Correlation of Audiovisual Features With Clinical Variables and Neurocognitive Functions in Bipolar Disorder, Mania

01.08.2017

Study Protocol

In order to gather sociodemographic and clinical information, all patients were assessed with semi-structured interviews based on the SKIP-TURK. This form includes: identity, sociodemographic personal and family information, age at disease onset, severity, clinical presentation and used treatments.

Thirty-five male and sixteen female patients who were diagnosed as BD type I , manic episode according to DSM-5 in the mental health service of Erenkoy Mental State Hospital and 24 male and 16 female healthy controls matched for age, gender, educational status were included. During hospitalization in every follow up day (0th- 3rd- 7th- 14th- 28th day) and after discharge on the 3rd month, presence of depressive and manic features for patients was evaluated using YMRS and MADRS.

Video recording is done by a presentation guide including seven tasks such as explaining the reason to come to hospital/participate in the activity, describing happy and sad memories, counting up to thirty, explaining two emotion eliciting pictures. To increase the challenge of automatic discrimination, the Control Subjects were recorded with two additional conditions, where they are asked to portray mania and depression conditions.

To provide a benchmark system and insight about the data, we investigated two different approaches. The first is a direct approach to classify video sessions into BD and normal classes, using audio-visual features. The second is an indirect approach, where we use emotion predictions for Pearson correlation analysis and regression of YMRS drop.

Speech Analysis: Acoustic Feature Extraction

Speech utterances are first separated from the video signal and then a standard set of acoustic descriptors are extracted from the audio signal using the open-source openSMILE tool (Eyben et al, 2013). In our preliminary work, we apply functionals over the whole utterance and compare performances of two sets of functionals. The first set of functionals are those used in INTERSPEECH 2010 baseline features. The second is our proposed set of 10 functionals, which we apply on all LLDs and obtain a supra-segmental feature set of 76 x 10 = 760 dimensions.

Three affect classifiers are trained, one with the original 4-classes, one for valence (showing the pleasantness of emotion) and one for arousal (reflecting the activity/vitality of the speaker). The arousal and valence labels are obtained by clustering the original 4-classes into corresponding binary labels (e.g. Happy to positive valence and high arousal, Anger to negative valence and high arousal). The predictions of these affect classifiers are then used as mid-level features to classify bipolar disorder.

Video Processing and Visual Features

The faces are detected, cropped and registered using the method proposed by Xiong and De la Torre (2013) with additional procrustes analysis for frontalization as in (Kaya et al., 2015). Along with the faces, we record both the original and the aligned landmark points for subsequent geometric feature extraction. Also, we extract appearance descriptors from registered faces, using a pre-trained Deep Convolutional Neural Network (DCNN) fine-tuned on a face based emotion corpus (Kaya et al, 2017). From DCNN, we extracted 4096 dimensional features from a convolutional layer and 7 dimensional features (outputs for 7 basic emotions) from the last layer.

Neurocognitive Evaluation

Cambridge Neurophysiological Assessment Battery (CANTAB) are administered to both groups) to assess neurocognitive functions. For patients both in the manic phase(in the first week of hospitalization) and in the remission (at 3rd month after beginning of mania treatment), tests are applied. Applied four different tests and their outcome measures are explained below :

Cambridge Gambling Test: Deliberation time, quality of decision making , risk taking, aversion delay

Emotion Recognition Test: Total rate of happiness, sadness, anger,disgust, fear, surprise

Stop Signal Test: Rate of succesful stops, reaction time, delay of stop signal, reaction time for stop signal, total of true stops

Rapid Visual Processing: Outcome measures cover latency (speed of response), probability of false alarms and sensitivity.

Study groups' sociodemographic and clinical features are defined. Bipolar manic phase patients' first YMRS ratings are correlated with first week neurocognitive functions of the same group. The comparison of bipolar mania neurocognitive functions with both bipolar remission and healthy controls are evalauted. The voice analysis results correlation with neurocognitive functions in bipolar mania is examined.