, N. M., Cadely, F. A., Peters, M. G., Dahm, C. R. & Neely, K. A. Speech motor control and orofacial point pressure sensation in adults with ADHD. *Neurosci. Lett.* **744**, 135592. <https://doi.org/10.1016/j.neulet.2020.135592> (2021).
- Echternach, M., Burk, F., Burdumy, M., Traser, L. & Richter, B. Morphometric differences of vocal tract articulators in different loudness conditions in singing. *PLOS ONE.* **11**, e0153792. <https://doi.org/10.1371/journal.pone.0153792> (2016).
- Li, S., Nair, R. & Naqvi, S. M. Acoustic and text features analysis for adult ADHD screening: A Data-Driven approach utilizing DIVA interview. *IEEE J. Transl. Eng. Health Med.* **12**, 359–370. <https://doi.org/10.1109/jthm.2024.3369764> (2024).
- American psychiatric association. *diagnostic and statistical manual of mental disorders*. 4th edition, text rev. edn. (2000).
- Rösler, M. et al. Instrumente Zur diagnostik der Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung (ADHS) Im erwachsenenalter. *Der Nervenarzt.* **75**, 888–895. <https://doi.org/10.1007/s00115-003-1622-2> (2004).
- Quatieri, T. F. et al. in *Sixteenth Annual Conference of the International Speech Communication Association*. 452–455 (Keele University, Department of Psychology).
- Pyfrom, M., Lister, J. & Anand, S. Influence of cognitive load on voice production: A scoping review. *J. Voice.* <https://doi.org/10.1016/j.jvoice.2023.08.024> (2023).
- Bracewell, R. N. *The Fourier transform and its applications: McGraw-Hill series in electrical and computer engineering: Circuits and systems*. 3. ed., internat. ed. edn, 616Mc, (Graw Hill Higher Education, 2000).
- Langner, J. & Auhagen, W. Method for determining acoustic features of acoustic signals for the analysis of unknown acoustic signals and for modifying sound generation (2007).
- Kuny, S. & Stassen, H. H. Speaking behavior and voice sound characteristics in depressive patients during recovery. *J. Psychiatr. Res.* **27**, 289–307. [https://doi.org/10.1016/0022-3956\(93\)90040-9](https://doi.org/10.1016/0022-3956(93)90040-9) (1993).
- Langner, J., Kopiez, R., Stoffel, C. & Wilz, M. in *Proceedings of the 6th International Conference on Music Perception and Cognition*. 452–455 (Keele University, Department of Psychology).
- Langner, J. in *Proceedings of the Third Triennial ESCOM Conference, Uppsala, Sweden*. 713–719.
- Fastl, H. & Zwicker, E. *Psychoacoustics: Facts and Models*. 3rd Edition edn, 475Springer, (2007).
- Langner, J. Method of determining characteristics and quality factors of a piece of music Germany patent (2000).
- Nadeau, C. & Bengio, Y. Inference for the generalization error. *Mach. Learn.* **52**, 239–281. <https://doi.org/10.1023/A:1024068626366> (2003).

35. Mathworks. *Predictor importance estimates by permutation of out-of-bag predictor observations for random forest of regression trees - MATLAB*, < <https://www.mathworks.com/help/stats/regressionbaggedensemble.oobpermutedpredictorimportance.html>. (.
36. Hamdan, S. et al. Confound-leakage: confound removal in machine learning leads to leakage. *Gigascience* **12** <https://doi.org/10.1093/gigascience/giad071> (2022).
37. Cummins, N. et al. A review of depression and suicide risk assessment using speech analysis. *Speech Commun.* **71**, 10–40 (2015).
38. Low, D. M., Bentley, K. H. & Ghosh, S. S. Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investig Otolaryngol.* **5**, 96–116. <https://doi.org/10.1002/lio2.354> (2020).
39. Cordova, M. et al. Heterogeneity of executive function revealed by a functional random forest approach across ADHD and ASD. *Neuroimage Clin.* **26**, 102245. <https://doi.org/10.1016/j.nicl.2020.102245> (2020).
40. Spencer, T. J. et al. Lower speech connectedness linked to incidence of psychosis in people at clinical high risk. *Schizophr Res.* <https://doi.org/10.1016/j.schres.2020.09.002> (2020).
41. Slobodin, O., Yahav, I., Berger, I. A. & Machine-based prediction model of ADHD using CPT data. *Front. Hum. Neurosci.* **14**, 560021. <https://doi.org/10.3389/fnhum.2020.560021> (2020).
42. Wang, J. et al. Acoustic differences between healthy and depressed people: a cross-situation study. *BMC Psychiatry.* **19**, 300. <https://doi.org/10.1186/s12888-019-2300-7> (2019).
43. Christiansen, H. et al. Use of machine learning to classify adult ADHD and other conditions based on the Conners' adult ADHD rating scales. *Sci. Rep.* **10**, 18871. <https://doi.org/10.1038/s41598-020-75868-y> (2020).
44. Jung, M. et al. Surface-based shared and distinct resting functional connectivity in attention-deficit hyperactivity disorder and autism spectrum disorder. *Br. J. Psychiatry.* **214**, 339–344. <https://doi.org/10.1192/bjp.2018.248> (2019).
45. Hult, N., Kadesjo, J., Kadesjo, B., Gillberg, C. & Billstedt, E. ADHD and the qbtest: diagnostic validity of qbtest. *J. Atten. Disord.* **22**, 1074–1080. <https://doi.org/10.1177/1087054715595697> (2018).
46. 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.* **4**, 108–120. <https://doi.org/10.1016/j.bpsc.2018.06.003> (2019).
47. Andersson, A. et al. Research review: the strength of the genetic overlap between ADHD and other psychiatric symptoms – a systematic review and meta-analysis. *J. Child Psychol. Psychiatry.* **61**, 1173–1183. <https://doi.org/10.1111/jcpp.13233> (2020).
48. Katzman, M. A., Bilkey, T. S., Chokka, P. R., Fallu, A. & Klassen, L. J. Adult ADHD and comorbid disorders: clinical implications of a dimensional approach. *BMC Psychiatry.* **17**, 302. <https://doi.org/10.1186/s12888-017-1463-3> (2017).
49. Nussbaum, N. L. ADHD and female specific concerns: A review of the literature and clinical implications. *J. Atten. Disord.* **16**, 87–100. <https://doi.org/10.1177/1087054711416909> (2011).
50. Bedi, G. et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophr.* **1**, 15030. <https://doi.org/10.1038/npjshcz.2015.30> (2015).
51. Guidi, A., Gentili, C., Scilingo, E. P. & Vanello, N. Analysis of speech features and personality traits. *Biomed. Signal Process. Control.* **51**, 1–7. <https://doi.org/10.1016/j.bspc.2019.01.027> (2019).
52. Uddin, M. A., Hossain, M. S., Pathan, R. K. & Biswas, M. in *2020 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*. 1–7.
53. Williams, O. O. F., Coppolino, M., George, S. R. & Perreault, M. L. Sex differences in dopamine receptors and relevance to neuropsychiatric disorders. *Brain Sci.* **11**, 1199 (2021).
54. Sadjadi, S. O., Ganapathy, S. & Pelecanos, J. W. in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 5040–5044. (2016).
55. Rojas, S., Kefalianos, E. & Vogel, A. How does our voice change as we age?? A systematic review and Meta-Analysis of acoustic and perceptual voice data from healthy adults over 50 years of age?. *J. Speech Lang. Hear. Res.* **63**, 533–551. https://doi.org/10.1044/2019_JSLHR-19-00099 (2020).
56. Dehqan, A., Scherer, R. C., Dashti, G., Ansari-Moghaddam, A. & Fanaie, S. The effects of aging on acoustic parameters of voice. *Folia Phoniatr. Et Logopaedica.* **64**, 265–270. <https://doi.org/10.1159/000343998> (2012).
57. Lewandowski, A. et al. Adult normative data for phonatory aerodynamics in connected speech. *Laryngoscope* **128**, 909–914. <https://doi.org/10.1002/lary.26922> (2018).
58. Hamdan, A. L. et al. Vocal characteristics in children with attention deficit hyperactivity disorder. *J. Voice.* **23**, 190–194. <https://doi.org/10.1016/j.jvoice.2007.09.004> (2009).
59. Bloch, Y. et al. Methylphenidate mediated change in prosody is specific to the performance of a cognitive task in female adult ADHD patients. *World J. Biol. Psychiatry.* **16**, 635–639. <https://doi.org/10.3109/15622975.2015.1036115> (2015).
60. Yalcin, O., Aslan, A. A., Sari, B. A. & Turkbay, T. Possible methylphenidate related hoarseness and disturbances of voice quality: two pediatric cases. *Klinik Psikofarmakoloji Bülteni-Bulletin Clin. Psychopharmacol.* **22**, 278–282. <https://doi.org/10.5455/bcp.20120731061626> (2012).
61. Breznitz, Z. The speech and vocalization patterns of boys with ADHD compared with boys with dyslexia and boys without learning disabilities. *J. Genet. Psychol.* **164**, 425–452. <https://doi.org/10.1080/00221320309597888> (2003).
62. Turgay, A. & Ansari, R. Major depression with ADHD. *Psychiatry (Edmont)*. **3**, 20–32 (2006).
63. Alghowinem, S. et al. Detecting depression: A comparison between spontaneous and read speech. *Int. Conf. Acoust. Spee*, 7547–7551 (2013).
64. Cohen, A. S., Kim, Y. & Najolia, G. M. Psychiatric symptom versus neurocognitive correlates of diminished expressivity in schizophrenia and mood disorders. *Schizophr. Res.* **146**, 249–253. <https://doi.org/10.1016/j.schres.2013.02.002> (2013).
65. Ward, M. F., Wender, P. H. & Reimherr, F. W. The Wender Utah Rating-Scale - an aid in the retrospective diagnosis of childhood Attention-Deficit hyperactivity disorder. *Am. J. Psychiat.* **150**, 885–890 (1993).
66. Retz-Junginger, P. et al. Wender Utah rating scale (WURS-k): die Deutsche Kurzform Zur retrospektiven erfassung des hyperkinetischen syndroms Bei erwachsenen. [Wender Utah rating scale: The short-version for the assessment of the attention-deficit hyperactivity disorder in adults]. *Der Nervenarzt.* **73**, 830–838. <https://doi.org/10.1007/s00115-001-1215-x> (2002).

Acknowledgements

The authors wish to thank Michael Colla and Laura Gentschow, who enabled and conducted data collection together with DL. We thank Paula Kunze and Inga Leerhoff for their great support in data collection and study coordination. DL wishes to thank Rolf Dietmar Wolf, Gerlafingen, Switzerland, for the inspiration and support to design the study. We extend our gratitude to all participants for their efforts in taking part.

Author contributions

D.L. and J.L. conceived and planned the experiments; G.P., J.V. and E.A. conducted literature research; J.L. calculated audio features; G.P., K.P., S.E. and F.H. designed the model and the computational framework and analyzed the data. G.P., E.A., and J.V. wrote the manuscript. E.A. and D.L. supervised and verified the data acquisition, patient enrolment, patient consent, and patient characteristics. S.E., K.P., D.L., J.L., F.H., A.K. and G.P. revised the final manuscript. All authors discussed the results and contributed to the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing interests

EA and GP participated and received payments in the national advisory board ADHD of Shire/Takeda. JL is co-founder and CTO of PeakProfiling GmbH. He created audio-features used in this study, that are intellectual property of PeakProfiling GmbH. FH and AK received payments by PeakProfiling GmbH. JV became an employee at PeakProfiling after finishing the manuscript. KR, SE and DL declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-01989-x>.

Correspondence and requests for materials should be addressed to G.G.P.

Reprints and permissions information is available at www.nature.com/reprints.

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

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025