Psychological Medicine

cambridge.org/psm

Original Article

Cite this article: Schultebraucks K, Yadav V, Shalev AY, Bonanno GA, Galatzer-Levy IR (2020). Deep learning-based classification of posttraumatic stress disorder and depression following trauma utilizing visual and auditory markers of arousal and mood. *Psychological Medicine* 1–11. https://doi.org/10.1017/S0033291720002718

Received: 20 January 2020
Revised: 23 June 2020
Accepted: 10 July 2020

Key words:
Computer vision; deep learning; depression; digital biomarker; emergency department; landmark feature; posttraumatic stress; resilience; voice analysis

Author for correspondence:
Katharina Schultebraucks,
E-mail: ks3796@cumc.columbia.edu

Deep learning-based classification of posttraumatic stress disorder and depression following trauma utilizing visual and auditory markers of arousal and mood

Katharina Schultebraucks[1,2,3], Vijay Yadav[4], Arieh Y. Shalev[2], George A. Bonanno[5] and Isaac R. Galatzer-Levy[2,4]

[1]Department of Emergency Medicine, Vagelos School of Physicians and Surgeons, Columbia University Irving Medical Center, New York, New York, USA; [2]Department of Psychiatry, New York University Grossman School of Medicine, New York, New York, USA; [3]Data Science Institute, Columbia University, New York, New York, USA; [4]AiCure, New York, New York, USA and [5]Department of Counseling and Clinical Psychology, Teachers College, Columbia University, New York, New York, USA

Abstract

Background. Visual and auditory signs of patient functioning have long been used for clinical diagnosis, treatment selection, and prognosis. Direct measurement and quantification of these signals can aim to improve the consistency, sensitivity, and scalability of clinical assessment. Currently, we investigate if machine learning-based computer vision (CV), semantic, and acoustic analysis can capture clinical features from free speech responses to a brief interview 1 month post-trauma that accurately classify major depressive disorder (MDD) and posttraumatic stress disorder (PTSD).

Methods. $N = 81$ patients admitted to an emergency department (ED) of a Level-1 Trauma Unit following a life-threatening traumatic event participated in an open-ended qualitative interview with a para-professional about their experience 1 month following admission. A deep neural network was utilized to extract facial features of emotion and their intensity, movement parameters, speech prosody, and natural language content. These features were utilized as inputs to classify PTSD and MDD cross-sectionally.

Results. Both video- and audio-based markers contributed to good discriminatory classification accuracy. The algorithm discriminates PTSD status at 1 month after ED admission with an AUC of 0.90 (weighted average precision = 0.83, recall = 0.84, and f1-score = 0.83) as well as depression status at 1 month after ED admission with an AUC of 0.86 (weighted average precision = 0.83, recall = 0.82, and f1-score = 0.82).

Conclusions. Direct clinical observation during post-trauma free speech using deep learning identifies digital markers that can be utilized to classify MDD and PTSD status.

Introduction

Treatment for psychiatric disorders is predicated on the identification of discrete psychiatric outcomes. Yet, such outcomes say little about the mechanisms that might govern behavioral and physiological functioning underlying the disorders. A greater understanding of such clinical characteristics and their prognostic value would improve diagnostic precision, help tailor treatment choice, and enhance risk detection and treatment outcome monitoring. A promising avenue to this end is digital phenotyping, the direct, moment-to-moment, objective measurement of clinical characteristics using digital data sources (Huckvale, Venkatesh, & Christensen, 2019).

Visual and vocal characteristics represent a compelling direction in digital phenotyping as signs and symptoms of diverse central nervous system (CNS) disorders have known behavioral signatures, such as vigilance, arousal, fatigue, agitation, psychomotor retardation, flat affect, inattention, compulsive repetition, and negative affective biases, to name a few (American Psychiatric Association, 2013). Advances in both deep learning and computational power now allow for rapid and accurate measurement of myriad markers that have already demonstrated robust effects in clinical populations. For example, facial expressions of emotion, which have demonstrated effects in multiple clinical populations (Cohn et al., 2009; Ekman & Friesen, 1978; Ekman, Matsumoto, & Friesen, 1997; Gaebel & Wölwer, 1992; Gehricke & Shapiro, 2000; Girard, Cohn, Mahoor, Mavadati, & Rosenwald, 2013; Renneberg, Heyn, Gebhard, & Bachmann, 2005) can be coded using computer vision (CV) based open-source software (Amos, Ludwiczuk, & Satyanarayanan, 2016; Baltrusaitis, Zadeh, Lim, & Morency, 2018; Bradski & Kaehler, 2008) and utilized in real-time to measure clinical functioning (Bao & Ma, 2014; Cohn et al., 2009; Gaebel & Wölwer, 1992; Girard et al., 2013; Wang, 2016; Xing & Luo, 2016; Zhong, Chen, & Liu, 2014). Further, measurements of facial emotion intensity provide a direct index for flat affect, fatigue, and pain (Ekman, Freisen, & Ancoli, 1980; Kohler et al., 2008; Simon, Craig, Gosselin, Belin, & Rainville, 2008). Similarly, pitch, tone, rate of speech, and valence of language, all of which are quantifiable based on either prosaic or natural language models, index motor, mood, and cognitive functioning (Bernard & Mittal, 2015; Cannizzaro et al., 2004; Cohn et al., 2009; Eichstaedt et al., 2018; France, Shiavi, Silverman, Silverman, & Wilkes, 2000; He, Veldkamp, & de Vries, 2012; Kleim, Horn, Kraehenmann, Mehl, & Ehlers, 2018; Leff & Abberton, 1981; Lu et al., 2012; Marmar et al., 2019; Pestian, Nasrallah, Matykiewicz, Bennett, & Leenaars, 2010; Quatieri & Malyska, 2012; Sobin & Sackeim, 1997; van den Broek, van der Sluis, & Dijkstra, 2010; Yang, Fairbairn, & Cohn, 2013). Previous studies were able to link digital biomarker with well-known symptoms of posttraumatic stress disorder (PTSD) and major depressive disorder (MDD) that can be measured via facial markers such as decreased flexibility of emotion expression in PTSD (Rodin et al., 2017), decreased positive affect and higher anger expression (Blechert, Michael, & Wilhelm, 2013; Kirsch & Brunnhuber, 2007) and higher facial affect intensity in PTSD (McTeague et al., 2010) as well as decreased facial expressivity in patients with MDD (Davies et al., 2016; Gaebel & Wölwer, 1992; Girard et al., 2013; Renneberg et al., 2005; Sloan, Strauss, Quirk, & Sajatovic, 1997). In addition, voice markers including volume, fundamental frequency, jitter, shimmer, and harmonics-to-noise ratio have been associated with PTSD (Scherer, Stratou, Gratch, & Morency, 2013; Xu et al., 2012) and MDD (Asgari, Shafran, & Sheeber, 2014; Breznitz, 1992; Cummins, Sethu, Epps, Schnieder, & Krajewski, 2015b; Hönig, Batliner, Nöth, Schnieder, & Krajewski, 2014; Kiss, Tulics, Sztahó, Esposito, & Vicsi, 2016; Nilsonne, Sundberg, Ternström, & Askenfelt, 1988; Ozdas, Shiavi, Silverman, Silverman, & Wilkes, 2004; Porritt, Zinser, Bachorowski, & Kaplan, 2014; Quatieri & Malyska, 2012; Scherer et al., 2013). Previous studies also identified relevant markers of speech content. For instance, the speech rate was negatively correlated with both PTSD and depression symptom severity (Scherer, Lucas, Gratch, Rizzo, & Morency, 2015) and narrative coherence in PTSD (He, Veldkamp, Glas, & de Vries, 2017). Also, unique patterns of speech content were identified as indicators of MDD such as the rate of speech, lexical diversity, pauses between words and the sentiment of speech content (Alghowinem et al., 2013; Calvo, Milne, Hussain, & Christensen, 2017; Cummins, Epps, Breakspear, & Goecke, 2011; Cummins et al., 2015a; Marge, Banerjee, & Rudnicky, 2010; McNally, Otto, & Hornig, 2001; Mowery, Smith, Cheney, Bryan, & Conway, 2016; Nilsonne, 1988, 1987; Sturim, Torres-Carrasquillo, Quatieri, Malyska, & McCree, 2011). Digital biomarkers of movement in PTSD revealed an association with suppressed motor activity to neutral stimuli (Litz, Orsillo, Kaloupek, & Weathers, 2000) and heightened arousal (Blechert et al., 2013) and increased eye blink (McTeague et al., 2010) and increased fixation on trauma-related stimuli (Felmingham, Rennie, Manor, & Bryant, 2011). Digital biomarkers of movement in MDD have also been examined (Anis, Zakia, Mohamed, & Jeffrey, 2018; Bhatia, Goecke, Hammal, & Cohn, 2019; Dibeklioğlu, Hammal, & Cohn, 2017; Shah, Sidorov, & Marshall, 2017), such as psychomotor retardation (Syed, Sidorov, & Marshall, 2017).

Deep learning is an emerging tool to bridge the gap between empirical findings and explanatory theories of psychology and cognitive neuroscience (Hasson, Nastase, & Goldstein, 2020). While simple correlational analyses are limited to discern informative associations from spurious effects (Meehl, 1990), deep neural networks are impressively successful in learning to mimic human cognitive processes such as face recognition in a data-driven way (LeCun, Bengio, & Hinton, 2015). Based on higher-order representations of multivariate dependencies, deep learning can achieve near-perfect accuracy in face recognition (>99.7%) (Grother, Ngan, & Hanaoka, 2020) without making theoretical assumptions that explain how humans perform such tasks (Hasson et al., 2020). Moreover, existing theories of emotion processing in PTSD such as the early Bio-Informational Processing Theory (Lang, 1979) can provide theoretical motivation for Digital Phenotyping based on deep learning without the reverse being true. Informed by existing theories, deep learning can attempt to emulate sensory imagery and text comprehension that link and activate conceptual networks that are directly coupled with overt behavioral expression. However, the aim of applying deep learning is not to accurately model such theories but, more modestly, to find stable probabilistic patterns in a data-driven way (Valiant, 1984). With the focus on prediction rather than explanation (Shmueli, 2010), existing theories of PTSD, such as the Emotional Processing Theory (Foa, Huppert, & Cahill, 2006) are still highly valuable and informative for the selection of candidate predictors, but the successful predictive model will be theoretically agnostic and neither corroborate nor disprove any particular explanatory theory. Emotional Processing Theory is particularly informative for Digital Phenotyping as it explains how the brain dynamically integrates multidimensional information resulting in rich context-dependent emotional, cognitive, and behavioral reactions. Deep learning allows integrating multiple empirical associations, including subtle ones, into a computational framework. The study of face, voice, and speech content as indicators of human psychology and psychopathology such as PTSD has a long tradition. As one of the first and most well-known examples, Charles Darwin postulated that biologically 'hard-wired' facial expressions of emotions (Darwin, 1872/1965) signal and convey important information about the emotional and mental states of a person (Ekman, 2006). Decades of neuropsychological research showed that emotional expression and valence contain predictive probabilistic information of diverse forms of psychopathology (Gaebel & Wölwer, 1992; Gehricke & Shapiro, 2000; Renneberg et al., 2005). Speech and voice are additional channels conveying probabilistic information about mental health (Cannizzaro et al., 2004; Cohn et al., 2009; France et al., 2000; Leff & Abberton, 1981). Physical movements represent a further behavioral output that can be used to characterize clinical functioning across a spectrum from psychomotor retardation to agitation (Bernard & Mittal, 2015; Sobin & Sackeim, 1997). However, dimensions of facial expressivity, speech, and movement are most likely not univocal categorical indicators of mutually exclusive disorders, but rather vary and overlap across clinical presentations making these dimensions transdiagnostic indicators of clinical functioning more generally.

A shift in focus away from psychiatric diagnostic classifications, that are known to be heterogeneous and lack a biological basis (Galatzer-Levy & Bryant, 2013), to directly observable dimensions of behavior and physiology, may improve diagnosis and treatment based on the underlying neurobiological functioning (Insel, 2014). For example, motor functioning is both affected across a wide variety of psychiatric disorders (e.g. psychomotor retardation in schizophrenia and depression, agitation in anxiety disorders, and tremor in Parkinson's disease and essential tremor) and has known treatment targets (e.g.

dopaminergic pathways; motor cortex activity). The direct and frequent measurement of motor deficits can facilitate the modulation of motor functioning across diverse disorders using known treatment options.

In the current study, we examined if the direct digital measurement of facial features and their intensity, head movement and eye movement, prosaic and natural language features can accurately identify clinical functioning in a population at heightened risk for MDD and PTSD. Core features of PTSD and depression include variability in arousal, mood, and vigilance (American Psychiatric Association, 2013). Further, patients with PTSD and MDD have demonstrated individual differences compared to healthy controls in the expression of facial features of emotion, prosaic vocal features, and speech content (Cannizzaro et al., 2004; Cohn et al., 2009; Gaebel & Wölwer, 1992; Gehricke & Shapiro, 2000; He et al., 2012; Kleim et al., 2018; Marmar et al., 2019; Quatieri & Malyska, 2012; Renneberg et al., 2005; van den Broek et al., 2010; Yang et al., 2013).

We capitalized on recent developments in deep learning (Goodfellow, Bengio, Courville, & Bengio, 2016) that have facilitated groundbreaking advances in affect detection, movement modeling, and speech/language analysis (Baltrusaitis et al., 2018; Cannizzaro et al., 2004; Cohn et al., 2009; Gaebel & Wölwer, 1992; He et al., 2012; Kleim et al., 2018; Pestian et al., 2010; Quatieri & Malyska, 2012; van den Broek et al., 2010; Yang et al., 2013). Convolutional Neural Networks and Deep Neural Networks can be utilized to identify face, voice, language, and movement characteristics from audio and video data (Amos et al., 2016; Baltrusaitis et al., 2018; Jadoul, Thompson, & de Boer, 2018) and can be utilized to integrate features to build and validate a predictive model. Intuitively, this modeling approach matches human clinical decision making where multiple aspects of the patient's presentation are integrated to identify risk.

Our aim was to use CV and neural networks to label facial landmark features of emotions as well as landmark features for voice prosody and to identify prognostic features of speech content using natural language processing (NLP) and to use them as labels for the classification of mental wellbeing. We hypothesized that deep learning would uncover unique probabilistic information on the integration of those information channels (multimodal fusion) that would yield discriminatory accuracy for the prediction of post-traumatic stress and MDD status ('proof-of-concept'). Such models can facilitate a more robust, accurate, ecologically valid, and ultimately automated and scalable method of risk identification based on unstructured data sources. Remote assessment of these models has particular relevance in the context of trauma exposure, as such events are ubiquitous, can occur rapidly and unexpectedly, and can affect individuals who are remote from appropriate clinical services (Carmi, Schultebraucks, & Galatzer-Levy, 2020).

Methods

Participants

Trauma survivors who were admitted to the emergency department (ED) of a Level-1 Trauma Center after experiencing a DSM-5 criterion A trauma were enrolled into a prospective longitudinal study cohort ($n = 221$) from 2012 to 2017 at Bellevue Hospital Center, New York City, NY (Schultebraucks et al., 2020). To be included in the study, participants had to be between 18 and 70 years of age and fluent in English, Spanish, or Mandarin. In addition, only participants who did not have an ongoing traumatic exposure such as domestic violence, no evidence of homicidal or suicidal behavior, and who were no prisoners were included in the study. Exclusion criteria included present or past psychotic symptoms, open head injury, coma, or evidence of traumatic brain injury [Glasgow Coma Scale score <13 (Teasdale et al., 2014)] or no reliable access to electronic mail or telephone. All procedures were reviewed, approved, and monitored by the NYU Institutional Review Board.

Procedure

Two primary outcomes were used in this analysis: (a) provisional PTSD diagnosis and (b) provisional depression diagnosis (yes/no) at 1 month following ED admission. PTSD status was evaluated using the PTSD Checklist for DSM-5 (PCL-5) (Weathers et al., 2013). Depression severity was evaluated using the Center for Epidemiologic Studies of Depression Scale (CES-D) (Eaton, Smith, Ybarra, Muntaner, & Tien, 2004). A PCL-5 total score ⩾33 and CES-D score ⩾23 was defined as the cut-off for screening positive for a provisional diagnosis of PTSD (Weathers et al., 2013) and provisional depression diagnosis (Henry, Grant, & Cropsey, 2018). We used the qualifier 'provisional diagnosis' according to DSM-5: 'when there is a strong presumption that the full criteria will ultimately be met for a disorder but not enough information is available to make a firm diagnosis' (American Psychiatric Association, 2013). The PCL-5 shows a 'good diagnostic utility for predicting a CAPS-5 PTSD diagnosis' and 'good structural validity, and sensitivity to clinical change comparable to that of a structured interview' (Weathers, 2017). In the population of trauma survivors, studies found that 'CAPS-5 and PCL-5 total scores correlated strongly ($r = 0.94$)' (Geier, Hunt, Nelson, Brasel, & de Roon-Cassini, 2019). Both measures have good reliability, convergent, concurrent, discriminant, and structural validity (Weathers, 2017).

Candidate predictors were extracted from a brief qualitative interview that was conducted along with other procedures under laboratory conditions at Bellevue Hospital 1 month following hospital discharge. Patients were asked to respond however they saw fit to the following five questions within a 3 min predetermined time limit for each question: (1) Tell me about your life before the event that brought you to the hospital; (2) Tell me about the event that brought you to the hospital; (3) Tell me about your hospital experience; (4) Tell me about your life since leaving the hospital; (5) What are your expectations about life in the future. Interviewers only asked brief pre-determined follow-up questions when patients stopped responding such as 'tell me more about that'. Interviews were audio and video recorded with a high-resolution camera mounted behind the interviewer's shoulder to provide a face-on view of the research subject.

Statistical analysis

Initial unsupervised video data processing

Images: For initial processing, each frame was extracted and then broken down to $3 \times m \times n$ matrices of m columns and n rows where three matrices represent red, blue, green spectrum extracted from the image using OpenCV (Bradski & Kaehler, 2008) in Python. Each value in each $m{\times}n$ matrix represents a pixel value from light to dark on the corresponding color spectrum.

Data labeling: visual and auditory markers of arousal and mood

Facial features of arousal and mood: Facial expressions of emotion were coded based on visible facial movements. Facial features corresponding to action units (AUs) identified by the Facial Action Coding System (FACS) (Ekman & Friesen, 1978) were labeled from raw MP4 video files using the OpenFace package in python (Amos et al., 2016), which has a confidence score that is more than 75% for face detection. The extracted raw features were used to compute a Facial Expressivity Score for each emotion (Happiness, Sadness, Anger, Disgust, Surprise, Fear, Contempt), and Peak Expressivity (1, 3, 6, 9, 12, 15 s windows). We also analyzed normalized emotions according to the Emotional Facial Action Coding System (EMFACS), Facial Expressivity Index, and Expressivity Peak Count.

Voice prosody features of arousal: For verbal analysis, PRAAT Software python library Parsel-mouth (Jadoul et al., 2018) was used. We analyzed the following parameters: Audio Expressivity Index, Audio Intensity (dB), Fundamental Frequency, Harmonic Noise Ratio, Glottal to Noise Excitation Ratio, Voice Frame Score, Formant Frequency Variability, Intensity Variability, Pitch Variability, Normalized Amplitude Quotient.

Speech content features of arousal and mood: Speech content was extracted with NLP using Receptiviti which uses the LIWC 2015 dictionary (Pennebaker, Boyd, Jordan, & Blackburn, 2015). Extracting features include, for instance, summary language variables, linguistic dimensions, psychological, social, cognitive, perceptual, and biological processes. We further extracted content using DeepSpeech (Hannun et al., 2014), which is an open-source pre-trained neural network model to extract text from speech. This software identifies features like rate of speech, intent expressivity, emotion label, and word repetition.

Movement features: Movement variables were extracted from raw MP4 video files using the OpenFace package in python (Amos et al., 2016; Baltrusaitis et al., 2018). We analyzed head movement, attentiveness, and pupil dilation rate.

For further information on the extracted features, please see online Supplementary Information.

Model development and model validation

Data were preprocessed using R package caret (Breiman, 1996; Kuhn, 2008; Kuhn & Johnson, 2013). Categorical variables were dummy coded to binary numerical values (‘one-hot encoding’) and numerical variables were normalized to the range of [0;1]. Variables with near-zero variance were removed. We had $\leqslant 1\%$ missing values. Those missing values were imputed using the *k*-nearest neighbor algorithm (knnImpute in caret) (Beretta & Santaniello, 2016).

To evaluate the model on data not used to select the model (Hastie, Tibshirani, & Friedman, 2009), we split the total sample into a training (75%) and test set (25%) (see online Supplementary Table S1). We used *k*-fold cross-validation with 10 folds in the training set to decrease the risk of overfitting (Stone, 1974).

For outcome prediction of provisional PTSD and provisional depression caseness at 1 month, supervised classification used a deep neural network with two hidden layers with Rectified Linear Unit (‘relu’) activation (Hahnloser, Sarpeshkar, Mahowald, Douglas, & Seung, 2000) and 20 units and an output layer with ‘sigmoid’ binary classification using the Keras library in Python (Chollet, 2018). Optimal weights were determined using ‘adam’ optimization of binary cross-entropy as loss function (De Boer, Kroese, Mannor, & Rubinstein, 2005) and precision as the evaluation metric for binary classification.

The pipeline of data analysis is visualized in Fig. 1.

To examine the stability of our results, we additionally used two times repeated nested cross-validation with a 10-fold inner loop and a 10-fold outer loop for prediction of provisional PTSD and depression diagnostic status at 1 month after ED admission.

Additionally, we predicted PTSD and MDD symptom severity using two deep neural networks with two hidden layers with ‘relu’ activation and 20 units and an output layer. Optimal weights were determined using ‘adadelta’ optimization of ‘mean squared error’ as the loss function and mean absolute error (MAE) as an evaluation metric.

Predictive importance ranking

We used Explainable Machine Learning using SHAP (SHapley Additive exPlanation) to identify those features that are mainly responsible for driving the individual outcome prediction. It is an additive feature attribution method that uses kernel functions and currently the gold standard to interpret deep neural networks (Lundberg & Lee, 2017).

Results

We extracted 247 features in $N = 81$ trauma survivors ($N = 34$, 42.5% female; mean age 37.86 ± 13.99; $N = 20$, 25% were Hispanic) as shown in Table 1.

Predictive model performance

The neural networks achieved good predictive power in the internal test set for predicting the provisional diagnosis (see Fig. 2). The algorithm achieved high discriminatory accuracy to classify PTSD status (AUC = 0.9, weighted average precision = 0.83, weighted average recall = 0.84, weighted average f1-score = 0.83) and MDD status (AUC = 0.86, weighted average precision = 0.83, weighted average recall = 0.82, weighted average f1-score = 0.82) in the internal test set.

The neural network for predicting PTSD symptom severity obtained a root-mean-squared-error (RMSE) of 10.31, MAE of 6.38, and $R^2 = 0.60$. For predicting MDD symptom severity, we attained an RMSE of 7.23, MAE of 5.58, and $R^2 = 0.62$.

Using the classifier obtained using a nested cross-validation approach, we achieved an AUC = 0.88 (weighted average precision = 0.89, weighted average recall = 0.87, weighted average f1-score = 0.87) for predicting MDD status and an AUC = 0.9 (weighted average precision = 0.9, weighted average recall = 0.89, weighted average f1-score = 0.9) for predicting PTSD.

Ranking the features for predictive value

Figures 3 and 4 display the variable importance using SHAP feature ranking. All four domains (face, voice, speech content, and movement) were ranked highly among the 20 most important predictors. The most important predictors for predicting PTSD and MDD status were NLP features, but also features of voice prosody such as audio intensity (PTSD and MDD status), pitch (PTSD status), facial features of emotion (PTSD and MDD status), and movement features, such as pupil dilation rate (MDD status). The most important predictor for PTSD was NPL LIWC ‘self-assured’ followed by NLP LIWC ‘compare’ with the

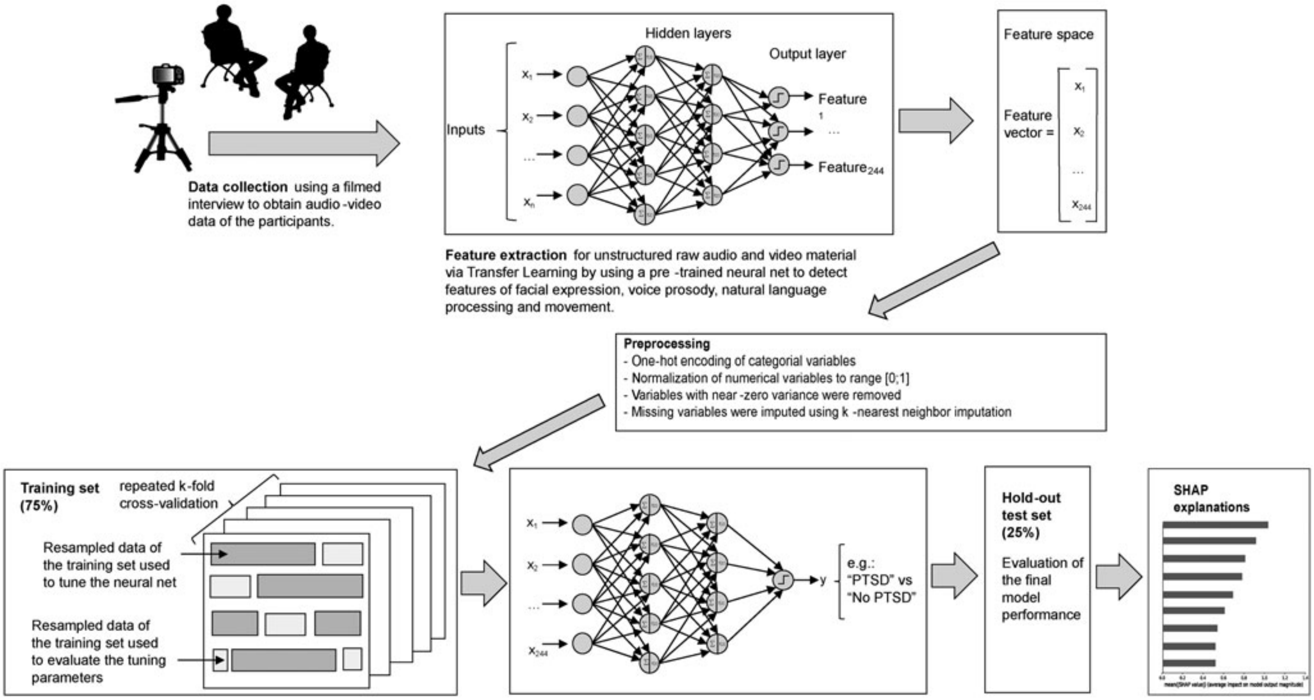
Data collection using a filmed interview to obtain audio-video data of the participants.
Hidden layers
Output layer
Inputs
x_1
x_2
…
x_n
Feature 1
…
Feature 244
Feature space
Feature vector =
x_1
x_2
…
x_{244}
Feature extraction for unstructured raw audio and video material via Transfer Learning by using a pre-trained neural net to detect features of facial expression, voice prosody, natural language processing and movement.
Preprocessing
- One-hot encoding of categorial variables
- Normalization of numerical variables to range [0;1]
- Variables with near-zero variance were removed
- Missing variables were imputed using k-nearest neighbor imputation
Training set (75%)
repeated k-fold cross-validation
Resampled data of the training set used to tune the neural net
Resampled data of the training set used to evaluate the tuning parameters
x_1
x_2
…
x_{244}
y
e.g.: "PTSD" vs "No PTSD"
Hold-out test set (25%)
Evaluation of the final model performance
SHAP explanations

Fig. 1. The pipeline of data analysis.

Table 1. Sample characteristics

	All participants (N = 81)	'PTSD' *v.* 'no PTSD'	'MDD' *v.* 'no MDD'
Age (mean ± S.D.)	37.86 ± 13.99	$t(74) = 0.89$, $p = 0.38$	$t(65) = 2.04$, $p = 0.05$ (younger age in MDD)
Gender (% Female)	42.5%	$\chi^2(1) = 0.003$, $p = 0.96$	$\chi^2(1) = 0.194$, $p = 0.66$
Trauma types (%)		$\chi^2(9) = 7.64$, $p = 0.57$	$\chi^2(10) = 5.76$, $p = 0.84$
Gunshot wound	1.2%		
Pedestrian *v.* car	16.0%		
Motor vehicle collision	13.6%		
Motorcycle collision	2.5%		
Bike accident	32.1%		
Fall	17.3%		
Non-sexual assault	8.6%		
Others	8.7%		

other predictors with similar variable importance (see Fig. 3*a*). The most important predictor for MDD status was age, followed by NLP LIWC 'workhorse' and NLP LIWC 'organized' with similar variable importance ranking for the following predictors (see Fig. 3*b*). We found similar predictive features when predicting PTSD and depression symptom severity at 1 month after ED admission (see Fig. 3*c* and *d*). A description of the definition of each feature shown in the variable importance ranking (Figs 3 and 4) can be found in online Supplementary Table S2.

Discussion

We utilized CV and NLP, and audio analysis to measure features associated with mood and arousal during free and continuous speech. In keeping with our underlying hypothesis that the integration of multiple sources of information will provide a stronger prediction than one source independently (Schultebraucks & Galatzer-Levy, 2019), we utilized a deep learning neural network approach. By analogy, a clinician interviewing a patient will integrate visual, auditory, and linguistic information to assess a patient. Experienced clinicians will process many more channels of information, can make use of context-dependent prior clinical experience, and will be able to form an empathetic therapeutic alliance. Although no algorithm is able to capture this level of skilled clinical expertise, there are common and much more fundamental clues of overt behavior that can be objectively encoded using digital methods and the development of such tools can further support clinicians by providing objective access to behavioral clues that are otherwise automatically processed by humans and often not perceived with deliberate

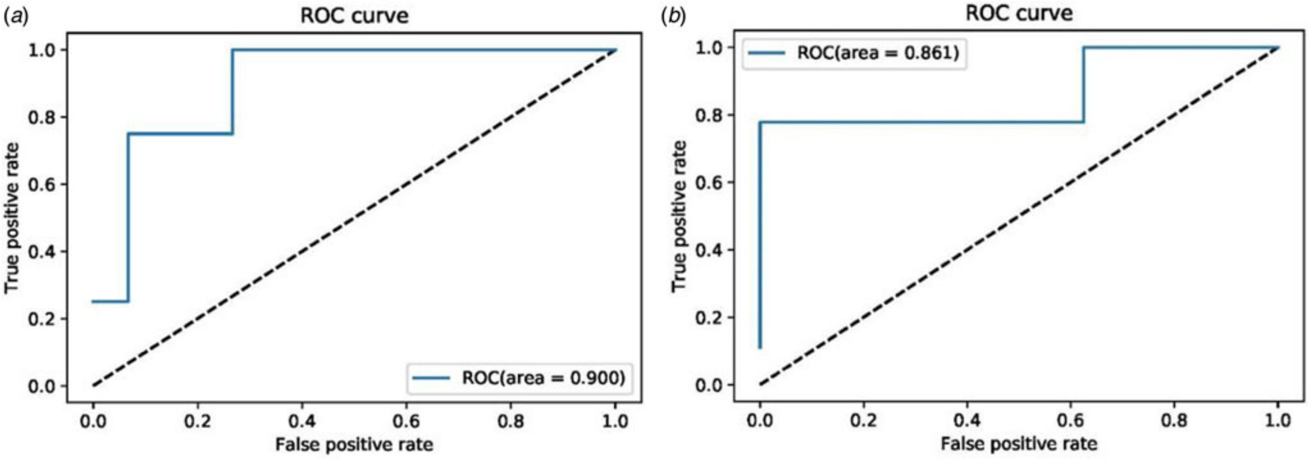

Fig. 2. Receiver operating characteristic (ROC) curve of the internal test set for predicting (*a*) PCL-5 cut-off ⩾33 (AUC = 0.90) and (*b*) CES-D cut-off ⩾23 (AUC = 0.86).

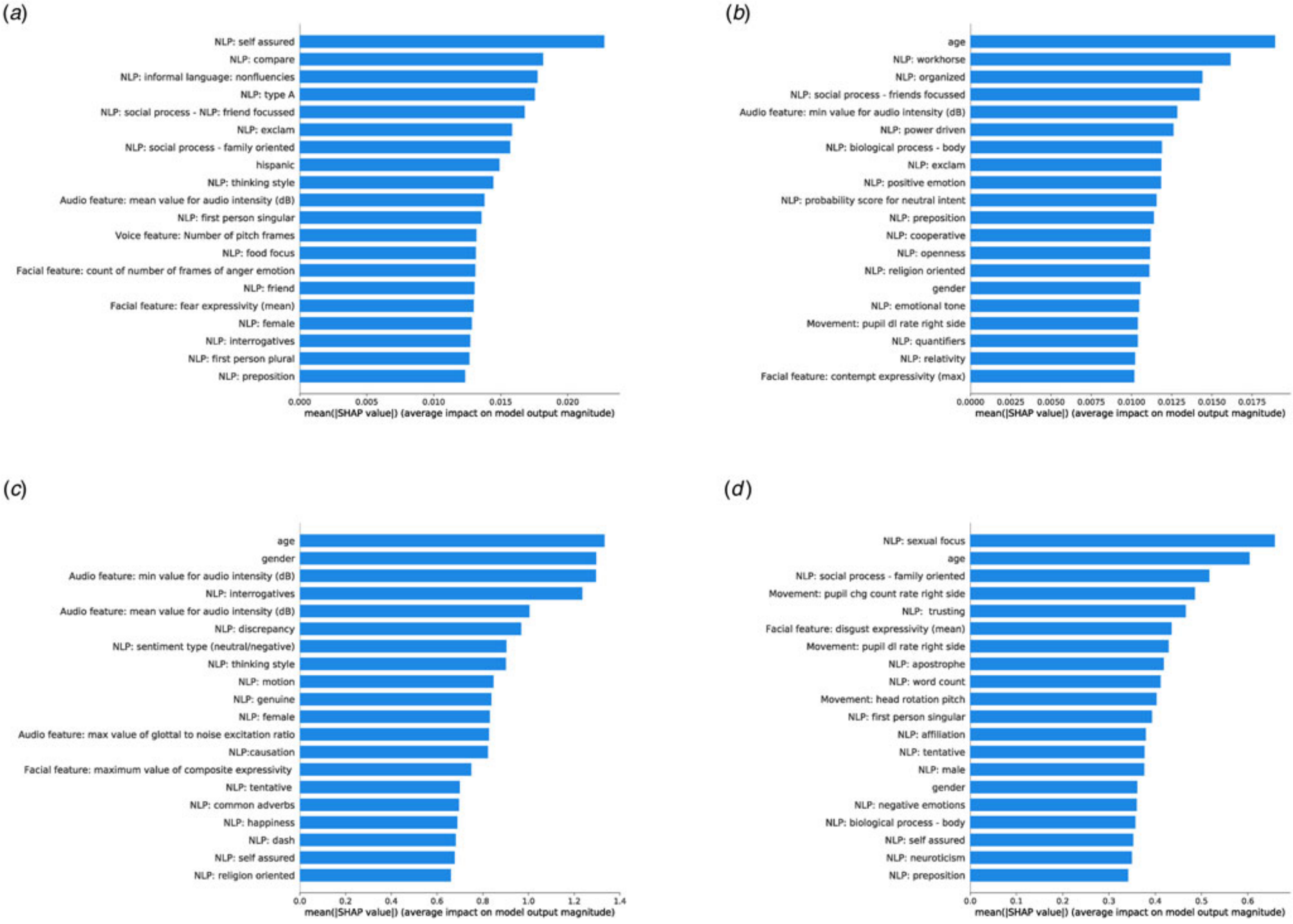

Fig. 3. SHAP (SHapley Additive exPlanations) variable importance (Lundberg & Lee, 2017) of the Neural Network for the internal test set for predicting (*a*) PCL-5 cut-off ⩾33, (*b*) CES-D cut-off ⩾23, (*c*) PCL-5 symptom severity, and (*d*) CES-D symptom severity. The mean absolute SHAP value per feature is presented in the bar plot with larger bar plots displaying higher importance of the feature in discriminating between the 'provisional PTSD diagnosis' and 'no PTSD'/'provisional depression diagnosis' and 'no depression'. The variable importance based on SHAP values is calculated by evaluating the model performance with and without each feature included in the model in every possible order.

attention. The encoding of such objective information about overt behavioral clues of visual, auditory, and linguistic information can be formalized and deployed in a reproducible manner using neural network architecture that encodes high dimensional representations of the relationship between multiple features (i.e. face, movement, speech, and language). The features that we found to be most important for the classification of provisional PTSD and MDD corroborated pre-existing findings reported in the current literature. We extended those findings to demonstrate that integrated features from different modalities, i.e. face, voice prosody,

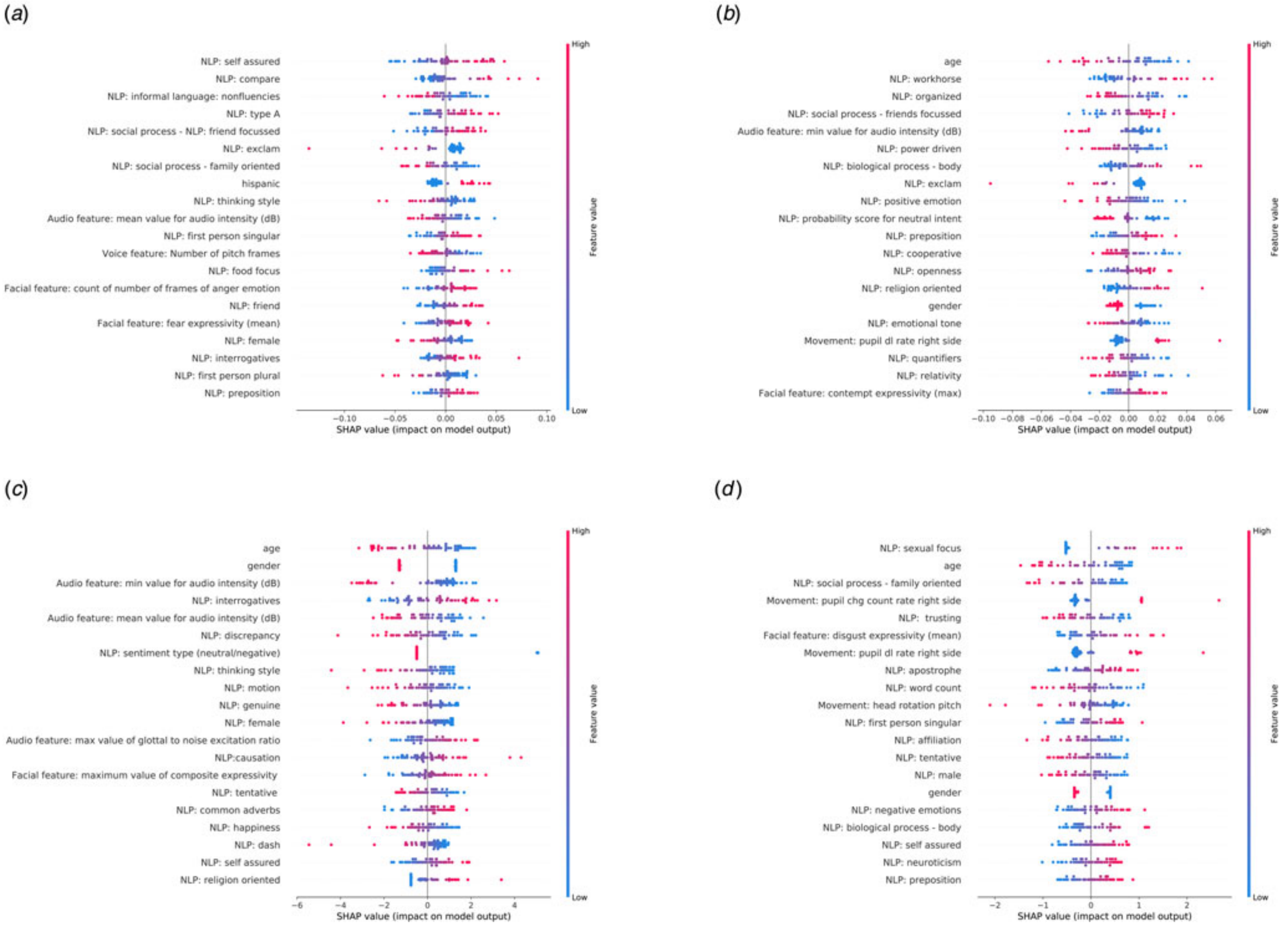

Fig. 4. SHAP (Lundberg & Lee, 2017) summary dot plot of the Neural Network for the internal test set for predicting (*a*) PCL-5 cut-off ⩾33 and (*b*) CES-D cut-off ⩾23, (*c*) PCL-5 symptom severity and (d) CES-D symptom severity. The higher the SHAP value of a feature, the higher the log odds of the 'provisional PTSD diagnosis'/ 'provisional depression diagnosis'. On the *y*-axis, the features are sorted by their general feature importance (see Fig. 4). The dots represent, for each variable value of each participant in the sample, how the variable value influences the attribution of the participant to one of the two outcome classes. Dots that are on the left side shift the classification of participants to the class 'no PTSD/no MDD', whereas dots on the right side of the *x*-axis shift the classification of participants to the class 'PTSD/MDD'. The color represents the range of the feature values from low (blue) to high (red). For instance, the lower the score for the feature 'positive emotion' (NLP), the higher the odds for 'provisional depression diagnosis' (Fig. 4*b*).

speech content, and movement, all contribute uniquely to the classification and prediction of both MDD and PTSD.

A significant body of research has identified facial, vocal, and motor movement markers of neuro-psychiatric functioning (Bernard & Mittal, 2015; Cannizzaro et al., 2004; Cohn et al., 2009; Eichstaedt et al., 2018; Gaebel & Wölwer, 1992; Gehricke & Shapiro, 2000; He et al., 2012; Kleim et al., 2018; Lu et al., 2012; Pestian et al., 2010; Quatieri & Malyska, 2012; Renneberg et al., 2005; Sobin & Sackeim, 1997; van den Broek et al., 2010; Yang et al., 2013). These characteristics relate to core symptomatology across diverse disorder including posttraumatic stress and depression as well as resilience (Cannizzaro et al., 2004; Cohn et al., 2009; France et al., 2000; Gaebel & Wölwer, 1992; Gehricke & Shapiro, 2000; He et al., 2012; Kleim et al., 2018; Leff & Abberton, 1981; Pestian et al., 2010; Quatieri & Malyska, 2012; Renneberg et al., 2005; van den Broek et al., 2010; Yang et al., 2013). In addition to informing psychopathology, these markers also provide information about CNS mechanisms that may affect clinical functioning and identify treatable targets for intervention. Visual and auditory markers have long been associated with mood and arousal, which in turn, represent core features of posttraumatic stress pathology including PTSD and depression (Otte et al., 2016; Shalev, Liberzon, & Marmar, 2017).

Our classification algorithm, based on participants' free discussion of their trauma experience, identified many of the features previously found to be predictive. For example, consistent with the literature, we observed that higher fear-expressivity and anger-expressivity were important for the classification of PTSD while higher contempt-expressivity was predictive of MDD (Ekman & Friesen, 1978; Ekman et al., 1997). Similarly, consistent with the literature, we found that the increased use of first-person singular pronouns provided probabilistic information in classifying PTSD (Kleim et al., 2018) and that reduced frequency of positive words predicted depression (Pennebaker, Mehl, & Niederhoffer, 2003; Rude, Gortner, & Pennebaker, 2004) while lowered audio intensity and reduced pitches per frame was relevant to the classification of PTSD (Marmar et al., 2019). This concordance with existing literature provides important validation of the probabilistic information used by our classification algorithm.

Extracting features from unstructured video data sources offers several important advantages. Beginning with the introduction of the research diagnostic criteria (Spitzer, Endicott, & Robins, 1978), which aimed to align clinical research and practice around objective criteria, visual and auditory signs were either measured through expert rating or through introspective self-report. While this created standardization, it was limited by the need for a rigid structure to assess symptoms, challenges of inter-rater reliability, subjective error in assessment, and significant assessment time burden. Our results successfully demonstrate 'proof-of-concept' by showing that clinical features measured without a rigid assessment structure yield discriminatory accuracy to classify provisional PTSD or MDD diagnostic status. The results are based on clinical signs captured from free exchanges with a para-professional. Further, multiple signs can be assessed simultaneously, reducing the assessment burden. The use of algorithms to code clinical signs also obviates issues of inter-rater reliability as the algorithm performs identically each time. The use of algorithms rather than rating scales provides a real number metric rather than a ranking of severity. The use of real numbers, by definition, increases the sensitivity of the metric. Finally, the use of audio and video data sources is scalable as it can be integrated into cellphones and web-based telemedicine applications. This can greatly increase the reach of assessment of clinical functioning.

There are also some limitations to note. Most importantly, the sample size will caution against the direct generalization to other samples without replication. Future studies might benefit from incorporating additional contextual information and feedback from experienced clinicians into the analysis. Our approach successfully combined multiple information channels such as facial emotion expression with NLP sentiment analysis based on word frequencies. This already provides contextual information across different modalities since facial expression complement characteristics of speech and audio modalities. However, with the benefits of larger samples, it will be useful to go beyond word frequencies by identifying predictive features of sentence-level meaning units and to directly test for cross-modal interactions of facial expressions and verbal expressions of emotion. While the current algorithm internally accounts for possible non-linear dependence between modalities in a data-driven way, the current approach is limited by not explicitly testing potential interactions between features and modalities. The variable importance ranking highlights the features that were, on average, most important to discriminate between 'PTSD' or 'MDD' and 'no PTSD' or 'no MDD' respectively. However, since the classification is only achieved by the combination of all variables together, the interpretation of univariate associations is limited and should not be interpreted causally. Larger samples are required to corroborate our results and also to directly test for interactions between facial and verbal modalities that provide an important opportunity to incorporate the rich clinical expertise of experienced clinicians. Moreover, the current study was focused on the classification of 'PTSD' *v.* 'no PTSD' and 'MDD' *v.* 'no MDD' while it would also be clinically relevant to discriminate between 'PTSD' and 'MDD' which remains an important desideratum for further studies. Another limitation is the reliance on pre-trained models for feature extraction. While we used state-of-the-art methods, there are known limitations and risk of bias that need to be pointed out with regard to facial expression recognition (Buolamwini & Gebru, 2018; Klare, Burge, Klontz, Bruegge, & Jain, 2012) and NLP (Caliskan, Bryson, & Narayanan, 2017).

The next step is to further gauge the predictive performance of the digital biomarkers in larger samples and, most importantly, in diverse and heterogeneous patient populations. To go beyond 'proof-of-concept', rigorous testing in a large confirmatory study design is warranted for extensive clinical validation (Mathews et al., 2019).

Conclusion

This study presented an approach that robustly and accurately predicted mental well-being in trauma survivors using an automated, scalable and ecologically valid method. Our proof-of-concept analysis requires further development and validation in independent samples. Nonetheless, the results demonstrated that construct-valid features such as facial affect, movement, speech content, and prosody can be captured in minimally structured contexts to accurately quantify clinical functioning. These results hold significant implications for how deep learning-based methods can automate and scale clinical assessment. Our results also suggest implications for the fields' ability to put a clinical focus on discrete behavioral and physiological dimensions as metrics of risk and treatment response consistent with the research domain criteria approach (Cuthbert & Insel, 2013; Insel, 2014; Insel et al., 2010). The emphasis on directly observable behavior and physiology shifts the attention away from a narrow focus on psychiatric diagnostic classifications that are known to be heterogeneous and lack a biological basis (Galatzer-Levy & Bryant, 2013). Remote assessment based on digital markers is, for instance, important in the context of trauma exposure in inaccessible areas after natural catastrophes or unsafe terrain such as warzone or areas of humanitarian crises, which often affect individuals who are distant from appropriate clinical services (Carmi et al., 2020). There is a high potential for the future use of remote assessment using digital biomarkers in these circumstances and the here presented proof-of-principle demonstration of digital biomarkers for PTSD and MDD warrants further investigation in larger samples and diverse clinical contexts. Ultimately, digital biomarkers bear great promise to improve current telemedicine services to provide digital diagnostic screening at scale.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291720002718.

Financial support. K.S. was supported by the German Research Foundation (SCHU 3259/1-1). The study was also supported by K01MH102415 (I.R.G.-L.).

Conflict of interest. I.R.G.-L. and V.Y. receive salary from AiCure. I.R.G.-L. and V.Y. have no stocks. All other authors declare no potential conflict of interest that is relevant to this article.

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 2013.

References

Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Breakspear, M., & Parker, G. (2013). *Detecting depression: a comparison between spontaneous and read speech*. IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, pp. 7547–7551.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (*DSM-5*®) (Fifth Edition). Arlington, VA: American Psychiatric Association.

Amos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*, *6*, 1–18.

Anis, K., Zakia, H., Mohamed, D., & Jeffrey, C. (2018). *Detecting depression severity by interpretable representations of motion dynamics*. 13th IEEE

International Conference on Automatic Face & Gesture Recognition (FG 2018), IEEE, pp. 739–745.

Asgari, M., Shafran, I., & Sheeber, L. B. (2014). *Inferring clinical depression from speech and spoken utterances*. IEEE International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, pp. 1–5.

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). *Openface 2.0: Facial behavior analysis toolkit*. 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), IEEE, pp. 59–66.

Bao, H., & Ma, T. (2014). *Feature extraction and facial expression recognition based on bezier curve*. IEEE International Conference on Computer and Information Technology, IEEE, pp. 884–887.

Beretta, L., & Santaniello, A. (2016). Nearest neighbor imputation algorithms: A critical evaluation. *BMC Medical Informatics and Decision Making*, *16* (Suppl. 3), 74–74.

Bernard, J. A., & Mittal, V. A. (2015). Updating the research domain criteria: The utility of a motor dimension. *Psychological Medicine*, *45*, 2685–2689.

Bhatia, S., Goecke, R., Hammal, Z., & Cohn, J. F. (2019). *Automated measurement of head movement synchrony during dyadic depression severity interviews*. 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), IEEE, pp. 1–8.

Blechert, J., Michael, T., & Wilhelm, F. H. (2013). Video-based analysis of bodily startle and subsequent emotional facial expression in posttraumatic stress disorder. *Journal of Experimental Psychopathology*, *4*, 435–447.

Bradski, G., & Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. Sebastopol, CA: O'Reilly Media, Inc.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*, 123–140.

Breznitz, Z. (1992). Verbal indicators of depression. *The Journal of General Psychology*, *119*, 351–363.

Buolamwini, J., & Gebru, T. (2018). *Gender shades: Intersectional accuracy disparities in commercial gender classification*. Conference on Fairness, Accountability and Transparency, pp. 77–91.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science (New York, N.Y.)*, *356*, 183–186.

Calvo, R. A., Milne, D. N., Hussain, M. S., & Christensen, H. (2017). Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering*, *23*, 649–685.

Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., & Snyder, P. J. (2004). Voice acoustical measurement of the severity of major depression. *Brain and Cognition*, *56*, 30–35.

Carmi, L., Schultebraucks, K., & Galatzer-Levy, I. (2020). Identification, prediction, and intervention via remote digital technology: Digital phenotyping & deployment of clinical interventions following terror and mass casualty events. In E. Vermetten, I. Frankova, L. Carmi, O. Chaban & J. Zohar (Eds.), *Management of terrorism induced stress – guideline for the golden hours* (pp. 175–181). Amsterdam: IOS Press BV.

Chollet, F. (2018). Keras: The python deep learning library. Astrophysics Source Code Library.

Cohn, J. F., Kruez, 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*. 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009, IEEE, pp. 1–7.

Cummins, N., Epps, J., Breakspear, M., & Goecke, R. (2011). *An investigation of depressed speech detection: Features and normalization*. Twelfth Annual Conference of the International Speech Communication Association.

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015a). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, *71*, 10–49.

Cummins, N., Sethu, V., Epps, J., Schnieder, S., & Krajewski, J. (2015b). Analysis of acoustic space variability in speech affected by depression. *Speech Communication*, *75*, 27–49.

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, *11*, 126.

Darwin, C. (1872/1965). *The expression of the emotions in man and animals*. Chicago: University of Chicago Press.

Davies, H., Wolz, I., Leppanen, J., Fernandez-Aranda, F., Schmidt, U., & Tchanturia, K. (2016). Facial expression to emotional stimuli in non-psychotic disorders: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *64*, 252–271.

De Boer, P.-T., Kroese, D. P., Mannor, S., & Rubinstein, R. Y. (2005). A tutorial on the cross-entropy method. *Annals of Operations Research*, *134*, 19–67.

Dibeklioğlu, H., Hammal, Z., & Cohn, J. F. (2017). Dynamic multimodal measurement of depression severity using deep autoencoding. *IEEE Journal of Biomedical and Health Informatics*, *22*, 525–536.

Eaton, W. W., Smith, C., Ybarra, M., Muntaner, C., & Tien, A. (2004). Center for Epidemiologic Studies Depression Scale: Review and Revision (CESD and CESD-R). In Maruish E (Ed.), *The use of psychological testing for treatment planning and outcomes assessment: Instruments for adults* (pp. 363–377). Lawrence Erlbaum Associates Publishers.

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoțiuc-Pietro, D., … Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, 11203–11208.

Ekman, P. (2006). *Darwin and facial expression: A century of research in review*. Los Alton, CA: Malor Books.

Ekman, P., Freisen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology*, *39*, 1125.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: Investigator's guide*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Matsumoto, D., & Friesen, W. V. (1997). Facial expression in affective disorders. In Ekman E, & Rosenberg EL (Eds.), *Series in affective science. What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS)* (pp. 429–440). Oxford University Press.

Felmingham, K. L., Rennie, C., Manor, B., & Bryant, R. A. (2011). Eye tracking and physiological reactivity to threatening stimuli in posttraumatic stress disorder. *Journal of Anxiety Disorders*, *25*, 668–673.

Foa, E., Huppert, J., & Cahill, S. (2006). *Pathological anxiety: Emotional processing in etiology and treatment*. New York, NY: Guilford Press.

France, D. J., Shiavi, R. G., Silverman, S., Silverman, M., & Wilkes, D. M. (2000). Acoustical properties of speech as indicators of depression and suicidal risk. *IEEE Transactions on Biomedical Engineering*, *47*, 829–837.

Gaebel, W., & Wölwer, W. (1992). Facial expression and emotional face recognition in schizophrenia and depression. *European Archives of Psychiatry and Clinical Neuroscience*, *242*, 46–52.

Galatzer-Levy, I. R., & Bryant, R. A. (2013). 636120 Ways to have posttraumatic stress disorder. *Perspectives on Psychological Science*, *8*, 651–662.

Gehricke, J.-G., & Shapiro, D. (2000). Reduced facial expression and social context in major depression: Discrepancies between facial muscle activity and self-reported emotion. *Psychiatry Research*, *95*, 157–167.

Geier, T. J., Hunt, J. C., Nelson, L. D., Brasel, K. J., & de Roon-Cassini, T. A. (2019). Detecting PTSD in a traumatically injured population: The diagnostic utility of the PTSD Checklist for DSM-5. *Depression and Anxiety*, *36*, 170–178.

Girard, J. M., Cohn, J. F., Mahoor, M. H., Mavadati, S., & Rosenwald, D. P. (2013). *Social risk and depression: Evidence from manual and automatic facial expression analysis*. 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), IEEE, pp. 1–8.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*. Cambridge: MIT press.

Grother, P. J., Ngan, M., & Hanaoka, K. (2020). Face recognition vendor test (FRVT). Part 2: Identification. *National Institute of Standards and Technology* (*NIST*) *Interagency Report 8271*. U.S. Department of Commerce.

Hahnloser, R. H., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., & Seung, H. S. (2000). Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, *405*, 947.

Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., … Coates, A. (2014). Deep speech: Scaling up end-to-end speech recognition. *arXiv preprint arXiv*:1412.5567.

Hasson, U., Nastase, S. A., & Goldstein, A. (2020). Direct fit to nature: An evolutionary perspective on biological and artificial neural networks. *Neuron*, *105*, 416–434.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York, NY: Springer Science & Business Media.

He, Q., Veldkamp, B. P., & de Vries, T. (2012). Screening for posttraumatic stress disorder using verbal features in self narratives: A text mining approach. *Psychiatry Research*, *198*, 441–447.

He, Q., Veldkamp, B. P., Glas, C. A., & de Vries, T. (2017). Automated assessment of patients' self-narratives for posttraumatic stress disorder screening using natural language processing and text mining. *Assessment*, *24*, 157–172.

Henry, S. K., Grant, M. M., & Cropsey, K. L. (2018). Determining the optimal clinical cutoff on the CES-D for depression in a community corrections sample. *Journal of Affective Disorders*, *234*, 270–275.

Hönig, F., Batliner, A., Nöth, E., Schnieder, S., & Krajewski, J. (2014). *Automatic modelling of depressed speech: relevant features and relevance of gender*. Fifteenth Annual Conference of the International Speech Communication Association.

Huckvale, K., Venkatesh, S., & Christensen, H. (2019). Toward clinical digital phenotyping: A timely opportunity to consider purpose, quality, and safety. *NPJ Digital Medicine*, *2*, 1–11.

Insel, T. R. (2014). The NIMH research domain criteria (RDoC) project: Precision medicine for psychiatry. *American Journal of Psychiatry*, *171*, 395–397.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., … Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry*, *167*, 748–751. doi:10.1176/appi.ajp.2010.09091379.

Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, *71*, 1–15.

Kirsch, A., & Brunnhuber, S. (2007). Facial expression and experience of emotions in psychodynamic interviews with patients with PTSD in comparison to healthy subjects. *Psychopathology*, *40*, 296–302.

Kiss, G., Tulics, M. G., Sztahó, D., Esposito, A., & Vicsi, K. (2016). Language independent detection possibilities of depression by speech. In Esposito A (Ed.), *Recent advances in nonlinear speech processing. Smart Innovation, Systems and Technologies* (Vol. *48*, pp. 103–114). Cham.

Klare, B. F., Burge, M. J., Klontz, J. C., Bruegge, R. W. V., & Jain, A. K. (2012). Face recognition performance: Role of demographic information. *IEEE Transactions on Information Forensics and Security*, *7*, 1789–1801.

Kleim, B., Horn, A. B., Kraehenmann, R., Mehl, M. R., & Ehlers, A. (2018). Early linguistic markers of trauma-specific processing predict post-trauma adjustment. *Frontiers in Psychiatry*, *9*, 1–7.

Kohler, C. G., Martin, E. A., Milonova, M., Wang, P., Verma, R., Brensinger, C. M., … Gur, R. C. (2008). Dynamic evoked facial expressions of emotions in schizophrenia. *Schizophrenia Research*, *105*, 30–39.

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28*(5), 1–26. doi:http://dx.doi.org/10.18637/jss.v028.i05.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. New York, NY: Springer.

Lang, P. J. (1979). A bio-informational theory of emotional imagery. *Psychophysiology*, *16*, 495–512.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444.

Leff, J., & Abberton, E. (1981). Voice pitch measurements in schizophrenia and depression. *Psychological Medicine*, *11*, 849–852.

Litz, B. T., Orsillo, S. M., Kaloupek, D., & Weathers, F. (2000). Emotional processing in posttraumatic stress disorder. *Journal of Abnormal Psychology*, *109*, 26.

Lu, H., Frauendorfer, D., Rabbi, M., Mast, M. S., Chittaranjan, G. T., Campbell, A. T., … Choudhury, T. (2012). *Stresssense: Detecting stress in unconstrained acoustic environments using smartphones*. Proceedings of the 2012 ACM conference on ubiquitous computing, ACM, pp. 351–360.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*, 4765–4774.

Marge, M., Banerjee, S., & Rudnicky, A. I. (2010). *Using the Amazon Mechanical Turk for transcription of spoken language*. IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, pp. 5270–5273.

Marmar, C. R., Brown, A. D., Qian, M., Laska, E., Siegel, C., Li, M., … Smith, J. (2019). Speech-based markers for posttraumatic stress disorder in US veterans. *Depression and Anxiety*, *36*, 607–616.

Mathews, S. C., McShea, M. J., Hanley, C. L., Ravitz, A., Labrique, A. B., & Cohen, A. B. (2019). Digital health: A path to validation. *NPJ Digital Medicine*, *2*, 1–9.

McNally, R. J., Otto, M. W., & Hornig, C. D. (2001). The voice of emotional memory: Content-filtered speech in panic disorder, social phobia, and major depressive disorder. *Behaviour Research and Therapy*, *39*, 1329–1337.

McTeague, L. M., Lang, P. J., Laplante, M.-C., Cuthbert, B. N., Shumen, J. R., & Bradley, M. M. (2010). Aversive imagery in posttraumatic stress disorder: Trauma recurrence, comorbidity, and physiological reactivity. *Biological Psychiatry*, *67*, 346–356.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*, 195–244.

Mowery, D., Smith, H. A., Cheney, T., Bryan, C., & Conway, M. (2016). Identifying depression-related tweets from Twitter for public health monitoring. *Online Journal of Public Health Informatics*, *8*, 1.

Nilsonne, Å. (1987). Acoustic analysis of speech variables during depression and after improvement. *Acta Psychiatrica Scandinavica*, *76*, 235–245.

Nilsonne, A. (1988). Speech characteristics as indicators of depressive illness. *Acta Psychiatrica Scandinavica*, *77*, 253–263.

Nilsonne, Å, Sundberg, J., Ternström, S., & Askenfelt, A. (1988). Measuring the rate of change of voice fundamental frequency in fluent speech during mental depression. *The Journal of the Acoustical Society of America*, *83*, 716–728.

Otte, C., Gold, S. M., Penninx, B. W., Pariante, C. M., Etkin, A., Fava, M., … Schatzberg, A. F. (2016). Major depressive disorder. *Nature Reviews Disease Primers*, *2*, 16065.

Ozdas, A., Shiavi, R. G., Silverman, S. E., Silverman, M. K., & Wilkes, D. M. (2004). Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk. *IEEE Transactions on Biomedical Engineering*, *51*, 1530–1540.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*, 547–577.

Pestian, J., Nasrallah, H., Matykiewicz, P., Bennett, A., & Leenaars, A. (2010). Suicide note classification using natural language processing: A content analysis. *Biomedical Informatics Insights*, *3*, 19–28.

Porritt, L. L., Zinser, M. C., Bachorowski, J.-A., & Kaplan, P. S. (2014). Depression diagnoses and fundamental frequency-based acoustic cues in maternal infant-directed speech. *Language Learning and Development*, *10*, 51–67.

Quatieri, T. F., & Malyska, N. (2012). *Vocal-source biomarkers for depression: A link to psychomotor activity*. Thirteenth Annual Conference of the International Speech Communication Association.

Renneberg, B., Heyn, K., Gebhard, R., & Bachmann, S. (2005). Facial expression of emotions in borderline personality disorder and depression. *Journal of Behavior Therapy and Experimental Psychiatry*, *36*, 183–196.

Rodin, R., Bonanno, G. A., Rahman, N., Kouri, N. A., Bryant, R. A., Marmar, C. R., & Brown, A. D. (2017). Expressive flexibility in combat veterans with posttraumatic stress disorder and depression. *Journal of Affective Disorders*, *207*, 236–241.

Rude, S., Gortner, E.-M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, *18*, 1121–1133.

Scherer, S., Lucas, G. M., Gratch, J., Rizzo, A. S., & Morency, L.-P. (2015). Self-reported symptoms of depression and PTSD are associated with reduced vowel space in screening interviews. *IEEE Transactions on Affective Computing*, *7*, 59–73.

Scherer, S., Stratou, G., Gratch, J., & Morency, L.-P. (2013). Investigating voice quality as a speaker-independent indicator of depression and PTSD. In *Proceedings of Interspeech* (pp. 847–851). Lyon, France: ISCA.

Schultebraucks, K., & Galatzer-Levy, I. R. (2019). Machine learning for prediction of posttraumatic stress and resilience following trauma: An overview of basic concepts and recent advances. *Journal of Traumatic Stress*, *32*, 215–225. doi:10.1002/jts.22384 .

Schultebraucks, K, Shalev, AY, & Michopoulos, V, et al. (2020). A validated predictive algorithm of post-traumatic stress course following emergency department admission after a traumatic stressor. *Nat Med*, *26*, 1084–1088. https://doi.org/10.1038/s41591-020-0951-z.

Shah, Z. S., Sidorov, K., & Marshall, D. (2017). *Psychomotor cues for depression screening*. 22nd International Conference on Digital Signal Processing (DSP), IEEE, pp. 1–5.

Shalev, A., Liberzon, I., & Marmar, C. (2017). Post-traumatic stress disorder. *New England Journal of Medicine*, *376*, 2459–2469.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*, 289–310.

Simon, D., Craig, K. D., Gosselin, F., Belin, P., & Rainville, P. (2008). Recognition and discrimination of prototypical dynamic expressions of pain and emotions. *Pain*, *135*, 55–64.

Sloan, D. M., Strauss, M. E., Quirk, S. W., & Sajatovic, M. (1997). Subjective and expressive emotional responses in depression. *Journal of Affective Disorders*, *46*, 135–141.

Sobin, C., & Sackeim, H. A. (1997). Psychomotor symptoms of depression. *American Journal of Psychiatry*, *154*, 4–17.

Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research diagnostic criteria: Rationale and reliability. *Archives of General Psychiatry*, *35*, 773–782.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*, 111–133.

Sturim, D., Torres-Carrasquillo, P. A., Quatieri, T. F., Malyska, N., & McCree, A. (2011). *Automatic detection of depression in speech using gaussian mixture modeling with factor analysis*. Twelfth Annual Conference of the International Speech Communication Association.

Syed, Z. S., Sidorov, K., & Marshall, D. (2017). *Depression severity prediction based on biomarkers of psychomotor retardation*. Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge, pp. 37–43.

Teasdale, G., Maas, A., Lecky, F., Manley, G., Stocchetti, N., & Murray, G. (2014). The Glasgow Coma Scale at 40 years: Standing the test of time. *The Lancet Neurology*, *13*, 844–854.

Valiant, L. G. (1984). A theory of the learnable. *Communications of the ACM*, *27*, 1134–1142.

van den Broek, E. L., van der Sluis, F., & Dijkstra, T. (2010). Telling the story and re-living the past: How speech analysis can reveal emotions in post-traumatic stress disorder (PTSD) patients. In Westerink J, Krans M, & Ouwerkerk M (Eds.), *Sensing emotions*. Philips Research Book Series, (Vol. *12*, pp. 153–180). Dordrecht: Springer.

Wang, G. (2016). *Facial expression recognition method based on Zernike moments and MCE based HMM*. 9th International Symposium on Computational Intelligence and Design (ISCID), IEEE, pp. 408–411.

Weathers, F. W. (2017). Redefining posttraumatic stress disorder for DSM-5. *Current Opinion in Psychology*, *14*, 122–126.

Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2013). The PTSD Checklist for DSM-5 (PCL-5). *Scale available from the National Center for PTSD at* http://www.ptsd.va.gov.

Xing, Y., & Luo, W. (2016). *Facial expression recognition using local Gabor features and adaboost classifiers*. International Conference on Progress in Informatics and Computing (PIC), IEEE, pp. 228–232.

Xu, R., Mei, G., Zhang, G., Gao, P., Judkins, T., Cannizzaro, M., & Li, J. (2012). A voice-based automated system for PTSD screening and monitoring. In MMVR, pp. 552–558.

Yang, Y., Fairbairn, C., & Cohn, J. F. (2013). Detecting depression severity from vocal prosody. *IEEE Transactions on Affective Computing*, *4*, 142–150.

Zhong, S., Chen, Y., & Liu, S. (2014). *Facial expression recognition using local feature selection and the extended nearest neighbor algorithm*. Seventh International Symposium on Computational Intelligence and Design, IEEE, pp. 328–331.