Intelligent Advanced User Interfaces for Monitoring Mental Health Wellbeing

Anna Esposito[1], Zoraida Callejas[2], Matthias L. Hemmje[3], Michael Fuchs[4], Mauro N. Maldonato[5], Gennaro Cordasco[1]

[1]Università della Campania "Luigi Vanvitelli", Dept of Psychology and IIASS, Italy, iiass.annaesp@tin.it, gennaro.cordasco@unicampania.it
[2]Dept. Software Engineering, University of Granada, Spain, zoraida@ugr.es
[3] Research Inst. for Telecommunication and Cooperation (FTK), Germany, mhemmje@ftk.de
[4] Wilhelm Büchner University of Applied Science Hessen, Germany, m.fuchs@globit.com
[5]Dipartimento di Neuroscience, Naples University "Federico", Italy, m.maldonato@gmail.com

Abstract. It has become pressing to develop objective and automatic measurements integrated in intelligent diagnostic tools for detecting and monitoring depressive states and enabling an increased precision of diagnoses and clinical decision-makings. The challenge is to exploit behavioral and physiological biomarkers and develop Artificial Intelligent (AI) models able to extract information from a complex combination of signals considered key symptoms. The proposed AI models should be able to help clinicians to rapidly formulate accurate diagnoses and suggest personalized intervention plans ranging from coaching activities (exploiting for example serious games), support networks (via chats, or social networks), and alerts to caregivers, doctors, and care control centers, reducing the considerable burden on national health care institutions in terms of medical, and social costs associated to depression cares.

Keywords: Artificial Intelligence, Customer Care, Biometric Data, Social Signal Processing, Social Behavior; Intelligent Human-Computer Interfaces.

1 Introduction

Depression is the most pervasive world population mental disorder, leading to significant social, occupational, and cognitive impairments in individuals' life. Depression is a source of social disabilities, and a main risk factor for suicide in older and young people (Dehaye et al. 2018, Groholt & Ekeberg 2009). Depression rarely occurs alone and is often accompanied by anxiety and stress (Horowitz 2010), or is in comorbidity with other mental disorders such as schizophrenia and dementia (Warson et al. 2019, Grover et al. 2017). The most obvious manifestations of depressive disorders are loss of interest, anhedonia, feelings of shame and guilt, low self-esteem, loss of sleep, and appetite, pervasive sadness not related to specific motivations (mood disorders), feelings of tiredness, poor concentration, reduced social contacts and emotional feelings, apathy (Yates et al. 2007, Freeman et al., 2017).

According to the World Health Organization (WHO) at the least 25% of people visiting family doctors live with depression (http://www.euro.who.int/en/health-topics/noncommunicable-diseases/mental-health/news/news/2012/10/depression-in-europe), and this number, projected to increase, is placing considerable burdens on national health care institutions in terms of medical, and social costs associated to depression cares.
Prevention, accurate diagnosis, appropriate treatments and constantly monitoring are crucial to reduce and limit the negative consequences of the disorder on patients' quality of life and the encumbrance of costs for public healthcare institutions (Olesen et al., 2012).

2 Questionnaires assessing depressive states

Research in depression is driven by the need of identifying indicators of the disease able to accurately predict its development. However, depression is difficult to detect, difficult to measure, and it is nontrivial to quantify the effectiveness of the suggested cares during the disorder's monitoring. Routinely screenings of depression are performed by means of semi-structured interviews (SCID, Structured Clinical Interview, https://www.appi.org/products/structured-clinical-interview-for-dsm-5-scid-5) recommended by the Diagnostic and Statistical Manual of Mental Disorders (recently updated to version 5, DSM-5). These interviews are extremely time consuming, may deliver heterogeneous measures (accounting of different mental disorders that may co-exist in comorbidity with depressive disorders), do not provide a clinically cut off in the outcome rates, require training to ensure reliable and valid interpretations of the results, and can be affected by biases resulting either from clinicians' theoretical orientations and training or clinicians overestimation of patient's progresses (that may occur when treatments' success is encouraged), or both. To avoid these pitfalls in the implementation of structured interviews for depression diagnoses, clinicians and researchers increasingly use self-report screening questionnaires, which do not require cumbersome time and special training for being administered, and whose ratings are highly correlated with those offered by professionals (Zimmerman et al. 2018).
Among the many clinically endorsed self report questionnaires, it is worth to mention the Beck Depression Inventory (BDI-II) (Beck et al. 1996), the Depression, Stress, and Anxiety Scales (DASS) (Loviband & Loviband, 1995), the Patient Health Questionnaire for depression (PHQ) (Kroenke et al., 2001), the Geriatric Depression Scale (GDS) (Yeasavage et al. 1983), the Hamilton Depression Rating Scale (HDRS) (Hamilton 1960, 1967), the Major Depression Inventory (MDI) (Bech et al. 2015), the Center for Epidemiologic Studies Depression Scale (CES-D) (Lewinsohn et al. 1997), the Zung Self-Rating Depression Scale (SDS) (Zung 1965), the Montgomery-Äsberg Depression Rating Scale (MADRS) (Montgomery&Asberg 1979), the Quick Inventory of Depressive Symptomatology (QIDS) (Rush et al 2003, Rush et al 2006), the Wechsler Depression Rating Scale (WDRS) (Wechsler et al. 1963); the Clinically Useful Depression Outcome Scale (CUDOS) (Zimmerman et al. 2008). In addition,

since depression may co-occur with other mental disorders such as schizophrenia and dementia, specific self reports have been proposed for its assessment such as the Calgary Depression Scale for Schizophrenia (CDSS) (Addington et al., 1990) and the Cornell Scale for Depression in Dementia (CSDD) (Alexopoulos et al. 1988). These questionnaires are available in long, short, and even shorter forms both as clinician-rated and self report scales.
The validity of these clinician-rated and self report scales has been assessed by several authors (Torbey et al. 2015, Rabinowitz et al. 2019, da Silva et al. 2019, Rohan et al. 2016, Cunningham et al. 2011, among others) and criticisms have emerged. Some self reports have shown poor accord with clinicians' ratings (Moller 2000). Clinicians' expertise and restrictive inclusion criterions applied to assess the efficacy of treatments may affect clinicians' ratings (Landin et al. 2000). Patients' self reports may be biased by their interpretation of symptoms and disagreements with clinicians may reflect differences caused by language interpretation, importance assigned to the different scale's items, degree of severity assigned to symptoms, illiteracy or poor reading and comprehension abilities of patients, physical debility, compromised cognitive functioning and symptoms overlapping with other mental disorders (Hershenberg et al. 2020, Tolton et al. 2019, Ahmed et al. 2018, Grover et al. 2017, Shirazian et al. 2017, Dunlop et al. 2014, among others).

To overcome these drawbacks, it has become pressing to develop objective and automatic measurements with the intention of favoring fast, non-invasive, economical, and automated on-demand assistance. These decision tools should integrate available and/or emerging diagnostic means enabling an increased precision of diagnoses and clinical decision-makings.
The automatic detection of depressive states requires the identification of behavioral and physiological biomarkers which undergo to subtle changes (e.g., retardation in speech production) because of the disease and develop Artificial Intelligent (AI) models able to extract information from a complex combination of signals considered key symptoms. This will help medical professionals to rapidly formulate accurate diagnoses. An accurate detection is also important for implementing therapeutic interventions at an early stage of the disease, i.e., when treatments may be more effective. The proposed AI models should be able to recognize the detected depressive patterns and suggest personalized pre-modelled intervention plans ranging from coaching activities (exploiting for example serious games), support networks (via chats, or social networks), and alerts to caregivers, doctors, and care control centers.

3 Behavioral and physiological parameters indicating depressive states

Research in depression is driven by the need of identifying indicators of the disease able to accurately predict its development.

Among behavioral indicators of depression are:

- Scores of depressed patients to emotional facial expressions' decoding tests (Esposito et al. 2016a, Scibelli et al. 2016, Troncone et al. 2014)
- Speech analytics: For this purpose, different characteristics of verbal (such as acoustic variation of F0) and paralinguistic information (such as empty and filled pauses) can be used (Ringeval et al. 2017, Mendiratta et al. 2017; Esposito et al. 2016b, Alghowinem et al., 2013; Cummins et al. 2015, Mundt et al. 2012).
- On-line handwriting and drawing features (Cordasco et al. 2019, Likforman et al. 2017)
- Sentiment analysis, or psychological content analysis based on linguistic text-based algorithms, such as topic modeling (Gong & Poellabauer,2017), natural language processing (Dang et al.,2017) to identify links between produced words and psychological states
- Faces, head, and gestural analysis (Almeida et al, 2014)

Among physiological indicators of depression are:

- Gaze points, eye movements, pupil dilation, and fixations' times, which provide measures of where someone is looking, how visually she/he scan the environment, the interest and the like for what is seen (eye tracking measurements). It has been shown that eye tracking measurements change significantly between depressed and control subjects (Sanchez-Lopez et al. 2019, Ypsilanti et al. 2020, Moirand et al, 2019)
- EEG signals describe the electrical activity of the brain over a period of time, as recorded from multiple electrodes placed on the scalp. Frontal EEG asymmetries refer to differences in the brain activity of the two frontal hemispheres and have been associated with individual's motivational and affective states. Relatively greater right (left) brain frontal activation is linked to negatively (positively) valenced emotions, and therefore frontal EEG asymmetries have been associated both with trait-like (such as personality traits) and state-dependent processes such as moods (Cai et al. 2020, Tolgay et al 2019, Thibodeau et al. 2006). In addition, it has been shown that absolute EEG power in theta, alpha, beta, and gamma bands in depressed patients is significantly lower than healthy controls (Xinfang et al. 2019).
- Biofeedback markers such as skin conductance, and heart rates, have been investigated in relation to depressive states. Previous studies have found for example that depressed patients have lower or flat levels of skin conductance response than healthy controls (Vahey & Becerra, 2015). Hearth rate variability (HRV) is a beat-to-beat variation in heart rate and is measured through electrocardiograms (ECGs). HRV is described both in time (using the standard deviation of normal heartbeats) and frequency domains distinguishing between high (HF - 0.15–0.4 Hz) and low frequencies (LF - 0.04–0.15 Hz) HRV. It was found that HF-HRV and not LF-HRV is reduced in depressed children and young depressed adults relative to healthy controls (Kemp et al., 2010) while LF-HRV is reduced in older depressed adults

(Brown et al. 2018). These discordances in HRV were unified by Koch et al. (2019) which proved that both HF and LF-HRV measures were lower in depressed than in healthy subjects.

4 The MENHIR (Mental health monitoring through interactive conversations) project

Depression and anxiety are common disorders. Establishing how to treat them and when specific interventions may be necessary is quite difficult. A virtual assistant can help to monitor the symptoms and facilitate the definition of appropriate interventions. The MENHIR project (https://menhir-project.eu/) aims to develop conversational technologies to promote and assist people with mental health problems (depression and anxiety) and help them to manage their condition. The project partners (Spain, Germany, Italy and United Kingdom) intend to create a network to exchange and share their expertise for developing conversational assistants to support depressed individuals.

Such assistants should offer personalized social support through mental health education and continuous symptoms' monitoring providing valuable feedback and promoting physical and intellectual activities to improve patients' abilities to manage their conditions and prevent recurrence and relapse of depressive patterns. By providing automatic on-demand assistance, MENHIR aims to reduce social inequalities due to the cost of treating such disorders. In the proposed context, artificial intelligence is the theoretical structure from which to derive algorithms for information processing able to produce new representations of depressive disorders and offer generalized solutions for the nonstationary and non-linear input-output relations describing the disease.

Limitations due to the fact that algorithms of this type may not converge towards adequate solutions and produce meaningless results are surpassed by biologically inspired machine learning models that exploit behavioral analyses performed on communication signals (in the case of MENHIR essentially expressed by voice) instantiated in different scenarios and automated through audio/video processing, synthesis, detection and recognition algorithms. To realize an artificial intelligence aimed at humans it is necessary to understand the nature with which humans interact with each other through the multimodal exchange of signals that include voice, facial expressions, gestures, gaze, and body movements into a single perception. MENHIR aims to use speech behavioral analyses and mathematically structure symbolic and functional speech concepts to define computational models able to understand and synthesize the human capacity to communicate through voice individual choices, perceptions and actions. In addition, MENHIR establishes the need of data visualization for users and developers in order to guarantee that the proposed AI models will preserve human rights, human well-being, and privacy and security of their personal data. At the same time it will ensure that the proposed conversational systems are effective, competent, explainable, interpretable, and aware of their limitations, accounting of all relevant legal and compliance aspects as identified by the EU directives and the white paper (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016L2102, 2016, https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en, February 2020) on artificial intelligence. A data visualization model, called AI2VIS4BigData Reference Model, is under study by MENHIR's partners (Reis et al. 2020, Bornschlegl et al. 2016) and is going through modifications and re-assessment in order to satisfy the specific MENHIR context.

5 Automatic detection of depressive states exploiting speech

Speech is, among the behavioral data, the most exploited to detect depressive states, since depression causes psychomotor retardation, i.e. depressed speech exhibits a "slow" auditory dimension (Marazziti et al. 2010, Bennabi et al 2013) and is perceived as sluggish. This sluggishness is due to changes in prosodic and acoustic features of the produced speech. Automatic detection of depression through speech analysis aims to accomplish: a) the discrimination between depressed and healthy subjects and b) the discrimination among different degree of depression (e.g. typical, mild, moderate, severe). A typical automatic classifier consists of two parts: the speech processing module that extract prosodic, and spectral features from speech, and the application of a computational method to predict the belonging of the data to one of the two classes (e.g. depressed and non-depressed subjects) or to multiple severity categories (e.g. mild, moderate, and severe depression).

Generally, when learning algorithms are involved, the computational method requires a training phase to learn an association between the available data and categories to be learned, and a testing phase where an unseen speech sample is assigned to one of the learned categories. Support Vector Machine (SVM), Gaussian Mixture Models (GMM), K-Nearest Neighbors (K-NN), Artificial Neural Network (ANN) are among the most utilized classifiers to detect depression, and exploit several spectral and prosodic speech features (Jiang et al. 2017, Cumming et al. 2015). The performance of these classifiers is reported in terms of accuracy (percentage of subjects attributed to a class that actually belong to such a class) and/or recall (percentage of samples belonging to a given class that have actually been attributed to that class). Most of these studies reported an accuracy of more than 70% demonstrating that this automatic approach can be a useful tool for helping clinicians in the diagnosis of depression, in addition to traditional diagnostic instruments.

Experiments reported in literature on this line of investigation show noticeable achievements. Scherer et al. (2013), extracted glottal features from speech signals collected from a sample of 39 subjects (14 depressed) whose depression degree was assessed through the Patient Health Questionnaire (PHQ-9, Kroenke et al. 2001) and on these features trained a SVM classifier, obtaining an accuracy of 75%. Alghowinem et al. (2013) tested spontaneous and read speech of 30 depressed patients - diagnosed with severe depression through the HDRS questionnaire - and matched them with 30 healthy subjects. Speech was processed extracting several spectral features such as pitch, MFCCs, energy, intensity, loudness, formants, jitter shimmer, voice quality, Harmonic Noise Ratio. A SVM classifier trained on these features produced a recall of 68-69% on spontaneous and 61-64% on read speech respectively. On the same speech recordings, the performances of SVM, Hierarchical Fuzzy Signature (HFS), MLP, GMM and fusion methods were assessed (Alghowinem et al. 2013b). The fusion of GMM and SVM classifiers produced a recall of 81%, followed by a recall of 76% obtained with SVM.

Kiss et al. (2016) recruited a sample of 53 Hungarian and 11 Italian depressed patients (diagnosed by psychiatrists and assessed with BDI-II with respect to the depression degree) matched with 53 Hungarian and 11 Italian controls. From reading speech they extracted spectral (formants, Jitter and Shimmer) and prosodic (variance of intensity, range of fundamental frequency, total length of pauses, articulation and speech rate) features which were given as input to a SVM and ANN classifiers. Performance reached an accuracy of 75% for both classifiers when Hungarian speech were used on training and testing data, and 77% when Hungarian speech was used for training and Italian speech for testing. Jiang et al. (2017) extracted different low descriptors from recordings obtained in readings, interviews, and picture descriptions tasks obtained from 85 depressed subjects (diagnosed through PHQ-9 scores) matched with 85 controls. The performance of SVM, GMM and K-NN classifiers were compared on this data. Findings showed a slightly higher accuracy of SVM (65%) with respect to GMM and K-NN (60-62%) classifiers. When classifiers' performances were compared on tasks, speech from pictures' description provided better accuracy 68% than interview (63%) and reading (60%) tasks. Mendiratta et al. (2017) extracted F0 and MFCC features from the spontaneous speech of 12 depressed subjects (diagnosed by psychiatrists), and 12 matched controls and clustered the two groups through a self-organizing map (SOM) neural network, obtaining an accuracy of 80%. Scibelli et al. (2018) exploited recordings of spontaneous and read speech from 62 depressed patients (diagnosed by psychiatrists) and 54 healthy controls. Speech sample were transformed in input vectors having as components 384 low level descriptors (among those Root Mean Square (RMS) of the energy, MFCC coefficients (MFCC), Zero Crossing Rate (ZCR)). A SVM classifier was trained on these vectors obtaining a 75% discrimination accuracy, which was independent from gender, depression-related pathology, and length of the pharmacological treatment (if any).

Currently, new audio analyses have been proposed that rely on linguistic text-based methods, such as topic modeling (Gong & Poellabauer, 2017), natural language processing (Dang et al., 2017), articulatory features associated to vowel production (such as front and back vowels and phonetical markedness, Stasak et al. 2017a, 2017b), as well as, linguistic stress associated to vowel duration, loudness (Stasak et al. 2019) and affect based measures, all of them along with spectral selection measures (Williamson et al., 2016). These new approaches were presented, starting from 2013, at the Audio/Visual Emotion Challenges (AVEC), either as primary challenges AVEC 2013, AVEC 2014, AVEC 2015, or as sub-challenges AVEC 2016, AVEC 2017, AVEC 2018 (Valstar et al., 2016, 2014, 2013, Rigenval et al. 2018, 2017, and 2015). The AVEC 2019 edition, https://sites.google.com/view/avec2019/, is a challenge on states of mind where depression plays a fundamental role, and where signal processing and machine learning methodologies are called to compete as tools offering the most effective performances to accurately detect mental health and behavioral disorders from audio and video, and text data.

6 Conclusions

There is a huge demand to develop complex autonomous systems capable of assisting people with depression. One solution is to develop complex autonomous systems, in the form of socially and emotionally believable ICT interfaces capable of detecting the onset of such disorders, providing, where possible, initial on-demand support to patients, offering doctors' sustenance for diagnoses and treatments, suggesting strategies for favoring social inclusion and wellbeing and "meeting" users' expectations and demands. These interfaces should exploit data by sensors, and data analysis by artificial intelligent and user-friendly conversational interfaces which naturally integrated behavioral and psychological features derived for depressed people's daily life observations providing emotional and psychological support to them, their doctors and caregivers, enabling advance in diagnostics and healthcare support through non-obtrusive technologies that favor the physical, cognitive, social and mental wellbeing of depressed individuals. However, this task is intrinsically complex requiring the need of a holistic approach that accounts for the multiple factors affecting depression, including personality traits, social and contextual information, and cultural diversities (Esposito & Jain 2016). "The goal is to provide experimental and theoretical models of behaviors for developing a computational paradigm that should produce [ICT interfaces] equipped with a human level [of] automaton intelligence" (Esposito et al. 2015, p. 48). Such ICT interfaces can be exploited as automatic diagnostic tools for the diagnosis of different degrees of depressive states and in general for detecting from interactional exchanges reliable behavioral and contextual information. The challenge is how to make available the data collected from multiple sources and differently processed to intelligent computational devices for a cross-multimodal analysis and how endorse them of an effective ability to intelligently process behavioral and contextual information.

ACKNOWLEDGEMENTS
The research leading to these results has received funding from the EU H2020 under grant agreement N. 769872 (EMPATHIC) and 823907 (MENHIR), and from the Italian projects SIROBOTICS , MIUR, PNR 2015-2020, DD1735, 13/07/ 2017, and ANDROIDS, V:ALERE, UniCampania, D.R. 906 del 4/10/2019, prot. 157264, 17/10/2019. 7.

References

1. Addington D, Addington J, Schissel B (1990) A depression rating scale for schizophrenics. Schizophr. Res. 3 (4), 247–251.
2. Ahmed AT, Frye MA, Rush AJ, Biernacka JM, Craighead WE, McDonald WM, Bobo WV, Riva-Posse P, Tye SJ, Mayberg HS, Flavin DH, Skime MK, Jenkins GD, Wang L, Krishnan RR, Weinshilboum RM, Kaddurah-Daouk R, Dunlop BW (2018) Mapping

depression rating scale phenotypes onto Research Domain Criteria (RDoC) to inform biological research in mood disorders. J. Affect. Disord 238, 1–7.

3. Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Breakspear, M., & Parker, G. (2013). Detecting Depression: A comparison between spontaneous and read speech. In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 7547–7551
4. Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Gedeon, T., Breakspear, M., & Parker, G. (2013b). A comparative study of different classifiers for detecting Depression from spontaneous speech. In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 8022–8026.
5. Bornschlegl M.X., Berwind K., Kaufmann M., Engel F.C., Walsh P., Hemmje M.L., Riestra R 2016 IVIS4BigData: A reference model for advanced visual interfaces supporting big data analysis in virtual research environments, LNCS, vol. 10084 LNCS, pp. 1-18
6. Almeida E, Ferruzca M, del Pilar Morales Tlapanco M (2014) Design of a System for Early Detection and Treatment of Depression in Elderly Case Study. In: Cipresso P., Matic A., Lopez G. (eds) Pervasive Computing Paradigms for Mental Health. MindCare 2014. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 100, 115-124, Springer, Cham
7. Alexopoulos GS, Abrams RC, Young RC, Shamoian CA (1988) Cornell Scale for Depression in Dementia. Biol Psychiatry, 23, 27184.
8. Bech P, Timmerby N, Martiny K, Lunde M, SoendergaardS (2015) Psychometric evaluation of the Major Depression Inventory (MDI) as depression severity scale using the LEAD (Longitudinal Expert Assessment of All Data) as index of validity Psychiatry, 15:190.
9. Beck AT, Steer RA, Brown GK (1996). Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation
10. Bennabi D, Vandel P, Papaxanthis C, Pozzo T, Haffen E (2013) Psychomotor retardation in depression: A systematic review of diagnostic, pathophysiologic, and therapeutic implications. BioMed research international, vol. 2013, Article ID 158746, 18 pages, http://dx.doi.org/10.1155/2013/158746
11. Bernard D, Lançon,C, Auquier P, Reine,G, Alghowinem S, Goecke R, Wagner M, Epps J, Breakspear M, Parker G (2013a) Detecting Depression: A comparison between spontaneous and read speech. In proceeding of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 7547–7551
12. Brown L, Karmakar C, Gray R, Jindal R, Lim T, Bryant C (2018) Heart rate variability alterations in late life depression: A meta-analysis. Journal of Affective Disorders, 235, 456–466
13. Cai H, Qu Z, Li Z, Zhang Y, Hu X,Hu B (2020) Feature-level fusion approaches based on multimodal EEG data for depression recognition. Information Fusion, 59, 127-138
14. Cordasco G, Scibelli F, Faundez-Zanuy M, Likforman-Sulem L, Esposito A (2019) Handwriting and Drawing Features for Detecting Negative Moods. In Esposito A et al. (eds) Quantifying and Processing Biomedical and Behavioral Signals. Smart Innovation, Systems and Technologies, vol 103, 71-86. Springer International Publishing

15. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri T F (2015). A review of Depression and suicide risk assessment using speech analysis. Speech Communication, 71, 10–49.
16. Cunningham JL, Wernroth L, von Knorring L, Berglund L, Ekselius L (2011) Agreement between physicians' and patients' ratings on the Montgomery–Äsberg Depression Rating Scale. Journal of Affective Disorders 135, 148–153
17. Dang T, Stasak B, Huang Z, Jayawardena S, Atcheson M, Hayat M, Le P, Sethu V, Goecke R, Epps J (2017) Investigating word affect features and fusion of probabilistic predictions incorporating uncertainty in AVEC 2017. In: Proceedings of the Seventh Annual Workshop on Audio/Visual Emotion Challenge, Mountain View, CA. 27–35.
18. da Silva AK, Reche M, da Silva Lima AF, de Almeida Fleck MP, Edison Capp E, Milman Shansis F (2019) Assessment of the psychometric properties of the 17- and 6-item Hamilton Depression Rating Scales in major depressive disorder, bipolar depression and bipolar depression with mixed features. Journal of Psychiatric Research, 108, 84-89
19. Dehaye M, Leemans C, Loas G (2018) Elderly's suicide attempt. Rev Med Brux, 39, 15-21
20. Dunlop BW, McCabe B, Eudicone JM, Sheehan JJ, Baker RA (2014) How well do clinicians and patients agree on depression treatment outcomes? Implications for personalized medicine. Hum. Psychopharmacol. 29, 528–536.
21. Esposito A, Esposito AM, Likforman-Sulem L, Maldonato NM, Vinciarelli A (2016b) On the significance of speech pauses in depressive disorders: results on read and spontaneous narratives. In: Esposito, A, et al. (eds.) Springer SIST series on Recent Advances in Nonlinear Speech Processing, vol. 48, pp. 73–82
22. Esposito A, Esposito AM, Vogel C (2015). Needs and challenges in human computer interaction for processing social emotional information. Pattern Recognition Letters, 66, 41-51.
23. Esposito A, Jain LC (2016) Modeling social signals and contexts in robotic socially believable behaving systems. In Esposito A, Jain LC (eds.) Toward Robotic Socially Believable Behaving Systems Volume II "Modeling Social Signals" Springer International Publishing Switzerland, ISRL series 106, pp. 5–13
24. Esposito A, Scibelli S, Vinciarelli A (2016a) A Pilot study on the decoding of dynamic emotional expressions in major depressive disorder. In Bassis et al. (Eds): Advances in Neural Networks: Computational Intelligence for ICT. Series: Smart Innovation, Systems and Technologies, Vol. 34, 189-200.
25. Freeman D, Sheaves B, Goodwin GM, Yu LM, Nickless A, Harrison PJ., Hinds C (2017). The effects of improving sleep on mental health (OASIS): A randomized controlled trial with mediation analysis. The Lancet Psychiatry, 4(10), 749-758.
26. Gong Y, Poellabauer C (2017) Topic modeling based on multi-modal depression detection. In: Proceeding of the Seventh Annual Workshop on Audio/Visual Emotion Challenge, Mountain View, CA. 69–76.
27. Groholt B, Ekeberg Ø (20009) Prognosis after Adolescent Suicide Attempt: Mental Health, Psychiatric Treatment, and Suicide Attempts in a Nine-Year Follow-up Study. Suicide and Life-Threatening Behavior 39(2), 125-136

28. Grover S, Sahoo S, Dua D, Chakrabarti S, Avasthi A (2017) Scales for assessment of depression in schizophrenia: Factor analysis of Calgary depression rating scale and Hamilton depression rating scale. Psychiatry Research 252 (2017) 333–339
29. Hamilton M (1960) A rating scale for depression. J Neurol Neurosurg Psychiatry, 23:56–62.
30. Hamilton M (1967) Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 6:278 –296.
31. Helfer BS, et al. (2013) Classification of depression state based on articulatory precision. In: 14th Annual Conference of the International Speech Communication Association, 1–5
32. Hershenberg R, McDonald WM, Crowell A, Riva-Posse P, Craighead WE, Mayberg HS, Dunlop BW (2020) Concordance between clinician-rated and patient reported outcome measures of depressive symptoms in treatment resistant depression. Journal of Affective Disorders 266, 22–29
33. Horwitz AV (2010) How an Age of Anxiety Became an Age of Depression. Milbank Q. 2010 Mar; 88(1): 112–138.
34. Jiang, H., Hu, B., Liu, Z., Yan, L., Wang, T., Liu, F., … Li, X. (2017). Investigation of different speech types and emotions for detecting Depression using different classifiers. Speech Communication, 90, 39–46.
35. Kemp AH, Quintana DS, Gray MA, Felmingham KL, Brown K, Gatt JM (2010). Impact of depression and antidepressant treatment on heart rate variability: a review and Meta-Analysis. Biological Psychiatry, 67(11), 1067-1074.
36. Kiss, G., Tulics, M. G., Sztahó, D., Esposito, A., & Vicsi, K. (2016). Language independent detection possibilities of Depression by speech. In Recent Advances in Nonlinear Speech Processing (pp. 103–114). Springer.
37. Koenig J, Kemp AH, Beauchaine TP, Thayer JF, Kaess M (2016). Depression and resting state heart rate variability in children and adolescents—a systematic review and meta-analysis. Clin. Psychol. Rev. 46, 136–150.
38. Kroenke K, Spitzer R, Williams J (2001). The PHQ-9. Validity of a brief depression severity measure. J. Gen. Intern. Med. 16, 606–613.
39. Landin R, DeBrota DJ, DeVries TA, Potter WZ, Demitrack MA. (2000) The impact of restrictive entry criterion during the placebo lead-in period. Biometrics, 56(1), 271-8.
40. Lewinsohn PM, Seeley JR, Roberts RE, Allen NB (1997). Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. Psychology and Aging, 12(2), 277–287.
41. Likforman-Sulem L, Esposito A, Faundez-Zanuy M, Clémençon S, Cordasco G (2017) EMOTHAW: A Novel Database for Emotional State Recognition from Handwriting and Drawing. IEEE Transactions on Human-Machine Systems, 47(2):273-284, http://ieeexplore.ieee.org/document/7807324/
42. Lovibond PF, Lovibond SH (1995) The structure of negative emotional states: comparison of the depression anxiety stress scales (DASS) with the Beck Depression and Anxiety Inventories. Behavioral Research Therapy. Vol. 33, No. 3, pp. 335-343.
43. Marazziti, D., Consoli, G., Picchetti, M., Carlini, M., Faravelli, L.: Cognitive impairment in major depression. Eur. J. Pharmacol. 626, 83–86 (2010)
44. Mendiratta A, Scibelli F, Esposito AM, Capuano V, Likforman-Sulem L, Maldonato MN, Vinciarelli A, Esposito A (2017) Automatic detection of depressive states from

speech. In A. Esposito et al. (eds.), Multidisciplinary Approaches to Neural Computing, Smart Innovation, Systems and Technologies 69, 301-314

45. Moirand R, Galvao F, Brunelin J (2019) Early shifts of emotional attention as a possible predictor of remission in patients with depression receiving ECT: Preliminary results of an eye-tracker study. L'Encéphale, 45(2), 73
46. Möller HJ (2000) Rating depressed patients: observer- vs self-assessment. Eur Psychiatry, 15, 160-72
47. Montgomery SA, Äsberg M (1979): A new depression scale designed to be sensitive to change. Br J Psychiatry 134:382–389.
48. Mundt JC, Vogel AP, Feltner DE, Lenderking WR (2012) Vocal acoustic biomarkers of depression severity and treatment response. Biol. Psychiatry 72, 580–587
49. Olesen J, Gustavsson A, Svensson M, Wittchen HU, Jonsson B, (2012) The economic cost of brain disorders in Europe. European Journal of Neurology, 19(1), 155–162.
50. Rabinowitz J, Schoolerb NR, Brownc B,Dalsgaardd M, Engelhardte N, Friedbergerf G, Kinong BJ, Leeh D, Ockuni F, Mahableshwarkarj A, Tsaik J, Williams JBW, Sauderm C, Yavorskyo C (2019) Consistency checks to improve measurement with the Montgomery-Äsberg Depression Rating Scale (MADRS). Journal of Affective Disorders 256, 143–147.
51. Reis T., Bornschlegl M.X, Hemmje ML 2020 Towards a Reference Model for Artificial Intelligence Supporting Big Data Analysis. Proceedings of the 2020 International Conference on Data Science (ICDATA'20), 2020
52. Ringeval F, Schuller B, Valstar M, Gratch J, Cowie R, Scherer S, Mozgai S, Cummins N, Pantic M 2017. AVEC 2017 – Real-life depression, and affect recognition workshop and challenge. In Proceedings of the 7th International Workshop on Audio/Visual Emotion Challenge (AVEC), co-located with the 25th ACM International Conference on Multimedia (ACM MM). ACM, Mountain View, CA, USA, 3–9.
53. Ringeval F, Schuller B, Valstar M, Jaiswal S, Marchi E, Lalanne D, Cowie R, Pantic M 2015. AV+EC 2015 – The first affect recognition challenge bridging across audio, video, and physiological data. In Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge (AVEC), co-located with the ACM International Conference on Multimedia (ACM MM). ACM, Brisbane, Australia, 3–8.
54. Ringeval F, Schuller B, Valstar M, Cowie R, Kaya H, Schmitt M, Amiriparian S, Cummins N, Lalanne D, Michaud A, Çiftçi E, Güleç H, Salah AA, Pantic M 2018 AVEC 2018 Workshop and Challenge: Bipolar Disorder and Cross-Cultural Affect Recognition. In Proceedings of the 8th International Workshop on Audio/Visual Emotion Challenge (AVEC), co-located with the 26th ACM International Conference on Multimedia (ACM MM), October 22, 2018, Seoul, Republic of Korea
55. Rohan KJ, Rough JN, Evans M, Ho S-Y, Meyerhoff J, Roberts LM, Vacek PM (2016) A protocol for the Hamilton Rating Scale for Depression: Item scoring rules, Rater training, and outcome accuracy with data on its application in a clinical trial. Journal of Affective Disorders 200, 111-118
56. Rosser BA, Vowles KE, Keogh E, Eccleston C, Mountain GA (2009) Technologically assisted behaviour change: a systematic review of studies of novel technologies for the management of chronic illness. Telemed. Telecare 15(7), 327–338

57. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al (2003) The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. Biol Psychiatry 54, 573–583.
58. Rush AJ, Bernstein IH, Trivedi MH, Carmody TJ, Wisniewski S, Mundt JC, Shores-Wilson K, Biggs MM, Woo A, Nierenberg AA, Fava M (2006). An evaluation of the quick inventory of depressive symptomatology and the Hamilton Rating Scale for Depression: A sequenced treatment alternative to relieve depression trial report. Biol. Psychiatry 59, 493–501.
59. Sanchez-Lopez A, Koster EHW, Put JV, De Raedt R (2019) Attentional disengagement from emotional information predicts future depression via changes in ruminative brooding: A five-month longitudinal eye-tracking study. Behaviour Research and Therapy, 118, 30-42
60. Scibelli F, Troncone A, Likforman-Sulem L, Vinciarelli A, Esposito A (2016). How Major Depressive Disorder Affects the Ability to Decode Multimodal Dynamic Emotional Stimuli. FRONTIERS 10.3389/fict.2016.00016 IN ICT, vol. 3, ISSN: 2297-198X
61. Scibelli, F., Roffo, G., Tayarani, M., Bartoli, L., De Mattia, G., Esposito, A., & Vinciarelli, A. (2018) Depression speaks: automatic discrimination between depressed and non-depressed speakers based on nonverbal speech features. In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, 6842-6846, 15-20 April 2018, Calgary, AB, Canada
62. Scherer, S., Stratou, G., Gratch, J., Morency, L., 2013b. Investigating voice quality as a speaker-independent indicator of depression and PTSD. Proceedings of Interspeech. ISCA, Lyon, France, 847–851.
63. Shirazian S, Grant CD, Aina O, Mattana J, Khorassani F, Ricardo AC (2016) Depression in Chronic Kidney Disease and End-Stage Renal Disease: Similarities and Differences in Diagnosis, Epidemiology, and Management. Kidney Int Rep., 2(1):94–107. Published 2016 Sep 20. doi: 10.1016/j.ekir.2016.09.005
64. Stasak, B., Epps, J., Goecke, R., 2017a. Elicitation design for acoustic depression classification: an investigation of articulation effort, linguistic complexity, and word affect. In: Proceedings of INTERSPEECH Conference, Stockholm, Sweden. 834–838.
65. Stasak, B., Epps, J., Lawson, A., 2017b. Analysis of phonetic markedness and gestural effort measures for acoustic speech-based depression classification. In: Proceedings of the ACII Conference, Parramatta, Australia. 1–6.
66. Stasak, B., Epps, J., & GOECKE, R. (2019). An investigation of linguistic stress and articulatory vowel characteristics for automatic depression classification. Computer Speech and Language, 53, 140-155.
67. Thibodeau R, Jorgensen SR, Kim S (2006) Depression, anxiety, and resting frontal EEG asymmetry: a meta-analytic review. Journal of abnormal psychology, vol. 115, 4.
68. Tolgay B, Dell'Orco S, Maldonato MN, Vogel Carl Trojano L, Esposito A (2020) EEGs as potential predictors of virtual agent's' acceptance. In Proc. of 10th IEEE International Conference in Cognitive Infocommunication, Naples, 23-25 October 2019.
69. Tolton D, Steffens D, Chan G (2019) Patient versus clinician rated depression scores: a comparison of participant scores on the Carroll Depression Scale and the Hamilton

Depression Rating Scale - The American Journal of Geriatric Psychiatry, 27(3)S 124-125

70. Torbey E, Pachana NA,Dissanayaka NNW (2015) Depression rating scales in Parkinson's disease: A critical review updating recent literature. Journal of Affective Disorders, 184, 216–224
71. Troncone A, Palumbo D, Esposito A (2014) Mood effects on the decoding of emotional voices. In: Bassis, S., et al. (eds.) Recent Advances of Neural Network Models and Applications, SIST 26, pp. 325–332. International Publishing Switzerland
72. Vahey R, Becerra R (2015). Galvanic skin response in mood disorders: a critical review. Int. J. Psychol. Psychol. Ther. 15 (2), 275–304.
73. Valstar M, Schuller B, Krajewski J, Cowie R, Pantic M. 2013, AVEC'13, Workshop summary for the 3rd international audio/visual emotion challenge and workshop (AVEC'13). In Proc. MM, 1085–1086, Barcelona, Spain, October 2013. ACM.
74. Valstar M, Schuller B, Krajewski J, Cowie R, Pantic M 2014 AVEC 2014: the 4th international audio/visual emotion challenge and workshop. In Proc. MM, pages 1243–1244, Orlando (FL), USA, November 2014. ACM.
75. Valstar M, Gratch J, Schuller B, Ringeval F, Cowie R, Pantic M 2016. AVEC 2016. Summary for AVEC 2016: Depression, mood, and emotion recognition workshop and challenge. In Proceedings of the 24th ACM International Conference on Multimedia (ACM MM). ACM, Amsterdam, The Netherlands, 1483 - 1484.
76. Yates WR, Mitchell J, Rush AJ, Trivedi M, Wisniewski SR, Warden D, Bryan C, Fava M, Husain MM, Gaynes BN (2007). Clinical features of depression in outpatients with and without co-occurring general medical conditions in STAR* D: Confirmatory analysis. Primary care companion to the Journal of clinical psychiatry, 9(1), 7-15.
77. Ypsilanti A, Robson A, Lazuras L, Powell PA, Overton PG (2020) Self-disgust, loneliness and mental health outcomes in older adults: An eye-tracking study. Journal of Affective Disorders, 2661, 646-65
78. Yeasavage JA, Brink TL, Rose TL, Lum O, Huang V, Adley M, Leirer O (1983) Development and validation of a geriatric depression screening scale: A preliminary report. J Psychiatr Res, 17,3749
79. Watson B, Tatangelo G, McCabe M (2019) Depression and Anxiety Among Partner and Offspring Carers of People with Dementia: A Systematic Review. The Gerontologist, 59(5), e597–e610,
80. Wechsler H, Grosser GH, Busfield BL (1963) The Depression Rating Scale, Arch. Gen. Psychiatry, 9, 334-343.
81. Williamson,J.R.,Godoy,E.,Cha,M.,Schwarzentruber,A.,Khorrami,P.,Gwon,Y.,Kung,H.T.,Dagli,C.,Quatieri,T.F., 2016.Detecting depression using vocal, facial, and semantic communication cues. In Proceedings of the Sixth International Workshop on Audio/Visual Emotion Challenge (AVEC), Amsterdam, The Netherlands 11–18.
82. Zimmerman M, Chelminski I, McGlinchey JB, Posternak MA (2008). A clinically useful depression outcome scale. Compr. Psychiatry 49, 131–140.
83. Zimmerman M, Walsh E, Friedman M, Boerescu DA, Attiullah A (2018) Are self-report scales as effective as clinician rating scales in measuring treatment response in routine clinical practice? Juornal of Affective Disorders, 225, 449-452
84. Zung WWK (1965) Self-Rating Depression Scale, Arch. Gen. Psychiatry, 12, 63-70.