

Original Paper

Acoustic and Natural Language Markers for Bipolar Disorder: A Pilot, mHealth Cross-Sectional Study

Cristina Crocamo¹, PhD; Riccardo Matteo Cioni¹, MD; Aurelia Canestro¹, MD; Christian Nasti¹, MD; Dario Palpella¹, MD; Susanna Piacenti¹, MD; Alessandra Bartocetti¹, MD; Martina Re¹, MD; Valentina Simonetti², MSc; Chiara Barattieri di San Pietro^{2,3}, PhD; Maria Bulgheroni², MSc; Francesco Bartoli¹, MD, PhD; Giuseppe Carrà¹, MD, MSc, PhD

¹School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy

²Ab.Acus, Milan, Italy

³Laboratory of Neurolinguistics and Experimental Pragmatics (NEP), University School for Advanced Studies IUSS, Pavia, Italy

Corresponding Author:

Cristina Crocamo, PhD
School of Medicine and Surgery
University of Milano-Bicocca
via Cadore 48
Monza, 20900
Italy
Phone: 39 0264488483
Email: cristina.crocamo@unimib.it

Abstract

Background: Monitoring symptoms of bipolar disorder (BD) is a challenge faced by mental health services. Speech patterns are crucial in assessing the current experiences, emotions, and thought patterns of people with BD. Natural language processing (NLP) and acoustic signal processing may support ongoing BD assessment within a mobile health (mHealth) framework.

Objective: Using both acoustic and NLP-based features from the speech of people with BD, we built an app-based tool and tested its feasibility and performance to remotely assess the individual clinical status.

Methods: We carried out a pilot, observational study, sampling adults diagnosed with BD from the caseload of the Nord Milano Mental Health Trust (Italy) to explore the relationship between selected speech features and symptom severity and to test their potential to remotely assess mental health status. Symptom severity assessment was based on clinician ratings, using the Young Mania Rating Scale (YMRS) and Montgomery-Åsberg Depression Rating Scale (MADRS) for manic and depressive symptoms, respectively. Leveraging a digital health tool embedded in a mobile app, which records and processes speech, participants self-administered verbal performance tasks. Both NLP-based and acoustic features were extracted, testing associations with mood states and exploiting machine learning approaches based on random forest models.

Results: We included 32 subjects (mean [SD] age 49.6 [14.3] years; 50% [16/32] females) with a MADRS median (IQR) score of 13 (21) and a YMRS median (IQR) score of 5 (16). Participants freely managed the digital environment of the app, without perceiving it as intrusive and reporting an acceptable system usability level (average score 73.5, SD 19.7). Small-to-moderate correlations between speech features and symptom severity were uncovered, with sex-based differences in predictive capability. Higher latency time ($\rho=0.152$), increased silences ($\rho=0.416$), and vocal perturbations correlated with depressive symptomatology. Pressure of speech based on the mean intraword time ($\rho=-0.343$) and lower voice instability based on jitter-related parameters (ρ ranging from -0.19 to -0.27) were detected for manic symptoms. However, a higher contribution of NLP-based and conversational features, rather than acoustic features, was uncovered, especially for predictive models for depressive symptom severity (NLP-based: $R^2=0.25$, mean squared error [MSE]=110.07, mean absolute error [MAE]=8.17; acoustics: $R^2=0.11$, MSE=133.75, MAE=8.86; combined: $R^2=0.16$; MSE=118.53, MAE=8.68).

Conclusions: Remotely collected speech patterns, including both linguistic and acoustic features, are associated with symptom severity levels and may help differentiate clinical conditions in individuals with BD during their mood state assessments. In the future, multimodal, smartphone-integrated digital ecological momentary assessments could serve as a powerful tool for clinical purposes, remotely complementing standard, in-person mental health evaluations.

Keywords: digital mental health; remote assessment; mHealth; speech; NLP; natural language processing; acoustic; symptom severity; machine learning; markers; mental health; bipolar disorders; app; applications; multimodal; mobile health; voice; vocal; bipolar; verbal; emotion; emotional; psychiatry; psychiatric; mental illness

Introduction

Bipolar disorder (BD) is a lifelong, episodic illness characterized by mood recurrences, including manic or hypomanic, depressive, and mixed episodes [1-3]. The burden associated with BD, affecting families, carers, and mental health care systems, is heavy [4]. Community services often struggle in delivering regular monitoring of BD treatment needs, resulting in relapses that seem difficult to predict [4-6].

Language disturbances are among the core symptoms of acute episodes in BD, since speech patterns are modulated by the emotional and neurophysiological status [7,8]. Therefore, language may play a key role in the assessment of an individual's current experiences, emotions, thought patterns, and symptoms. While content analysis may reveal grandiosity associated with elevated mood, impulsivity, or changes in goal-directed activities, natural language may provide insights into mood fluctuations, cognitive processes, and behavioral patterns [9]. In particular, changes in the rate of speech are likely to indicate mood oscillations, including pressure of speech and increased verbosity during manic episodes [10] and poverty of speech and increased pause times during depressive episodes [11-13]. Clinicians are trained to recognize variations in language and voice, along with gestures and facial expressions, implicitly assessing both coherence and organization of speech and natural language features. However, this process is inevitably vulnerable to inconsistencies and biases.

Recent research in mental health and computer science has put forward computational approaches for speech analysis across a variety of mental disorders, proposing automated methods to assess and monitor the individual's mental state through speech patterns [14-18]. Promising techniques in speech acoustic signal processing [10,11,17,19-21], using mobile health (mHealth) technology, can bridge subjective and objective components across various stages, such as prediction of illness onset, diagnostic processes, assessment of severity, and forecast of treatment outcomes [22-25]. Indeed, natural language processing (NLP) techniques, exploring language resources (eg, lexical choices, syntax, and semantics) both qualitatively and quantitatively (eg, topic modeling, clustering, and classification), may produce deeper insights across different clinical conditions [9,26]. For example, observable linguistic traits (eg, increased use of both first-person pronouns and negative emotion expressions) can be identified among people with BD [23]. However, although linguistic features are informative, they are context-dependent and inferred according to word transcriptions [27]. Thus, speech analyses combining acoustic-dependent features (eg, speech prosody and voice quality) with NLP-based measures appear more promising in terms of model predictions,

possibly providing a more accurate mental health assessment [23,27,28].

Indeed, research has shown that acoustic features are markers of emotional states in BD [29], and that quantifiable speech differences can predict the scores of scales such as the Young Mania Rating Scale (YMRS) and the Montgomery-Åsberg Depression Rating Scale (MADRS) [13,27]. On the other hand, recent evidence has shown how smartphone-based voice data [30] can enhance BD monitoring in real time, detecting possible mood changes [31,32]. Thus, speech-based systems embedded in smartphones might be useful tools for complementary, continuous assessments of BD clinical states. We therefore built an app-based tool, jointly using acoustic and NLP-based features from the speech of people with BD who delivered a narrative, and carried out a pilot study aimed at testing its feasibility and performance to remotely assess the individual clinical status. Continuous, uninterrupted spoken accounts, as supplied by individuals, provided the unique opportunity to combine communication style information from an in-depth set of acoustic features and NLP-based scores as potential digital markers of symptom severity in speech. We rigorously chose to test the tool's performance against standard psychometric assessments of mania and depression in order to explore its potential for remote, complementary assessments.

Methods

The report of this study adheres to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement (checklist presented in [Multimedia Appendix 1](#)) [33].

Study Design and Sampling Strategies

We conducted a pilot, cross-sectional study involving adult participants (aged 18 years or older) from the caseload of the Nord Milano Mental Health Trust (Italy). The Trust includes 2 psychiatric intensive care units, with a total of 27 beds, and also provides community mental health care for the same 280,000 inhabitants of the northern area of the Metropolitan City of Milan through 4 community mental health teams with multidisciplinary staff. The relevant catchment area comprises highly urbanized, both deprived and affluent, districts.

Inclusion criteria comprised a diagnosis of BD and the willingness to participate in the study. People with physical impairments affecting their acoustic capabilities were excluded. Based on inclusion and exclusion criteria, eligible individuals were identified among individuals consecutively admitted to the Trust. Then, they were approached by the research team, explaining the purpose of the study and, if any, potential risks.

Ethical Considerations

Recruitment efforts were carried out in accordance with ethical guidelines to ensure the well-being and safety of all participants. Study participants signed a written informed consent and were not compensated for their involvement. The study received ethical approval (protocol number 172-17032023) from the local ethical committee. To maintain participant privacy and confidentiality, all study data were pseudonymized prior to analysis. No individual participants are identifiable in any images included in this manuscript or Multimedia Appendices.

Procedures

Acoustic data were retrieved by asking participants to self-administer verbal performance tasks through a mobile app on their smartphones (SPEAKapp; [34]). Clinical testing and app usage took place on the same day in the study setting (inpatient and outpatient services). Then, the System Usability Scale (SUS), a short 10-item questionnaire based on a 5-point Likert scale, was administered to assess the usability [35] of the app.

Verbal performance in terms of prose recall was based on the Babcock test [36], for which participants were asked to listen to a short story characterized by graphic and intense contents (eg, a death in a car crash) and then to repeat what she or he remembered from this narrative. This enabled to capture speech timing patterns based on sustained speech samples.

The app gathered participants' verbal production by using the smartphone-integrated microphone, recording and processing participants' speech by leveraging Google Speech-To-Text APIs [37] and Python libraries (eg, Parselmouth for the Praat software [38]). Recordings involved the use of one audio channel based on the participant's voice in a controlled environment with minimal acoustic conditions. Both the raw audio data and the transcribed text content were processed to extract acoustic and NLP-based features from speech outputs. NLP and acoustic signal models were embedded in the backend part of the mobile app.

Measures

Consistent with recent evidence, we assumed speech as verbal behavior, the spoken output of the mental system underlying the language [39]. Through speech recognition, acoustic and linguistic features were extracted. Then, based on both NLP and acoustic features, we considered a multidimensional framework in order to generate appropriate discriminative information for the potential use of speech patterns as digital markers in BD [27,31]. A full description of selected features is provided in Table S1 in [Multimedia Appendix 2](#).

NLP-Based, Semantic, and Conversational Indices

NLP-based scores were computed according to distributional semantic models, encompassing vectorial representations for the meaning of words in a multi-dimensional space.

Standard linguistic scores included both the number of words, indicative of poverty of speech, and the number of words produced that matched the story text. On the other hand, novel NLP-based scores integrated mean intraword time, estimating the average time taken to articulate or pronounce subsequent words, as an indicator of processing speed, as well as word mover's distance (WMD), capturing both lexical overlap and semantic similarity. In particular, WMD was estimated as the minimum cumulative distance between words required to exactly match the point cloud of the text of the full correct story (ie, the content distance between the full correct story and the story narrative produced by the participant), thus incorporating the semantic similarity between individual word pairs into the word distance metric [40]. In addition, latency time was calculated as a novel NLP-based score, taking into account the delay between the initiation of a spoken utterance or action and the production of its intended outcome or response when starting the task (ie, the first word).

Additional objective information was extracted from speech data. These quantitative measures included (1) speech duration, (2) speaking time (ie, phonation), (3) silence, (4) ratios of speaking time to speech duration as well as of silence to speaking time, and (5) speech rate.

Acoustic Indices From Vocal Signals (Prosodic Cues Indices)

Measures for prosodic cues (acoustic indices quantifying how people talk during conversations) were based on the signal's frequency and energy or amplitude. These were assumed to contribute to conveying paralinguistic meaning [41]. Based on nontextual data, acoustic components of speech were defined as the key phonetic elements, that is, objectively and reproducibly quantified speech sounds [27,42]. Fundamental frequency (F0) was measured by the frequency of phonation [43]. The short-term instability of the vibration of the vocal cords during phonation (ie, jitter-related indices) was also extracted (Table S1 in [Multimedia Appendix 2](#)). Higher jitter values indicated speech patterns likely characterized by irregularities or hesitations, thus mirroring potential underlying psychological distress or emotional instability. Furthermore, microperturbations of the amplitude of the signal (ie, how variable acoustic peaks refer to the period-to-period variability of the signal peak-to-peak amplitude) were identified as small fluctuations in the intensity of vocal sound waves by shimmer-related measures, with higher values indicating greater variability or instability, while lower ones suggesting more stable vocal intensity (ie, smoother and more regular speech production).

Since both periodic and nonperiodic sound waves may characterize the voice, the mean harmonics-to-noise ratio was used to measure the relationship between harmonic and nonharmonic voice elements. Noisier, more raucous voices (ie, not smooth or clear) were expected to show lower harmonics-to-noise ratios, indicating vocal cord tension or irritation, possibly suggesting emotional distress.

Psychometric Measures

Diagnosis of BD was confirmed by the Structured Clinical Interview for *DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; SCID-5)*. Based on clinician-rated assessments, depressive symptom severity was measured by the MADRS [44], while YMRS was used to assess manic symptoms [45]. Scores ranged from 0 to 60 for both MADRS [44] and YMRS [45]. In addition, cutoffs for severe mood symptoms were either a YMRS score ≥ 20 [46,47] or a MADRS score ≥ 19 [48].

Statistical Analyses

First, we summarized participants' characteristics, providing standard statistics for continuous and categorical variables. For both MADRS and YMRS, continuous scores were used. However, a supplementary analysis was performed based on clinically meaningful thresholds for symptom severity. A bivariate analysis was then carried out to measure the strength of the potential association between speech indices and psychometric measures. Features' summary statistics were plotted, and correlation coefficients (Pearson and Spearman, according to assumptions on data distribution, eg, normality) were estimated. Color gradient heat plots were also generated for data visualization. Taking into account potential sex differences in speech acoustic indices [49-51], subgroup analyses were performed. Statistical significance was set at $P < .05$.

Second, based on state-of-the-art algorithms, NLP and acoustic features extracted from natural language and audio streams (Table S1 in [Multimedia Appendix 2](#)) were used to train machine-learning models to detect depressive and manic states by means of scores from MADRS and YMRS. Data were randomly split using a 5-fold nested cross-validation approach for training and testing in order to provide an unbiased evaluation of the model's performance. In particular, random forest (RF) models, with the potential

to handle both linear and nonlinear relationships between features and the target variable, were implemented. The supervised learning algorithm, with no assumptions about the distribution of the target variable, was based on the ensemble learning method of different decision trees, whose predictions were aggregated using the scikit-learn library in Python. Exploiting the bagging techniques, building multiple decision trees, RF contributed to minimizing overfitting issues by randomizing the feature selection during each tree split. This was assumed to reduce sensitivity to noise and to make decision trees less correlated through the use of a unique subset of the initial data for every base model. Moreover, we deemed features scaling unnecessary due to both the properties of the RF model and the performance metrics of comparisons. Relevant models were trained to test final performance by metrics (ie, mean squared error [MSE], mean absolute error [MAE], and R -squared [R^2]). These tested overall performance, even controlling for sex. Shapley Additive Explanations values, showing features' impact, were plotted. Data were analyzed using Stata release 18 and Python (version 3.10.9).

Results

Sample Characteristics

We included 32 subjects with BD (mean age 49.6, SD 14.3 years; 50% [16/32] females). The mean (SD) age at onset was 24.4 (10) years. As a whole, participants experienced more manic (median 4, IQR 8) than depressive episodes (median 2, IQR 5). About 40% (12/32) of participants reported a previous mood episode within 1 year before study enrollment. The MADRS median (IQR) score was 13 (21), while the YMRS median (IQR) score was 5 (16). Considering the app usage, participants reported high SUS scores on average (mean 73.5, SD 19.7). Demographic and clinical details are fully provided in [Table 1](#).

Table 1. Sample^a characteristics.

Characteristics	BD ^b (N=32)
Sex, n (%)	
Female	16 (50)
Male	16 (50)
Age (years), mean (SD)	49.6 (14.3)
Marital status, n (%)	
In a relationship	12 (37)
Family situation, n (%)	
Living alone	11 (34)
Education, n (%) ^c	
Elementary	1 (3)
Middle	12 (37)
High	13 (41)
University or superior	5 (16)
Employment, n (%) ^c	
Employed	13 (41)

Characteristics	BD ^b (N=32)
Setting, n (%)	
Outpatient	13 (41)
Inpatient	19 (59)
Polarity of first episode, n (%) ^c	
Depressive	12 (38)
Hypomaniac or maniac	13 (41)
Unknown	1 (3)
Age of onset ^c (years), mean (SD)	24.4 (10)
Family history ^c	11 (34%)
Hospitalizations, median (IQR)	
Lifetime	3 (7.5)
12 months	1 (2)
Suicide attempts (lifetime), n (%)	10 (31)
Alcohol use disorder (lifetime), n (%)	3 (9)
Substance use disorder (lifetime), n (%)	6 (19)
Medication, n (%)	
FGA ^d	6 (19)
SGA ^e	28 (87)
Mood stabilizer	26 (81)
Antidepressant	8 (25)
Benzodiazepine	16 (50)
Psychometric assessment	
Depressive symptoms (MADRS ^f), median (IQR)	13 (21)
MADRS <19, n (%)	17 (53)
MADRS ≥19, n (%)	15 (47)
Manic symptoms (YMRS ^g), median (SD)	5 (16)
YMRS <20, n (%)	26 (81)
YMRS ≥20, n (%)	6 (19)
SUS ^h score, mean (SD)	73.5 (19.7)

^aThe sample is for a pilot, cross-sectional study in Italy.

^bBD: bipolar disorder.

^cMissing values: education (1), employment (2), age of onset (10), polarity of first episode (6), family history (10), alcohol use disorder (2), substance use disorder (1), FGA (2), SGA (1), mood stabilizer (2), antidepressant (3), benzodiazepine (4).

^dFGA: first-generation antipsychotics.

^eSGA: second-generation antipsychotics.

^fMADRS: Montgomery-Åsberg Depression Rating Scale.

^gYMRS: Young Mania Rating Scale.

^hSUS: System Usability Scale (range 0-100).

Associations Between Symptom Severity and Speech Features

For descriptive purposes, NLP-based, conversational, and acoustic features are summarized in Figures S1A-S1D and S2A-S2D in [Multimedia Appendix 3](#) by depressive and manic symptom severity, respectively.

In particular, grouping data into 2 categories ([Multimedia Appendix 3](#)), statistically significant differences by depressive symptoms' severity were found for many NLP-based and conversational-like measures, including word number, phonation (also as percentage over the speech duration), and mean intraword time. Correlation analyses, based on Spearman nonparametric analysis of symptom severity

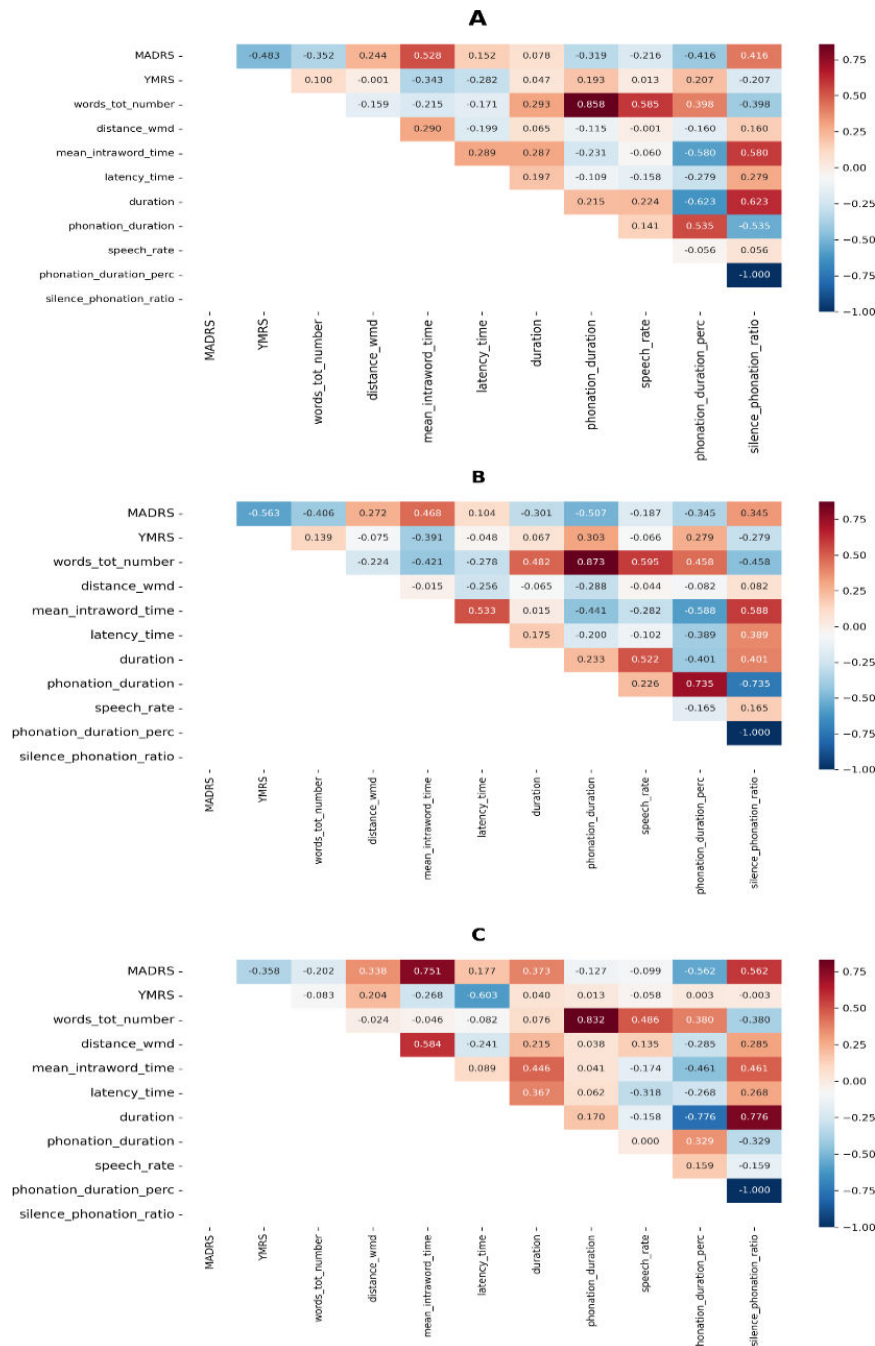
continuous scores, are displayed in [Figures 1A-C](#) and [2A-C](#). These showed that both the total number of words and the length of phonation, as well as the related percentage out of segment duration, were negatively correlated (coefficients=−0.35, −0.32, and −0.42) to depressive symptoms ([Figure 1A](#)). Consistent results were observed for the ratio between silence and phonation (coefficient=0.42), as well as for mean intraword time, which was positively correlated to depressive (coefficient=0.53) and negatively to manic (coefficient=−0.34) symptoms. Among items for depressive symptoms assessment, this correlation was particularly clear between acoustic features and suicidal thoughts (coefficients ranging from 0.18 to 0.51). In addition, latency time also showed a moderate, though obviously opposite, correlation

with manic and depressive symptoms, respectively (coefficients=-0.28 and 0.15).

Subgroup analyses for NLP-based and conversational features revealed more pronounced relationships in females (Figure 1C) as compared with males (Figure 1B), showing a high correlation between depressive symptoms and mean

intraword time (coefficient=0.75), phonation percentage (coefficient=-0.56), and, consequently, the silence-phonation ratio (coefficient=0.56). Similarly, latency time was negatively correlated to manic symptoms among females (coefficient=-0.60).

Figure 1. Correlation heatmap of NLP-based, semantic and conversational features in people with bipolar disorder. (A) Overall sample; (B) Male subgroup; (C) Female subgroup. MADRS: Montgomery-Åsberg Depression Rating Scale, YMRS: Young Mania Rating Scale.



On the other hand, a small positive correlation was uncovered between depressive symptoms and higher values of instability in speech patterns (jitter-related indices, with coefficients ranging from 0.10 to 0.16). In contrast, small-to-moderate negative correlations were observed between

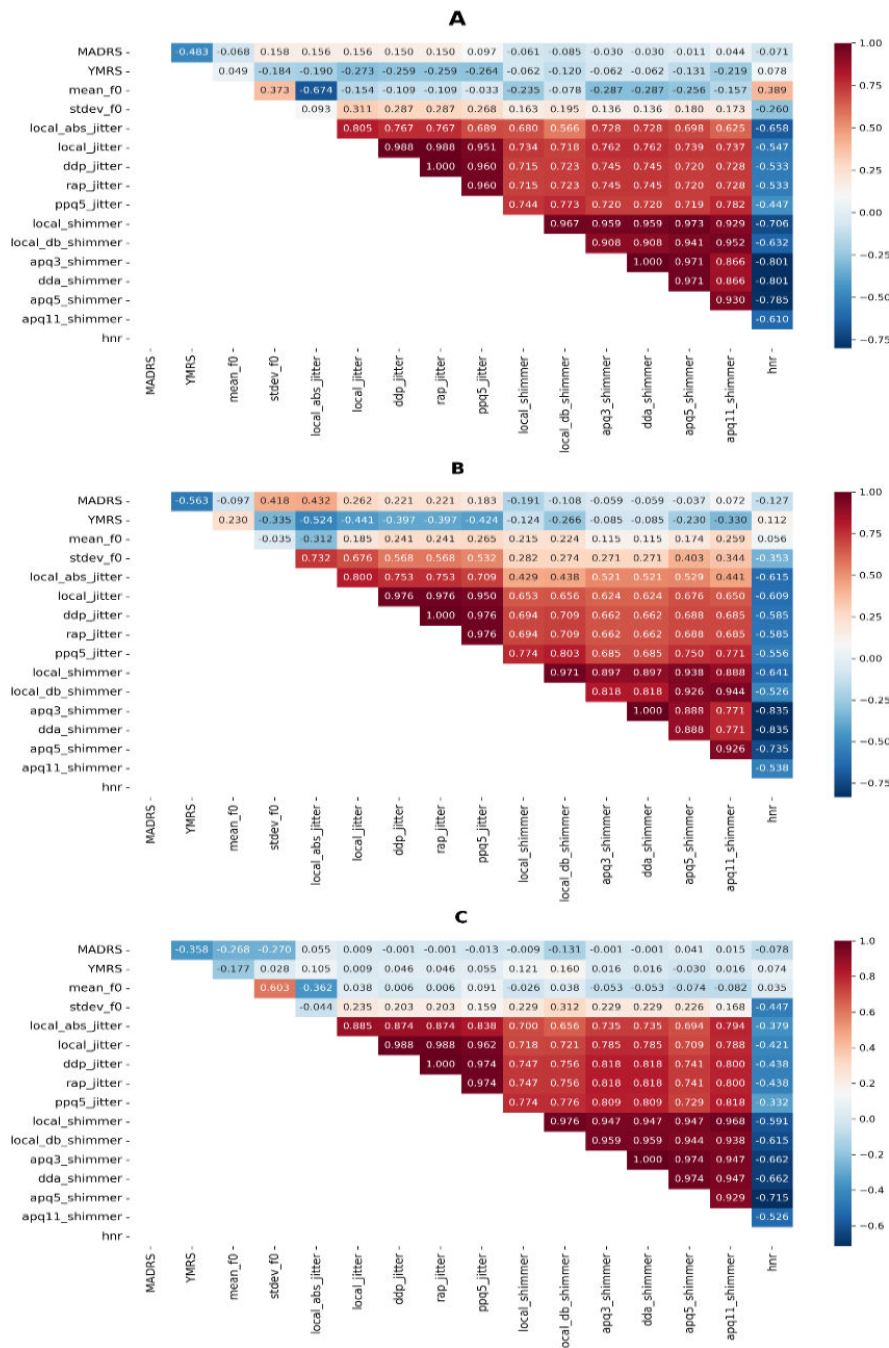
manic symptoms and lower values of instability (jitter-related indices, with coefficients ranging from -0.19 to -0.27). Small estimates were found for F0, respectively (coefficient=0.16 and -0.18; Figure 2A). Except for shimmer_apq11 (manic symptoms coefficient=-0.22), we did not find any substantial

relationship between shimmer-related indices (describing stable and unstable vocal intensity and speech production) and symptomatology.

connections in males as compared with females, especially in terms of F0 and jitter-related indices (Figure 2B and C).

Subgroup analyses suggested a role for sex also in influencing acoustic features. In particular, we found deeper

Figure 2. Correlation heatmap of acoustic features in people with bipolar disorder. (A) Overall sample; (B) Male subgroup; (C) Female subgroup. MADRS: Montgomery-Åsberg Depression Rating Scale, YMRS: Young Mania Rating Scale.



Predictive Models From Speech Features

Considering depressive symptoms, performance metrics showed a contribution of NLP-based and conversational features higher than what was attributable to acoustic ones

(Table 2). In particular, mean intraword time, silence-phonation ratio, ppq5 jitter (ie, perturbations in F0), WMD, and percentage of phonation over duration all ranked high in terms of relative importance.

Including sex into the analysis, a differential contribution of various features (NLP-based and conversational vs acoustics) to the predictive models for depressive (Figure 3A) and manic (Figure 3B) symptoms can be found. However, as for manic symptoms, although a relative contribution of

different NLP-based and acoustic (eg, F0 SD) features was recorded, we could not find any reliable estimates for the relevant model, even including sex. Table 2 shows detailed estimated performance metrics for testing for the trained RF regressors, even controlling for sex.

Table 2. Performance estimates for random forest regression models in people with bipolar disorder.

Performance ^a	Depressive symptoms		Manic symptoms	
	Unadjusted	Adjusted ^b	Unadjusted	Adjusted ^b
NLP^c				
R^2 average	0.26	0.25	— ^d	—
Fold 1	0.10	-0.55	-0.54	0.18
Fold 2	0.48	0.53	-0.13	0.02
Fold 3	0.06	0.37	0.25	0.12
Fold 4	0.54	0.64	0.23	-0.42
Fold 5	0.13	0.26	0.01	147.98
Mean squared error average	105.46	110.07	153.78	147.98
Fold 1	136.73	259.26	223.06	92.05
Fold 2	33.02	46.49	121.79	156.74
Fold 3	137.64	104.25	135.66	121.60
Fold 4	79.32	33.00	134.85	167.15
Fold 5	140.61	107.35	153.32	202.35
Mean absolute error average	8.08	8.17	10.58	10.13
Fold 1	9.58	13.64	12.47	7.79
Fold 2	3.36	5.57	9.28	10.90
Fold 3	10.31	8.56	10.40	9.26
Fold 4	7.59	4.34	9.82	9.71
Fold 5	9.26	8.74	10.96	13.00
Acoustics				
R^2 average	—	0.11	—	—
Fold 1	0.29	-0.22	0.002	-0.22
Fold 2	-0.83	-0.10	-0.02	-0.15
Fold 3	-0.59	0.03	-0.14	-0.14
Fold 4	0.23	0.18	-0.38	-0.44
Fold 5	0.36	0.64	-0.28	0.01
Mean squared error average	161.64	133.75	162.86	163.51
Fold 1	47.97	222.18	68.9	125.14
Fold 2	333.00	200.40	160.47	122.80
Fold 3	202.17	85.04	185.63	175.34
Fold 4	128.54	148.52	272.25	225.23
Fold 5	96.49	12.62	127.06	170.30
Mean absolute error average	10.02	8.86	10.35	10.73
Fold 1	5.27	11.76	7.09	9.77
Fold 2	16.43	13.8	10.94	8.82
Fold 3	11.77	6.00	12.48	11.70
Fold 4	9.69	10.13	14.05	12.34
Fold 5	6.95	2.62	7.21	11.01
Combined				
R^2 average	0.05	0.16	—	—
Fold 1	0.32	0.60	-0.56	-0.13

Performance ^a	Depressive symptoms		Manic symptoms	
	Unadjusted	Adjusted ^b	Unadjusted	Adjusted ^b
Fold 2	-0.09	0.11	0.24	0.07
Fold 3	-0.29	0.07	0.08	-0.54
Fold 4	0.10	0.04	0.06	0.18
Fold 5	0.22	0.22	-0.41	0.14
Mean squared error average	120.90	118.53	135.94	140.03
Fold 1	87.51	34.13	158.54	183.39
Fold 2	183.67	111.11	60.67	122.32
Fold 3	184.71	164.45	192.81	126.43
Fold 4	47.71	148.83	178.84	112.36
Fold 5	100.91	134.10	88.83	155.68
Mean absolute error average	8.65	8.68	9.61	10.00
Fold 5	6.69	4.37	11.21	11.50
Fold 5	11.49	7.33	6.95	8.86
Fold 5	11.04	11.29	11.40	10.30
Fold 5	5.57	10.15	11.19	8.26
Fold 5	8.46	10.26	7.28	11.08

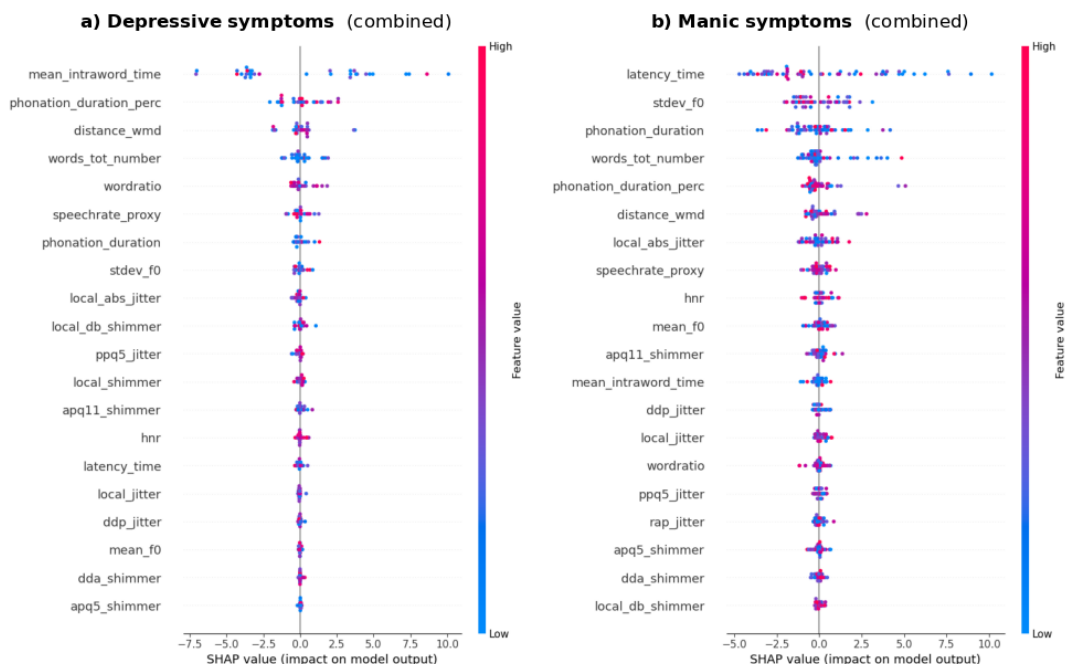
^aMetrics for testing based on a nested cross-validation approach (pilot, cross-sectional study, N=32). Range for symptom scores: 0-60.

^bIncluding sex.

^cNLP: natural language processing.

^dNot available.

Figure 3. Individual features contribution to depressive and manic symptoms predictions in sex-adjusted models among people with bipolar disorder.



Discussion

Main Findings

This study aimed at piloting the simultaneous use of speech acoustics, as well as natural language features, to glean insights into BD depressive and manic symptoms. Our findings corroborate evidence on the relationships between symptom severity and speech features, supporting

the potential predictive role for clinical purposes of digital mental health applications, embedded in a mHealth integrated system.

First, the speech of participants with BD showed that vocal perturbations (eg, higher instability and hesitations considering voice quality), latency time, and increased silences and pauses over time speaking all correlated to depressive symptoms. Consistently, increased depressive symptoms resulted in NLP-based features such as a smaller number

of words and longer mean intraword time, with lower pressure of speech. In our exploratory study, this relationship was particularly clear among females. This effect was corroborated by the predictive model, showing a contribution of NLP-based and conversational features higher than for acoustic ones. This finding aligns with prior evidence, advocating that text-based features contribute more to model accuracy than audio parameters [18]. However, also the latter component (ie, fundamental frequency, jitter- and shimmer-related indices) deserves a careful assessment, since our findings show that these indices might have an impact at least among males to predict future episodes. Indeed, recent evidence from healthy populations sheds light on sex differences in speech markers (eg, prosodic features) with different acoustic cues conveying various emotions [50]. A combination of inherent biological dissimilarities, socialization processes, influences of the social environment, and cultural expectations might contribute to these differences in both expression and perception of related emotional prosody [52,53]. Moreover, individuals may modulate their speech to align with the dominant pitch range within a specific linguistic community [54], and similar modulation may occur in conversational dialogues versus monologues and in spontaneous versus elicited speech. Thus, this criterion should be taken into account when designing apps with speech recognition and processing tasks for people with BD [31].

Second, voice instability and hesitations, as well as mean intraword time, were negatively correlated to manic symptoms. However, the interpretation of the relationship between manic features and vocal abnormalities is not straightforward. Mixed findings emerged on the relationships between speech features and manic symptoms, preventing us from supporting our original hypothesis. One plausible explanation may stem from the sample characteristics. Indeed, our participants were more likely to report depressive symptoms, and just a few had severe manic features.

However, the overall moderate correlations between speech markers and symptom severity were consistent with previous work that used speech smartphone data to discriminate between different mood states [20,21]. It has been argued that speech features may be useful to detect a trait [55] rather than a state [56] in BD. However, alterations in voice perturbations have been observed when assessing vocal markers of suicidal ideation [57], and this makes further research for vocal features reasonable, at least for depressive conditions.

Smartphone-Based Applications

Consistent with previous research on smartphone-based applications designed to record and analyze speech patterns in real time, our findings emphasize the feasibility of a simple, yet clinically useful, application of digital technology [13]. In particular, we developed the frontend of the app as a basic digital environment, freely managed by participants on their own smartphones. Participants reported a high level of engagement with the tool, showing an acceptable system usability level as assessed by SUS [35], without perceiving

intrusiveness of the recording of both elicited and spontaneous conversations.

Comparisons of the vocal performance of people with BD with unaffected relatives and healthy controls have shown a clear speech “fingerprint” of the clinical condition [58], suggesting the utility of multilevel inputs [59]. However, there is also the need for a wider understanding of fluctuations in symptom severity and mood states in this population [60]. The major strength of our study consists in the usefulness of different speech data (eg, linguistic, conversational, acoustics) to differentially identify symptoms of BD. Thus, for relapse prevention purposes, future research should possibly explore systems combining smartphone-based generated objective acoustics data with additional information, such as from facial expressions and gestures [61]. This would ultimately improve BD state prediction, even considering classification tasks [21,62-64].

Clinical Implications: Interdisciplinary Perspective

This pilot study represents a step forward in the identification and utilization of digital biomarkers for BD from natural language and audio streams, with implications for personalized mental health care and early intervention strategies. Our approach holds promise for complementary, remote assessments enhancing depressive and partly manic states prediction by exploiting participants’ speech. This would have significant implications, especially considering BD fluctuating symptomatology. Nonetheless, leveraging live speech recordings as a predictive tool, repeated assessments are needed to identify individuals at risk of transitioning to depressive and manic states.

Despite promising findings from automated assessments, mental health care heavily relies on participant interviews, yet with often subjective reports, cognitive limitations, and stigma [18]. Integrated systems, aiming at taking advantage of candidate digital markers from speech recognition, would possibly boost a care approach in which digital technology enhances, but does not replace, existing models from clinical assessment [30]. Indeed, automated assessment does not inherently lead to adherence and engagement of individuals with BD [65].

Finally, clinical, hypothesis-driven research on BD should not be dismissed, since algorithms may not be considered a black-box replacement for traditional data modeling, but they rather integrate with other systems, embedding a substantial clinical validation [66,67].

Limitations and Future Directions

We should acknowledge some limitations of this study. Analyzing speech and natural language in individuals with BD implies a challenge due to the nature of the disorder and to ethical considerations.

First, properties of chosen machine-learning models may hamper identification of unknown patterns based on values that fall outside the training set. Effective NLP and supervised learning models may require high-quality, annotated datasets.

While exploratory in nature, the study's limited sample size may have constrained the model's statistical power and the ability to capture the full complexity of the underlying data distribution, thereby hindering meaningful subgroup comparisons. Our preliminary findings should be replicated and extended in a larger, more diverse sample of people with BD to mitigate the risks associated with overfitting. Furthermore, future research should address classification approaches based on severity thresholds for both MADRS and YMRS. Accordingly, there is potential for alternative modeling approaches for regression tasks (eg, splines) that might be implemented in the future. While still considering the number of predictors, these may possibly enable a better understanding of the nature of the existing relationships and nonlinear patterns.

Consistently, the lack of standardized (linguistic and acoustic) markers represents a barrier when studying relationships with mood states. Indeed, the model may still learn to overfit to irrelevant or noisy features the data may contain, especially if they are informative in the training set by chance.

Furthermore, the speaker's identity may show a possible confounding role in a between-subject design. Therefore, studies with a longitudinal design (ie, within-subjects) should be recommended, deploying Ecological Momentary Assessment approaches [24,68]. In addition, speech patterns may generate misinterpretations if individual cultural and linguistic factors are not accounted for [69]. Similarly, speech during manic episodes may exhibit circumstantiality or tangentiality, where individuals provide excessive details or veer off-topic. Rapid speech, tangential thinking,

or unconventional language use pose challenges for automatic speech recognition systems. Analyzing such complex speech patterns requires a deep evaluation of language and context, achieving appropriate understanding of an individual's usual way of communicating in order to distinguish changes associated with BD episodes.

Furthermore, in our study, speech features were averaged over relevant duration, thus constraining the role of temporal variations across related measures in predicting symptom severity. Future research should endeavor to integrate dynamic aspects of speech on mood states transitioning.

Finally, other clinical variables, not investigated in our sample, are likely to influence the individual's speech. For instance, it should be noted that anxiety and anxious distress, often co-occurring with bipolar depression [70], may significantly influence speech features [71], as well as medication prescribed [72-74] and drug or alcohol comorbid conditions [75].

Conclusions

Speech patterns, underlying both linguistic and acoustic features, are able to yield quantifiable differences, thus embodying digital markers of symptom severity. Multimodal, smartphone-integrated digital assessments could serve as powerful tools for clinical purposes to remotely complement standard mental health evaluations, potentially contributing to distinguish clinical conditions in people with BD. Feasibility of similar systems seems promising, though issues related to privacy, intrusiveness, and clinical therapeutic relationships should be carefully considered.

Acknowledgments

This research was supported by the FSE REACT-EU Competitive Research Grant Axis-IV DM 1062/2021: "Natural Language Processing in Digital Mental Health." The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The datasets including audio data streams supporting the conclusions of this article are not publicly available as the original source has not granted permission to share that information but may be available from the corresponding author on reasonable request.

Authors' Contributions

CC, FB, and GC handled conceptualization. RMC, AC, CN, DP, SP, AB, and MR performed investigation. CC, FB, VS, CB, and MB contributed to methodology. VS and MB assisted with software. CC conducted formal analysis. GC performed supervision. CC contributed to writing – original draft. CC, RMC, AC, CN, DP, SP, AB, MR, VS, CB, MB, FB, and GC contributed to writing – review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Checklist.

[\[DOC File \(Microsoft Word File\), 87 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Features.

[\[DOCX File \(Microsoft Word File\), 16 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Supplementary analyses.

[\[DOCX File \(Microsoft Word File\), 330 KB-Multimedia Appendix 3\]](#)

References

1. Bartoli F, Crocamo C, Carrà G. Clinical correlates of DSM-5 mixed features in bipolar disorder: a meta-analysis. *J Affect Disord*. Nov 1, 2020;276:234-240. [doi: [10.1016/j.jad.2020.07.035](https://doi.org/10.1016/j.jad.2020.07.035)] [Medline: [32697704](https://pubmed.ncbi.nlm.nih.gov/32697704/)]
2. Grande I, Berk M, Birmaher B, Vieta E. Bipolar disorder. *Lancet*. Apr 9, 2016;387(10027):1561-1572. [doi: [10.1016/S0140-6736\(15\)00241-X](https://doi.org/10.1016/S0140-6736(15)00241-X)] [Medline: [26388529](https://pubmed.ncbi.nlm.nih.gov/26388529/)]
3. McIntyre RS, Berk M, Brietzke E, et al. Bipolar disorders. *Lancet*. Dec 5, 2020;396(10265):1841-1856. [doi: [10.1016/S0140-6736\(20\)31544-0](https://doi.org/10.1016/S0140-6736(20)31544-0)] [Medline: [33278937](https://pubmed.ncbi.nlm.nih.gov/33278937/)]
4. Karambelas GJ, Filia K, Byrne LK, Allott KA, Jayasinghe A, Cotton SM. A systematic review comparing caregiver burden and psychological functioning in caregivers of individuals with schizophrenia spectrum disorders and bipolar disorders. *BMC Psychiatry*. Jun 23, 2022;22(1):422. [doi: [10.1186/s12888-022-04069-w](https://doi.org/10.1186/s12888-022-04069-w)] [Medline: [35733174](https://pubmed.ncbi.nlm.nih.gov/35733174/)]
5. Fajutrao L, Locklear J, Priaux J, Heyes A. A systematic review of the evidence of the burden of bipolar disorder in Europe. *Clin Pract Epidemiol Ment Health*. Jan 23, 2009;5:3. [doi: [10.1186/1745-0179-5-3](https://doi.org/10.1186/1745-0179-5-3)] [Medline: [19166608](https://pubmed.ncbi.nlm.nih.gov/19166608/)]
6. Ogilvie AD, Morant N, Goodwin GM. The burden on informal caregivers of people with bipolar disorder. *Bipolar Disord*. 2005;7 Suppl 1(Suppl 1):25-32. [doi: [10.1111/j.1399-5618.2005.00191.x](https://doi.org/10.1111/j.1399-5618.2005.00191.x)] [Medline: [15762866](https://pubmed.ncbi.nlm.nih.gov/15762866/)]
7. Goodwin FK, Jamison KR. *Manic-Depressive Illness: Bipolar Disorders and Recurrent Depression*. Oxford University Press, USA; 2007. ISBN: 9780195135794
8. Karam ZN, Provost EM, Singh S, et al. Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech. *Proc IEEE Int Conf Acoust Speech Signal Process*. May 2014;2014:4858-4862. [doi: [10.1109/ICASSP.2014.6854525](https://doi.org/10.1109/ICASSP.2014.6854525)] [Medline: [27630535](https://pubmed.ncbi.nlm.nih.gov/27630535/)]
9. Harvey D, Lobban F, Rayson P, Warner A, Jones S. Natural language processing methods and bipolar disorder: scoping review. *JMIR Ment Health*. Apr 22, 2022;9(4):e35928. [doi: [10.2196/35928](https://doi.org/10.2196/35928)] [Medline: [35451984](https://pubmed.ncbi.nlm.nih.gov/35451984/)]
10. Birnbaum ML, Abrami A, Heisig S, et al. Acoustic and facial features from clinical interviews for machine learning-based psychiatric diagnosis: algorithm development. *JMIR Ment Health*. Jan 24, 2022;9(1):e24699. [doi: [10.2196/24699](https://doi.org/10.2196/24699)] [Medline: [35072648](https://pubmed.ncbi.nlm.nih.gov/35072648/)]
11. Gideon J, Provost EM, McInnis M. Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder. *Proc IEEE Int Conf Acoust Speech Signal Process*. Mar 2016;2016:2359-2363. [doi: [10.1109/ICASSP.2016.7472099](https://doi.org/10.1109/ICASSP.2016.7472099)] [Medline: [27570493](https://pubmed.ncbi.nlm.nih.gov/27570493/)]
12. Guidi A, Schoentgen J, Bertschy G, Gentili C, Scilingo EP, Vanello N. Features of vocal frequency contour and speech rhythm in bipolar disorder. *Biomed Signal Process Control*. Aug 2017;37:23-31. [doi: [10.1016/j.bspc.2017.01.017](https://doi.org/10.1016/j.bspc.2017.01.017)]
13. Maxhuni A, Muñoz-Meléndez A, Osmani V, Perez H, Mayora O, Morales EF. Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients. *Pervasive Mob Comput*. Sep 2016;31:50-66. [doi: [10.1016/j.pmcj.2016.01.008](https://doi.org/10.1016/j.pmcj.2016.01.008)]
14. Matton K, McInnis MG, Provost EM. Into the wild: transitioning from recognizing mood in clinical interactions to personal conversations for individuals with bipolar disorder. Presented at: Interspeech 2019; Sep 15-19, 2019:1438-1442; Graz, Austria. 2019.URL: https://www.isca-archive.org/interspeech_2019 [doi: [10.21437/Interspeech.2019-2698](https://doi.org/10.21437/Interspeech.2019-2698)]
15. Arevian AC, Bone D, Malandrakis N, et al. Clinical state tracking in serious mental illness through computational analysis of speech. *PLoS ONE*. 2020;15(1):e0225695. [doi: [10.1371/journal.pone.0225695](https://doi.org/10.1371/journal.pone.0225695)] [Medline: [31940347](https://pubmed.ncbi.nlm.nih.gov/31940347/)]
16. Girard JM, Vail AK, Liebenthal E, et al. Computational analysis of spoken language in acute psychosis and mania. *Schizophr Res*. Jul 2022;245:97-115. [doi: [10.1016/j.schres.2021.06.040](https://doi.org/10.1016/j.schres.2021.06.040)] [Medline: [34456131](https://pubmed.ncbi.nlm.nih.gov/34456131/)]
17. Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: a systematic review. *Laryngoscope Investig Otolaryngol*. Feb 2020;5(1):96-116. [doi: [10.1002/lio2.354](https://doi.org/10.1002/lio2.354)] [Medline: [32128436](https://pubmed.ncbi.nlm.nih.gov/32128436/)]
18. Malgaroli M, Hull TD, Zech JM, Althoff T. Natural language processing for mental health interventions: a systematic review and research framework. *Transl Psychiatry*. Oct 6, 2023;13(1):309. [doi: [10.1038/s41398-023-02592-2](https://doi.org/10.1038/s41398-023-02592-2)] [Medline: [37798296](https://pubmed.ncbi.nlm.nih.gov/37798296/)]
19. Cummins N, Baird A, Schuller BW. Speech analysis for health: current state-of-the-art and the increasing impact of deep learning. *Methods*. Dec 1, 2018;151:41-54. [doi: [10.1016/j.ymeth.2018.07.007](https://doi.org/10.1016/j.ymeth.2018.07.007)] [Medline: [30099083](https://pubmed.ncbi.nlm.nih.gov/30099083/)]
20. Faurholt-Jepsen M, Busk J, Frost M, et al. Voice analysis as an objective state marker in bipolar disorder. *Transl Psychiatry*. Jul 19, 2016;6(7):e856. [doi: [10.1038/tp.2016.123](https://doi.org/10.1038/tp.2016.123)] [Medline: [27434490](https://pubmed.ncbi.nlm.nih.gov/27434490/)]

21. Faurholt-Jepsen M, Rohani DA, Busk J, et al. Discriminating between patients with unipolar disorder, bipolar disorder, and healthy control individuals based on voice features collected from naturalistic smartphone calls. *Acta Psychiatr Scand*. Mar 2022;145(3):255-267. [doi: [10.1111/acps.13391](https://doi.org/10.1111/acps.13391)] [Medline: [34923626](https://pubmed.ncbi.nlm.nih.gov/34923626/)]
22. Daus H, Bloecher T, Egeler R, De Klerk R, Stork W, Backenstrass M. Development of an emotion-sensitive mHealth approach for mood-state recognition in bipolar disorder. *JMIR Ment Health*. Jul 3, 2020;7(7):e14267. [doi: [10.2196/14267](https://doi.org/10.2196/14267)] [Medline: [32618577](https://pubmed.ncbi.nlm.nih.gov/32618577/)]
23. Dikaïos K, Rempel S, Dumpala SH, Oore S, Kieft M, Uher R. Applications of speech analysis in psychiatry. *Harv Rev Psychiatry*. 2023;31(1):1-13. [doi: [10.1097/HRP.0000000000000356](https://doi.org/10.1097/HRP.0000000000000356)] [Medline: [36608078](https://pubmed.ncbi.nlm.nih.gov/36608078/)]
24. Dunster GP, Swendsen J, Merikangas KR. Real-time mobile monitoring of bipolar disorder: a review of evidence and future directions. *Neuropsychopharmacology*. Jan 2021;46(1):197-208. [doi: [10.1038/s41386-020-00830-5](https://doi.org/10.1038/s41386-020-00830-5)] [Medline: [32919408](https://pubmed.ncbi.nlm.nih.gov/32919408/)]
25. Marzano L, Bardill A, Fields B, et al. The application of mHealth to mental health: opportunities and challenges. *Lancet Psychiatry*. Oct 2015;2(10):942-948. [doi: [10.1016/S2215-0366\(15\)00268-0](https://doi.org/10.1016/S2215-0366(15)00268-0)] [Medline: [26462228](https://pubmed.ncbi.nlm.nih.gov/26462228/)]
26. Le Glaz A, Haralambous Y, Kim-Dufor DH, et al. Machine learning and natural language processing in mental health: systematic review. *J Med Internet Res*. May 4, 2021;23(5):e15708. [doi: [10.2196/15708](https://doi.org/10.2196/15708)] [Medline: [33944788](https://pubmed.ncbi.nlm.nih.gov/33944788/)]
27. Farrús M, Codina-Filbà J, Escudero J. Acoustic and prosodic information for home monitoring of bipolar disorder. *Health Informatics J*. 2021;27(1):1460458220972755. [doi: [10.1177/1460458220972755](https://doi.org/10.1177/1460458220972755)] [Medline: [33438502](https://pubmed.ncbi.nlm.nih.gov/33438502/)]
28. Gong Y, Poellabauer C. Topic modeling based multi-modal depression detection. Presented at: MM '17; Oct 23-27, 2017:2017; Mountain View California USA. Oct 23, 2017. URL: <https://dl.acm.org/doi/proceedings/10.1145/3133944> [doi: [10.1145/3133944.3133945](https://doi.org/10.1145/3133944.3133945)]
29. Muaremi A, Gravenhorst F, Grünerbl A, Arnrich B, Tröster G. Assessing bipolar episodes using speech cues derived from phone calls. In: *Lect Notes Inst Comput Sci Soc Inform Telecommun Eng*. 2014:103-114. [doi: [10.1007/978-3-319-11564-1_11](https://doi.org/10.1007/978-3-319-11564-1_11)]
30. Bond RR, Mulvenna MD, Potts C, O'Neill S, Ennis E, Torous J. Digital transformation of mental health services. *Npj Ment Health Res*. Aug 22, 2023;2(1):13. [doi: [10.1038/s44184-023-00033-y](https://doi.org/10.1038/s44184-023-00033-y)] [Medline: [38609479](https://pubmed.ncbi.nlm.nih.gov/38609479/)]
31. Flanagan O, Chan A, Roop P, Sundram F. Using acoustic speech patterns from smartphones to investigate mood disorders: scoping review. *JMIR Mhealth Uhealth*. Sep 17, 2021;9(9):e24352. [doi: [10.2196/24352](https://doi.org/10.2196/24352)] [Medline: [34533465](https://pubmed.ncbi.nlm.nih.gov/34533465/)]
32. de Oliveira L, Portugal LCL, Pereira M, et al. Predicting bipolar disorder risk factors in distressed young adults from patterns of brain activation to reward: a machine learning approach. *Biol Psychiatry Cogn Neurosci Neuroimaging*. Aug 2019;4(8):726-733. [doi: [10.1016/j.bpsc.2019.04.005](https://doi.org/10.1016/j.bpsc.2019.04.005)] [Medline: [31201147](https://pubmed.ncbi.nlm.nih.gov/31201147/)]
33. von Elm E, Altman DG, Egger M, et al. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*. Oct 20, 2007;335(7624):806-808. [doi: [10.1136/bmj.39335.541782.AD](https://doi.org/10.1136/bmj.39335.541782.AD)] [Medline: [17947786](https://pubmed.ncbi.nlm.nih.gov/17947786/)]
34. Ab.acus Srl. URL: <https://www.ab-acus.eu/index.php/portfolio-items/speakapp/> [Accessed 2024-03-19]
35. Hyzy M, Bond R, Mulvenna M, et al. System Usability Scale benchmarking for digital health apps: meta-analysis. *JMIR Mhealth Uhealth*. Aug 18, 2022;10(8):e37290. [doi: [10.2196/37290](https://doi.org/10.2196/37290)] [Medline: [35980732](https://pubmed.ncbi.nlm.nih.gov/35980732/)]
36. Italian standardization and classification of neuropsychological tests. The Italian Group on the Neuropsychological Study of Aging. *Ital J Neurol Sci*. Dec 1987;Suppl 8(1-120):1-120. [Medline: [3330072](https://pubmed.ncbi.nlm.nih.gov/3330072/)]
37. Google speech-to-text apis. Google Cloud. URL: <https://cloud.google.com/speech-to-text> [Accessed 2025-04-09]
38. Jadoul Y, Thompson B, de Boer B. Introducing Parselmouth: a Python interface to Praat. *J Phon*. Nov 2018;71:1-15. [doi: [10.1016/j.wocn.2018.07.001](https://doi.org/10.1016/j.wocn.2018.07.001)]
39. de Boer JN, Brederoo SG, Voppel AE, Sommer IEC. Anomalies in language as a biomarker for schizophrenia. *Curr Opin Psychiatry*. May 2020;33(3):212-218. [doi: [10.1097/YCO.0000000000000595](https://doi.org/10.1097/YCO.0000000000000595)] [Medline: [32049766](https://pubmed.ncbi.nlm.nih.gov/32049766/)]
40. Kusner M, Sun Y, Kolkin N, Weinberger K. From word embeddings to document distances. Presented at: Proceedings of the 32nd International Conference on Machine Learning, PMLR; Jul 5-7, 2015:957-966; Lille, France. 2015.
41. Nadeu M, Prieto P. Pitch range, gestural information, and perceived politeness in Catalan. *J Pragmat*. Feb 2011;43(3):841-854. [doi: [10.1016/j.pragma.2010.09.015](https://doi.org/10.1016/j.pragma.2010.09.015)]
42. Smith SW. *Digital Signal Processing: A Practical Guide for Engineers and Scientists*. California Technical Publishing; 2002:978. [doi: [10.1016/B978-0-7506-7444-7.X5036-5](https://doi.org/10.1016/B978-0-7506-7444-7.X5036-5)] ISBN: 978-0-7506-7444-7
43. Ververidis D, Kotropoulos C. Emotional speech recognition: resources, features, and methods. *Speech Commun*. Sep 2006;48(9):1162-1181. [doi: [10.1016/j.specom.2006.04.003](https://doi.org/10.1016/j.specom.2006.04.003)]
44. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. Apr 1979;134:382-389. [doi: [10.1192/bjp.134.4.382](https://doi.org/10.1192/bjp.134.4.382)] [Medline: [444788](https://pubmed.ncbi.nlm.nih.gov/444788/)]

45. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry*. Nov 1978;133:429-435. [doi: [10.1192/bjp.133.5.429](https://doi.org/10.1192/bjp.133.5.429)] [Medline: [728692](https://pubmed.ncbi.nlm.nih.gov/728692/)]
46. Lukasiewicz M, Gerard S, Besnard A, et al. Young Mania Rating Scale: how to interpret the numbers? Determination of a severity threshold and of the minimal clinically significant difference in the EMBLEM cohort. *Int J Methods Psychiatr Res*. Mar 2013;22(1):46-58. [doi: [10.1002/mpr.1379](https://doi.org/10.1002/mpr.1379)] [Medline: [23526724](https://pubmed.ncbi.nlm.nih.gov/23526724/)]
47. Samara MT, Levine SZ, Leucht S. Linkage of Young Mania Rating Scale to Clinical Global Impression Scale to enhance utility in clinical practice and research trials. *Pharmacopsychiatry*. Jan 2023;56(1):18-24. [doi: [10.1055/a-1841-6672](https://doi.org/10.1055/a-1841-6672)] [Medline: [35896419](https://pubmed.ncbi.nlm.nih.gov/35896419/)]
48. Thase ME, Harrington A, Calabrese J, Montgomery S, Niu X, Patel MD. Evaluation of MADRS severity thresholds in patients with bipolar depression. *J Affect Disord*. May 1, 2021;286:58-63. [doi: [10.1016/j.jad.2021.02.043](https://doi.org/10.1016/j.jad.2021.02.043)] [Medline: [33677183](https://pubmed.ncbi.nlm.nih.gov/33677183/)]
49. Eichhorn JT, Kent RD, Austin D, Vorperian HK. Effects of aging on vocal fundamental frequency and vowel formants in men and women. *J Voice*. Sep 2018;32(5):644. [doi: [10.1016/j.jvoice.2017.08.003](https://doi.org/10.1016/j.jvoice.2017.08.003)] [Medline: [28864082](https://pubmed.ncbi.nlm.nih.gov/28864082/)]
50. Ertürk A, Gürses E, Kulak Kayıkcı ME. Sex related differences in the perception and production of emotional prosody in adults. *Psychol Res*. Mar 2024;88(2):449-457. [doi: [10.1007/s00426-023-01865-1](https://doi.org/10.1007/s00426-023-01865-1)] [Medline: [37542581](https://pubmed.ncbi.nlm.nih.gov/37542581/)]
51. Mendoza E, Valencia N, Muñoz J, Trujillo H. Differences in voice quality between men and women: use of the long-term average spectrum (LTAS). *J Voice*. Mar 1996;10(1):59-66. [doi: [10.1016/s0892-1997\(96\)80019-1](https://doi.org/10.1016/s0892-1997(96)80019-1)] [Medline: [8653179](https://pubmed.ncbi.nlm.nih.gov/8653179/)]
52. Chaplin TM. Gender and emotion expression: a developmental contextual perspective. *Emot Rev*. Jan 2015;7(1):14-21. [doi: [10.1177/1754073914544408](https://doi.org/10.1177/1754073914544408)] [Medline: [26089983](https://pubmed.ncbi.nlm.nih.gov/26089983/)]
53. Lin Y, Ding H, Zhang Y. Gender differences in identifying facial, prosodic, and semantic emotions show category- and channel-specific effects mediated by encoder's gender. *J Speech Lang Hear Res*. Aug 9, 2021;64(8):2941-2955. [doi: [10.1044/2021_JSLHR-20-00553](https://doi.org/10.1044/2021_JSLHR-20-00553)] [Medline: [34310173](https://pubmed.ncbi.nlm.nih.gov/34310173/)]
54. Aung T, Puts D. Voice pitch: a window into the communication of social power. *Curr Opin Psychol*. Jun 2020;33:154-161. [doi: [10.1016/j.copsy.2019.07.028](https://doi.org/10.1016/j.copsy.2019.07.028)] [Medline: [31445439](https://pubmed.ncbi.nlm.nih.gov/31445439/)]
55. Zhang J, Pan Z, Gui C, et al. Analysis on speech signal features of manic patients. *J Psychiatr Res*. Mar 2018;98:59-63. [doi: [10.1016/j.jpsychires.2017.12.012](https://doi.org/10.1016/j.jpsychires.2017.12.012)] [Medline: [29291581](https://pubmed.ncbi.nlm.nih.gov/29291581/)]
56. Guidi A, Salvi S, Ottaviano M, et al. Smartphone application for the analysis of prosodic features in running speech with a focus on bipolar disorders: system performance evaluation and case study. *Sensors (Basel)*. Nov 6, 2015;15(11):28070-28087. [doi: [10.3390/s151128070](https://doi.org/10.3390/s151128070)] [Medline: [26561811](https://pubmed.ncbi.nlm.nih.gov/26561811/)]
57. Ozdas A, Shiavi RG, Silverman SE, Silverman MK, Wilkes DM. Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk. *IEEE Trans Biomed Eng*. Sep 2004;51(9):1530-1540. [doi: [10.1109/TBME.2004.827544](https://doi.org/10.1109/TBME.2004.827544)] [Medline: [15376501](https://pubmed.ncbi.nlm.nih.gov/15376501/)]
58. Faurholt-Jepsen M, Rohani DA, Busk J, Vinberg M, Bardram JE, Kessing LV. Voice analyses using smartphone-based data in patients with bipolar disorder, unaffected relatives and healthy control individuals, and during different affective states. *Int J Bipolar Disord*. Dec 1, 2021;9(1):38. [doi: [10.1186/s40345-021-00243-3](https://doi.org/10.1186/s40345-021-00243-3)] [Medline: [34850296](https://pubmed.ncbi.nlm.nih.gov/34850296/)]
59. Torous J, Bucci S, Bell IH, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry*. Oct 2021;20(3):318-335. [doi: [10.1002/wps.20883](https://doi.org/10.1002/wps.20883)] [Medline: [34505369](https://pubmed.ncbi.nlm.nih.gov/34505369/)]
60. Balcombe L, De Leo D. Digital mental health challenges and the horizon ahead for solutions. *JMIR Ment Health*. Mar 29, 2021;8(3):e26811. [doi: [10.2196/26811](https://doi.org/10.2196/26811)] [Medline: [33779570](https://pubmed.ncbi.nlm.nih.gov/33779570/)]
61. Soenksen LR, Ma Y, Zeng C, et al. Integrated multimodal artificial intelligence framework for healthcare applications. *NPJ Digit Med*. Sep 20, 2022;5(1):149. [doi: [10.1038/s41746-022-00689-4](https://doi.org/10.1038/s41746-022-00689-4)] [Medline: [36127417](https://pubmed.ncbi.nlm.nih.gov/36127417/)]
62. Grünerbl A, Muaremi A, Osmani V, et al. Smartphone-based recognition of states and state changes in bipolar disorder patients. *IEEE J Biomed Health Inform*. Jan 2015;19(1):140-148. [doi: [10.1109/JBHI.2014.2343154](https://doi.org/10.1109/JBHI.2014.2343154)] [Medline: [25073181](https://pubmed.ncbi.nlm.nih.gov/25073181/)]
63. Osmani V. Smartphones in mental health: detecting depressive and manic episodes. *IEEE Pervasive Comput*. 2015;14(3):10-13. [doi: [10.1109/MPRV.2015.54](https://doi.org/10.1109/MPRV.2015.54)]
64. Cohen J, Richter V, Neumann M, et al. A multimodal dialog approach to mental state characterization in clinically depressed, anxious, and suicidal populations. *Front Psychol*. 2023;14:1135469. [doi: [10.3389/fpsyg.2023.1135469](https://doi.org/10.3389/fpsyg.2023.1135469)] [Medline: [37767217](https://pubmed.ncbi.nlm.nih.gov/37767217/)]
65. Or F, Torous J, Onnela JP. High potential but limited evidence: using voice data from smartphones to monitor and diagnose mood disorders. *Psychiatr Rehabil J*. Sep 2017;40(3):320-324. [doi: [10.1037/prj0000279](https://doi.org/10.1037/prj0000279)] [Medline: [28891659](https://pubmed.ncbi.nlm.nih.gov/28891659/)]

66. Garcia-Ceja E, Riegler M, Nordgreen T, Jakobsen P, Oedegaard KJ, Tørresen J. Mental health monitoring with multimodal sensing and machine learning: a survey. *Pervasive Mob Comput.* Dec 2018;51:1-26. [doi: [10.1016/j.pmcj.2018.09.003](https://doi.org/10.1016/j.pmcj.2018.09.003)]
67. McCoy LG, Brenna CTA, Chen SS, Vold K, Das S. Believing in black boxes: machine learning for healthcare does not need explainability to be evidence-based. *J Clin Epidemiol.* Feb 2022;142:252-257. [doi: [10.1016/j.jclinepi.2021.11.001](https://doi.org/10.1016/j.jclinepi.2021.11.001)] [Medline: [34748907](https://pubmed.ncbi.nlm.nih.gov/34748907/)]
68. Yerushalmi M, Sixsmith A, Pollock Star A, King DB, O'Rourke N. Ecological momentary assessment of bipolar disorder symptoms and partner affect: longitudinal pilot study. *JMIR Form Res.* Sep 2, 2021;5(9):e30472. [doi: [10.2196/30472](https://doi.org/10.2196/30472)] [Medline: [34473069](https://pubmed.ncbi.nlm.nih.gov/34473069/)]
69. Clark EL, Easton C, Verdon S. The impact of linguistic bias upon speech-language pathologists' attitudes towards non-standard dialects of English. *Clin Linguist Phon.* Jun 3, 2021;35(6):542-559. [doi: [10.1080/02699206.2020.1803405](https://doi.org/10.1080/02699206.2020.1803405)] [Medline: [32781853](https://pubmed.ncbi.nlm.nih.gov/32781853/)]
70. Bartoli F, Bachi B, Callovin T, et al. Anxious distress in people with major depressive episodes: a cross-sectional analysis of clinical correlates. *CNS Spectr.* Feb 2024;29(1):49-53. [doi: [10.1017/S1092852923002377](https://doi.org/10.1017/S1092852923002377)] [Medline: [37489522](https://pubmed.ncbi.nlm.nih.gov/37489522/)]
71. Malgaroli M, Hull TD, Calderon A, Simon NM. Linguistic markers of anxiety and depression in somatic symptom and related disorders: observational study of a digital intervention. *J Affect Disord.* May 1, 2024;352:133-137. [doi: [10.1016/j.jad.2024.02.012](https://doi.org/10.1016/j.jad.2024.02.012)] [Medline: [38336165](https://pubmed.ncbi.nlm.nih.gov/38336165/)]
72. Bartoli F, Crocamo C, Clerici M, Carrà G. Allopurinol as add-on treatment for mania symptoms in bipolar disorder: systematic review and meta-analysis of randomised controlled trials. *Br J Psychiatry.* Jan 2017;210(1):10-15. [doi: [10.1192/bjp.bp.115.180281](https://doi.org/10.1192/bjp.bp.115.180281)] [Medline: [27856422](https://pubmed.ncbi.nlm.nih.gov/27856422/)]
73. Bartoli F, Cavaleri D, Bachi B, et al. Repurposed drugs as adjunctive treatments for mania and bipolar depression: a meta-review and critical appraisal of meta-analyses of randomized placebo-controlled trials. *J Psychiatr Res.* Nov 2021;143:230-238. [doi: [10.1016/j.jpsychires.2021.09.018](https://doi.org/10.1016/j.jpsychires.2021.09.018)] [Medline: [34509090](https://pubmed.ncbi.nlm.nih.gov/34509090/)]
74. Bartoli F, Cavaleri D, Nasti C, et al. Long-acting injectable antipsychotics for the treatment of bipolar disorder: evidence from mirror-image studies. *Ther Adv Psychopharmacol.* 2023;13:20451253231163682. [doi: [10.1177/20451253231163682](https://doi.org/10.1177/20451253231163682)] [Medline: [36994116](https://pubmed.ncbi.nlm.nih.gov/36994116/)]
75. Carrà G, Scioli R, Monti MC, Marinoni A. Severity profiles of substance-abusing patients in Italian community addiction facilities: influence of psychiatric concurrent disorders. *Eur Addict Res.* 2006;12(2):96-101. [doi: [10.1159/000090429](https://doi.org/10.1159/000090429)] [Medline: [16543745](https://pubmed.ncbi.nlm.nih.gov/16543745/)]

Abbreviations

BD: bipolar disorder

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition

MADRS: Montgomery-Åsberg Depression Rating Scale

MAE: mean absolute error

mHealth: mobile health

MSE: mean squared error

NLP: natural language processing

RF: random forest

SCID-5: Structured Clinical Interview for DSM-5

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

WMD: word mover's distance

YMRS: Young Mania Rating Scale

Edited by Amaryllis Mavragani; peer-reviewed by Denny Meyer, Vincent Martin; submitted 20.08.2024; final revised version received 29.01.2025; accepted 12.02.2025; published 16.04.2025

Please cite as:

Crocamo C, Cioni RM, Canestro A, Nasti C, Palpella D, Piacenti S, Bartocetti A, Re M, Simonetti V, Barattieri di San Pietro C, Bulgheroni M, Bartoli F, Carrà G

Acoustic and Natural Language Markers for Bipolar Disorder: A Pilot, mHealth Cross-Sectional Study

JMIR Form Res 2025;9:e65555

URL: <https://formative.jmir.org/2025/1/e65555>

doi: [10.2196/65555](https://doi.org/10.2196/65555)

© Cristina Crocamo, Riccardo Matteo Cioni, Aurelia Canestro, Christian Nasti, Dario Palpella, Susanna Piacenti, Alessandra Bartocchetti, Martina Re, Valentina Simonetti, Chiara Barattieri di San Pietro, Maria Bulgheroni, Francesco Bartoli, Giuseppe Carrà. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 16.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.