E-DAIC Depression Database

The Extended Distress Analysis Interview Corpus (E-DAIC) (DeVault et al., 2014) is an extended version of WOZ-DAIC (Gratch et al.,2014) that contains semi-clinical interviews designed to support the diagnosis of psychological distress conditions such as anxiety, depression, and post-traumatic stress disorder. These interviews were collected as part of a larger effort to create a computer agent that interviews people and identifies verbal and nonverbal indicators of mental illnesses.
The interviews are conducted by an animated virtual interviewer called Ellie.
A subset of the sessions are collected in a wizard-of-Oz (WoZ) setting, where the virtual agent is controlled by a human interviewer (wizard) in another room.
The other subset are collected using an AI-controlled agent, who acts in a fully autonomous way using different automated perception and behavior generation modules.

The dataset is partitioned into training, development, and test sets while preserving the overall speaker diversity -- in terms of age, gender distribution, and the eight-item Patient Health Questionnaire (PHQ-8) scores -- within the partitions. Whereas the training and development sets include a mix of WoZ and AI scenarios, the test set is solely constituted from the data collected by the autonomous AI.

Sessions with IDs in the range [300,492] are collected with WoZ-controlled agent and sessions with IDs [600,718] are collected with an AI-controlled agent.

The data includes 219 participant directories, each following the below structure:

XXX_P

- XXX_AUDIO.wav
- XXX_Transcript.csv
- features
    - XXX_BoAW_openSMILE_2.3.0_eGeMAPS.csv
    - XXX_BoAW_openSMILE_2.3.0_MFCC.csv
    - XXX_BoVW_openFace_2.1.0_Pose_Gaze_AUs.csv
    - XXX_CNN_ResNet.mat
    - XXX_CNN_VGG.mat
    - XXX_densenet201.csv

- XXX_OpenFace2.1.0_Pose_gaze_AUs.csv
- XXX_OpenSMILE2.3.0_egemaps.csv
- XXX_OpenSMILE2.3.0_mfcc.csv
- XXX_vgg16.csv

More information regarding each feature set is provided in the table below.

Feature Set	Modality	Feature Type	Description
Bag-of-audio-words eGeMAPS (Schmitt et al, 2017)	Audio	Bag-of-words	eGeMAPS features processed and summarized over a block of 4-second length duration for each step of 1 second
Bag-of-audio-words MFCCs (Schmitt et al, 2017)	Audio	Bag-of-words	MFCC features processed and summarized over a block of 4-second length duration for each step of 1 second
Bag-of-visual-words Pose Gaze AUs (Schmitt et al, 2017)	Visual	Bag-of-words	Pose/Gaze/AU features processed and summarized over a block of 4-second length duration for each step of 1 second
CNN ResNet (He at al, 2016)	Visual	Deep Representations	Aligned face images are fed to the pretrained ResNet-50 model with frozen weights, and the output of the first FC layer is extracted as representation.
CNN VGG (Simonyan et al, 2014)	Visual	Deep Representations	Aligned face images are fed to the pretrained VGG-16 model with frozen weights, and the output of the global average pooling layer is extracted as representation.

Densenet (Huang et al, 2017)	Audio	Deep Representations	The speech files are first transformed into mel-spectrogram images with 128 mel-frequency bands, a window width of 4 seconds and a hop size of 1 second. The spectral-based images are fed to the densenet 201 pretrained network, and a feature vector is obtained from activations of the last average pooling layer of DenseNet.
OpenFace - Pose, Gaze, AUs (Baltrusaitis et al, 2018)	Visual	Expert Knowledge	The intensities of 17 FAUs for each video frame, along with a confidence measure are extracted using OpenFace
extended Geneva Minimalistic Acoustic Parameter Set (eGeMaPS) (Eyben et al, 2016)	Audio	Expert Knowledge	Contains 88 measures covering spectral, cepstral, prosodic, and voice quality information
MFCCs (Eyben et al, 2013)	Audio	Expert Knowledge	MFCCs 1-13, including their first and second order derivatives (deltas and double-deltas) are computed as a set of acoustic LLDs, using the OpenSMILE toolkit
VGG-16 (Simonyan et al, 2014)	Audio	Deep Representations	The speech files are first transformed into mel-spectrogram images with 128 mel-frequency bands, a window width of 4 seconds and a hop size of 1 second. The spectral-based images are fed to the densenet 201 pretrained network, and a feature vector is obtained from activations of the second fully connected layer in VGG16.

The Train/dev/test splits are also provided in the labels directory. Each split file includes:

Participant_ID, gender, PHQ_Binary, PHQ_Score, PCL-C(PTSD), PTSD Severity

Additionally, the file Detailed_PHQ8_Labels.csv includes detailed answers to each question on the PHQ8 questionnaire. The detailed PHQ scores are the responses given to every single question on the PHQ8 questionnaire. This is useful in case a particular symptom is being studied, for example, you can have the rating given in response to “difficulty sleeping”.