Can an Unobtrusive, Multimodal Mixed-Effects Regressor Based on Open-Ended Interviews Predict OCD Severity?

Saurabh Hinduja*, Ali Darzi*, Itir Onal Ertugrul, Nicole Provenza, Ron Gadot, Eric A Storch, Sameer A Sheth, Wayne K Goodman, and Jeffrey F Cohn#

Abstract—Reliable, valid, efficient measurement of symptom severity in internalizing disorders is critical to gauge treatment response. Self-report and clinical interview are subjective and difficult to standardize, impose patient burden, and lack granularity. We tested the hypothesis that comprehensive sampling of audio and visual modalities during open-ended interviews can reveal severity of obsessive-compulsive disorder (OCD) and comorbid depression. Participants were six patients with chronic, refractory OCD that were treated with deep brain stimulation (DBS). They were recorded during open-ended interviews at pre- and post-surgery baselines and at 3-month intervals following activation of the DBS. Ground-truth severity was assessed by clinical interview and self-report. Visual and auditory modalities included facial action units, head and facial landmarks, speech behavior and content, and voice acoustics. Using mixed-effects random forest regression with Shapley feature reduction strongly predicted severity of OCD, severity of comorbid depression, and total electrical energy delivered by the DBS electrodes (ICC = 0.83, 0.87, and 0.81, respectively). Multimodal measures of behavior outperformed ones from single modalities. The approach could contribute to closed-loop DBS that would automatically titrate DBS based on affect measures.

Index Terms—Obsessive-Compulsive Disorder (OCD), Depression, Deep Brain Stimulation (DBS), Mixed-effects, multimodal machine learning, Shapley feature reduction

1 Introduction

INTERNALIZING disorders (e.g., obsessive-compulsive disorder and depression) are characterized by anxiety, depressive, and somatic symptoms [1]. Advances in the development and provision of effective treatments for internalizing disorders depend on patient self-report and clinical interview. Self-report is limited by patients' reading ability, idiosyncratic use, inconsistent metric properties across scale dimensions, reactivity, and differences between clinicians' and patients' conceptualization of symptoms. Clinician interviews enable more consistent use, but are time-intensive, difficult to standardize across settings, inherently subjective, and susceptible to reactivity effects, rater drift, and bias. Neither self-report nor clinical interview have the granularity necessary to measure moment-to-moment response to intervention or enable brain-behavior quantification. To assess quantitative changes in treatment response, objective measures are needed.

Extant assessment methods fail to consider that internalizing disorders have marked observable influence on psychomotor functioning (e.g., agitation), expression of affect (reductions in positive affect and increases in negative), and interpersonal communication (lack of synchrony). Behavioral signal processing of audio and video recorded behavior has shown great potential to objectively measure symptoms of depression and to a lesser extent anxiety [2], [3], [4], [5], [6].

Further advances depend in part on three challenges. One is greater emphasis on severity rather than detection. While detection matters for screening purposes, to inform treatment and assess outcomes precise measurement of severity is what matters. For instance, percentage reduction in severity is a common measure of treatment response. Unless severity is measured, treatment response cannot be quantified. Two is attention to internalizing disorders beyond depression. Depression is only one of many internalizing disorders that are cause for significant distress and disability and often are inter-related or comorbid. In the following work, we focus on obsessive-compulsive disorder (OCD) with comorbid depression.

And three, previous work on computational approaches to clinical measurement is limited to measures at single points in time. In clinical treatment and research, what matters more is quantitative change (degree to which patients are getting better or worse) over the course of treatment. With exception of [7], when databases have included repeated assessments, investigators have treated interviews from the same persons as if they were independent [4]. Failure to model the correlation of observations within persons ignores individual differences that can present serious confounds

* *Denotes equal contribution*
# *Corresponding author*

- *Saurabh Hinduja[1], Ali Darzi[2] and Jeffrey F Cohn[3] are with the Department of Psychology at the University of Pittsburgh, PA, 15213. Email:{[1]sah273, [2]ald260,[3]jeffcohn}@pitt.edu*
- *Itir Onal Ertugrul is with the Department of Information and Computing Sciences, Utrecht University, The Netherlands. Email: i.onalertugrul@uu.nl*
- *Nicole Provenza[4], Ron Gadot[5] and Sameer A Sheth[6] are with the Department of Neurosurgery at Baylor College of Medicine, TX, 77090. Email:{[4]nprovenz, [5]ron.gadot, [6]sasheth}@bcm.edu*
- *Eric A Storch[7] and Wayne K Goodman[8] are with the Menninger Department of Psychiatry and Behavioral Science at Baylor College of Medicine, TX, 77090. Email:{[7]storch, [8]wayne.goodman }@bcm.edu*

Manuscript received April 19, 2005; revised August 26, 2015.

if not taken into account. When serial observations within subjects are combined, trends may disappear or even reverse; an effect known as Simpson's paradox [8].

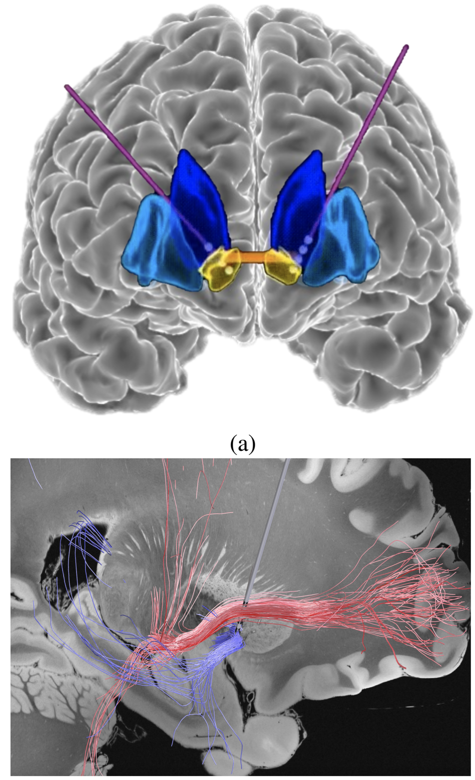
We used mixed-effects multimodal random forest regression to objectively measure change in severity within patients over the course of their treatment for chronic, severe, obsessive-compulsive disorder. Obsessive-compulsive disorder (OCD) is a persistent, oftentimes disabling condition that is characterized by obsessive thoughts and compulsive behavior. Obsessions are repetitive and intrusive thoughts (e.g., contamination), images (violent scenes), or urges (e.g., to stab someone) that can be highly disturbing. Individuals with OCD attempt to ignore or suppress obsessions or to neutralize them with other thoughts or actions (American Psychiatric Association, 2015). Compulsions are repetitive behaviors that an individual feels driven to perform in effort to reduce or avoid obsessions. Obsessions and compulsions are time-consuming (many hours per day), result in clinically significant impairment, and often are comorbid with depression, especially in more severe cases [9]. Both disorders entail high levels of negative affectivity, and related brain networks have been associated with each [10], [11].

Participants met research criteria for treatment-resistant, or refractory, OCD. Treatment-resistant OCD is defined as repeated failure to respond to front- or second-line treatments. Frontline treatments for OCD are exposure and response prevention (ERP), a cognitive-behavior therapy, and serotonin reuptake inhibitors with or without clomipramine, a tricyclic antidepressant [12], [13]. Second-line treatments may include anti-psychotics [14]. About 25% of patients with OCD fail to respond to front- or second-line treatments or have difficulty with adherence or tolerance, respectively, and are considered treatment-resistant.

Participants were treated with deep brain stimulation (DBS) of or close to the ventral capsule/ventral striatum (VC/VS). The VC/VS is in a subcortical circuit involved in error detection, habit formation, and motivational processes [15], [16]. In studies by our group and others, DBS using implanted electrodes targeting nodes of this circuitry (Fig. 1 and 2) has proven highly effective in relieving treatment-resistant OCD. The most comprehensive and up to date review of DBS outcomes found that 66% of patients fully responded to treatment [17]. DBS also proved effective in treating comorbid depression; 50% of patients fully recovered from comorbid depression and another 16% partially recovered.

We measured change in severity of OCD and comorbid depression over the course of an 18-month clinical trial for treatment-resistant OCD. We tested the hypothesis that an unobtrusive AI-based system deployed in open-ended interviews can effectively yield biomarkers of OCD and comorbid depression severity as well as total electrical energy delivered (TEED) by the DBS electrodes. Participants undergoing DBS treatment for refractory OCD were recorded in open-ended interviews at regular intervals over the course of the trial. Modalities included facial expression, eye movement, head pose, voice acoustics and timing, and linguistic measures of speech. Because each participant was seen on a variable number of occasions, we used a mixed-effects random forest regression with feature reduction and cross-validation to control for individual differences and overfitting. We seek to objectively measure response to treatment.

We first briefly review multimodal measures of affect related to internalizing disorders and novelties of the research and the research questions



(a)

(b)

Fig. 1: (a) Frontal view of an OCD patient's brain. Implanted DBS leads and their electrodes are shown with purple lines and white circles, respectively. The ventral striatum (target area) is in yellow. (b) Saggital view of the DBS electrodes in relation to the cortico-striatal-thalamo-cortical circuit that is implicated in OCD.

1.1 Multimodal measures of affect

Extensive evidence in psychology and affective computing supports the view that affective communication is multimodal [2], [18], [19], [20], [21]. We briefly review literature relevant to both unimodal and multimodal communication of emotion and internalizing disorders such as OCD and depression.

Visual features: The Facial Action Coding System (FACS) affords description of nearly all-possible visually discernible facial movement [22]. Movements for which the anatomic basis is known are referred to as Action units (AUs). Examples of AUs include AU 1 (medial strand of the frontalis, which raises the inner brow), AU 2 (lateral frontalis, which raises the outer brow), AU 6 (orbicularis oculi, which raises the cheeks, narrows the eye aperture and may cause "crows-feet" wrinkles at the lateral eye corners), and AU 12 (zygomatic major, which pulls the lip corners obliquely in smiling). While not without controversy, strong evidence suggests that specific combinations of actions are strongly related to specific emotions and intentions [23], [24], [25], [26], [27]. Automatic detection of AU occurrence and intensity and continuous measurement of some action descriptors has become possible [28], [29], [30], [31]. Velocity of automatically detected action units and head motion has been strongly related to emotional distress, depression, mania, and autism spectrum disorder [4], [21], [32], [33], [34], [35], [36]. Preliminary evidence suggests that facial AUs and head dynamics may differentiate between different levels of DBS stimulation [37] and predict OCD severity [38].

Acoustic features: Affective states strongly influence voice production [39], [40]. Change in subglottal pressure, transglottal airflow, and vocal fold vibration can be seen in acoustic features of affective speech. Additional features that have proven informative include vocal fundamental frequency (intonation and rhythm) [41], energy (volume or intensity) [42], utterance duration [43], and intra- and inter-speaker pause duration [42], [44]. Due to the effectiveness of acoustic and temporal features, they are frequently used in mental health studies: Anxiety [45], Distress Assessment [46], and depression and suicide [3]. Hence, acoustic and related temporal features are good candidates to assess DBS treatment in OCD patients. Several packages are available to analyze voice acoustics and behavior. They include OpenSMILE [47], COVAREP [48], and GeMAP [49].

Linguistic features: Linguistic features reveal sentiment and interests [50]. Prior to analysis, pre-processing is typically required, which includes localization of speakers' audio, speech recognition [51], and speech-to-text conversion [52]. To calculate linguistic features, several natural language processing techniques and models can be used. Notable examples include BERT [53], RoBERTa [54], PALM [55], cTAKES [56], and LIWC [50]. The instances of well-known linguistic features are syntax parsing using dependency trees, Chomsky transformational grammars, and statistical methods (e.g., word counting) [57]. Language-based deficits are common symptoms of psychiatric disorders [58]. Linguistic features are frequently used to detect depression and suicidal ideation [59], [60], [61], [62], [63], addiction [64], [65], [66], anxiety [67], and bipolar disorder [68].

Multimodal features: In social interaction, affective states are expressed multimodally. Because modalities may carry different messages, attention to a single modality can result in ambiguous or misleading results. To increase precision and accuracy, multimodal fusion can be performed. Feature-level fusion (or early fusion) [69], decision-level fusion (or late fusion), and hybrid-level fusion all may be useful. In early fusion, all features across modalities are placed together; and all or subsets are used to train a desired model. In decision-level fusion, separate modality-specific models may be developed and then fused using majority voting. Multimodal affective analysis can vary in the combination of modalities used to detect affective states. Several studies investigated how different modalities may complement each other to increase the performance of an ensemble model. For instance, combinations of acoustic-visual [46], acoustic-linguistic, or all three [4], [59] may be used. Most multimodal affective computing methods, using either early- or late fusion, typically outperform unimodal models.

1.2 Machine learning for internalizing disorders

Machine learning has been increasingly used to detect depression [70], [71], [3], [72], [73]. Machine learning has been used less often to infer symptom severity [4], [7], [74]. Conventional machine learning approaches are based on designing and selecting hand-crafted features and training classifiers to detect disorders. Previous research has trained models including support vector machines (SVMs) [75], logistic regression [4], and decision trees [76] mostly with the aim of achieving high prediction performance. Deep learning approaches that automatically learn important features from the data often realize superior performance in detecting depressive [77] and manic episodes [78] compared to conventional approaches. However, a major drawback of deep learning based approaches is that large numbers of participants are required and features typically lack interpretability that is important for clinical science and treatment.

In clinical fields, a common goal is to develop a system that informs assessment, treatment, and mechanisms. To achieve a machine learning model with good performance in each of these areas, it is crucial to understand why a model has given a particular decision and which features are critical in evaluating the degree to which patients are improving or not. For that reason, recent works have revisited the use of hand-crafted features. They afford interpretable results and high predictive performance.

Shapley analysis has been especially informative in interpreting feature contributions to model performance [79]. Recent examples include mothers' depression in dyadic interactions with their adolescent offspring [21], mania prediction in bipolar disorder [59], and differentiation of apathy and depression in older adults [80].

As noted above, prior work has failed to consider repeated assessments over time of the same individuals. When repeated assessments have been available, they have been treated as if they were independent [4]. When longitudinal assessments are available, attention to within-subject correlation is important to control for individual differences. For observations nested within individuals, mixed-effects models are needed. In mixed-effects models, each individual has their own, unique slope and intercept. Mixed-effects models are well known in behavioral statistics [81], [82] as multilevel models, but less so in machine learning. When multilevel structure is ignored, statistical artifacts can emerge [83].

In recent work, mixed effect random forests (MERFs) have been used to predict depression severity from physiological measures in a longitudinal study [7]. With their ability to personalize model parameters, mixed-effects models improve performance compared with standard random forests. We extend mixed-effects random forests three ways. First is to include multimodal features; second is to predict OCD as well as depression severity; third is to predict total electrical energy delivered by the DBS electrodes (TEED).

1.3 Novelties and Research Questions

This paper extends our preliminary work [38] in several ways:

1) Predict OCD severity, comorbid depression severity, and total electrical energy delivered (TEED) by DBS from voice acoustics and timing, linguistic features, head and face dynamics, and facial action units; evaluate relative contributions of each set of features.
2) Train mixed-effects random forests (MERFs) that account for the nesting of observations within participants.
3) Use Shapley analysis to reduce the number of features, optimize prediction, and afford interpretable parameters.
4) Include additional participants and assessments.

To our knowledge, this paper presents the first use of multimodal affective computing to assess OCD severity or DBS stimulation and one of very few to consider longitudinal measures. In a clinical trial of 18 months duration with a single-subjects-with-replications design, each participant served as their own control. Participants are a highly select group with treatment-resistant

OCD that have been implanted with a deep brain stimulator and a connected battery pack implanted in the chest cavity. The repeated measures design addresses the longitudinal demands of clinical research and treatment. A multidisciplinary team of psychiatrists, neurosurgeons, clinical psychologists, neuro-scientists, and engineers are actively involved in all phases of the study. Given the nature of the research, program officials from the U.S. National Institutes of Health (NIH) and Food and Drug Administration (FDA) are closely involved as well.

DBS stimulation as measured from contact sensors in the VC/VS region was operationalized as total electrical energy delivered (TEED) as defined in section 2. Visual features were measured using AFAR (AUs, eye behavior, and head dynamics) [28]. Acoustic and vocal timing features were measured using OpenSMILE [47] and COVAREP [48]. Linguistics features were measured using LIWC [50].

OCD severity was operationalized using the Yale-Brown Obsessive-Compulsive Scale - Second Edition (YBOCS-II), which is the gold standard for assessing OCD severity [84], [85]. Depression severity was operationalized using the self-report Beck Depression Interview - Second Edition, BDI-II [86]. Total electrical energy delivered by the DBS electrodes is defined below.

We address three research questions.

RQ1- To what extent can severity of OCD, comorbid depression severity, and total electrical energy delivered (TEED) by the DBS electrodes be predicted using visual, acoustic, and linguistic modalities?

RQ2- Which features within and across modalities are most predictive of symptom severity and TEED?

RQ3- What is the relative predictability from multimodal features of OCD severity, depression severity, and total energy delivered?

2 Study Setup and Protocol

This study is from an ongoing clinical trial of DBS for treatment-resistant OCD. Inclusion criteria are: 1) Repeated failure to respond to evidence-based treatments (cognitive behavioral therapy and medication); and 2) severe symptoms as measured by a score greater than 27 on the Yale-Brown Obsession Compulsion Scale-I (YBOCS-I)) (scale of 0-40). The first two participants were implanted with the Medtronic Activa PC+S DBS device; the other four were implanted with the Medtronic Summit RC+S Percept. With one exception, the three men and three women have completed at least 15 months of the study (Table I). A brief description of the study protocol follows.

Participants underwent a 1-month pre-implantation baseline evaluation followed by implantation of bilateral DBS electrodes in or near the VC/VS. A second baseline was observed prior to initial activation and programming of the DBS device. Patients then were seen for in-person or virtual visits monthly for open-loop programming of the DBS to optize treatment. Each visit started with an open-ended interview with a clinician. The interviews were 3 to 8 minutes in duration and were followed by assessment of symptom severity using the YBOCS-II for OCD [87] and the BDI-II [86] for depression.

Interviews were recorded using a GoPro camera and high-resolution microphone positioned about 10 to 15 degrees of frontal view of the patient. A separate GoPro camera recorded the interviewer. On the same or following day, the stimulation parameters were titrated as needed in what is referred to as a

TABLE 1: Sessions available for analysis. Baselines 1 and 2 occurred before and after implantation of DBS electrodes, respectively, and prior to DBS activation.

Participant	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**
Baseline1	✓	✓	✓	✓	✓	✓
Baseline2	✓	✓	✓	✓	✓	✓
3rd Month	✓	✓	✓	✓	✓	✓
6th Month	✓	✓	✓	✓	✓	✓
9th Month	✓	✓	✓	✓	✓	NA
12th Month	✓	✓	✓	✓	✓	NA
15th month	✓	✓	✓	✓	✓	NA
18th month	NA	✓	✓	✓	NA	NA

programming session. At approximately six months from study start, patients received Exposure and Response Prevention (ERP) therapy, a form of Cognitive Behavior Therapy, for two months. Over the course of the trial, we analyzed pre- and post-baseline interviews and interviews approximately every 3 months (Table 1). To consider possible within-session differences, each session was divided into two halves.

Parameters relevant to DBS include: amplitude, pulse duration, and frequency. The total electrical energy delivered per second, or power, was computed using the formula [88]:

$$TEED(W * 1s) = I(A)^2.PW(sec).f(Hz).R(\Omega), \quad (1)$$

where power is expressed in Watts, current in Amperes, pulse width in seconds, frequency in Hertz, and resistance in Ohms. Throughout the clinical trial, the stimulation frequency was held constant at 150.6 Hz. Since the purpose of the study was to predict variations in total electrical energy delivered, the constant term was omitted. Because measurement may be affected by transient fluctuations in battery output, pulse shape, or resistance, actual delivered energy may differ slightly from calculated values.

3 Methods

Figure 2 depicts the analysis pipeline. Facial action units, head and face dynamics, and eye motion are extracted from video; acoustic and linguistic features are extracted from audio. Using individual sets of extracted features, mixed-effects random forests (MERFs) are trained to predict OCD severity, depression severity, and total electrical energy delivered (TEED). SHAP analysis is used to evaluate the most informative unimodal features. We then aggregate the top-*k* features from each set and train a multimodal MERF. Finally, with a SHAP analysis on multimodal features, we identify the most-informative *k* multimodal features and the corresponding model.

3.1 Feature extraction

We extract four sets of features namely action units, head and face dynamics, acoustic features, and linguistic features.

Action units: Faces in the video are tracked and normalized using the Zface module of AFAR [28]. ZFace [89] is a real-time face alignment software that accomplishes dense 3D registration from 2D videos and images without requiring person-specific training. Faces are normalized to have an interocular distance (iod) of 80 pixels. AU detector module of AFAR is used to detect facial action units (AUs) in the normalized faces.

The version of AFAR, used in this study, was trained on the EB+ dataset (an expanded version of BP4D+ [90]), in which

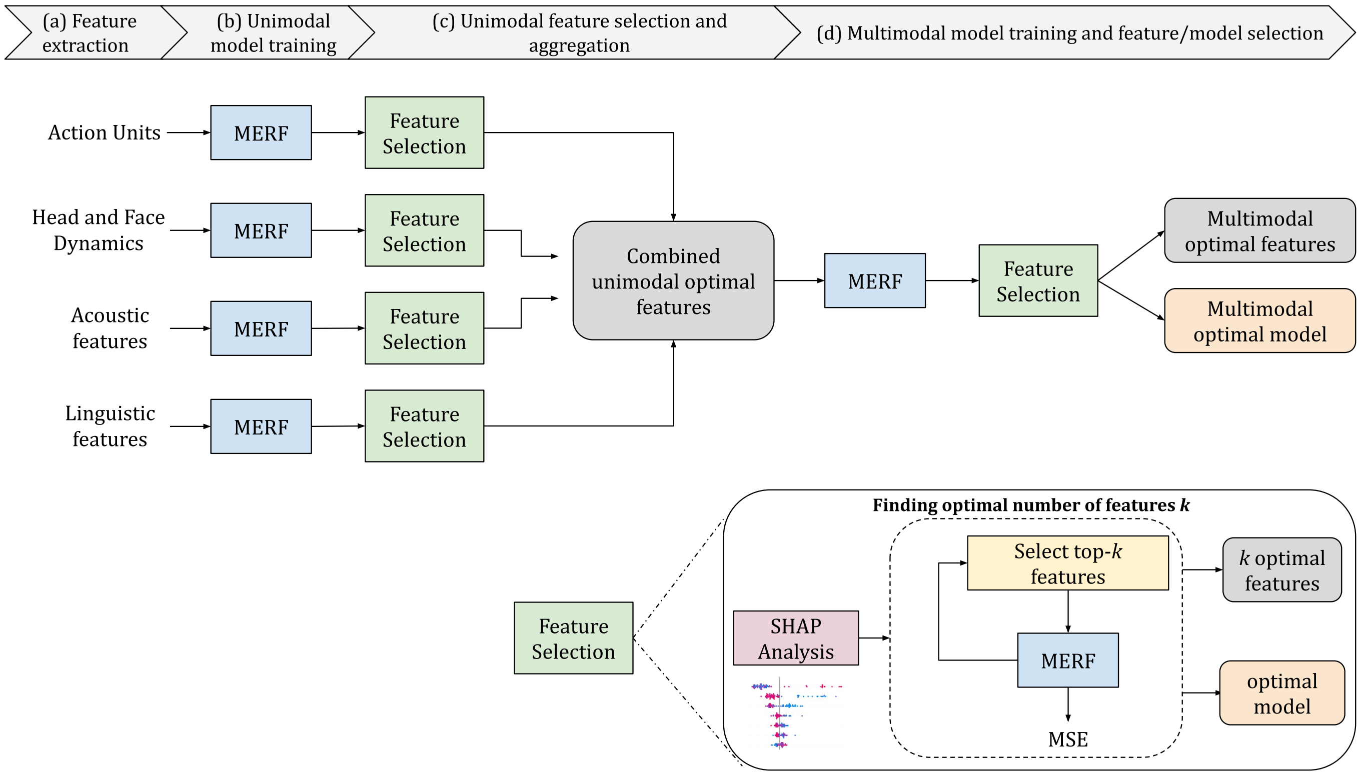


Fig. 2: Overview of the pipeline. (a) Action units, head and face dynamics, acoustic features, and linguistic features are extracted. (b) Mixed effects random forests (MERFs) are trained using each of the four feature sets separately. (c) The top-k features are selected from each of the four sets of features. Feature selection consists of SHAP analysis to identify important features, rank them based on their Shapley values, choose top-k optimal features for each feature set, and aggregate them into a single feature vector. (d) Mixed effects random forests (MERFs) are trained with a combined multimodal set of optimal features. (e) Multimodal optimal features and corresponding model are found with a multimodal feature selection step.

participants interact with an experimenter in a variety of emotion related tasks. Reliability of AFAR in EB+ was tested using k-fold cross validation. Average free-margin kappa was 0.75 and AUC 0.73 [29]. Cross-domain generalization was assessed by testing AFAR in Sayette GFT [91]. Average free-margin kappa was 0.49 and AUC 0.66, which represent moderate cross-domain generalizability. Because test results in GFT were likely attenuated by the larger head motion and lower video resolution in GFT, these comparisons provide a conservative estimate of the cross-domain generalizability in the current study. EB+ and the clinical trial were more alike than EB+ and GFT. EB+ and the clinical trial both used higher resolution video and were more similar in their more limited head motion.

AFAR was used to asses intensity of 6 facial AUs: AU1 (inner brow raiser), AU6 (cheek raiser), AU10 (upper lip raiser), AU12 (lip corner puller), AU14 (dimpler), AU17 (chin raiser). AU 1+2 is typically seen in surprise and affective engagement. An additional feature was average intensity of AU 6+12, which comprises the Duchenne smile, a marker of positive affect. For each of these, we extract time-series featues using tsfresh [92]. TsFresh outputs 794 time series characteristics for each feature for a total of 5,558 (7×794) features. In case of tracking failure or an AU feature fails to vary throughout the video, all related TsFresh features are set to 0. While the number of features is initially large, the analysis plan greatly reduces the number in a number of unimodal and multimodal steps.

Head and face dynamics: Head dynamics is defined using the time series of the 3 degrees of freedom of out-of-plane rigid head movement, which correspond to head nods (i.e., pitch), head turns (i.e., yaw), and lateral head inclinations (i.e., roll). Face dynamics is defined as time series of per frame eye and mouth openings. Eye opening is calculated using the Eye Aspect Ratio (EAR) [93], which is a normalized measure that divides the distance between landmarks on the upper and lower eyelids to the distance between inner and outer eye corners. Average of left and right EAR is used. Similarly, mouth opening is calculated using the Mouth Aspect Ratio (MAR), which is a normalized measure that divides the distance between landmarks on the upper and lower lips to the distance between left and right mouth corners. After head (pitch, yaw, and roll) and face (EAR, MAR) dynamics are calculated, time series characteristics are extracted using TsFresh [92], yielding a total of 3,970 features.

Acoustic Features: Audio for each speaker is localized and transcribed using TranscribeMe [94]. Audio and text are aligned using the Montreal-Forced-Aligner [95]. The openSMILE [96] toolkit and Collaborative Voice Analysis Repository (COVAREP) [48] are used to extract acoustic features. For openSMILE, we use the Geneva minimalistic acoustic parameter set (eGeMAPS [49]), which is a subset of audio features chosen for their ability to represent affective physiological changes in voice production.

eGeMAPS contains 62 features: arithmetic mean and coefficient of variation of 18 low-level descriptors (LLD), 8 functionals applied to loudness and pitch LLD, and 6 temporal features. COVAREP provides 72 low-level speech acoustic features, which are derived from the speech signal, that include pitch, energy, spectral envelope, loudness, voice quality and other characteristics.

Linguistic features: We use Linguistic Inquiry and Word Count (LIWC) [97], [98], which is a text analysis tool that determines the percentage of words in a text that fall into one or more linguistic, psychological, and topical categories. We extract 92 features from the verbal content of each interview. Approximately 93% of the words used in each interview were present in the LIWC dictionary and analyzed. We drop the coverage variable (referred to as "Dic" in LIWC) and normalize the word count variable with the duration of interaction.

3.2 Unimodal model training with mixed effects random forests (MERF)

When measures cluster within persons, as in a longitudinal study, mixed effects models are used in statistics and econometrics [99], [100]. In addition to the fixed effect terms, they include random effect parameters, which change the model's assumptions to accommodate heterogeneous data with many sources of random variability (e.g., both intra- and inter-individual). As a result, mixed effects methods allow for more accurate statistical inferences about the factors that connect with observed variance.

Motivated by two previous studies that used mixed effects models to infer depression severity in a longitudinal design [7], [38], we use MERF to infer OCD severity, related depression severity, and TEED.

MERF [100] is defined as:

$$Y_{ij} = f(X_{ij}) + b_i + \varepsilon_i \tag{2}$$

where: $i = 1, ..., m$ are the *clusters* (participants) each with n_i observations ($j = 1, ..., n_i$); X_{ij} is the input feature matrix; $f(X_{ij})$ is the *fixed effects random forest*; b_i is the *random effect* parameter; ε_{ij} is the measurement error; and Y_{ij} is the regression target variable. In our unimodal experiments, the fixed effect parameters are the features derived from individual modalities and the random effect parameter is the participant ID. The model is trained using expectation maximisation (EM) with a generalised log-likelihood (GLL) function to monitor convergence. For each cross- training / testing, the mixed effects random forests were trained for 50 iterations.

3.3 Unimodal feature selection

Shapley values [101] were introduced in game theory to gauge each player's participation in cooperative games. The machine learning and explainable AI communities recently have shown interest in Shapley. Shapley value for the j^{th} feature is defined as the weighted average of differences in predictions in the presence of the j^{th} feature and when it is marginalized, given the i^{th} data instance with m features represented by $X_i{}^m$. Marginalization is accomplished by leveraging predictions from several feature subsets. Calculating Shapley value is computationally expensive due to the laborious marginalization procedure. However, the SHAP (SHapley Additive exPlanations) framework can be used to estimate Shapley values [102]. Shapley value φ_j of feature j can be computed as:

$$\varphi_j(v) = \sum_{S \subseteq \{1,2,...,m\} \setminus \{j\}} \frac{|S|!\,(m - |S| - 1)!}{m!} (v(S \cup \{j\}) - v(S)),$$

where v is the model, m is the total number of features and S is a subset of features.

We use kernel-based LIME, which combines Shapley values with Local Interpretable Model-agnostic Explanations (LIME) [103]. LIME has been widely used to interpret model decisions in the explainable AI field. While LIME provides local correctness, the SHAP framework improves upon that by ensuring feature consistency and robustness to missing features. Missing features have no impact on the contribution of features of interest.

We use SHAP values to rank characteristics in terms of their relative contribution to prediction performance. We then choose the top-k features, where the optimal value of k is found iteratively based on the mean square error of the MERF trained with top-k features. Optimal k may differ for each set of features (e.g. action units, head and face dynamics, acoustic features, and linguistic features). Optimal features for all individual modalities are concatenated to obtain combined unimodal optimal features (see Figure 2c)

3.4 Multimodal model training and feature/model selection

Following unimodal modeling, we train MERF using multimodal features. Multimodal features comprise the combined top-k features selected from action units, head and face dynamics, acoustics, and linguistics. By training MERF with selected multimodal features, we seek to reveal the relative contribution of each modality to performance. We use combined features as the fixed effects and participant ID as the random effect parameter.

We use SHAP values to rank the multimodal features based on their relative contribution to the prediction performance. Similarly to unimodal feature selection, we choose top-k multimodal features and optimize the value of k using an iterative approach.

3.5 Model training and evaluation

We trained separate models to predict OCD severity, depression severity, and TEED. To validate our approach we performed leave-one-session-out cross-validation. For the multimodal model and each of the unimodal models, we optimized the number of features k in the set $k \in \{6, 11, 16, ... 46\}$.

To evaluate relative performance of the models we used the following performance metrics:

1) *Mean Absolute Error (MAE)* is one of the most commonly used performance metric for continuous labels. It is defined as the sum of the absolute errors divided by the number of observations.
$$MAE = \frac{\sum_{i=1}^{D} |x_i - y_i|}{D} \tag{3}$$
where D is the number of observations, x_i is the ground truth score, y_i is the predicted score.
2) *Root Mean Squared Error (RMSE)* is the root of the mean of the square of the errors. RMSE score can never be zero. It is a frequently used metric for continuous observations.
$$RMSE = \sqrt{\frac{\sum_{i=1}^{D} (x_i - y_i)^2}{D}} \tag{4}$$

3) *R Square (R^2)* is also known as the coefficient of determination. It is always in a range (0,1).

$$R^2 = 1 - \sqrt{\frac{\sum_{i=i}^{D}(x_i - y_i)}{\sum_{i=i}^{D}(x_i - \bar{y})}} \tag{5}$$

where $\bar{y}$ is mean ground truth score.

4) *Intraclass Correlation (ICC)* is commonly used to determine the correlation between raters. In our case, "raters" are represented by ground-truth and predicted scores. ICC may be computed for agreement or for consistency. We use consistency.
5) *Normalized mean absolute error* is the ratio of MAE to the range of measure (ROM), which is the difference between the possible maximum and minumum values of the measure. It is calculated as:

$$Norm\_MAE = \frac{MAE}{ROM} \tag{6}$$

As the ranges of the OCD, depression, and TEED measures differ, directly comparing MAEs obtained with each of them would not be meaningful. By dividing the MAEs by the range of each measure, we obtain a normalized measure that is comparable across all of them.

6) *Contribution* is used to find the importance of a modality in the prediction of the model. It is based on SHAP values. We define contribution for a particular modality as:

$$Contribution = \frac{\sum_{j=1}^{F}\sum_{i=1}^{D} SM_{ij}}{\sum_{k=1}^{M}\sum_{j=1}^{F}\sum_{i=1}^{D} S_{ijk}} \times 100 \tag{7}$$

where SM is SHAP value for a particular modality, S are all SHAP values, F is number of features in a modality, M is number of modalities. It is the ratio of sum of the all the SHAP values for a particular modality by the sum of all SHAP values across all modalities.

4 Results

In Section 4.1, we report MERF results for OCD severity, comorbid depression severity, and total energy delivered (i.e., YBOCS II, BDI II, and TEED, respectively). These include the test statistics (e.g., ICC) for each modality before and after feature selection by SHAP analysis. In Section 4.2, we present the most important SHAP identified features within each modality.

4.1 Prediction results

OCD severity: The left side of Table 2 shows performance for each of the unimodal models when all features are used. Among unimodal models trained using all features, the one trained with acoustic features gave the best performance on each of the performance metrics. ICC for the unimodal acoustic model was 0.76. ICCs for the other unimodal models ranged from 0.45 - 0.48.

The right-hand side of Table 2 shows the performance for each of the unimodal models after SHAP reduction; the multimodal model that includes all SHAP-reduced features for each modality (i.e., "Combined"); and the SHAP-reduced multimodal model (i.e., "Best").

SHAP reduction improved each of the unimodal models. The SHAP-reduced acoustic model was the best among them and required only six features.

By comparing the "Combined" and unimodal SHAP-reduced models, one can evaluate whether SHAP-reduced multimodal models improved performance relative to unimodal. For acoustics, the multimodal model afforded no advantage. The SHAP-reduced acoustics model was equal to or outperformed the multimodal model on all four performance metrics. For the other SHAP-reduced unimodal models, the differences between unimodal and multimodal were mixed.

By comparing "Combined" and "Best" one can evaluate whether further SHAP reduction is valuable. The best SHAP-reduced multimodal model achieved the highest performance with an ICC of 0.83 and a large reduction in the number of features. SHAP-reduction in the multimodal model optimized prediction of OCD severity.

Depression severity: Table 3 shows the corresponding results for depression (BDI II). As for OCD, the unimodal model for acoustics again was the best performing unimodal model with an ICC of .80. ICCs for the other unimodal models were much lower and similar trends were found for the other test metrics.

SHAP reduction again strongly improved performance for each of the unimodal models. Among the SHAP-reduced unimodal models, the model for acoustics again outperformed the other unimodal models. The number of features required in the SHAP-reduced model for acoustics, however, was larger than that for OCD prediction (Table 2).

As in Table 2, by comparing "Combined" and unimodal SHAP-reduced models, one can evaluate whether SHAP-reduced multimodal models improved performance relative to unimodal. In contrast to the results for OCD, the SHAP-reduced multimodal model for depression outperformed that for acoustics and other modalities.

Comparing "Combined" and "Best", we again see that further SHAP reduction improves perfomance on all metrics. The relative improvement is greater for MAE and RMSE; for R sq and ICC, the gain is minimal.

Total Electrical Energy Delivered (TEED): Table 4 shows the corresponding performance for TEED. Similar to what was found for OCD and depression, prediction of TEED was highest for the acoustics model. The relative advantage of acoustics, however, was smaller than that found for OCD and depression.

Relative to the unimodal SHAP-reduced model for acoustics, the SHAP-reduced multimodal model resulted in only small improvement in two of four performance metrics. The SHAP-reduced multimodal model ("Best") optimized prediction of TEED. The SHAP-reduced multimodal model was consistently best.

Performance comparison across labels: To further compare performance across OCD, depression, and TEED, we compute the normalized MAE values for each individual set of features and their combination as shown in Figure 3. For MAE, lower scores are better. For individual modalities, acoustic features yielded the lowest MAE for OCD, depression, and TEED. Linguistic features yielded the highest. Multimodal models with SHAP reduction yielded the smallest MAEs for depression and TEED although not OCD.

Prediction of OCD severity and depression severity (Figure 3a and Figure 3b) are similar for all feature types except for acoustic features, which perform better. TEED prediction performances given in Figure 3c are much lower compared to other two for all feature sets.

TABLE 2: Prediction of OCD servity using unimodal and multimodal features. OCD was measured using the YBOCS-II

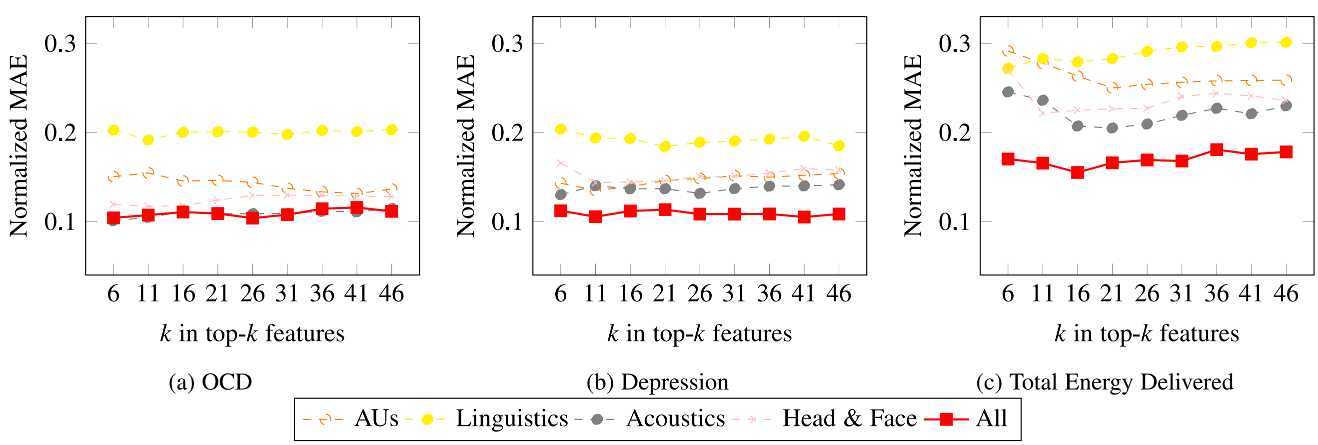
	All Features				SHAP Reduced Features					
	Acoustics	Linguistics	AUs	Head & Face	Acoustics	Linguistics	AUs	Head & Face	Combined	Best
Number of Features					6	11	36	16	69	6
MAE	5.93	8.67	8.67	8.46	5.05	7.85	6.7	5.86	7.65	5.2
RMSE	7.29	10.26	10.29	10.25	6.22	9.58	8.1	7.67	9.82	6.47
R sq	0.61	0.2	0.22	0.23	0.72	0.32	0.5	0.57	0.67	0.7
ICC	0.76	0.45	0.46	0.48	0.84	0.57	0.7	0.74	0.81	0.83

TABLE 3: Prediction of comorbid depression severity using unimodal and multimodal features. Comorbid depression was measured using the BDI-II

	All Features				SHAP Reduced Features					
	Acoustics	Linguistics	AUs	Head & Face	Acoustics	Linguistics	AUs	Head & Face	Combined	Best
Number of Features					25	20	10	15	81	11
MAE	8.14	8.83	11.65	11.83	7.9	11.05	8.1	8.67	6.55	6.30
RMSE	10.29	10.40	13.88	14.05	10.29	11.05	10.0	10.76	8.59	8.28
R sq	0.64	0.18	0.33	0.32	0.64	0.39	0.60	0.60	0.75	0.76
ICC	0.80	0.44	0.57	0.56	0.80	0.62	0.80	0.75	0.86	0.87

TABLE 4: Prediction of Total Electrical Energy Delivered (TEED) using unimodal and multimodal features

	All Features				SHAP Reduced Features					
	Acoustics	Linguistics	AUs	Head & Face	Acoustics	Linguistics	AUs	Head & Face	Combined	Best
Number of Features					20	5	20	10	55	16
MAE	3.22	3.71	4.00	3.66	2.67	3.34	3.05	2.78	2.70	2.33
RMSE	3.88	4.67	4.7	4.74	3.28	4.35	4.00	3.54	3.30	2.79
R sq	0.34	0.02	0.00	0.02	0.53	0.24	0.36	0.45	0.60	0.69
ICC	0.31	0.27	0.30	0.24	0.71	0.46	0.54	0.62	0.70	0.81



(a) OCD (b) Depression (c) Total Energy Delivered

Fig. 3: Cross-validation Normalized MAE performance of top-k features derived from SHAP analysis

4.2 Relative contribution of features within modalities for the best-performing model

For OCD severity, depression severity, and OCD, Table 5 presents the top-k multimodal features in predicting their respective values. Red denotes that an increase in the value of the feature leads to an increase in the predicted value. Blue indicates that an increase in the value of the feature leads to a decrease in the predicted value.

For OCD and depression severity, the majority of the top-k multimodal features are acoustic. Of these, Mel Frequency Cepstrum Coefficients from the lower registers were especially informative. This finding is consistent with previous research that suggests this set of features is strong predictor of depression [104]. In addition to MFCC, harmonic and phase distortion coefficients were negatively correlated with depression. Linguistic features were related to TEED only. Head and face dynamics were related to depression and TEED but not OCD. Facial action units contributed to OCD and TEED.

4.3 Individual differences among participants

To visualize individual differences, we plotted MERF predicted values by ground truth for OCD, depression, and TEED (Figure 4a, Figure 4b, and Figure 4c). With exception of S6, the slopes

TABLE 5: Rank ordering of the top-k features across all modalities in predicting symptom severity of OCD, symptom severity of depression, and total electrical energy delivered (TEED) by the DBS electrodes. *Note: The color indicates the sign of the correlation; red for positive correlation and blue for negative correlation.*

Modality	Feature	OCD	Depression	TEED
Acoustic	MFCC 4	1	4	2
	Loudness	2	-	-
	HMPDM 13	3	1	-
	MFCC 6	5	8	-
	MFCC 8	6	3	-
	MFCC 18	-	2	1
	HMPDM 8	-	6	-
	MFCC 24	-	7	5
	MFCC 13	-	11	-
	NAQ	-	-	16
	HMPDM 10	-	-	4
Linguistic	Words with greater than six letters	-	-	6
	Work	-	-	8
	Tentative	-	-	11
	Auxiliary verbs	-	-	15
Head and Face Dynamics	Yaw spectral Welch density	-	-	9
	Mouth approximate entropy	-	-	10
	Eyes spectral Welch density	-	-	12
	Eyes permutation entropy	-	5	-
	Eye change in quantile of mean	-	10	-
	Eye velocity of change in opening	-	9	-
	Mouth longest strike below mean	-	-	3
	Yaw partial auto correlation	-	-	13
	Yaw permutation entropy	-	-	7
Action Units	AU1 Benford correlation	4	-	-
	AU1 augmented dickey fuller	-	-	14

for OCD were closely related in intercepts and slopes across participants. For S6, a factor may have the smaller number of observations and attenuated variability of OCD for them.

For depression and especially for TEED, there was more variability across participants in slopes and intercepts. Attenuated variability may have contributed to this finding especially in S6. These findings support the importance of mixed-effects modeling for longitudinal machine learning.

Figure 4c shows the regression plots for TEED. Slopes and intercepts are more variable than for the symptom scores.

5 Discussion

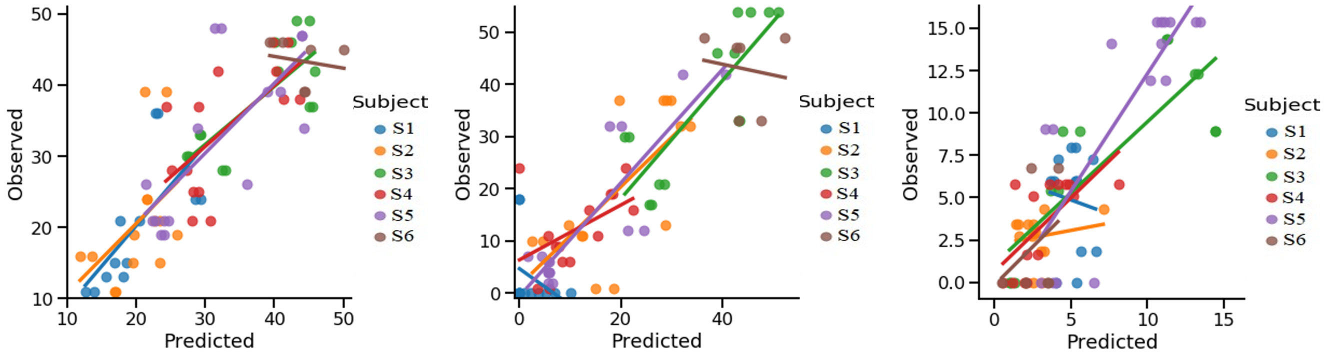
An unobtrusive, multimodal mixed-effects regressor based on open-ended interviews measured severity of OCD and comorbid depression over the course of the clinical trial with good reliability. For OCD, the ICC for the SHAP-reduced ICC model was 0.83; for depression, the corresponding ICC was 0.87. These ICCs rival the interrater reliabilities of trained clinicians. If supported by further research, these findings suggest that OCD and comorbid depression severity could be obtained without the need for formal clinical interviews. Routine use of this approach could reduce patient burden and clinical costs in treatment and clinical trial settings and eliminate error due to judgement differences between raters and drift in criteria over time. Consistency within and between treatment settings could be achieved by multimodal assessment of spontaneous patient behavior.

In the neuroscience community, there is increasing interest in brain-behavior quantification and synchronization [105]. That is, how changes in neural activity relate to synchronous changes in behavior. We found strong correlation between total energy delivered by the DBS electrodes in or near the VC/VS and multimodal behavior in the open-ended interviews. ICC for the SHAP-reduced model was 0.81. This ICC is for summary measures of behavior and TEED over several minutes duration. Further work will be needed to discover whether moment-to-moment changes in multimodal behavior reveal activity in the VC/VS on same time scale.

The current state of the art in DBS for treatment refractory OCD is open-loop programming. That is, patients return to the clinical setting at frequent intervals to evaluate recent symptoms and adjust DBS parameters as needed. In these sessions, clinical interviews and observations inform determination of DBS parameters through trial and error. Judgments are subjective and vary within and between clinicians and over time. A multimodal regressor could greatly reduce or eliminate subjective judgment and enable more accurate titration of the DBS.

Because exposure to OCD triggers and the severe symptoms they elicit can vary within and across days, more frequent evaluation and adjustment of DBS parameters than is possible in periodic office visits would be beneficial. In DBS-treatment for essential tremor, effective closed-loop programming has been achieved. The same is a current goal of research in DBS treatment for refractory OCD. The current findings suggest that multimodal behavior acquired via audio and video could be an effective component of a self-titrating DBS system. It also could be effective in detecting hypomania or mania, which are side effects of DBS, and automatically down-regulating DBS energy to reduce or eliminate this unwanted and potentially dangerous side effect.

An unexpected discovery was made when voice acoustics alone approached the accuracy of the best multimodal model. This finding highlights the potential of voice in effectively revealing affective states, particularly those associated with OCD and de-



(a) OCD (b) Depression (c) TEED

Fig. 4: Observed and MERF Predicted values

pression. The inherent dynamism of the voice, coupled with its connection to the vagus nerve—the longest nerve in the autonomic nervous system and the primary nerve of the parasympathetic nervous system—renders it well-suited to capturing variations in arousal and stress levels. Notably, voice has been recognized as a reliable predictor of depression. Similarly, we observed comparable results for OCD and the total electrical energy delivered to the DBS electrodes. Nevertheless, multimodal models that include facial action units, face and head dynamics, and linguistic features maximized prediction. As well, they may afford robustness to signal loss from any one modality.

In behavioral and clinical science, explanation has historically been more important than prediction. Especially when dealing with health, explanation is critical. Recent work has called for an integration of explanation and prediction [106]. Our models were informative in detecting relative contribution of each modality and of key features within each modality.

Shapley analysis contributed to enabling more powerful models. By using SHAP-reduction, the number of features could be reduced while eliminating those may have contributed error. SHAP analysis is potentially a powerful tool for optimizing model performance.

In machine learning we typically assume that labels or outcomes are independent. Most often, this assumption is warranted (i.e., each participant provides only a single label) and thus the assumption of independence is not violated. In a clinical trial, on the other hand, participants are assessed longitudinally and labels are clustered within participants. In our study, labels came from as many as 8 longitudinal assessments. To ignore the dependency of labels within participants would have seriously violated the assumption of independence and risked Simpson's paradox. For this reason, mixed-effects models were used. Because mixed-effects models are subject specific, however, they are unable to guide prediction for unseen participants. To the extent that individual differences are limited to intercepts, however, it may be possible to make valid predictions to new participants from relatively few data points. In the case of chronic, severe, treatment-resistant disorders, it is likely that intercepts would be reliably estimated from baseline assessments. It is likely, as well, that predictions would become more accurate as the number of observations increases. This would be particularly advantageous for a closed-loop DBS. For now, these hypotheses are open research question.

A limitation of the study was the small number of participants. The participant pool from which to draw was small. Participants had to meet stringent criteria for severe and chronic treat-resistant DBS, additional psychiatric and medical criteria, opt for surgical implantation of electrodes deep in their brain, and participate in an 18-month trial. The within/participants (longitudinal) design with up to 8 assessments from each participant provided some offset to the limited numbers of participants. Supporting the validity of the findings was the consistency of the findings for depression. Comparable to previous research that had access to larger samples of participants, our findings for depression were quite consistent. For OCD and total energy delivered to the DBS, comparative data are unavailable. OCD and DBS energy are new research topics in affective computing.

To protect participant confidentiality, the audio-video data used in this study cannot be shared with other investigators. We will seek from IRB permission to distribute deidentified and anonymized features for use by other researchers.

6 Conclusion

An unobtrusive, multimodal mixed-effects regressor based on open-ended interviews measured severity of OCD, severity of comorbid depression, and TEED over the course of a clinical trial for treatment resistant OCD. The regressor achieved strong consistency with state-of-the art clinical measures. With further validation, the proposed system could greatly reduce subjective variation in clinical judgment within- and between clinicians and eliminate drift over time in assessments for refractory DBS. An unexpected finding was the strength of acoustic features in inferring symptom severity and TEED. Because lower vocal track parameters may be recorded from contact sensors on the throat, they may be especially advantageous in a closed-loop DBS system. Facial action units and head and face dynamics contributed further predictive power. Linguistic features contributed relatively little. A key contributor to the modeling results was use of SHAP reduction in selecting most informative features and use of mixed-effects modeling. Most prior work has considered only single assessments or has ignored clustering of longitudinal assessments within participants. Mixed-effects models enabled predictions robust to individual differences in participants.

Acknowledgments

This research was supported in part by the U.S. National Institutes of Health through NINDS BRAIN Initiative award UH3

NS100549 and U.S. National Institute of Mental Health award MH096951 and by the McNair Foundation. DBS devices were donated by Medtronic as part of the BRAIN Initiative Public-Private Partnership Program.

REFERENCES

[1] American Psychiatric Association, *DSM-5*, Washington, DC, 2015.

[2] S. Alghowinem, T. Gedeon, R. Goecke, J. Cohn, and G. Parker, "Depression detection model interpretation via feature selection methods," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 133–151, 2023.

[3] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, "A review of depression and suicide risk assessment using speech analysis," *Speech Communication*, vol. 71, pp. 10–49, 7 2015.

[4] H. Dibeklioglu, Z. Hammal, and J. F. Cohn, "Dynamic multimodal measurement of depression severity using deep autoencoding," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 2, pp. 525–536, 2018.

[5] M. Fang, S. Peng, Y. Liang, C.-C. Hung, and S. Liu, "A multimodal fusion model with multi-level attention mechanism for depression detection," *Biomedical Signal Processing and Control*, vol. 82, p. 104561, 2023.

[6] S. Scherer, G. Stratou, M. Mahmoud, J. Boberg, J. Gratch, A. S. Rizzo, and L.-P. Morency, "Automatic behavior descriptors for psychological disorder analysis," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 1–8, 2013.

[7] R. A. Lewis, A. Ghandeharioun, S. Fedor, P. Pedrelli, R. Picard, and D. Mischoulon, "Mixed effects random forests for personalised predictions of clinical depression severity," *Computational Approaches to Mental Health Workshop*, 2021.

[8] E. H. Simpson, "The interpretation of interaction in contingency tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 13, no. 2, pp. 238–241, 1951.

[9] L. C. Quarantini, A. R. Torres, A. S. Sampaio, V. Fossaluza, M. A. de Mathis, M. C. Do Rosário, L. F. Fontenelle, Y. A. Ferrão, A. V. Cordioli, K. Petribu *et al.*, "Comorbid major depression in obsessive-compulsive disorder patients," *Comprehensive psychiatry*, vol. 52, no. 4, pp. 386–393, 2011.

[10] J. J. Taylor, C. Lin, D. Talmasov, M. A. Ferguson, F. L. Schaper, J. Jiang, M. Goodkind, J. Grafman, A. Etkin, S. H. Siddiqi *et al.*, "A transdiagnostic network for psychiatric illness derived from atrophy and lesions," *Nature Human Behaviour*, pp. 1–10, 2023.

[11] S. K. Peters, K. Dunlop, and J. Downar, "Cortico-striatal-thalamic loop circuits of the salience network: A central pathway in psychiatric disease and treatment," *Frontiers in Systems Neuroscience*, vol. 10, 2016. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnsys.2016.00104

[12] R. J. Romanelli, F. M. Wu, R. Gamba, R. Mojtabai, and J. B. Segal, "Behavioral therapy and serotonin reu Behavioral therapy and serotonin reuptake inhibitor pharmacotherapy in the treatment of obsessive-compulsive disorder: a systematic review and meta-analysis of head-to-head randomized controlled trials," *Depression and Anxiety*, vol. 31, no. 8, pp. 641–652, 8 2014.

[13] L.-G. Öst, A. Havnen, B. Hansen, and G. Kvale, "Cognitive behavioral treatments of obsessive–compulsive disorder. A systematic review and meta-analysis of studies published 1993–2014," *Clinical Psychology Review*, vol. 40, pp. 156–169, 8 2015.

[14] Y. C. Janardhan Reddy, A. S. Sundar, J. C. Narayanaswamy, and S. B. Math, "Clinical practice guidelines for obsessive-compulsive disorder," *Indian J Psychiatry*, vol. 59, no. Suppl 1, pp. S74–s90, 2017.

[15] P. J. Karas, S. Lee, J. Jimenez-Shahed, W. K. Goodman, A. Viswanathan, and S. A. Sheth, "Deep brain stimulation for obsessive compulsive disorder: evolution of surgical stimulation target parallels changing model of dysfunctional brain circuits," *Frontiers in neuroscience*, p. 998, 2019.

[16] A. Rădulescu, J. Herron, C. Kennedy, and A. Scimemi, "Global and local excitation and inhibition shape the dynamics of the cortico-striatal-thalamo-cortical pathway," *Scientific Reports*, vol. 7, no. 1, p. 7608, 2017.

[17] R. Gadot, R. Najera, S. Hirani, A. Anand, E. Storch, W. K. Goodman, B. Shofty, and S. A. Sheth, "Efficacy of deep brain stimulation for treatment-resistant obsessive-compulsive disorder: systematic review and meta-analysis," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 93, no. 11, pp. 1166–1173, 2022.

[18] B. Beebe and L. J. Gerstman, "The "packaging" of maternal stimulation in relation to infant facial-visual engagement: A case study at four months," *Merrill-Palmer Quarterly of Behavior and Development*, vol. 26, no. 4, pp. 321–339, 1980.

[19] M. V. McCall, P. Riva-Possea, S. J. Garlow, H. S. Mayberg, and A. L. Crowell, "Analyzing non-verbal behavior throughout recovery in a sample of depressed patients receiving deep brain stimulation," *Neurology, Psychiatry, and Brain Research*, vol. 37, pp. 33–40, 2020.

[20] R. O. Cotes, M. Boazak, E. Griner, Z. Jiang, B. Kim, W. Bremer, S. Seyedi, A. B. Rad, and G. D. Clifford, "Multimodal assessment of schizophrenia and depression utilizing video, acoustic, locomotor, electroencephalographic, and heart rate technology: protocol for an observational study," *JMIR Research Protocols*, vol. 11 —, no. 7, p. e36417, 2022.

[21] M. Bilalpur, S. Hinduja, L. A. Cariola, L. B. Sheeber, N. Alien, L. A. Jeni, L.-P. Morency, and J. F. Cohn, "Multimodal feature selection for detecting mothers' depression in dyadic interactions with their adolescent offspring," in *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*. IEEE, 2023, pp. 1–8.

[22] P. Ekman, W. V. Friesen, and J. C. Hager, "Facial action coding system: Research Nexus," in *Network Research Information, Salt Lake City, UT.*, 2002.

[23] A. S. Cowen and D. Keltner, "Universal facial expressions uncovered in art of the ancient americas: A computational approach," *Science advances*, vol. 6, no. 34, p. eabb1005, 2020.

[24] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychological science in the public interest*, vol. 20, no. 1, pp. 1–68, 2019.

[25] D. Keltner, D. Sauter, J. Tracy, and A. Cowen, "Emotional expression: Advances in basic emotion theory," *Journal of nonverbal behavior*, vol. 43, pp. 133–160, 2019.

[26] D. T. Cordaro, R. Sun, D. Keltner, S. Kamble, N. Huddar, and G. McNeil, "Universals and cultural variations in 22 emotional expressions across five cultures." *Emotion*, vol. 18, no. 1, p. 75, 2018.

[27] R. E. Mattson, R. D. Rogge, M. D. Johnson, E. K. Davidson, and F. D. Fincham, "The positive and negative semantic dimensions of relationship satisfaction," *Personal Relationships*, vol. 20, no. 2, pp. 328–355, 2013.

[28] I. O. Ertugrul, L. A. Jeni, W. Ding, and J. F. Cohn, "AFAR: A deep learning based tool for automated facial affect recognition," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 5 2019, pp. 1–1.

[29] I. O. Ertugrul, J. F. Cohn, L. A. Jeni, Z. Zhang, L. Yin, and Q. Ji, "Crossing domains for AU coding: perspectives, approaches, and measures," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2020.

[30] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. IEEE, 2018, pp. 59–66.

[31] N. I. Technology, "Facereader v6.1," Report, 2015.

[32] Z. Hammal, J. F. Cohn, and D. T. George, "Interpersonal coordination of headmotion in distressed couples," *IEEE transactions on affective computing*, vol. 5, no. 2, pp. 155–167, 2014.

[33] Z. Hammal, J. F. Cohn, C. Heike, and M. L. Speltz, "Automatic measurement of head and facial movement for analysis and detection of infants' positive and negative affect," *Frontiers in ICT*, vol. 2, 12 2015.

[34] M. Gavrilescu and N. Vizireanu, "Predicting depression, anxiety, and stress levels from videos using the facial action coding system," *Sensors*, vol. 19, no. 17, p. 3693, 8 2019.

[35] T.-H. Yang, C.-H. Wu, M.-H. Su, and C.-C. Chang, "Detection of mood disorder using modulation spectrum of facial action unit profiles," in *2016 International Conference on Orange Technologies (ICOT)*. IEEE, 12 2016, pp. 5–8.

[36] K. B. Martin, Z. Hammal, G. Ren, J. F. Cohn, J. Cassell, M. Ogihara, J. C. Britton, A. Gutierrez, and D. S. Messinger, "Objective measurement of head movement differences in children with and without autism spectrum disorder," *Molecular Autism*, vol. 9, no. 1, p. 14, 12 2018.

[37] Y. Ding, I. Onal Ertugrul, A. Darzi, N. Provenza, L. A. Jeni, D. Borton, W. Goodman, and J. Cohn, "Automated detection of optimal DBS Ddevice settings," in *Companion Publication of the 2020 International Conference on Multimodal Interaction*. New York, NY, USA: ACM, 10 2020, pp. 354–356.

[38] A. Darzi, N. R. Provenza, L. A. Jeni, D. A. Borton, S. A. Sheth, W. K. Goodman, and J. F. Cohn, "Facial action units and head dynamics in longitudinal interviews reveal ocd and depression severity and dbs

energy," in *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*. IEEE, 2021, pp. 1–6.
[39] J. Sundberg, S. Patel, E. Bjorkner, and K. R. Scherer, "Interdependencies among voice source parameters in emotional speech," *IEEE Transactions on Affective Computing*, vol. 2, no. 3, pp. 162–174, 2011.
[40] D. T. Cordaro, D. Keltner, S. Tshering, D. Wangchuk, and L. M. Flynn, "The voice conveys emotion in ten globalized cultures and one remote village in bhutan." *Emotion*, vol. 16, no. 1, p. 117, 2016.
[41] C. Busso, S. Lee, and S. Narayanan, "Analysis of emotionally salient aspects of fundamental frequency for emotion detection," *IEEE transactions on audio, speech, and language processing*, vol. 17, no. 4, pp. 582–596, 2009.
[42] M. Alpert, E. R. Pouget, and R. R. Silva, "Reflections of depression in acoustic measures of the patient's speech," *Journal of affective disorders*, vol. 66, no. 1, pp. 59–69, 2001.
[43] J. C. Mundt, A. P. Vogel, D. E. Feltner, and W. R. Lenderking, "Vocal acoustic biomarkers of depression severity and treatment response," *Biological psychiatry*, vol. 72, no. 7, pp. 580–587, 2012.
[44] Y. Yang, C. Fairbairn, and J. F. Cohn, "Detecting depression severity from vocal prosody," *IEEE transactions on affective computing*, vol. 4, no. 2, pp. 142–150, 2012.
[45] T. Özseven, M. Düğenci, A. Doruk, and H. I. Kahraman, "Voice traces of anxiety: acoustic parameters affected by anxiety disorder," *Archives of Acoustics*, pp. 625–636, 2018.
[46] S. Scherer, G. Stratou, and L.-P. Morency, "Audiovisual behavior descriptors for depression assessment," in *Proceedings of the 15th ACM on International conference on multimodal interaction*, 2013, pp. 135–140.
[47] F. Eyben, M. Wöllmer, and B. Schuller, "Openear—introducing the munich open-source emotion and affect recognition toolkit," in *2009 3rd international conference on affective computing and intelligent interaction and workshops*. IEEE, 2009, pp. 1–6.
[48] G. Degottex, J. Kane, T. Drugman, T. Raitio, and S. Scherer, "Covarep—a collaborative voice analysis repository for speech technologies," in *2014 ieee international conference on acoustics, speech and signal processing (icassp)*. IEEE, 2014, pp. 960–964.
[49] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan *et al.*, "The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing," *IEEE transactions on affective computing*, vol. 7, no. 2, pp. 190–202, 2015.
[50] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates*, vol. 71, no. 2001, p. 2001, 2001.
[51] M. Malik, M. K. Malik, K. Mehmood, and I. Makhdoom, "Automatic speech recognition: a survey," *Multimedia Tools and Applications*, vol. 80, no. 6, pp. 9411–9457, 2021.
[52] G. Dimauro, V. Di Nicola, V. Bevilacqua, D. Caivano, and F. Girardi, "Assessment of speech intelligibility in parkinson's disease using a speech-to-text system," *IEEE Access*, vol. 5, pp. 22 199–22 208, 2017.
[53] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
[54] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.
[55] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, H. W. Chung, C. Sutton, S. Gehrmann *et al.*, "Palm: Scaling language modeling with pathways," *arXiv preprint arXiv:2204.02311*, 2022.
[56] G. K. Savova, J. J. Masanz, P. V. Ogren, J. Zheng, S. Sohn, K. C. Kipper-Schuler, and C. G. Chute, "Mayo clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications," *Journal of the American Medical Informatics Association*, vol. 17, no. 5, pp. 507–513, 2010.
[57] C. Manning and H. Schutze, *Foundations of statistical natural language processing*. MIT press, 1999.
[58] A. S. Cohen, K. R. Mitchell, and B. Elvevåg, "What do we really know about blunted vocal affect and alogia? a meta-analysis of objective assessments," *Schizophrenia research*, vol. 159, no. 2-3, pp. 533–538, 2014.
[59] P. Baki, H. Kaya, H. Güleç, A. A. Salah *et al.*, "A multimodal approach for mania level prediction in bipolar disorder," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 2119–2131, 2022.
[60] N. J. Carson, B. Mullin, M. J. Sanchez, F. Lu, K. Yang, M. Menezes, and B. L. Cook, "Identification of suicidal behavior among psychiatrically hospitalized adolescents using natural language processing and machine learning of electronic health records," *PloS one*, vol. 14, no. 2, p. e0211116, 2019.
[61] M.-H. Metzger, N. Tvardik, Q. Gicquel, C. Bouvry, E. Poulet, and V. Potinet-Pagliaroli, "Use of emergency department electronic medical records for automated epidemiological surveillance of suicide attempts: a french pilot study," *International journal of methods in psychiatric research*, vol. 26, no. 2, p. e1522, 2017.
[62] G. Coppersmith, R. Leary, P. Crutchley, and A. Fine, "Natural language processing of social media as screening for suicide risk," *Biomedical informatics insights*, vol. 10, p. 1178222618792860, 2018.
[63] A. Bittar, S. Velupillai, A. Roberts, and R. Dutta, "Text classification to inform suicide risk assessment in electronic health records." in *MedInfo*, 2019, pp. 40–44.
[64] M. Tanana, K. A. Hallgren, Z. E. Imel, D. C. Atkins, and V. Srikumar, "A comparison of natural language processing methods for automated coding of motivational interviewing," *Journal of substance abuse treatment*, vol. 65, pp. 43–50, 2016.
[65] M. J. Baggott, M. G. Kirkpatrick, G. Bedi, and H. de Wit, "Intimate insight: Mdma changes how people talk about significant others," *Journal of Psychopharmacology*, vol. 29, no. 6, pp. 669–677, 2015.
[66] D. To, B. Sharma, N. Karnik, C. Joyce, D. Dligach, and M. Afshar, "Validation of an alcohol misuse classifier in hospitalized patients," *Alcohol*, vol. 84, pp. 49–55, 2020.
[67] M. Hoogendoorn, T. Berger, A. Schulz, T. Stolz, and P. Szolovits, "Predicting social anxiety treatment outcome based on therapeutic email conversations," *IEEE journal of biomedical and health informatics*, vol. 21, no. 5, pp. 1449–1459, 2016.
[68] R. Patel, T. Lloyd, R. Jackson, M. Ball, H. Shetty, M. Broadbent, J. R. Geddes, R. Stewart, P. McGuire, and M. Taylor, "Mood instability is a common feature of mental health disorders and is associated with poor clinical outcomes," *BMJ open*, vol. 5, no. 5, p. e007504, 2015.
[69] T. Banerjee, M. Kollada, P. Gersberg, O. Rodriguez, J. Tiller, A. E. Jaffe, and J. Reynders, "Predicting mood disorder symptoms with remotely collected videos using an interpretable multimodal dynamic attention fusion network," *arXiv preprint arXiv:2109.03029*, 2021.
[70] G. Stratou, S. Scherer, J. Gratch, and L.-P. Morency, "Automatic nonverbal behavior indicators of depression and ptsd: The effect of gender," *Journal on Multimodal User Interfaces*, 2014.
[71] pp. 11–18, 2014.
[72] Z. Huang, J. Epps, D. Joachim, and M. Chen, "Depression detection from short utterances via diverse smartphones in natural environmental conditions," *Interspeech*, 2018.
[73] C. W. Espinola, J. C. Gomes, J. M. S. Pereira, and W. P. d. Santos, "Detection of major depressive disorder, bipolar disorder, schizophrenia and generalized anxiety disorder using vocal acoustic analysis and machine learning," *Research on Biomedical Engineering*, vol. 38, p. 813–829, 2022.
[74] E. A. Stepanov, S. Lathuiliere, S. A. Chowdhury, A. Ghosh, R.-L. Vieriu, N. Sebe, and G. Riccardi, "Depression severity estimation from multiple modalities," *IEEE International Conference on e-Health Networking, Applications, and Services*, vol. 20, 2018.
[75] J. Joshi, R. Goecke, S. Alghowinem, A. Dhall, M. Wagner, J. Epps, G. Parker, and M. Breakspear, "Multimodal assistive technologies for depression diagnosis and monitoring," *Journal on Multimodal User Interfaces*, vol. 7, pp. 217–228, 2013.
[76] L. Yang, D. Jiang, L. He, E. Pei, M. C. Oveneke, and H. Sahli, "Decision tree based depression classification from audio video and language information," in *Proceedings of the 6th international workshop on audio/visual emotion challenge*, 2016, pp. 89–96.
[77] S. Sardari, B. Nakisa, M. N. Rastgoo, and P. Eklund, "Audio based depression detection using convolutional autoencoder," *Expert Systems with Applications*, vol. 189, p. 116076, 2022.
[78] Z. Zhang, W. Lin, M. Liu, and M. Mahmoud, "Multimodal deep learning framework for mental disorder recognition," in *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*. IEEE, 2020, pp. 344–350.
[79] S. Lipovetsky and M. Conklin, "Analysis of regression in game theory approach," *Applied Stochastic Models in Business and Industry*, vol. 17, no. 4, pp. 319–330, 2001.
[80] Y. Zhou, X. Yao, W. Han, Y. Wang, Z. Li, and Y. Li, "Distinguishing apathy and depression in older adults with mild cognitive impairment using text, audio, and video based on multiclass classification and shapely additive explanations," *International Journal of Geriatric Psychiatry*, vol. 37, no. 11, 2022.
[81] J. J. Hox, *Multilevel analysis: Techniques and applications*. New York, NY: Routledge, 2010.

[82] B. G. Tabachnick and L. S. Fidell, *Multilevel liniear modeling*, 5th ed., 2007, book section 15, pp. 781–857.

[83] E. H. Simpson, "The interpretation of interaction in contingency tables," *Journal of the Royal Statistical Society, Series B*, vol. 13, p. 238–241, 1951.

[84] P. Castro-Rodrigues, M. Camacho, S. Almeida, M. M, C. Soares, J. Barahona-Corrêa, and A. Oliveira-Maia, "Criterion validity of the yale-brown obsessive-compulsive scale second edition for diagnosis of obsessive-compulsive disorder in adults," *Frontiers in Psychiatry*, vol. 11, no. 9, 2018.

[85] E. A. Storch, M. Larson, L. Price, S. Rasmussen, T. Murphy, and W. K. Goodman, "Psychometric analysis of the yale-brown obsessive–compulsive scale second edition symptom checklist," *Journal of Anxiety Disorders*, vol. 24, no. 6, pp. 650–656, 2010.

[86] A. T. Beck, R. A. Steer, and G. Brown, *Manual for the Beck Depression Inventory-II*. San Antonio: Psychological Corporation, 1996.

[87] E. A. Storch, S. A. Rasmussen, L. H. Price, M. J. Larson, T. K. Murphy, and W. K. Goodman, "Development and psychometric evaluation of the Yale–Brown Obsessive-Compulsive Scale—Second Edition." *Psychological Assessment*, vol. 22, no. 2, pp. 223–232, 2010.

[88] M. D. McAuley, "Incorrect calculation of total electrical energy delivered by a deep brain stimulator," *Brain Stimulation*, vol. 13, no. 5, pp. 1414–1415, 9 2020.

[89] L. A. Jeni, J. F. Cohn, and T. Kanade, "Dense 3D face alignment from 2D video for real-time use," *Image and Vision Computing*, 2017.

[90] Z. Zhang, J. M. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, M. Reale, A. Horowitz, H. Yang, J. F. Cohn, Q. Ji, and L. Yin, "Multimodal spontaneous emotion corpus for human behavior analysis," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016.

[91] J. M. Girard, W.-S. Chu, L. A. Jeni, and J. F. Cohn, "Sayette group formation task (gft) spontaneous facial expression database," in *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)*. IEEE, 2017, pp. 581–588.

[92] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a python package)," *Neurocomputing*, vol. 307, pp. 72–77, 2018.

[93] C. Dewi, R.-C. Chen, X. Jiang, and H. Yu, "Adjusting eye aspect ratio for strong eye blink detection based on facial landmarks," *PeerJ Computer Science*, vol. 8, p. e943, 2022.

[94] "TranscribeMe! - fast & accurate human transcription services." [Online]. Available: https://www.transcribeme.com/

[95] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal forced aligner: trainable text-speech alignment Using kaldi," in *Proc. Interspeech 2017*, 2017, pp. 498–502.

[96] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor," in *Proceedings of the 21st ACM international conference on Multimedia*, 2013, pp. 835–838.

[97] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of liwc2015," Tech. Rep., 2015.

[98] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: Liwc and computerized text analysis methods," *Journal of language and social psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[99] H. Wu, *Nonparametric regression methods for longitudinal data analysis [mixed-effects modeling approaches]*, ser. Wiley series in probability and statistics. Hoboken, N.J: Wiley-Interscience, 2006.

[100] A. Hajjem, F. Bellavance, and D. Larocque, "Mixed effects regression trees for clustered data," *Statistics & Probability Letters*, vol. 81, no. 4, pp. 451–459, Apr. 2011.

[101] L. S. Shapley, "A value for n-person games," *Classics in game theory*, vol. 69, 1997.

[102] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[103] M. T. Ribeiro, S. Singh, and C. Guestrin, "" why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.

[104] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, G. Parker, M. Breakspear *et al.*, "Characterising depressed speech for classification," 2013.

[105] "Brain Behavior Quantification & Synchronization Workshop," https://event.roseliassociates.com/bbqs-workshop, 2023, Accessed on 21st May 2023.

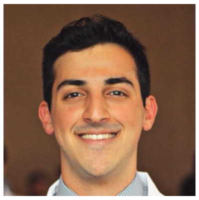
[106] J. M. Hofman, D. J. Watts, S. Athey, F. Garip, T. L. Griffiths, J. Kleinberg, H. Margetts, S. Mullainathan, M. J. Salganik, S. Vazire, A. Vespignani, and T. Yarkoni, "Integrating explanation and prediction in computational social science," *Nature*, vol. 595, pp. 181–188, 2021.

Saurabh Hinduja is a Post Doctorate Research Associate at the Affect Analysis Group, University of Pittsburgh, PA, USA. He received his PhD in Computer Science from University of South Florida. He has an MBA from the Symbiosis Center for Management and Human Resource Development (India) and a bachelors degree in engineering from the Birla Center for Management and Human Resource Development (India). His areas of interests include affective computing, artificial intelligence and machine learning. His work is in understanding contextual and self reported emotions. He is a member of the IEEE.

Ali Darzi is a visiting scholar and a former Postdoctorate Research Associate at the Affect Analysis Group, University of Pittsburgh, PA, USA. He received his Ph.D. in Electrical Engineering from the University of Wyoming, where his research focused on psychophysiology, rehabilitation robotics, and human-machine interaction. His interests lie in the fields of multimodal affective computing, digital signal and image processing, statistics, and machine learning. His current work revolves around the application of affective computing to develop a closed-loop adaptation paradigm for machines.

Itir Onal Ertugrul is an Assistant Professor at Social and Affective Computing Group, Department of Information and Computing Sciences at Utrecht University. Prior to joining Utrecht University, she was an Assistant Professor at Tilburg University, a postdoctoral researcher at the Robotics Institute at Carnegie Mellon University and Affect Analysis Group at University of Pittsburgh. She received B.Sc., M.Sc. and Ph.D. degrees from the Department of Computer Engineering at Middle East Technical University. Her research interests are in the broad areas of computer vision, machine learning, and affective computing, with a specific focus on automated analysis and synthesis of facial actions to understand human behavior, emotion, pain, and psychopathology.

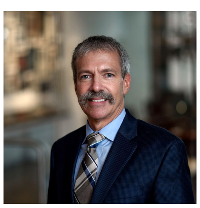
Nicole Provenza is a Postdoctoral Fellow at Baylor college of Medicine, TX, USA. She earned her PhD in Biomedical Engineering Brown University where she was a Draper Fellow. Her research seeks to develop closed-loop technologies to treat neurological and neuropsychiatric disorders. Areas of interest include neural data analysis, neural feature identification & classification, and real-time algorithm development.



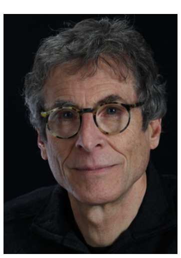
Ron Gadot is a medical student at Baylor College of Medicine and also a teaching assistant for the nervous system course at Baylor College of Medicine1. He is also affiliated with the Department of Neurosurgery at Baylor College of Medicine.

Eric A. Storch is a professor and the McIngvale Presidential Endowed Chair in the Department of Psychiatry and Behavioral Sciences at Baylor College of Medicine. He specializes in the cognitive-behavioral treatment of adult and childhood OCD, as well as other anxiety and OC-related disorders. His research interests involve the presentation and mechanisms and treatment of disorders, with a particular focus on driving innovation in the treatment of mental health disorders through the integration of academic research, evidence-based practice, and novel technological approaches.

Sameer A. Sheth is an Associate Professor of Neurosurgery and Vice-Chair of Clinical Research at Baylor College of Medicine. He specializes in the treatment of patients with Movement Disorders, Epilepsy, Brain Tumors, Trigeminal Neuralgia, Hydrocephalus, and certain Psychiatric Disorders. His translational research interests include developing and studying neuromodulation techniques for emerging neuropsychiatric conditions, including OCD, depression, addiction, schizophrenia.



Wayne K. Goodman is the D.C and Irene Ellwood Professor and chair of the Menninger Department of Psychiatry and Behavioral Sciences at Baylor College of Medicine. He is the principal developer, along with his colleagues, of the Yale-Brown Obsessive Compulsive Scale (Y-BOCS), which is considered to be the gold standard for assessing OCD. His research interests include OCD, deep brain stimulation, depression, and habenula.



Jeffrey F. Cohn is a professor of psychology, psychiatry, and intelligent systems at the University of Pittsburgh. He has led interdisciplinary and inter-institutional efforts to develop advanced methods of automatic analysis of facial expression, body motion, and prosody and applied those tools to research in human emotion, interpersonal processes, social development, and psychopathology. He has co-developed influential databases (Cohn-Kanade, MultiPIE DISFA, Pain Archive, and the BP4D series), co-edited special issues on facial expression analysis, and chaired international conferences in automatic face and gesture recognition, multimodal interaction, and affective computing.