Prediction of Depression Symptoms in Individual Subjects with Face and Eye Movement Tracking

Supplementary Materials

Aleks Stolicyn[12], J. Douglas Steele[3], Peggy Seriès[2]

1. Division of Psychiatry
   Centre for Clinical Brain Sciences, University of Edinburgh
   Kennedy Tower, Royal Edinburgh Hospital
   Morningside Park, Edinburgh EH10 5HF, UK.

2. Institute for Adaptive and Neural Computation
   School of Informatics, University of Edinburgh
   10 Crichton Street, Edinburgh EH8 9AB, UK.

3. Division of Imaging Science and Technology
   School of Medicine, Dundee University
   Ninewells Hospital & Medical School, Dundee DD1 9SY, UK.

Corresponding author:

Aleks Stolicyn
Division of Psychiatry, University of Edinburgh
Kennedy Tower, Royal Edinburgh Hospital
Morningside Park, Edinburgh EH10 5HF, UK.
Email: a.stolicyn@ed.ac.uk

S1. Background and Rationale

S1.1 Face and Eye Movement Tracking Rationale

Although in our study we do not claim that face and eye movement tracking is the best alternative for detecting depression, there are several arguments – from both theoretical and practical perspectives – which support these methods. We provide them below.

1) Some of the alternative methods for non-invasive assessment include pupillometry, electroencephalography (EEG), facial electromyography (EMG), skin conductance response (SCR), heart rate measurement and cognitive testing. Although some evidence indicates changes in pupil reactivity in depression (e.g. Siegle et al., 2001, 2011), pupillometers are relatively expensive and may not be as accessible as other methods. EEG and facial EMG can be less expensive, and EEG / MEG has been shown to have potential for classifying depression (e.g. Lu, Jiang, Bi, Liu, & Yao, 2014; Tekin Erguzel, Tas, & Cebi, 2015). These techniques, however, are more logistically complicated as they require precise placement of electrodes either on the scalp or on the face, and require specific training. SCR and heart rate measurement have been used to show changes in emotional reactivity in depression (e.g. Garcia et al., 2016; Rottenberg et al., 2005; Schneider et al., 2012), but little evidence indicates that these measures are sufficient on their own for diagnostic classification. Finally, cognitive testing is arguably the cheapest, and ample studies indicate cognitive deficits in depression (Rock, Roiser, Riedel, & Blackwell, 2014; Snyder, 2013), however again no evidence shows that these measures are sufficient on their own to accurately classify depression. Compared to the above methods, face-tracking and eye-tracking are 1) Easier to set up and administer, 2) Less expensive, dependent on the applied eye-tracker (possibly with an exception for heart rate measurement and cognitive testing), and 3) Provide potentially sufficient data to

classify depression, based on some of the previous literature (e.g. Alghowinem et al., 2013; Pampouchidou et al., 2019).

2) On the neural level, some of the brain areas involved in producing involuntary facial expressions include the amygdala, cingulate cortex, and the wider limbic system (Müri, 2016). Eye movements are controlled by an array of prefrontal areas including the frontal eye fields and the dorsolateral prefrontal cortex (Vernet, Quentin, Chanes, Mitsumasu, & Valero-Cabré, 2014). Notably, activity in these areas – the amygdala, cingulate and prefrontal cortices – is altered in depression, considered to be related to both emotional and cognitive aspects of the disorder (e.g. Disner et al., 2011; Drevets et al., 2008; Roiser & Sahakian, 2013). From a theoretical perspective, it is therefore plausible that the main neurocognitive changes in depression may be related directly to altered face and eye movements on the behavioural level.

3) Existing evidence indicates altered face movements (e.g. Falkenberg et al., 2012; Girard et al., 2013; Mergl et al., 2005; Rottenberg et al., 2005) and eye movements (Armstrong & Olatunji, 2012; Carvalho et al., 2015) in depression, with some support that these measures can be used for diagnostic classification in clinical interview settings (Alghowinem et al., 2013; Pampouchidou et al., 2019). This motivates investigation of feasibility of such measures in other more standardised settings (such as cognitive assessment).

S1.2 Cognitive Testing Rationale

There were three main factors which motivated combining face and eye movement tracking with cognitive tasks in our study. We briefly outline them below.

1) Classification of depression based on face and eye movements has previously been attempted in clinical interview settings (Alghowinem et al., 2013; Pampouchidou et al.,

2019). Although these attempts have been relatively successful, clinical interview settings allow for additional degrees of freedom for participant behaviour. For example, participant's face movements may be differentially influenced by 1) The social situation, i.e. having to answer face-to-face questions related to mental wellbeing, and 2) The way the questions are delivered by the interviewer. In contrast, cognitive testing is standardised and avoids impact of these factors on behaviour, and thus should make any behavioural signs related to depression more robust across different experimental settings.

2) Ample evidence supports presence of cognitive deficits in depression, with affected cognitive domains including attention and working memory (McIntyre et al., 2013; Rock et al., 2014; Snyder, 2013). Some evidence also indicates altered physiological response to performance feedback (Weinberg, Dieterich, & Riesel, 2015). Based on this evidence we hypothesised that performance at cognitive tests in depression, especially in the presence of affective distractions, may be characterised also by altered face and eye movements, which may in turn be useful for diagnostic classification.

3) During the cognitive tests the participants' attention is typically engaged – they have to either maintain information in working memory or follow and assess the stimuli which appear on the screen. This attentional engagement makes it more difficult to simultaneously maintain voluntary control over facial expressions, and makes it more likely that any face movements which occur during the task performance are involuntary. We hypothesised that because some brain areas which affect involuntary facial expression are affected in depression (Disner et al., 2011; Müri, 2016; Roiser & Sahakian, 2013), face movements during cognitive performance might also be altered, which may represent measurable diagnostic signs.

S1.3 Cost Comparison

A typical used or refurbished 3 Tesla (3T) scanner can cost between $200K and $700K (or between €200K and €500K in Europe), while new 3T scanners can be up to five times more expensive and cost between $1M and $2.1M (Hough, 2019; LBN Medical Team, 2019). 1.5T scanners are less expensive (used cost $175K – $350K), but provide lower resolution which may impact diagnostic accuracy. Additional costs come from allocating and maintaining a sufficiently large space for keeping the scanner. A single structural MRI scan is estimated to cost between £110 and £160 ($125 to $185) within the National Health Service in the UK (NHS Improvement, 2019), but can be up to £400 ($460) or higher from private healthcare providers (e.g. Spire Healthcare, 2020). This typically includes the radiographer's time, administrative costs (e.g. ensuring scan safety and absence of any metallic items on the patient), and contribution towards essential maintenance of the scanner. In the United States, a medical MRI scan can cost between $250 and $4000, depending on the provider, with an average cost estimated at around $2600 (GE Healthcare, 2019). In research settings in the UK, a scan with multiple modalities (e.g. functional MRI and diffusion-tensor imaging together with T1 structural) can cost up to £500 or more.

In contrast to MRI, the technical setup in our study can be deployed rapidly in any quiet room with sufficiently good lighting; deployment typically takes less than one day. Equipment in our study consisted of three components: computing workstation unit with a 21.5" monitor and keyboard (approx. $800 / £700), a Gazepoint GP3 eye-tracker ($695 / £600), and an Intel RealSense SR300 colour camera ($175 / £150) – with a total cost of £1450 ($1660). Running the system does not require significant training for the experimenter, apart from becoming familiar with the assessment sequence and instructions to the participants, and learning to adjust the eye-tracker for individual participants. The cost of a single assessment could be estimated at approximately £30 ($35), based on commitment of 1

and ½ hours of a research assistant time with an annual salary of £29K ($33K), including a small contribution towards electricity and maintenance of the experiment space (room).

S2. Cognitive Task Details

S2.1 Delayed Match to Sample Task Trials

Each trial at the DMS task started with a fixation cross at the centre of the screen which lasted 1500 ms. This was followed by presentation of the *sample* pattern, consisting of four coloured quadrants with variable numbers of white marks in each quadrant. The sample was present on the screen for 2500 ms. The sample was followed by a delay with a blank screen, lasting 1000 ms. After the delay, four *distraction* words flashed on the screen, one after another, at random locations. Each word was equidistant from the previous one and faded away for 2500 ms after an initial flash. Distraction words were followed by a second 1000 ms blank-screen delay. The delay / distractions stage lasted 12000 ms in total (2 x 1000 ms + 4 x 2500 ms). After the distractions stage, four patterns appeared on the screen and the participant was required to select the initial sample, using one of the four marked keys on the keyboard. There was no time limit for response. An additional distraction word was present in the lower part of the screen during the selection stage. Trial feedback was presented 500 ms after the response and lasted 2500 ms. The feedback consisted of a single word ‘Correct’ or ‘Wrong’, displayed at the centre of the screen. Figure 1A in the main text illustrates the task trial structure, Figure S1 illustrates patterns at the selection stage of the task.

To be 100% correct at a single DMS trial, the participant had to memorise colours and numbers of marks in at least two quadrants of the sample pattern. The pattern composition strategy was similar to the one in the original DMS task (Owen, Sahakian, Semple, Polkey, & Robbins, 1995; Sahakian et al., 1988). Specifically, one pattern at the selection stage had different colours and different numbers of marks from the sample. One pattern had different

colours but same marks in quadrants, and one had same colours but different marks. All four patterns shared one quadrant which had the same colour and the same number of marks. The task is a more standardised version of the original DMS because random abstract pattern shapes (distortions) are replaced with specific numbers of marks in each quadrant (between one and six).

S2.2 Rapid Detection Task Trials

Each trial in the RD task started with a fixation cross at the centre of the screen, displayed for 2000 ms. After the fixation cross, instruction for a target letter was displayed for 2500 ms. The target instruction consisted of the words ‘Target letter is’ and one of the eight possible target letters. The eight letters used in the task were ‘E’, ‘F’, ‘C’, ‘G’, ‘H’, ‘K’, ‘B’, and ‘R’. The letters were chosen due to pair-wise visual similarity. The instruction was followed by a 2000 ms delay with a fixation cross, after which five distraction words flashed at the centre of the screen, one after another (introductory distraction stage). Each distraction word faded away for 1500 ms. A blank-screen delay of 1000 ms followed the last word. After the delay, 44 letters flashed at the centre of the upper half of the screen, fading away, one after another. The participant was required to press space key when they saw the target letter. Each letter blinked initially, and then faded away for 1000 ms. Of the 44 trial letters, 5 were target, distributed uniformly among others. First four letters were introductory and always non-target. When participant correctly detected the target letter, the letter was highlighted in green and the word ‘Correct’ appeared in the lower half of the screen for 1500 ms. If the participant missed the target letter or made an erroneous response, words ‘Missed’ or ‘Wrong’ appeared instead of the letter for 3000 ms, alongside a reminder of the target letter in the lower half of the screen. Throughout the target detection stage, five distraction words appeared alongside the flashing letters, in the central lower part of the screen. Each word

flashed initially, and stayed on the screen for 5000 ms. Distractions were distributed uniformly throughout the detection stage, with a minimal interval of 1000 ms between any two consecutive words. No two trials had the same target letter in one block of trials. Figure 1B in the main text illustrates the task trial structure, Figure S2 illustrates letter and distraction positions during the detection stage, as well as correct and missed-response feedback screens.

S2.3 Affective Distractions

Distraction words were selected from the Warriner database (Warriner, Kuperman, & Brysbaert, 2013). The database contains 13,915 English lemmas, characterised with valence, arousal, and dominance ratings on a scale from 1 to 9. The words were selected mainly based on their valence ratings and categorical relevance. 60 selected neutral words had mean valence 5.05 and mean arousal 3.45. 60 selected positive words had mean valence 7.66 and mean arousal 5.28. Finally, 60 selected negative words had mean valence 2.27 and mean arousal 5.40. Neutral words were related to categories of physical materials (e.g. words ‘wood’ or ‘stone’), building types, furniture, stationery, or printed materials. Positive words were related to social relationships, career, finance, recreation, or general wellbeing. Negative words were related to categories of work, health conditions, social attitudes, deprivation, or general adversity. Specific applied affective distraction words can be obtained from the study authors.

S3. Recorded Metric Details

S3.1 Face-tracking Metrics

Recordings for the DMS task were segmented to epochs for 7 trial stages: sample stage, delay stage with neutral distractions, delay stage with positive distractions, delay stage with

negative distractions, selection stage, correct feedback, and error feedback. For the RD task, recordings were segmented to epochs for 7 trial stages: target (instruction) stage, neutral, positive, and negative introductory distraction stages, correct feedback, missed feedback, and error feedback. For each participant, epochs were thus grouped into 14 collections, each related to a different task stage.

Because 13 out of the 48 participants (9 symptomatic) did not make any errors at the RD task, and because average error count at the task was less than 2, we excluded face movement features related to *error* feedback at the RD task (but not for *missed*-target feedback). This reduced the set of face-tracking features from 714 to 663. In addition, 9 out of 48 participants did not miss any target letters at the RD task (8 symptomatic and 1 control), while 2 participants did not make any errors at the DMS task (1 symptomatic and 1 control). Face-tracking features respectively for the RD *missed-target feedback* stage or the DMS *error feedback* stage in these cases were replaced with value -1 (methods section in the main text).

S3.2 Eye-tracking Metrics

A 1.4 cm padded eye-tracking area was defined around each element to allow for limited accuracy of the eye-tracker. Eye-tracking area padding was based on the assumption that the eye-tracker accuracy is approximately 1 degree and participant's eyes are 70 cm from the screen. No two elements at the two tasks had overlapping eye-tracking areas.

S4. Classification Method Details

S4.1 Feature Selection

Features were selected within each (LOOCV) cross-validation fold using simple statistical filter based on two-sample *t*-test. *P*-value threshold in each fold was optimised using grid search within a *nested* LOOCV scheme – inner (nested) LOOCV accuracy was

used as the criterion for optimal threshold. Search grid for the *p*-value threshold was coarse to prevent overfitting and consisted of five values between 0.05 and 0.01 with a step of 0.01.

S4.2 Classifier Hyperparameters

Support vector machine (SVM) with a Gaussian (radial basis function) kernel was used in the study. *Regularisation* (box constraint) parameter of the classifier was set to the canonical value of 1. *Kernel scale* parameter was set to 9, following a simple heuristic to approximate the value equal to the square root of the number of features, with an expectation that up to around 10% of features would be selected. No exhaustive hyperparameter optimisation was performed.

S4.3 Additional Classification Models

We applied SVM with a Gaussian kernel in the main classification analysis due to the promising performance of this model in previous studies (Johnston, Tolomeo, et al., 2015; Johnston, Steele, Tolomeo, Christmas, & Matthews, 2015; Mwangi, Ebmeier, Matthews, & Steele, 2012). Additional *post hoc* classification analyses were completed with an SVM with a linear kernel, linear discriminant analysis (LDA, Hastie, Tibshirani, & Friedman, 2009), simple decision tree (Kingsford & Salzberg, 2008) and the robustly boosted decision tree learner (RBT, Freund, 2009).

For SVM with a linear kernel, two values of regularisation parameter were investigated – 0.01 (low) and 1 (standard). For decision tree, maximum number of splits was set to 20, minimum parent size was set to 10, and minimum leaf size was set to 4 – following heuristics for a medium-sized tree. The regularisation parameter in the LDA classification model was specified automatically in MATLAB R2018a as the minimal regularisation value necessary to invert the predictor covariance matrix. For the RBT classifier, the number of learned trees (learners) was set to 10 and each tree had the same constraints as above (maximum number of

splits 20, minimum parent size 10, and minimum leaf size 4).

S5. Behavioural Results

In our study we observed a trend towards better performance at the RD task in symptomatic participants – they tended to detect on average 1.15% more target letters (p = 0.057), and tended to make on average 0.66 fewer errors (p = 0.077, Table S7). This was unexpected as we hypothesised that symptomatic participants may have generally decreased cognitive performance akin to that seen in clinical depression (McIntyre et al., 2013; Rock et al., 2014; Snyder, 2013). Two explanations could be given for these trends. First, our sample was non-clinical and it is possible that moderately elevated symptoms in non-clinical participants (mean CES-D score 26.6, Table 1 in the main text) may lead to marginally improved rather than decreased performance. On the other hand, increased accuracies at attention tasks have previously been reported in some studies of depression (e.g. Chiu & Deldin, 2007; Dillon et al., 2015), and it is possible that the trend towards better performance was due to some aspects of the RD task design. Future studies with clinical participants should clarify whether the trends observed in our study might be related to task design or the participant clinical status.

S6. Classification Results

None of the alternative classification models (section S3.3 above) performed better compared to the SVM with a Gaussian kernel, although performance of SVM with a linear kernel and reduced regularisation was similar (accuracy 77.08%, sensitivity 72%, specificity 82.61%). Accuracies, sensitivities and specificities for each alternative classification model and each feature set (combined, face-tracking features only, or eye-tracking features only) are presented in Table S8.

Table S1

Delayed Match to Sample task eye-tracked elements

DMS trial stage	Eye-tracked elements
Sample stage	Sample pattern (all conditions)
Delay stage	Distraction word (all conditions) Neutral distraction word Positive distraction word Negative distraction word
Selection stage	Sample pattern (all conditions) Non-sample pattern (all conditions) Distraction word (all conditions) Neutral-block sample pattern Neutral-block non-sample pattern Neutral-block distraction word Positive-block sample pattern Positive-block non-sample pattern Positive-block distraction word Negative-block sample pattern Negative-block non-sample pattern Negative-block distraction word

Table S2

Delayed Match to Sample task differences between eye-tracking metric means

DMS trial stage	Eye-tracking metric differences
Delay stage	Positive to neutral distraction word Negative to neutral distraction word
Selection stage	Positive to neutral sample pattern Positive to neutral non-sample pattern Positive to neutral distraction word Negative to neutral sample pattern Negative to neutral non-sample pattern Negative to neutral distraction word

Table S3

Rapid Detection task eye-tracked elements

RD trial stage	Eye-tracked elements
Target stage	Target letter (all conditions)
Introductory distractions stage	Distraction word (all conditions) Neutral distraction word Positive distraction word Negative distraction word
Detection stage	Distraction word (all conditions) Neutral distraction word Positive distraction word Negative distraction word
Feedback stages	Correct feedback target letter Correct feedback notification Missed feedback target letter Missed feedback notification Error feedback target letter Error feedback notification

Table S4

Rapid Detection task differences between eye-tracking metric means

RD trial stage	Eye-tracking metric differences
Introductory distractions stage	Positive to neutral distraction word Negative to neutral distraction word
Detection stage	Positive to neutral distraction word Negative to neutral distraction word

Table S5

Characteristics of the sample used for analysis of behavioural measures

	Group		*P* value
	Control	**Symptomatic**	
Size (male / female)	38 (20 / 18)	34 (16 / 18)	-
Age	25.7 (6.6)	23.4 (3.0)	$p = 0.0564$
NART	35.7 (3.9)	36.9 (3.4)	n.s.
AUDIT	5.9 (4.5)	7.6 (7.1)	n.s.
Caffeine	0.9 (0.8)	1.5 (1.2)	$p = 0.0225$
CES-D	8.8 (3.8)	26.4 (7.0)	$p < 0.00001$

***Note*.** Caffeine is in cups of coffee per day. Standard deviations are in parentheses. P value defined according to two-sample independent t-tests.

Table S6

DMS task mean reaction time and accuracy measures

	Group		***P* value**
	Control	**Depressive**	
Accuracy	90.35% (8.32%)	89.54% (8.25%)	$p = 0.6807$ (n.s.)
Correct reaction time	3575 ms (1045 ms)	3887 ms (1682 ms)	$p = 0.3563$ (n.s.)
Error reaction time	5102 ms (1998 ms)	5323 ms (2529 ms)	$p = 0.6904$ (n.s.)

***Note*.** Standard deviations are in parentheses.

Table S7

RD task mean reaction time and accuracy measures

	Group		***P* value**
	Control	**Depressive**	
Correct reaction time	454 ms (50 ms)	444 ms (35 ms)	$p = 0.3394$ (n.s.)
Detection rate	96.74% (2.92%)	97.88% (2.05%)	$p = 0.0565$ (n.s.)
Error count	1.92 (1.78)	1.26 (1.31)	$p = 0.0769$ (n.s.)

***Note*.** Standard deviations are in parentheses.

Table S8

Accuracies, sensitivities and specificities for alternative tested classification models

	SVM Linear LR	SVM Linear SR	LDA	DT	RBT
Combined features	77.08% Sens. 72% Spec. 82.61%	68.75% Sens. 60% Spec. 78.26%	68.75% Sens. 64% Spec. 73.91%	60.42% Sens. 60% Spec. 60.87%	56.25% Sens. 52% Spec. 60.87%
Face-tracking features	62.50% Sens. 60% Spec. 65.22%	68.75% Sens. 72% Spec. 65.22%	64.58% Sens. 68% Spec. 60.87%	68.75% Sens. 68% Spec. 69.57%	68.75% Sens. 64% Spec. 73.91%
Eye-tracking features	60.42% Sens. 68% Spec. 52.17%	45.83% Sens. 40% Spec. 52.17%	50% Sens. 48% Spec. 52.17%	45.83% Sens. 52% Spec. 39.13%	47.92% Sens. 44% Spec. 52.17%

Abbreviations: SVM Linear – support vector machine with linear kernel; LR – low regularisation; SR – standard regularisation; LDA – linear discriminant analysis; DT – decision tree; RBT – robustly boosted decision tree learner.

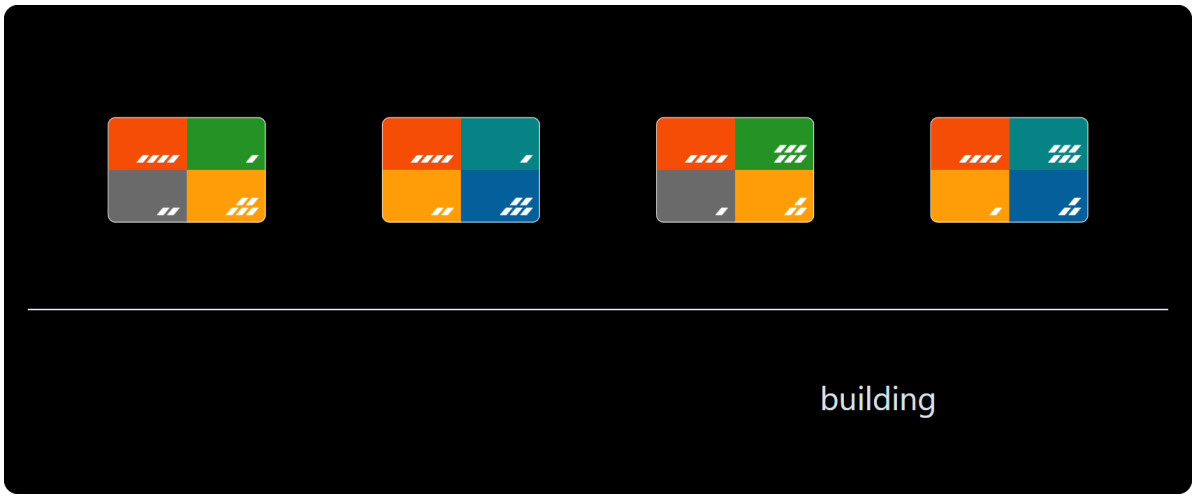

Figure S1. DMS task selection stage example. The participant has to select the initial sample pattern from the four alternatives.

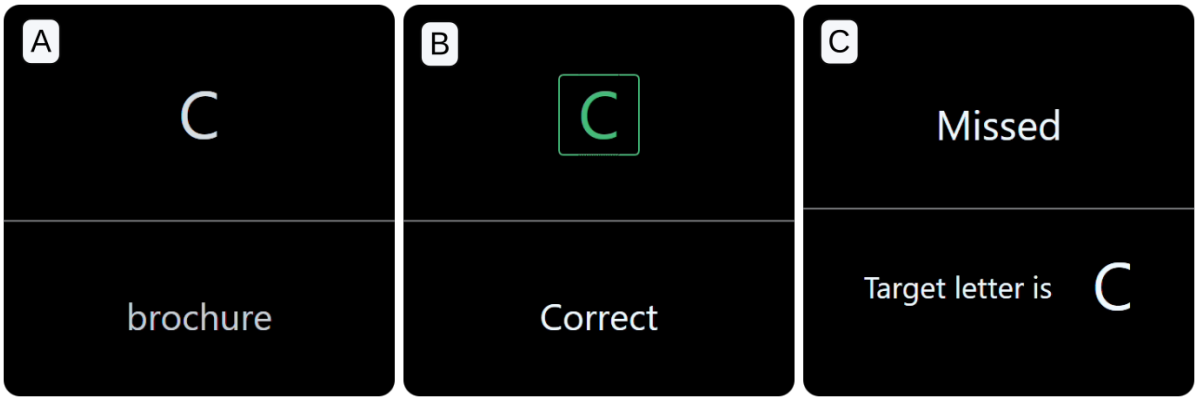

Figure S2. RD task target detection stage illustration. Letters flash in the upper part of the screen, while distraction words appear in the lower part (A). Target letter is highlighted in green when it is correctly detected (B). Reminder appears when a target letter is missed (C), or when an erroneous response is committed (not shown, similar to C).

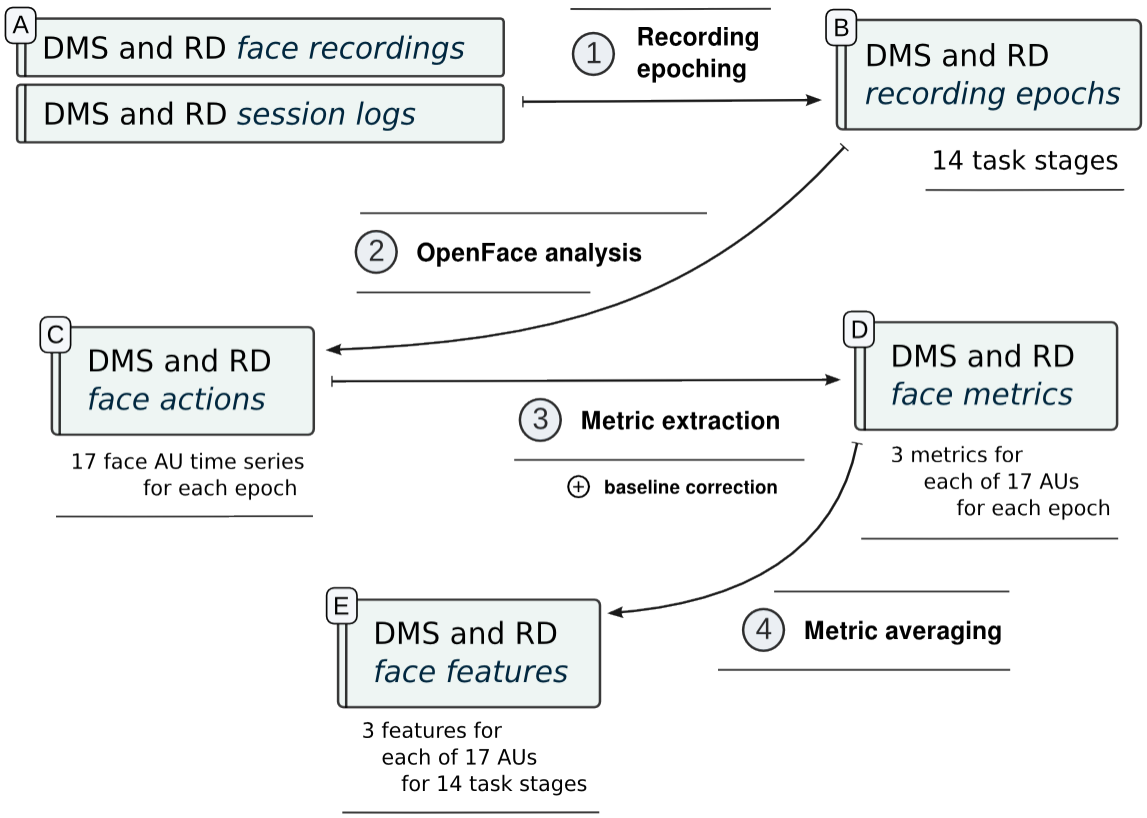

Figure S3. Facial movement feature extraction for a single participant. Facial recordings (A) are first epoched into segments for each task stage (B). OpenFace is then used to extract facial AU time series for each epoch (C). Facial movement metrics (D) are then extracted from each AU time series. Finally, each metric for each task stage is averaged across epochs to obtain facial movement features (E).

SUPPLEMENTARY REFERENCES

Alghowinem, S., Goecke, R., Wagner, M., Parker, G., & Breakspear, M. (2013). Eye movement analysis for depression detection (pp. 4220–4224). IEEE.

Armstrong, T., & Olatunji, B. O. (2012). Eye tracking of attention in the affective disorders: a meta-analytic review and synthesis. *Clinical Psychology Review*, *32*(8), 704–723.

Carvalho, N., Laurent, E., Noiret, N., Chopard, G., Haffen, E., Bennabi, D., & Vandel, P. (2015). Eye Movement in Unipolar and Bipolar Depression: A Systematic Review of the Literature. *Frontiers in Psychology*, *6*, 1809.

Chiu, P. H., & Deldin, P. J. (2007). Neural evidence for enhanced error detection in major depressive disorder. *The American Journal of Psychiatry*, *164*(4), 608–616.

Dillon, D. G., Wiecki, T., Pechtel, P., Webb, C., Goer, F., Murray, L., … Pizzagalli, D. A. (2015). A computational analysis of flanker interference in depression. *Psychological Medicine*, *45*(11), 2333–2344.

Disner, S. G., Beevers, C. G., Haigh, E. A. P., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews. Neuroscience*, *12*(8), 467–477.

Drevets, W. C., Savitz, J., & Trimble, M. (2008). The subgenual anterior cingulate cortex in mood disorders. *CNS Spectrums*, *13*(8), 663–681.

Falkenberg, I., Kohn, N., Schoepker, R., & Habel, U. (2012). Mood induction in depressive patients: a comparative multidimensional approach. *PloS One*, *7*(1), e30016.

Freund, Y. (2009). A more robust boosting algorithm. Retrieved from https://arxiv.org/abs/0905.2138

Garcia, R. G., Valenza, G., Tomaz, C. A., & Barbieri, R. (2016). Relationship between cardiac vagal activity and mood congruent memory bias in major depression. *Journal of Affective Disorders*, *190*, 19–25.

GE Healthcare. (2019, May 13). How Much Does An MRI Cost? [Company]. Retrieved from https://www.gehealthcare.co.uk/feature-article/how-much-does-an-mri-cost

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. *Proceedings of the ... International Conference on Automatic Face and Gesture Recognition. International Conference on Automatic Face and Gesture Recognition*, 1–8.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. New York, NY: Springer New York.

Hough, J. (2019, January 25). MRI Machine Buyers Guide: Options and Pricing [Company]. Retrieved from https://www.meridianleasing.com/blog/medical-equipment-blog/mri-machine-buyers-guide-options-and-pricing

Johnston, B. A., Steele, J. D., Tolomeo, S., Christmas, D., & Matthews, K. (2015). Structural MRI-Based Predictions in Patients with Treatment-Refractory Depression (TRD). *PloS One*, *10*(7), e0132958.

Johnston, B. A., Tolomeo, S., Gradin, V., Christmas, D., Matthews, K., & Steele, J. D. (2015). Failure of hippocampal deactivation during loss events in treatment-resistant depression. *Brain: A Journal of Neurology*, *138*(Pt 9), 2766–2776.

Kingsford, C., & Salzberg, S. L. (2008). What are decision trees? *Nature Biotechnology*, *26*(9), 1011–1013.

LBN Medical Team. (2019, May 1). How Much Does an MRI Machine Cost? [Company]. Retrieved from https://lbnmedical.com/how-much-does-an-mri-machine-cost/

Lu, Q., Jiang, H., Bi, K., Liu, C., & Yao, Z. (2014). Discriminative analysis with a limited number of MEG trials in depression. *Journal of Affective Disorders*, *167*, 207–214.

McIntyre, R. S., Cha, D. S., Soczynska, J. K., Woldeyohannes, H. O., Gallaugher, L. A., Kudlow, P., … Baskaran, A. (2013). Cognitive deficits and functional outcomes in major depressive disorder: determinants, substrates, and treatment interventions. *Depression and Anxiety*, *30*(6), 515–527.

Mergl, R., Mavrogiorgou, P., Hegerl, U., & Juckel, G. (2005). Kinematical analysis of emotionally induced facial expressions: a novel tool to investigate hypomimia in patients suffering from depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *76*(1), 138–140.

Müri, R. M. (2016). Cortical control of facial expression. *The Journal of Comparative Neurology*, *524*(8), 1578–1585.

Mwangi, B., Ebmeier, K. P., Matthews, K., & Steele, J. D. (2012). Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder. *Brain: A Journal of Neurology*, *135*(Pt 5), 1508–1521.

NHS Improvement. (2019, December 19). National Tariff Payment System 2020/21 [Public]. Retrieved from https://improvement.nhs.uk/resources/national-tariff-2021-consultation/

Owen, A. M., Sahakian, B. J., Semple, J., Polkey, C. E., & Robbins, T. W. (1995). Visuo-spatial short-term recognition memory and learning after temporal lobe excisions, frontal lobe excisions or amygdalo-hippocampectomy in man. *Neuropsychologia*, *33*(1), 1–24.

Pampouchidou, A., Simos, P. G., Marias, K., Meriaudeau, F., Yang, F., Pediaditis, M., & Tsiknakis, M. (2019). Automatic Assessment of Depression Based on Visual Cues: A Systematic Review. *IEEE Transactions on Affective Computing*, *10*(4), 445–470.

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: a systematic review and meta-analysis. *Psychological Medicine*, *44*(10), 2029–2040.

Roiser, J. P., & Sahakian, B. J. (2013). Hot and cold cognition in depression. *CNS Spectrums*, *18*(3), 139–149.

Rottenberg, J., Gross, J. J., & Gotlib, I. H. (2005). Emotion context insensitivity in major depressive disorder. *Journal of Abnormal Psychology*, *114*(4), 627–639.

Sahakian, B. J., Morris, R. G., Evenden, J. L., Heald, A., Levy, R., Philpot, M., & Robbins, T. W. (1988). A comparative study of visuospatial memory and learning in Alzheimer-type dementia and Parkinson's disease. *Brain: A Journal of Neurology*, *111 (Pt 3)*, 695–718.

Schneider, D., Regenbogen, C., Kellermann, T., Finkelmeyer, A., Kohn, N., Derntl, B., … Habel, U. (2012). Empathic behavioral and physiological responses to dynamic stimuli in depression. *Psychiatry Research*, *200*(2–3), 294–305.

Siegle, G. J., Granholm, E., Ingram, R. E., & Matt, G. E. (2001). Pupillary and reaction time measures of sustained processing of negative information in depression. *Biological Psychiatry*, *49*(7), 624–636.

Siegle, G. J., Steinhauer, S. R., Friedman, E. S., Thompson, W. S., & Thase, M. E. (2011). Remission prognosis for cognitive therapy for recurrent depression using the pupil: utility and neural correlates. *Biological Psychiatry*, *69*(8), 726–733.

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. *Psychological Bulletin*, *139*(1), 81–132.

Spire Healthcare. (2020). MRI scan [Private healthcare company]. Retrieved from https://www.spirehealthcare.com/spire-edinburgh-hospitals-murrayfield-and-shawfair-park/treatments/a-z/mri-scan-magnetic-resonance-imaging-scan/

Tekin Erguzel, T., Tas, C., & Cebi, M. (2015). A wrapper-based approach for feature selection and classification of major depressive disorder-bipolar disorders. *Computers in Biology and Medicine*, *64*, 127–137.

Vernet, M., Quentin, R., Chanes, L., Mitsumasu, A., & Valero-Cabré, A. (2014). Frontal eye field, where art thou? Anatomy, function, and non-invasive manipulation of frontal regions involved in eye movements and associated cognitive operations. *Frontiers in Integrative Neuroscience*, *8*, 66.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207.

Weinberg, A., Dieterich, R., & Riesel, A. (2015). Error-related brain activity in the age of RDoC: A review of the literature. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *98*(2 Pt 2), 276–299.