


Digital assessment of nonverbal behaviors forecasts first onset of depression

Sekine Ozturk¹ , Scott Feltman², Daniel N. Klein¹, Roman Kotov³ and Aprajita Mohanty¹

Original Article

Cite this article: Ozturk S, Feltman S, Klein DN, Kotov R, Mohanty A (2024). Digital assessment of nonverbal behaviors forecasts first onset of depression. *Psychological Medicine* 1–12. <https://doi.org/10.1017/S0033291724002010>

Received: 22 March 2024

Revised: 29 July 2024

Accepted: 12 August 2024

Keywords:

depression; early identification; FaceReader; facial recognition; nonverbal behaviors

Corresponding author:

Sekine Ozturk;

Email: sekine.ozturk@stonybrook.edu

¹Department of Psychology, Stony Brook University, Stony Brook, NY, USA; ²Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY, USA and ³Department of Psychiatry and Behavioral Science, Stony Brook University, Stony Brook, NY, USA

Abstract

Background. Adolescence is marked by a sharp increase in the incidence of depression, especially in females. Identification of risk for depressive disorders (DD) in this key developmental stage can help prevention efforts, mitigating the clinical and public burden of DD. While frequently used in diagnosis, nonverbal behaviors are relatively understudied as risk markers for DD. Digital technology, such as facial recognition, may provide objective, fast, efficient, and cost-effective means of measuring nonverbal behavior.

Method. Here, we analyzed video-recorded clinical interviews of 359 never-depressed adolescents females via commercially available facial emotion recognition software.

Results. We found that average head and facial movements forecast future first onset of depression (AUC = 0.70) beyond the effects of other established self-report and physiological markers of DD risk.

Conclusions. Overall, these findings suggest that digital assessment of nonverbal behaviors may provide a promising risk marker for DD, which could aid in early identification and intervention efforts.

Depressive disorders (DD) are a debilitating public health problem and one of the leading causes of disability worldwide (Kessler & Bromet, 2013). The incidence of DD rises dramatically in adolescence (Birmaher et al., 2007), with an especially marked increase in females who are at two times higher risk for experiencing depression compared to their male counterparts (Salk, Hyde, & Abramson, 2017). Therefore, adolescence, particularly for females, is an important developmental stage to identify objective, cost-effective and non-invasive markers of risk and resilience, allowing for better prevention and treatment of DD (Leventhal, Pettit, & Lewinsohn, 2008; Mittal & Wakschlag, 2017).

In addition to evaluations of self-reported symptoms, clinical scientists have long been studying biological markers of vulnerability to DD using methods such as neuroimaging and genetics; however, the complexity and high costs of these methods limit their use for diagnostic and preventative purposes in the real world. Utilization of artificial intelligence (AI) and digital technology measuring client's verbal and non-verbal behavior in clinical settings offers significant promise (Cohn et al., 2018; Eichstaedt et al., 2018), due to the objective, efficient, and cost-effective nature of these techniques. In this domain, non-verbal signs of DD become a particular opportunity for digital technology. Clinicians have long relied on non-verbal signs such as psychomotor retardation or agitation and affective expressions (Fairbanks, McGuire, & Harris, 1982) in evaluating risk and making diagnostic decisions regarding depression; however, these methods lack reliability (Lanyon & Wershba, 2013; Levin-Aspenson & Watson, 2018; Snowden, 2003). Digital phenotyping or the use of digital technology to comprehensively measure behavior, has been shown to quantify nonverbal signs of DD (Abbas et al., 2021a) and is postulated to tap into the underlying biological phenotypes related to the clinical dysfunction (Insel, 2017; Abbas, Schultebrucks, & Galatzer-Levy, 2021b). These methods have been demonstrated to successfully distinguish nonverbal behaviors, such as head movements and facial expressions, of clinically depressed individuals from those of control participants (Abbas et al., 2021a, 2021c; Pampouchidou et al., 2020), however the power of these digital measures in forecasting depression from adolescence remains unknown.

Nonverbal behaviors encompass a range of observable actions, such as gross motor activity, posture, movements of head, hands, torso and limbs, facial mobility, and eye glances (Friedman et al., 1974; Sobin, 1997). Numerous studies have highlighted distinctions in non-verbal behaviors between depressed individuals and other groups (non-clinical and other clinically diagnosed). These distinctions encompass a variety of nonverbal behaviors, including gross motor behaviors such as psychomotor retardation or agitation, as well as finer facial movements (Kring & Stuart, 2008; Sobin, 1997; Woody et al., 2019).

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Early electromyography (EMG) studies demonstrate that depression is linked with reduced facial muscle activity (Gehricke & Shapiro, 2000; Greden, Genero, Price, Feinberg, & Levine, 1986). Particularly, there is diminished movement in zygomatic major muscles producing smiles (Chentsova-Dutton, Tsai, & Gotlib, 2010; Gaebel & Wölwer, 2004; Trémeau et al., 2005) and increased tension in the corrugator muscle that lowers the eyebrows and contributes to frowning facial expressions (Greden, Genero, & Price, 1985; Schwartz et al., 1978). These changes in EMG activity are shown to predict treatment outcome (Greden, Price, Genero, Feinberg, & Levine, 1984). However, a drawback lies in their limited ecological validity due to the fact that this data is obtained in highly controlled settings.

In contrast, employing digital tools for nonverbal behavior analysis offers advantages in naturalistic settings. Recent research utilizing automatic facial movement detection through digital methods revealed that individuals with higher levels of depression exhibit distinct differences in facial expressions. These differences include reduced smiling, more frequent displays of contempt and embarrassment (Girard et al., 2014), decreased synchrony in facial expressions (Altmann, Brummel, Meier, & Strauss, 2021), and less intense movements around the mouth and eyelids (Stolicyn, Steele, & Seriès, 2022). Additionally, models using the Facial Action Coding System (FACS; Ekman, Friesen, & Hager, 2002) have achieved 79% accuracy in automatically detecting depression through facial movements (Cohn et al., 2009). Therefore, preliminary evidence suggests that automatic detection of facial expressions using digital technology may offer insights into predicting the clinical course and severity of depression (Dibeklioglu, Hammal, & Cohn, 2018; Gavrilescu & Vizireanu, 2019; Kacem, Hammal, Daoudi, & Cohn, 2018).

Research on digital phenotyping of nonverbal behavior in depression extends beyond facial expressions. Previous studies indicate that depression is associated with diminished change in head position and slower movements (Alghowinem, Goecke, Wagner, Parker, & Breakspear, 2013; Joshi, Goecke, Parker, & Breakspear, 2013). AI models predicted depression up to 65% with head motion alone, and up to 78–85% when combined with other nonverbal behaviors including facial movements (Dibeklioglu et al., 2018), or with acoustic features. This relationship was shown to be dose-dependent such that the movement characteristics were related to the severity of illness as well as treatment progress and remission (Girard et al., 2014; Kacem et al., 2018). Notably, a smartphone-based video analysis by Abbas et al. demonstrated that positive response to treatment was linked with an increase in head movements in depressed participants (2021a). While consistently identifying depression in presently depressed adults, the research exhibits limitations, primarily relying on cross-sectional designs, with the exception of only a handful of studies delving into treatment response, and constrained by small sample sizes.

Rates of depression increase dramatically in adolescence, and the female preponderance emerges around this age, therefore it is critical to identify markers of risk in this period for targeted intervention efforts. Based on the current literature on adults, it is unknown if the behavioral abnormalities captured by facial recognition are present in adolescents. Furthermore, it is unknown whether these are risk factors or correlates of depression, as prior digital assessment studies investigated nonverbal behaviors in depressed individuals. The current evidence has not explored whether these behavioral abnormalities are limited to a depressed

state, or if digital tools can reveal risk markers that are evident even before the onset of depression. Most important of all, with the present evidence, it is unclear whether digital tools have a clinical utility that is as good as, or better than established clinical, self-report, and physiological indices of risk for depression. It is of utmost importance to compare digital methods to established predictors of depression in order to demonstrate their incremental validity beyond what is currently available. The present study aims to fill these significant gaps in the literature by examining digitally assessed nonverbal behaviors during a clinical interview in a large community sample of adolescents up to 3 years prior to the onset of depression. In this longitudinal study, never-depressed, healthy adolescent females were followed for 3 years and were evaluated for a wide range of depression risk factors including self- and parent-reported mood and personality measures, family history and physiological variables. Thereby, the present study has two aims. First, it seeks to investigate whether digitally assessed nonverbal behaviors during a clinical interview, measured by a state of the art commercially available software (FaceReader; Noldus Information Technology, 2010), can forecast depression longitudinally over 3 years in a community sample of adolescent females. Specifically, here we focus on head motions and fine facial movements, quantified by FACS, implementing a data-driven approach. We hypothesize that, consistent with previous research, reduced average facial and head movements will predict a DD at 3-year follow-up. Second, we aim to demonstrate whether digitally assessed nonverbal behaviors have incremental validity along with better established predictors of future onset depression. Based on the prediction accuracy reported in the previous literature, we hypothesize that nonverbal behaviors will demonstrate incremental validity.

Method

Participants

Participants were recruited from the Suffolk County area in New York, USA as part of the longitudinal Adolescent Development of Emotions and Personality Traits (ADEPT) study which aims to examine factors predicting the onset of DD in adolescent females. At the initial enrollment, a total of 550 adolescent females participated in the study. Both adolescents and their parents provided written informed consent prior to participation and the study was approved by Stony Brook University's Institutional Review Board. Exclusion criteria in the study included history of Diagnostic and Statistical Manual (DSM-IV) Major Depressive Disorder (MDD) or Dysthymia before baseline, an intellectual disability, absence of a participating biological parent, inability to read and/or comprehend questionnaires, and a lack of proficiency in English. As part of the larger study, participants were followed at multiple time points throughout the following 3 years. For the purposes of the current study, we have used baseline and the 3-year follow-up data in the analyses. Thirty five participants were excluded due to presence of a DD not otherwise specified (DD-NOS) before baseline, 120 participants were excluded due to missing analyzable video data from the baseline assessment and 36 participants were excluded due to missing diagnostic data at the 3-year follow-up, resulting in a final sample of 359 adolescents (mean age = 14.38, *s.d.* = .63, range = 13–15 years, 90.3% Caucasian, 65.5% with at least one college educated

parent; see online Supplementary Table S1 for further demographic breakdown).

Measures

Diagnostic assessment

DSM-IV diagnoses were assessed at baseline and 3-year follow-up, using the semi-structured diagnostic interview Kiddie Schedule for Affective Disorders and Schizophrenia for School Aged Children, Present and Lifetime Version (KSADS-PL) (Kaufman et al., 1997). Extensively trained research personnel, overseen by clinical psychologists, conducted the interviews. The KSADS-PL has excellent reliability and validity in diagnosing adolescent psychopathology (Kaufman et al., 1997). Diagnostic status at the 3-year follow-up was a dichotomous variable indicating presence or absence of any DD, which included DSM-IV diagnoses of major depressive episode (MDE), dysthymic disorder, and DD-NOS. We operationalized DD-NOS as a clinically significant depressive episode, characterized by presence of depressed mood, loss of interest or pleasure, or suicidality and clinically significant impairment or treatment that did not meet the full criteria for MDE or dysthymic disorder. The inter-rater reliability for any DD diagnosis across study time points was high based on 48 audio-recorded interviews ($\kappa = 0.81$; Michelini et al., 2021). Participants were video recorded while completing the baseline KSADS-PL diagnostic interview. The camera focused on participants' face.

Non-verbal baseline measures

The automatic detection of facial and head movements was conducted using FaceReader version 8.0, a commercially available software package developed by Noldus Information Technology (2010). FaceReader is an automated program that uses an Active Appearance Model (Cootes, Edwards, & Taylor, 2001) based approach to 3D model the face, identify key points in the face and facial texture, and use Convolutional Neural Networks for facial expression classification to calculate the Action Units (AUs) derived from the FACS, on a frame-by-frame basis. It has been shown to be a reliable measure of automated facial expression analysis (Clark et al., 2020; Dupré, Krumhuber, Küster, & McKeown, 2020). The software estimates activity of AUs from the face. For each frame, software calculates intensity of activation in each AU ranging from 0 (not present) to 1 (maximum). Mean intensity was calculated for each AU across all frames with valid data and used in subsequent analyses. It also captures other forms of non-verbal behavior including head movements along the X, Y, and Z axes. The current study obtained an average of 52 275 (s.d. = 33 813) video frames for each participant. Videos from diagnostic interviews were processed on standard office desktop computers, taking 125% to 150% longer than the original recording. Number of frames was used as a covariate in the analyses to mitigate any potential influence that variations in frame count might exert on the results. In order to quantify head movements, we calculated the frame-by-frame Euclidean distance in head position across the x, y, and z planes following Abbas et al. (for further details see Abbas et al., 2021c). For each participant, mean head movement values were calculated from frame-by-frame Euclidean distance. FaceReader's Action Unit module allows extraction of 20 AU's as described by Ekman et al. (2002, Table 1). FaceReader allowed examination of AU's both laterally and bilaterally. For the purposes of current analyses, only full bilateral values of AUs were

used, as lateral AU values were very highly correlated between left and right half.

Self-report, parent-report and physiological baseline measures

To determine the incremental value of non-verbal measures in uniquely forecasting DD, we also examined other baseline predictors of DD. These were selected based on a prior study in which they were found to forecast the first onset of DD in the ADEPT data set (Michelini et al., 2021). These measures included self-report questionnaires tapping depressive symptoms (Inventory of Depression and Anxiety Symptoms General Depression Scales, expanded version; Watson et al., 2012), irritability/hostility (Buss-Perry Aggression Scale; Buss & Perry, 1992), rumination (Response Styles Questionnaire; Nolen-Hoeksema & Morrow, 1991), self-criticism (Depressive Experiences Questionnaire; Blatt, Zohar, Quinlan, Zuroff, & Mongrain, 1995), dependency (Interpersonal Dependency Inventory; Hirschfeld et al., 1977), personality traits (neuroticism/negative affectivity, extraversion/positive affectivity, and conscientiousness) (Big Five Inventory; Soto & John, 2017), relationships with parents and best friend (Network of Relationship Inventory – Relationship Qualities Version; Furman & Buhrmester, 2009), perceived social support (Multidimensional Scale of Perceived Social Support; Zimet, Powell, Farley, Werkman, & Berkoff, 1990), parental warmth (Parental Bonding Instrument; Parker, 1979), and peer victimization (Revised Peer Experiences Questionnaire; De Los Reyes & Prinstein, 2004). History of an anxiety or behavioral disorder in the youth was assessed by KSADS-PL. Parental lifetime history of DSM-IV mood disorder was determined by interviews with participating parents about themselves (Structured Clinical Interview for DSM-IV Axis I Disorders; First, Spitzer, Gibbon, & Williams, 2002) and about the non-participating parents (Family History Screen; Weissman et al., 2000). Parental criticism of the adolescent was assessed through the Five-Minute Speech sample with participating parents (Magaña et al., 1986). See Michelini et al. (2021) for detailed descriptions and psychometric properties of these measures.

Results

Seventy (19.5%) participants experienced a first-onset DD over 3 years, an average of 14.67 (s.d. = 14.26) months after baseline assessment. Participants with and without a first onset DD did not differ in terms of demographics including baseline age, race, parental education, and household income (online Supplementary Table S1). Hence, these variables were not included in the following analyses.

First, we conducted individual bivariate regressions to assess whether each nonverbal behavior marker is associated with DD onset (see Table 1). Next, logistic regression analysis was conducted to examine non-verbal measures as predictors of DD at 3-year follow-up (Table 2). To avoid bias in the analyses, we used a regression model in which the data guided the determination of the key predictors in the model. This was implemented with a forward regression approach in which DD at 3-year follow-up was the dependent variable. In the forward regression, 3 AUs and head movements were significant in forecasting DD at 3-year follow-up: mean brow lowerer (AU4), jaw drop (AU26), eyes closed (AU43) and head movement (Table 2; Fig. 1). Among these significant predictors, head movements, AU4, and 26 survived %5 FDR correction, therefore were included in the further analyses, while AU43 was excluded. As a follow-up analysis, the number of analyzable valid frames from each

Table 1. Bivariate comparisons of 20 Action Units (AUs) and Head Movements predicting DD onset

	<i>M</i>	s.d.	OR	95% CI		<i>p</i>
AU 01 – Inner Brow Raiser	0.018	0.030	–0.06	–2.135	0.571	0.256
AU 02 – Outer Brow Raiser	0.0077	0.019	–0.047	–3.156	1.194	0.376
AU 04 – Brow Lowerer*	0.0115	0.0234	0.119	0.272	3.77	0.024
AU 05 – Upper Lid Raiser	0.0107	0.0247	–0.03	–2.295	1.278	0.576
AU 06 – Cheek Raiser	0.0102	0.0163	0.058	–1.088	3.818	0.274
AU 07 – Lid Tightener*	0.0077	0.0153	0.105	0.056	6.092	0.046
AU 09 – Nose Wrinkler	0.0001	0.0007	0.001	–71.383	73.284	0.979
AU 10 – Upper Lip Raiser	0.0058	0.0136	–0.018	–4.059	2.835	0.727
AU 12 – Lip Corner Puller	0.0574	0.0535	0.01	–0.689	0.829	0.857
AU 14 – Dimpler	0	0.0002	0.023	–147.01	228.64	0.669
AU 15 – Lip Corner Depressor	0.0019	0.0082	–0.076	–8.725	1.366	0.152
AU 17 – Chin Raiser	0.0123	0.0242	–0.049	–2.52	0.915	0.359
AU 18 – Lip Pucker	0.0005	0.0018	0.018	–18.345	26.127	0.731
AU 20 – Lip Stretcher	0.0007	0.0021	0.038	–11.762	25.066	0.478
AU 23 – Lip Tightener	0.0003	0.0014	0.004	–27.716	30.046	0.937
AU 24 – Lip Pressor	0.0005	0.0016	–0.056	–39.378	11.957	0.294
AU 25 – Lips Part	0.1007	0.0771	0.062	–0.211	0.843	0.239
AU 26 – Jaw Drop*	0.015	0.0212	0.168	1.177	4.844	0.001
AU 27 – Mouth Stretch	0.0008	0.0026	0.014	–12.907	17.039	0.786
AU 43 – Eyes Closed*	0.0194	0.0347	0.104	0.009	2.381	0.048
Head Movements*	0.8408	0.2914	0.147	0.06	0.342	0.005

*Variables that survived the 5% FDR correction.

participant's video was included in the analysis as a covariate, however it did not impact the significant findings.

Receiver operating characteristic (ROC) curve analyses were performed to calculate area under the curve (AUC), sensitivity, specificity, positive and negative predictive value for probability of DD onset estimated by the model (i.e. a weighted composite of the four behavioral predictors) (Fig. 2). The AUC was 0.70, suggesting that nonverbal head movements and facial AUs have moderate to low accuracy in discriminating between individuals who will and will not develop a first onset of a DD (Streiner & Cairney, 2007).

Next, we examined whether non-verbal movements have incremental value in forecasting future depression beyond previously established measures. We confirmed that 13 of the 16 traditional risk factors were selected in the present sample (see online Supplementary Table S2), and were included in a second forward entry binary logistic regression with mean head movements and significant AUs. Head movement, brow lowerer (AU4), and jaw drop (AU26) emerged as unique predictors of DD at 3-year follow-up, along with baseline IDAS and RSQ scores (Table 3). ROC analysis demonstrated a moderate accuracy with AUC value of 0.78 (Fig. 3). With lower cutoff, the risk composite showed high sensitivity (0.90), although with the downside of selecting 65% of the sample; thus, the risk composite can be used to screen out adolescents whose risk of DD onset is very low.

Finally, we have created a logistic regression model for the significant psychosocial predictors (IDAS Depression and RSQ)

predicting DD status at 3-year follow-up (see Fig. 4) to demonstrate the predictive value of these variables without nonverbal behavior markers.

Discussion

The clinical and public burden of DD can be mitigated by early prevention efforts that are shown to reduce rates and alleviate the course of the illness (Harrington & Clark, 1998; Ormel, Cuijpers, Jorm, & Schoevers, 2019). While studies show that prevention efforts for adolescents who are at high risk for developing DDs is effective, identification of these individuals remains a major challenge for the field (Kieling et al., 2019). The present study aimed to explore whether nonverbal behaviors, captured by digital assessment tools during a clinical interview in adolescence, may forecast the first onset of DD in 3 years. Our findings demonstrated that nonverbal behaviors, including greater head movement, AU4 (brow lowerer), AU26 (jaw drop), and AU43 (eyes closed) show promise in indexing future DD risk.

Critically, most of these digital measures showed incremental value in predicting depression beyond the previously established predictors of first onset depression in this sample (Michellini et al., 2021). Increased movements of the head as well as AU4 indexing brow lowering, and AU26 indexing jaw dropping were the non-verbal measures that entered the model predicting DD at 3-year follow-up, while many other clinical and psychosocial measures, such as baseline family psychiatric history, personality traits,

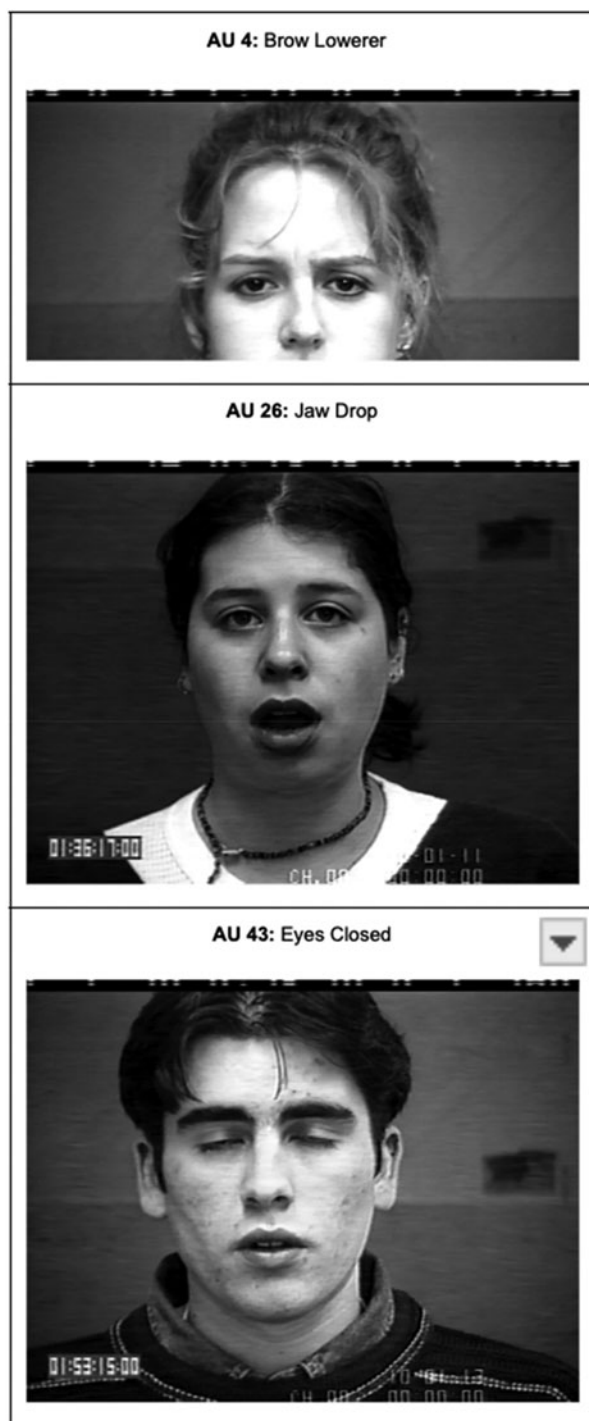


Figure 1. Picture descriptions of AUs that significantly forecast DD at 3-year follow-up.

and peer victimization, did not. Importantly, these non-verbal measures accounted for unique variance over and above baseline self-reported depression score and rumination, which are some of the strongest predictors of future depression (Michelinei et al., 2021). Our non-verbal measures are advantageous in providing an objective, scalable, non-invasive, and cost-effective approach for early risk assessment of DD. The findings from the present study can address various clinical needs depending on the cutoff values applied. For example, when cutoffs that enhance sensitivity (even with lower

specificity) are used, these digital assessments can effectively support widespread screening initiatives. In contrast, cutoffs that prioritize specificity (despite low sensitivity) make these assessments more suitable for confirmation following initial screenings or for use in specialized settings where the illness is more prevalent (Baldessarini, Finklestein, & Arana, 1983).

This is the first study on the digital assessment of nonverbal behaviors that has been conducted with a longitudinal community sample of adolescents. The literature so far has primarily used cross-sectional designs, attempting to detect the presence of DD within a sample, with a few studies examining prediction of the course of the illness over a few months (Dibeklioğlu et al., 2018; Kacem et al., 2018). The present study breaks new ground in forecasting the future first onset of DD during a critical developmental window, with a rich variety of clinical predictors along with nonverbal behavior.

The specific nonverbal behaviors that emerged as having predictive ability in the current study complement prior research. Our results showing that greater movement in AU4 (brow lowerer) predicts DD is in line with research showing that heightened intensity in this AU indexes the underlying activity of the corrugator muscle, which has been shown to be closely associated with depression (Kadison, Ragsdale, Mitchell, Cassisi, & Bedwell, 2015). Moreover, early EMG studies found that amplified corrugator muscle activity, which contributes to a frowning facial expression, may predict prognosis, as well as diagnosis of DD (Greden et al., 1984; Schwartz et al., 1978). On the other hand, although the examination of masseter muscle activity underlying AU26 (jaw dropping) is limited in clinical science literature, the dental literature extensively reports associations between higher self-reported depressive symptoms and EMG measured masseter muscle activity, particularly in women (Gonzalez, Nickel, Scott, Liu, & Iwasaki, 2018), and these findings are associated with jaw related temporomandibular disorders (Khawaja et al., 2015).

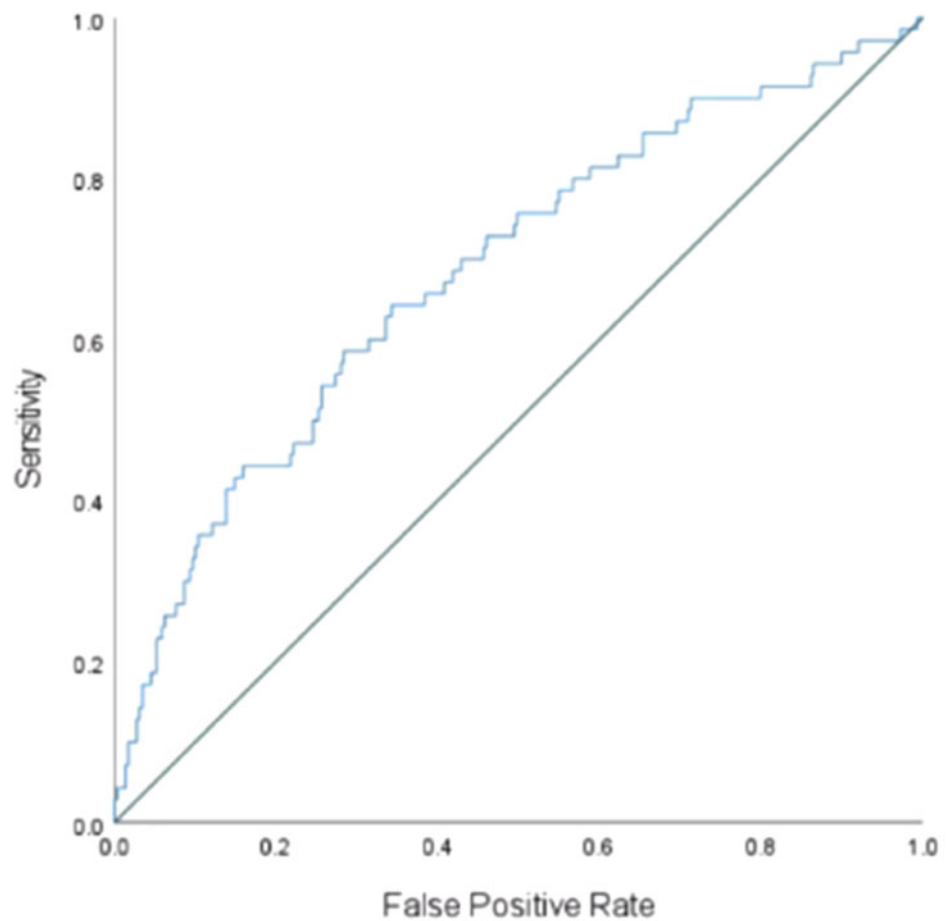
On the other hand, head movements have recently received considerable attention in studies that use digital assessment tools to identify current DD. However, our results depart from prior studies that predominantly linked DD with reduced head movement (Alghowinem et al., 2013; Joshi et al., 2013). Over time, head movement in patient videos increased as depression symptoms ameliorated in the laboratory (Girard et al., 2014) and remote assessments via smartphone-based video analyses (Abbas et al., 2021a). Several factors could explain the observed increase rather than decrease in head movement activity in the present study. Notably, previous research exclusively involved currently depressed adult participants during non-verbal data acquisition. In contrast, the present sample involves *healthy* adolescent participants who might be *at risk* for DD. In such a sample, the pattern of relationship between motor movement characteristics and DD might manifest differently.

Depression is characterized by both observable slowing of (psychomotor retardation) as well as increased (agitation) psychomotor movement (American Psychiatric Association, 2013; Sobin, 1997). Specifically, psychomotor agitation is defined as 'restless physical activity arising from mental disturbance' (APA Dictionary of Psychology, n.d.). The increased average head movements that are measured in 3D space is a potential method to quantify psychomotor agitation. In adolescence, self-reported psychomotor agitation has been demonstrated to couple with family loading of risk for depression, forecast future depression symptoms in 1-year (Damme et al., 2021), and index a transdiagnostic risk marker for both DD and psychotic-like experiences

Table 2. Forward-Entry Logistic Regression predicting DD at 3-year follow-up with nonverbal behaviors

	Beta	s.e.	Wald	df	p	OR	95% CI	
							Lower	Upper
AU 4 – Brow Lowerer	0.306	0.117	6.862	1	0.009	1.357	1.08	1.706
AU 26 – Jaw Drop	0.394	0.122	10.46	1	0.001	1.483	1.168	1.884
AU 43 – Eyes Closed	0.27	0.118	5.24	1	0.022	1.309	1.04	1.649
Head Movements	0.396	0.132	8.996	1	0.003	1.485	1.147	1.923

The entry of independent variables was set at 0.05 and variable removal was set at 0.10. Independent variables in the model included mean values of 20 AUs and head movements. All continuous variables were standardized.



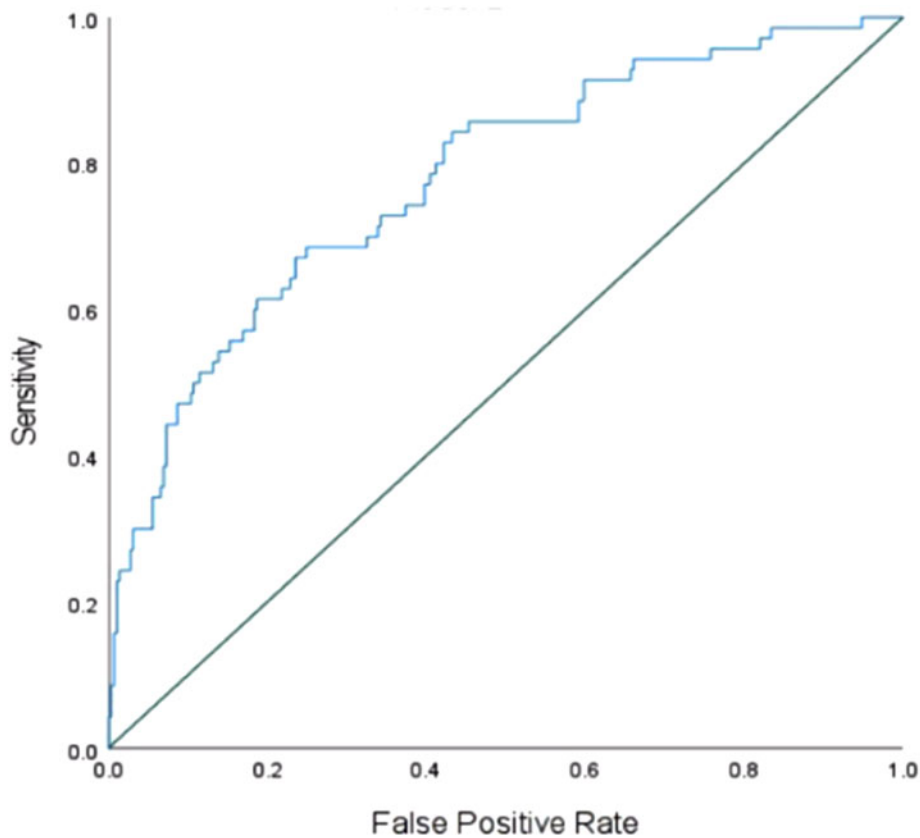
Area Under the Curve	Cutoff Value	Proportion Screen Positive	Sensitivity	Specificity	PPV	NPV
0.70	0.11	0.80	0.90	0.23	0.22	0.90
	0.13	0.63	0.80	0.41	0.25	0.89
	0.16	0.50	0.70	0.55	0.27	0.88
	0.18	0.39	0.61	0.67	0.31	0.87
	0.22	0.30	0.51	0.75	0.33	0.86
	0.26	0.30	0.41	0.85	0.39	0.85
	0.31	0.15	0.35	0.90	0.46	0.85

Figure 2. ROC curve for the logistic regression predicting DD at 3-year follow-up with nonverbal behaviors.

Table 3. Logistic Regression of head movements, significant action units (AUs) and baseline self- & parent-reported and biological measures predicting DD at 3-year follow-up

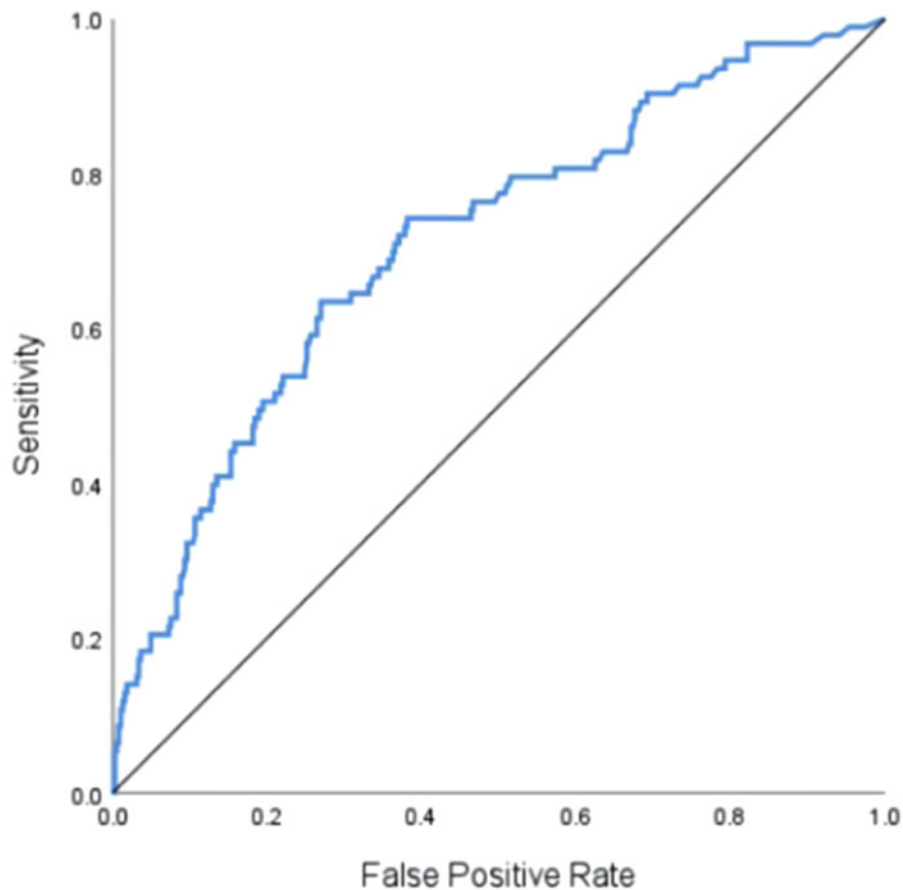
	Beta	s.e.	Wald	df	p	OR	95% CI	
							Lower	Upper
Depressive Symptoms	0.557	0.187	8.905	1	0.003	1.746	1.211	2.517
Rumination	0.443	0.178	6.201	1	0.013	1.557	1.099	2.206
Head Movements	0.342	0.141	5.86	1	0.015	1.408	1.067	1.858
AU 04 – Brow Lowerer	0.387	0.133	8.517	1	0.004	1.473	1.136	1.911
AU 26 – Jaw Drop	0.47	0.151	9.661	1	0.002	1.6	1.19	2.152

In this model, depression diagnosis at 3-year follow-up was again used as the dependent variable, and the same model parameters were used.



Area Under the Curve	Cutoff Value	Proportion Screen Positive	Sensitivity	Specificity	PPV	NPV
0.78	0.10	0.65	0.90	0.41	0.27	0.94
	0.14	0.50	0.80	0.58	0.31	0.92
	0.16	0.43	0.70	0.64	0.32	0.90
	0.22	0.31	0.60	0.77	0.38	0.89
	0.28	0.21	0.50	0.86	0.47	0.88
	0.40	0.13	0.40	0.94	0.61	0.87

Figure 3. ROC curve for the logistic regression demonstrating incremental validity of nonverbal behaviors along with previously established risk markers of DD at 3-year follow-up.



Area Under the Curve	Cutoff Value	Proportion Screen Positive	Sensitivity	Specificity	PPV	NPV
0.71	0.11	0.42	0.90	0.70	0.42	0.97
	0.12	0.48	0.81	0.60	0.33	0.93
	0.16	0.65	0.71	0.37	0.21	0.84
	0.19	0.71	0.60	0.26	0.16	0.73
	0.23	0.74	0.51	0.20	0.13	0.63
	0.28	0.77	0.41	0.14	0.10	0.49
	0.32	0.86	0.30	0.01	0.07	0.06

Figure 4. ROC curve for the logistic regression including IDAS-depression and RSQ predicting DD status at 3-year follow-up.

(Damme et al., 2022). Hence, it is possible that greater overall head movements reflect psychomotor agitation which might be a prominent marker of risk for DD that is unique to adolescence.

Second, phenotypic expressions of DD are highly heterogeneous and nuanced. Both psychomotor agitation and retardation may represent different subtypes (Leventhal et al., 2008; Schrijvers, Hulstijn, & Sabbe, 2008). Furthermore, there is weak evidence regarding gender differences in psychomotor movements of DD. Some early studies suggest potential sex-related differences in psychomotor abnormalities where some studies report more

pronounced psychomotor agitation in females than males (Avery & Silverman, 1984; Sobin, 1997; Winokur, Morrison, Clancy, & Crowe, 1973), while others report mixed or null results (Khan, Gardner, Prescott, & Kendler, 2002; Kornstein et al., 2000). However, questions regarding nonverbal behavior were not pursued extensively in more recent decades and psychomotor agitation of gross movements remained largely unexamined (Schrijvers et al., 2008).

Furthermore, it is crucial to emphasize that in the current study, the digital assessment of nonverbal behaviors was

conducted within the context of clinical diagnostic interviews. Consequently, the heightened head movement observed in participants who subsequently developed a DD might be a proxy of the significant distressing issues they shared during the interview, even though they had not yet experienced a DD at the time. In contrast, well-adjusted participants might have fewer issues to report, resulting in a more stable psychomotor movement during the interview.

Overall, nonverbal psychomotor behavior has long been viewed as a potential diagnostic and risk marker in clinical science, however research has been limited and constrained due to flaws with assessment methods (Schrijvers et al., 2008; Sobin, 1997). Digital phenotyping currently offers an exciting and novel approach for objective, scalable, cost-effective assessment of nonverbal behaviors to address the essential need for early identification of DD vulnerability. Furthermore, nonverbal behavior can be indicative of underlying mechanisms of depression (Girard & Cohn, 2015), as it is not limited to muscle contractions but involves perceptual processes and cognitive-control mechanisms that underlie the muscle activity (Schrijvers et al., 2008). In fact, electrical stimulation of facial musculature has been recently proposed as a potential intervention method for DD. As nonverbal psychomotor behavior may reflect underlying pathophysiology, changing facial muscle activity may have an impact on affect, in line with the facial feedback hypothesis (Demchenko et al., 2023).

There are a number of limitations of current study. The sample was predominantly white, mirroring the composition of the study sample and representing the demographic picture of Suffolk County, New York, and limited to female adolescents. To enhance the robustness of the findings and broaden applicability, replication of the study with diverse demographic profiles, including varying racial, ethnic, and cultural groups, is imperative (Barrett, Adolphs, Marsella, Martinez, & Pollak, 2019). On the other hand, there is recent evidence demonstrating a relatively high level of universality of nonverbal facial movements. By utilizing machine learning methods, evidence suggests 70% consistency of facial expressions in similar contexts, such as wedding as sports games, across 144 countries in 12 different world regions (Cowen et al., 2021). Future research should examine the applicability of findings related to nonverbal behaviors and DDs to broader and more diverse contexts, both encompassing demographic and clinical diversity. In addition to replication, it is crucial to assess the extent to which current facial recognition technology permits generalization to diverse populations and real-world applicability. Specifically, for the computer-vision-based facial recognition technology, there are concerns regarding the generalizability and equity of the methods (Buolamwini & Gebru, 2018), underscoring the need for carefully considering technology's potential biases. Algorithms rapidly adopt and reflect societal biases, including racism and sexism, which are now well-documented in current facial recognition technology. The high accuracy of facial identification is not universal; it is primarily effective for European white and male facial features. In contrast, error rates for darker-skinned females can reach up to 34% (Buolamwini & Gebru, 2018; Phillips, O'Toole, Jiang, Narvekar, & Ayadd, 2011). This issue has not been documented for facial emotion recognition algorithms. In particular, FaceReader was trained on a set of faces that included many African and Asian individuals (Spink, Barski, Brouwer, Riedel, & Sil, 2022). However, racial biases may be presented in emotion recognition models, even if not detected or will emerge in the future. Consequently,

deploying these technologies in real-world applications without extensive and meticulously conducted research and considerations can exacerbate existing inequalities in the mental health system (Maura & Weisman de Mamani, 2017), by potentially leading to inaccurate assessments, misdiagnoses, or overlooked symptoms in underrepresented populations.

Moreover, researchers and clinicians should exercise caution in employing digital methods in both research and real-world applications. While these methods hold great promise, it is crucial to acknowledge that the validation and regulation of digital measurement tools present a spectrum of ethical and privacy-related concerns. One primary issue is the potential for invasions of privacy, as facial recognition involves the collection and storage of sensitive biometric data, which can be prone to misuse or unauthorized access. There are also concerns about consent, as individuals may not fully understand how their data will be used or the implications of its collection. Finally, protection and regulations against mishandling and exposure of facial recognition data holds significant implications for further stigmatization and discrimination due to mental health conditions. Ensuring robust safeguards, transparency, and patient consent is essential prior to the application of facial recognition technology in mental health care. These concerns should be scrutinized closely in future research to ensure responsible and ethical use of such tools.

Overall, nonverbal psychomotor behavior has long been viewed as a potential diagnostic and risk marker in clinical science, however research has been limited and constrained due to flaws with assessment methods (Schrijvers et al., 2008; Sobin, 1997). For example, it can be integrated into the routine preventive care visits at a pediatrician's office, contributing to a broader evaluation that includes current symptoms of depression, anxiety, and rumination. Effect sizes that we observed for nonverbal behavior were too small to use them in isolation; these behavioral markers show potential to enhance prediction of future DD as part of a larger panel of risk factors. Nevertheless, additional research is essential to enhance precision of these markers. The current study is unique in presenting a multimodal assessment of longitudinal risk factors for DD in a large sample of adolescent girls. It is important to reiterate that most research studies in the literature that use digital methods to examine nonverbal behavior only predict concurrent depressive symptoms and the few longitudinal studies only follow participants for a limited duration. Our study stands out as following a non-depressed group of adolescents over 3 years, which covers a large span of the significant developmental period. By providing a comparison to other, better established, clinical and psychosocial risk markers, our study fills an important gap in the literature (Cohn et al., 2018). Thereby, it underscores the capacity of facial recognition as a promising avenue for future research. With further research support, it can be applied to diverse clinical settings with minimal effort, resources and training, to quantify nonverbal movements objectively, so that clinicians can access an efficient, easy, cost-effective tool for risk assessment that would aid in prevention and intervention of future DD. Future research should replicate these findings extensively with diverse demographic profiles and phenotypic expressions of DD, in combination with multimodal assessment of DD signs and risk.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291724002010>.

Funding statement. This study was supported by National Institute of Mental Health Grants R01 MH093479 (Roman Kotov) and R56 MH117116

(Roman Kotov and Daniel N. Klein). The authors have no disclosures to report.

Competing interests. None.

Ethical standards. 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.

References

- Birmaher, B., Brent, D., AACAP Work Group on Quality Issues, Bernet, W., Bukstein, O., Walter, H., ... Medicus, J. (2007). Practice parameter for the assessment and treatment of children and adolescents with depressive disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46(11), 1503–1526. <https://doi.org/10.1097/chi.0b013e318145ae1c>
- Abbas, A., Sauder, C., Yadav, V., Koesmahargyo, V., Aghjayan, A., Marecki, S., ... Galatzer-Levy, I. R. (2021a). Remote digital measurement of facial and vocal markers of major depressive disorder severity and treatment response: A pilot study. *Frontiers in Digital Health*, 3, 610006. <https://doi.org/10.3389/fgdh.2021.610006>
- Abbas, A., Schultebrucks, K., & Galatzer-Levy, I. R. (2021b). Digital measurement of mental health: Challenges, promises, and future directions. *Psychiatric Annals*, 51(1), 14–20. <https://doi.org/10.3928/00485713-20201207-01>
- Abbas, A., Yadav, V., Smith, E., Ramjas, E., Rutter, S. B., Benavidez, C., ... Galatzer-Levy, I. R. (2021c). Computer vision-based assessment of motor functioning in schizophrenia: Use of smartphones for remote measurement of schizophrenia symptomatology. *Digital Biomarkers*, 5(1), 29–36. <https://doi.org/10.1159/000512383>
- Alghowinem, S., Goecke, R., Wagner, M., Parkerx, G., & Breakspear, M. (2013). *Head pose and movement analysis as an indicator of depression. 2013 humane association conference on affective computing and intelligent interaction*, pp. 283–288. <https://doi.org/10.1109/ACII.2013.53>
- Altmann, U., Brummel, M., Meier, J., & Strauss, B. (2021). Movement synchrony and facial synchrony as diagnostic features of depression: A pilot study. *The Journal of Nervous and Mental Disease*, 209(2), 128–136. <https://doi.org/10.1097/nmd.0000000000001268>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. Washington, D.C.: American Psychiatric Association. <https://doi.org/10.1176/appi.books.9780890425596>
- APA Dictionary of Psychology. (n.d.). Retrieved November 2, 2023, from <https://dictionary.apa.org/>
- Avery, D., & Silverman, J. (1984). Psychomotor retardation and agitation in depression. Relationship to age, sex, and response to treatment. *Journal of Affective Disorders*, 7(1), 67–76. [https://doi.org/10.1016/0165-0327\(84\)90066-1](https://doi.org/10.1016/0165-0327(84)90066-1)
- Baldessarini, R. J., Finklestein, S., & Arana, G. W. (1983). The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry*, 40(5), 569–573. <https://doi.org/10.1001/archpsyc.1983.01790050095011>
- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1), 1–68. <https://doi.org/10.1177/1529100619832930>
- Blatt, S. J., Zohar, A. H., Quinlan, D. M., Zuroff, D. C., & Mongrain, M. (1995). Subscales within the dependency factor of the Depressive Experiences Questionnaire. *Journal of Personality Assessment*, 64(2), 319–339. https://doi.org/10.1207/s15327752jpa6402_11
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st conference on fairness, accountability and transparency*, pp. 77–91. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, 63(3), 452–459. <https://doi.org/10.1037/0022-3514.63.3.452>
- Chentsova-Dutton, Y. E., Tsai, J. L., & Gotlib, I. H. (2010). Further evidence for the cultural norm hypothesis: Positive emotion in depressed and control European American and Asian American women. *Cultural Diversity & Ethnic Minority Psychology*, 16(2), 284–295. <https://doi.org/10.1037/a0017562>
- Clark, E. A., Kessinger, J., Duncan, S. E., Bell, M. A., Lahne, J., Gallagher, D. L., & O’Keefe, S. F. (2020). The facial action coding system for characterization of human affective response to consumer product-based stimuli: A systematic review. *Frontiers in Psychology*, 11, 920. <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00920>
- Cohn, J. F., Krueez, T. S., Matthews, I., Yang, Y., Nguyen, M. H., Padilla, M. T., ... De La Torre, F. (2009). Detecting depression from facial actions and vocal prosody. In *3rd International conference on affective computing and intelligent interaction and workshops*.
- Cohn, J. F., Cummins, N., Epps, J., Goecke, R., Joshi, J., & Scherer, S. (2018). Multimodal assessment of depression from behavioral signals. In Monash University, S. Oviatt, B. Schuller, University of Augsburg and Imperial College London, P. R. C. Cohen, Monash University, D. Sonntag, German Research Center for Artificial Intelligence (DFKI), G. Potamianos, University of Thessaly, A. Krüger, & Saarland University and German Research Center for Artificial Intelligence (DFKI) (Eds.), *The handbook of multimodal-multisensor interfaces: Foundations, user modeling, and common modality combinations – volume 2* (pp. 375–417). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3107990.3108004>
- Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6), 681–685. <https://doi.org/10.1109/34.927467>
- Cowen, A. S., Keltner, D., Schroff, F., Jou, B., Adam, H., & Prasad, G. (2021). Sixteen facial expressions occur in similar contexts worldwide. *Nature*, 589(7841), 251–257. <https://doi.org/10.1038/s41586-020-3037-7>
- Damme, K. S. F., Park, J. S., Vargas, T., Walther, S., Shankman, S. A., & Mittal, V. A. (2021). Motor abnormalities, depression risk, and clinical course in adolescence. *Biological Psychiatry Global Open Science*, 2(1), 61–69. <https://doi.org/10.1016/j.bpsgos.2021.06.011>
- Damme, K. S. F., Park, J. S., Walther, S., Vargas, T., Shankman, S. A., & Mittal, V. A. (2022). Depression and psychosis risk shared vulnerability for motor signs across development, symptom dimensions, and familial risk. *Schizophrenia Bulletin*, 48(4), 752–762. <https://doi.org/10.1093/schbul/sbab133>
- De Los Reyes, A., & Prinstein, M. J. (2004). Applying depression-distortion hypotheses to the assessment of peer victimization in adolescents. *Journal of Clinical Child and Adolescent Psychology*, 33(2), 325–335. https://doi.org/10.1207/s15374424jccp3302_14
- Demchenko, I., Desai, N., Iwasa, S. N., Gholamali Nezhad, F., Zariffa, J., Kennedy, S. H., ... Bhat, V. (2023). Manipulating facial musculature with functional electrical stimulation as an intervention for major depressive disorder: A focused search of literature for a proposal. *Journal of NeuroEngineering and Rehabilitation*, 20(1), 64. <https://doi.org/10.1186/s12984-023-01187-8>
- Dibeklioğlu, H., Hammal, Z., & Cohn, J. F. (2018). Dynamic multimodal measurement of depression severity using deep autoencoding. *IEEE Journal of Biomedical and Health Informatics*, 22(2), 525–536. <https://doi.org/10.1109/JBHI.2017.2676878>
- Dupré, D., Krumhuber, E. G., Küster, D., & McKeown, G. J. (2020). A performance comparison of eight commercially available automatic classifiers for facial affect recognition. *PLoS ONE*, 15(4), e0231968. <https://doi.org/10.1371/journal.pone.0231968>
- Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preotiuc-Pietro, D., ... Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, 115(44), 11203–11208. <https://doi.org/10.1073/pnas.1802331115>
- Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system: Manual and investigator’s guide*. Salt Lake City, UT: Research Nexus.
- Fairbanks, L. A., McGuire, M. T., & Harris, C. J. (1982). Nonverbal interaction of patients and therapists during psychiatric interviews. *Journal of Abnormal Psychology*, 91(2), 109–119.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. (2002). *Structured Clinical Interview For DSM-IV-TR Axis I disorders, Research version, Non-Patient Edition (SCID-I/NP)*. New York: New York State Psychiatric Institute.
- Furman, W., & Buhrmester, D. (2009). The network of relationships inventory: Behavioral systems version. *International Journal of Behavioral Development*, 33(5), 470–478. <https://doi.org/10.1177/0165025409342634>

- Gaebel, W., & Wölwer, W. (2004). Facial expressivity in the course of schizophrenia and depression. *European Archives of Psychiatry and Clinical Neuroscience*, 254(5), 335–342. <https://doi.org/10.1007/s00406-004-0510-5>
- Gavrilescu, M., & Vizeanu, N. (2019). Predicting depression, anxiety, and stress levels from videos using the facial action coding system. *Sensors (Basel, Switzerland)*, 19(17), 3693. <https://doi.org/10.3390/s19173693>
- Gehricke, J., & Shapiro, D. (2000). Reduced facial expression and social context in major depression: Discrepancies between facial muscle activity and self-reported emotion. *Psychiatry Research*, 95(2), 157–167. [https://doi.org/10.1016/S0165-1781\(00\)00168-2](https://doi.org/10.1016/S0165-1781(00)00168-2)
- Girard, J. M., & Cohn, J. F. (2015). Automated audiovisual depression analysis. *Current Opinion in Psychology*, 4, 75–79. <https://doi.org/10.1016/j.copsy.2014.12.010>
- Girard, J. M., Cohn, J. F., Mahoor, M. H., Mavadati, S. M., Hammal, Z., & Rosenwald, D. P. (2014). Nonverbal social withdrawal in depression: Evidence from manual and automatic analysis. *Image and Vision Computing*, 32(10), 641–647. <https://doi.org/10.1016/j.imavis.2013.12.007>
- Gonzalez, Y. M., Nickel, J. C., Scott, J. M., Liu, H., & Iwasaki, L. R. (2018). Psychosocial scores and jaw muscle activity in women. *Journal of Oral & Facial Pain and Headache*, 32(4), 381–388. <https://doi.org/10.11607/ofph.2133>
- Greden, J. F., Price, H. L., Genero, N., Feinberg, M., & Levine, S. (1984). Facial EMG activity levels predict treatment outcome in depression. *Psychiatry Research*, 13(4), 345–352. [https://doi.org/10.1016/0165-1781\(84\)90082-9](https://doi.org/10.1016/0165-1781(84)90082-9)
- Greden, J. F., Genero, N., & Price, H. L. (1985). Agitation-increased electromyogram activity in the corrugator muscle region: A possible explanation of the 'Omega sign'? *The American Journal of Psychiatry*, 142(3), 348–351. <https://doi.org/10.1176/ajp.142.3.348>
- Greden, J. F., Genero, N., Price, H. L., Feinberg, M., & Levine, S. (1986). Facial electromyography in depression. Subgroup differences. *Archives of General Psychiatry*, 43(3), 269–274. <https://doi.org/10.1001/archpsyc.1986.01800030087009>
- Harrington, R., & Clark, A. (1998). Prevention and early intervention for depression in adolescence and early adult life. *European Archives of Psychiatry and Clinical Neuroscience*, 248(1), 32–45. <https://doi.org/10.1007/s004060050015>
- Hirschfeld, R. M., Klerman, G. L., Gough, H. G., Barrett, J., Korchin, S. J., & Chodoff, P. (1977). A measure of interpersonal dependency. *Journal of Personality Assessment*, 41(6), 610–618. https://doi.org/10.1207/s15327752jpa4106_6
- Insel, T. R. (2017). Digital phenotyping: Technology for a new science of behavior. *JAMA*, 318(13), 1215. <https://doi.org/10.1001/jama.2017.11295>
- Joshi, J., Goecke, R., Parker, G., & Breakspear, M. (2013). Can body expressions contribute to automatic depression analysis? In *Automatic face and gesture recognition (FG), 2013 10th IEEE international conference and workshops*. <https://doi.org/10.1109/FG.2013.6553796>
- Kacem, A., Hammal, Z., Daoudi, M., & Cohn, J. (2018). Detecting depression severity by interpretable representations of motion dynamics. *Proceedings of the ... international conference on automatic face and gesture recognition. IEEE international conference on automatic face & gesture recognition*, pp. 739–745. <https://doi.org/10.1109/FG.2018.00116>
- Kadison, L. S., Ragsdale, K. A., Mitchell, J. C., Cassisi, J. E., & Bedwell, J. S. (2015). Subtypes of anhedonia and facial electromyography response to negative affective pictures in non-psychiatric adults. *Cognitive Neuropsychiatry*, 20(1), 31–40. <https://doi.org/10.1080/13546805.2014.955172>
- Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., ... Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 980–988. <https://doi.org/10.1097/00004583-199707000-00021>
- Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. *Annual Review of Public Health*, 34, 119–138. <https://doi.org/10.1146/annurev-publhealth-031912-114409>
- Khan, A. A., Gardner, C. O., Prescott, C. A., & Kendler, K. S. (2002). Gender differences in the symptoms of major depression in opposite-sex dizygotic twin pairs. *The American Journal of Psychiatry*, 159(8), 1427–1429. <https://doi.org/10.1176/appi.ajp.159.8.1427>
- Khawaja, S. N., Iwasaki, L. R., Dunford, R., Nickel, J. C., McCall, W., Crow, H. C., & Gonzalez, Y. (2015). Association of masseter muscle activities during awake and sleep periods with self-reported anxiety, depression, and somatic symptoms. *Journal of Dental Health, Oral Disorders & Therapy*, 2(1), 00039. <https://doi.org/10.15406/jdhodt.2015.02.00039>
- Kieling, C., Adewuya, A., Fisher, H. L., Karmacharya, R., Kohrt, B. A., Swartz, J. R., & Mondelli, V. (2019). Identifying depression early in adolescence. *The Lancet, Child & Adolescent Health*, 3(4), 211–213. [https://doi.org/10.1016/S2352-4642\(19\)30059-8](https://doi.org/10.1016/S2352-4642(19)30059-8)
- Kornstein, S. G., Schatzberg, A. F., Thase, M. E., Yonkers, K. A., McCullough, J. P., Keitner, G. I., ... Keller, M. B. (2000). Gender differences in chronic major and double depression. *Journal of Affective Disorders*, 60(1), 1–11. [https://doi.org/10.1016/S0165-0327\(99\)00158-5](https://doi.org/10.1016/S0165-0327(99)00158-5)
- Kring, A. M., & Stuart, B. K. (2008). Nonverbal behavior and psychopathology. In J. Harrigan, R. Rosenthal, & K. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (pp. 313–339). New York, NY: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198529620.003.0008>
- Lanyon, R. I., & Wershba, R. E. (2013). The effect of underreporting response bias on the assessment of psychopathology. *Psychological Assessment*, 25(2), 331–338. <https://doi.org/10.1037/a0030914>
- Leventhal, A. M., Pettit, J. W., & Lewinsohn, P. M. (2008). Characterizing major depression phenotypes by presence and type of psychomotor disturbance in adolescents and young adults. *Depression & Anxiety (1091–4269)*, 25(7), 575–592. <https://doi.org/10.1002/da.20328>
- Levin-Aspenson, H. F., & Watson, D. (2018). Mode of administration effects in psychopathology assessment: Analyses of gender, age, and education differences in self-rated versus interview-based depression. *Psychological Assessment*, 30(3), 287–295. <https://doi.org/10.1037/pas0000474>
- Magaña, A. B., Goldstein, J. M., Karno, M., Miklowitz, D. J., Jenkins, J., & Falloon, I. R. (1986). A brief method for assessing expressed emotion in relatives of psychiatric patients. *Psychiatry Research*, 17(3), 203–212. [https://doi.org/10.1016/0165-1781\(86\)90049-1](https://doi.org/10.1016/0165-1781(86)90049-1)
- Maura, J., & Weisman de Mamani, A. (2017). Mental health disparities, treatment engagement, and attrition among racial/ethnic minorities with severe mental illness: A review. *Journal of Clinical Psychology in Medical Settings*, 24(3–4), 187–210. <https://doi.org/10.1007/s10880-017-9510-2>
- Michellini, G., Perlman, G., Tian, Y., Mackin, D. M., Nelson, B. D., Klein, D. N., & Kotov, R. (2021). Multiple domains of risk factors for first onset of depression in adolescent girls. *Journal of Affective Disorders*, 283, 20–29. <https://doi.org/10.1016/j.jad.2021.01.036>
- Mittal, V. A., & Wakschlag, L. S. (2017). Research domain criteria (RDoC) grows up: Strengthening neurodevelopment investigation within the RDoC framework. *Journal of Affective Disorders*, 216, 30–35. <https://doi.org/10.1016/j.jad.2016.12.011>
- Noldus Information Technology. (2010). *Facial expression recognition software | FaceReader*. Noldus. <https://www.noldus.com/facereader>
- Nolen-Hoeksema, S., & Morrow, J. (1991). A prospective study of depression and posttraumatic stress symptoms after a natural disaster: The 1989 Loma Prieta Earthquake. *Journal of Personality and Social Psychology*, 61(1), 115–121. <https://doi.org/10.1037/0022-3514.61.1.115>
- Ormel, J., Cuijpers, P., Jorm, A. F., & Schoevers, R. (2019). Prevention of depression will only succeed when it is structurally embedded and targets big determinants. *World Psychiatry*, 18(1), 111–112. <https://doi.org/10.1002/wps.20580>
- Pampouchidou, A., Padiaditis, M., Kazantzaki, E., Sfakianakis, S., Apostolaki, I. A., Argyraki, K., ... Simos, P. (2020). Automated facial video-based recognition of depression and anxiety symptom severity: Cross-corpus validation. *Machine Vision and Applications*, 31(4), 30. <https://doi.org/10.1007/s00138-020-01080-7>
- Parker, G. (1979). Parental characteristics in relation to depressive disorders. *The British Journal of Psychiatry*, 134(2), 138–147. <https://doi.org/10.1192/bjp.134.2.138>
- Phillips, P., O'Toole, A., Jiang, F., Narvekar, A., & Ayadd, J. (2011). An other-race effect for face recognition algorithms. *ACM Transactions on Applied Perception*, 8(2), 14. <https://doi.org/10.1145/1870076.187008>
- Salk, R. H., Hyde, J. S., & Abramson, L. Y. (2017). Gender differences in depression in representative national samples: Meta-analyses of diagnoses and symptoms. *Psychological Bulletin*, 143(8), 783–822. <https://doi.org/10.1037/bul0000102>

- Schrijvers, D., Hulstijn, W., & Sabbe, B. G. C. (2008). Psychomotor symptoms in depression: A diagnostic, pathophysiological and therapeutic tool. *Journal of Affective Disorders*, *109*(1–2), 1–20. <https://doi.org/10.1016/j.jad.2007.10.019>
- Schwartz, G. E., Fair, P. L., Mandel, M. R., Salt, P., Mieske, M., & Klerman, G. L. (1978). Facial electromyography in the assessment of improvement in depression. *Psychosomatic Medicine*, *40*(4), 355–360. <https://doi.org/10.1097/00006842-197806000-00008>
- Snowden, L. R. (2003). Bias in mental health assessment and intervention: Theory and evidence. *American Journal of Public Health*, *93*(2), 239–243.
- Sobin, C. (1997). Psychomotor symptoms of depression. *American Journal of Psychiatry*, *154*(1), 4–17.
- Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, *113*(1), 117–143. <https://doi.org/10.1037/pspp0000096>
- Spink, A., Barski, J., Brouwer, A.-M., Riedel, G., & Sil, A. (Eds.). (2022). *Proceedings of the joint 12th international conference on methods and techniques in behavioral research and 6th seminar on behavioral methods* (Vol. 2). Measuring Behavior. <https://doi.org/10.6084/m9.figshare.20066849>
- Stolicyn, A., Steele, J. D., & Seriès, P. (2022). Prediction of depression symptoms in individual subjects with face and eye movement tracking. *Psychological Medicine*, *52*(9), 1784–1792. <https://doi.org/10.1017/s0033291720003608>
- Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. *Canadian Journal of Psychiatry*. *Revue Canadienne De Psychiatrie*, *52*(2), 121–128. <https://doi.org/10.1177/070674370705200210>
- Trémeau, F., Malaspina, D., Duval, F., Corrêa, H., Hager-Budny, M., Coin-Bariou, L., ... Gorman, J. M. (2005). Facial expressiveness in patients with schizophrenia compared to depressed patients and nonpatient comparison subjects. *The American Journal of Psychiatry*, *162*(1), 92–101. <https://doi.org/10.1176/appi.ajp.162.1.92>
- Friedman, R. J., Katz, M. M., & United States (Eds.). (1974). *The psychology of depression: Contemporary theory and research*. Winston: [distributed by Halsted Press Division, Wiley].
- Watson, D., O'Hara, M. W., Naragon-Gainey, K., Koffel, E., Chmielewski, M., Kotov, R., ... Ruggero, C. J. (2012). Development and validation of new anxiety and bipolar symptom scales for an expanded version of the IDAS (the IDAS-II). *Assessment*, *19*(4), 399–420. <https://doi.org/10.1177/1073191112449857>
- Weissman, M. M., Wickramaratne, P., Adams, P., Wolk, S., Verdelli, H., & Olfson, M. (2000). Brief screening for family psychiatric history: the family history screen. *Archives of General Psychiatry*, *57*(7), 675–682. <https://doi.org/10.1001/archpsyc.57.7.675>
- Winokur, G., Morrison, J., Clancy, J., & Crowe, R. (1973). The Iowa 500: Familial and clinical findings favor two kinds of depressive illness. *Comprehensive Psychiatry*, *14*(2), 99–106. [https://doi.org/10.1016/0010-440x\(73\)90002-3](https://doi.org/10.1016/0010-440x(73)90002-3)
- Woody, M. L., Rosen, D., Allen, K. B., Price, R. B., Hutchinson, E., Amole, M. C., & Silk, J. S. (2019). Looking for the negative: Depressive symptoms in adolescent girls are associated with sustained attention to a potentially critical judge during in vivo social evaluation. *Journal of Experimental Child Psychology*, *179*, 90–102. <https://doi.org/10.1016/j.jecp.2018.10.011>
- Zimet, G. D., Powell, S. S., Farley, G. K., Werkman, S., & Berkoff, K. A. (1990). Psychometric characteristics of the Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment*, *55*(3–4), 610–617. <https://doi.org/10.1080/00223891.1990.9674095>