Behavioral Sentiment Analysis of Depressive States

Anna Esposito[*+], Gennaro Raimo[*], Mauro Maldonato[‡], Carl Vogel[§], Massimiliano Conson[*], and Gennaro Cordasco[*+]
[*]Department of Psychology, Universit`a degli Studi della Campania "L. Vanvitelli", and
[+]International Institute for Advanced Scientific Studies (IIASS), Italy,
Emails: *iiass.annaesp@tin.it; anna.esposito@unicampania.it.*
[‡] Dept of Neuroscience Università di Napoli "Federico II" Naples, Italy.
[§] School of Computer Science and Statistics Trinity College Dublin, Ireland.

***Abstract* — The need to release accurate and incontrovertible diagnoses of depression has fueled the search for new methodologies to obtain more reliable measurements than the commonly adopted questionnaires. In such a context, research has sought to identify non-biased measures derived from analyses of behavioral data such as voice and language. For this purpose, sentiment analysis techniques were developed, initially based on linguistic characteristics extracted from texts and gradually becoming more and more sophisticated by adding tools for the analyses of voice and visual data (such as facial expressions and movements). This work summarizes the behavioral features accounted for detecting depressive states and sentiment analysis tools developed to extract them from text, audio, and video recordings.**

Keywords - depression, sentiment analysis, speech, videos, texts, faces, language, behaviors.

I. INTRODUCTION

Depression, otherwise labelled as major depressive disorder (MDD) or clinical depression is classified as a mood disorder by the Diagnostic and Statistical Manual of Mental Disorders (DSM-5 [1]). Currently depression affect more than 300 million of people worldwide, and this number is projected to increase considering that the number of people living with depression has grown of about 18.4% between 2005 and 2015. The global burden of disease study (GBD 2015, [2]) estimated that depression ranked second among worldwide mental illnesses being responsible of years lived with disability (YLD), and major cause of suicides estimated to be approximately 800,000 per year [3,4].

Depressive symptoms and sequels encompass many aspects of human life, such as emotional states (feeling sad or hopeless, anhedonia), physical functioning (insomnia, low energy, changes in appetite or weight, headaches, pain), cognitive functions (low self-esteem, impairment of thinking or speaking, difficulty in goals pursuit, suicidal ideations, memory and concentration difficulties) and behaviors (social isolation, loss of interest in activities and social relationships, failure in working, suicide attempts) [2,3]. Depression causes social disabilities such as deficits in identifying emotional expressions of others engendering social relationships, causing loneliness and greatly impoverishing patients' quality of life [5]. Given the epidemiological nature of depression, the EU Community underlined the importance to develop accurate and cost-effective tools to identify early signs of the syndrome, so that large scale screenings and timely treatments would be possible, avoiding long suffering periods to patients and high costs associated to late cures [6-7]. Machine learning (ML) approaches, based on artificial agents able to learn how to adapt their actions to requested tasks analyzing behavioral data of depressed individuals are the most promising candidates to absolve this function. An automated approach "*may reduce multiple in-office clinical visits, facilitate accurate measurements and identification of the disease, and quicken the evaluation of treatments*" [8, p. 41]. Indeed, dedicated attempts have been made by the research community, through the international workshop on Audio/Visual Emotion Challenge (AVEC), starting from 2013, to detect depression using ML techniques analyzing several aspects of verbal (i.e., vocal dynamics of prosody) and nonverbal (i.e., gestures, and dynamics of facial appearance) and linguistic behaviors (i.e., word frequencies and vocabulary) of depressed patients [9-13].

This work summarizes the current results for automating the detection and prediction of depressive symptoms considering behavioral features related to them (section II) and how they have been incorporated in sentiment analysis tools that progressed from textual (section III) to multimodal (section IV). Conclusion (section V) discusses challenges and limitations of the multimodal approach (in order to develop faster diagnostics procedures for detecting mental disorders and provide objective automatic supports to clinicians), and its relevance to CoginfoCom scientific community.

II. BEHAVIORAL CHARACTERISTICS OF DEPRESSION

Various studies demonstrated that behavioral signals such as speech [14-17], gross motor activity and gesturing [18], self-touching [19], head movements [20], facial expressivity [21-22], smiling [18-19], and body expressions [23] can be used to distinguish between depressed and healthy individuals. Being able to identify changes in behavioral signals may promote more reliable diagnoses of depression than those currently derived from self-reports since behavioral measures are difficult for patients to manipulate, and mostly depend on reflex responses not completely under individuals' conscious control (as for example facial expressions and head movements during a clinical interview). The following text discusses both visual and vocal indicators of depressive disorders to put in context their mining by sentiment analysis tools.

A. Speech characteristics

Several acoustic and prosodic features of speech, such as pauses, speaking rate, loudness, intonation, intensity of overtones, and fundamental frequency (F0) have been proved to be sensitive markers of mood changes and hence of depression (see [15, 24] for a review). In particular, Honig et al. [25] reported that a reduced F0 frequency range and reduced F0 average values can be associated with increasing levels of depression. Different authors have shown that silent speech pause durations tend to be longer and number of pauses tends to be more frequent in acute depressed patients [15, 26]. These effects can be related to general psychomotor retardation symptoms extensively reported in depressive disorders [26-27]. A recent work by Stasak et al. [28] had shown that psychomotor retardation also affect vowel production: depressed speakers show shorter vowels' durations and less variance for 'low', 'back', and 'rounded' vowel positions. Taguchi et al. [29] exploited Mel-frequency cepstrum coefficients (MFCCs, [30]) and found significant differences between depressive patients and controls in the values of the second MFCC coefficient (MFCC 2) to the extent that it can discriminate between patients and controls with a sensitivity of 77.8% and a specificity of 86.1%. Mendiratta et al. [31] used only MFCC coefficients and a self-organizing map classifier (SOM, [32]) of 36 neurons obtaining 81% average accuracy. Stolar et al. [33] exploited an optimized set of 19 spectral roll-off parameters for automatically detecting depression from a clinical database of 63 male and female adolescents (29 depressed) interacting with parents. Using such optimized parameters as input to a support vector machine (SVM) classifier [32] provided a depression detection accuracy of 70.5% for females and 82.2% for males. Combining the roll-off parameters with other spectral measures (such as spectral flux, centroid, entropy, power spectral density and formants and bandwidths values) the accuracy increased to 97.5% for males and 92.3% for females. Recently, Williamson et al. [8] proposed to use articulatory coordination features described by eigenspectra derived from time series of formant frequencies and facial action units extracted from the Wyss Institute Biomarkers for Depression (WIBD) audio and video multi-modal dataset, and the audio dataset collected by Mundt et al. [34]. The accuracy of the proposed algorithm in terms of root mean squared error (RMSE) and Spearman correlation (r) was RMSE=5.49, r = 0.63 on WIBD and RMSE=5.99, r=0.49 on Mundt's databases. Kiss et al. [17] reported that the normalized values of the ratio of transient parts of speech used as only inputs to a support vector machine classifier allowed to discriminate between depressed and healthy subjects with an accuracy of 81%. Scibelli et al. [35] were the first to use a database of spontaneous and read speech obtained from a large number (62) of clinically diagnosed depressed patients appropriately matched by age and gender with a control group. From each speech sample were extracted 384 acoustic features (among those 12 MFCC coefficients, zero crossing rates, and voicing probabilities) on which a linear SVM neural net was trained. Accuracy was above 75% on both spontaneous and read speech and results were independent from patient's gender, depression-related pathology diagnose, and length of the pharmacological treatment (if any). These promising results can be considered more reliable than questionnaires and self reports for depression diagnoses since desirability and patients' individual trait biases are avoided. However, self–report may be preferred to behavioral signals because they are faster (generally it takes few minutes to complete the questionnaire) to be administered, and non-invasive since do not require patients to report stressful events or private aspects of their life.

B. Visual characteristics

In addition to speech, several visual behavioral features were exploited as indicators of depressive states such as gross motor activities, self-touching and facial expressions. Concerning gross motor activities (such as movements of arms, legs and trunk), Balsters et al. [18] reported that depressed patients compared to a control group show high levels of body activity during night-time and low levels of body activity during daytime. Scherer et al. [19] investigated the role of both automatic and manually annotated nonverbal features in describing mental illnesses and in particular depression exploiting the Distress Analysis Interview Corpus (DAIC) of human and computer interviews [36]. The DAIC is a databases of clinical interviews conducted in four conditions: a) by a human interviewer face to face; b) by a human interviewer in teleconference; c) by a virtual agent controlled by a human interviewer (Wizard of Oz conditions); d) by an autonomous virtual agent. Scherer et al. used for their investigation the face to face interviews and reported longer self-touching with both hands (e.g. rubbing, stroking) of depressed patients compared to a control group. Girard et al. [22] found that facial expressions and head motions of depressed patients significantly differ according to depression symptom's severity. Severe depressive states produce less AU 12 (Action Unit) -contractions of the zygomatic major muscle- and more AU 14 -contractions of the buccinator muscle- occurrences. In addition, mild depressive states produce more smiles and decreased expressions of contempt and embarrassment (confirmed also by [19]). Dibeklioğlu et al. [37] in addressing depression severity degrees through stacked de-noising autoencoder (SDAE, [32]) classifiers proved that facial dynamics and head movements produce accuracies similar to those obtained by adding vocalizations. Nasir et al. [38] using only facial landmarks as input to an SVM classifier obtained the same F1 scores (F1=0.63) reported by Pampouchidou et al. [39] (F1=0.62) which combined facial landmarks and audio features as input to a random forest algorithm [32]. Improved F1 scores (F1=0.5) were obtained by Valstar et al. [11] by using as input to an SVM classifier, either facial landmarks, or AUs features, or by adding to them audio features. Pampouchidou et al. [39] obtained the same F1 scores (F1=0.5) using only facial landmarks features as input to a random forest algorithm. Finally, Joshi et al. [23] and Alghowinem et al. [20] showed that body expressions and head poses respectively, are distinctive during.

C. Linguistic characteristics

In addition to non-verbal features, studies have carried out quantitative analyses of texts. The analyses were motivated by Collier et al. [40] which showed that sentiments affect several conversational variables such as lexical make-up, grammatical complexity, syntactic variables (i.e. adjectives, adverbs, negation, quantifiers), and types of sentences (e.g. statement, question, exclamation). Linguistic parameters mostly investigated in depression are associated to: a) linguistic content essentially *what it is said*, i.e. nouns; and b) linguistic style, *how it is said*, i.e. which pronouns, prepositions, articles are being used [41-42].

III. Sentiment Analysis

Initially, sentiment analysis (also opinion mining or emotion AI [43]) aimed at identifying and classifying selected text information related to opinions, and prevalence of positive and negative sentiments toward something (such as a product, etc.) and/or someone (such as a person, etc.), for social media monitoring, brand monitoring, market research, political actions' acceptance and so on. Sentiment analysis techniques can be split into different levels [44]: 1) document-level techniques establishing whether the mined document expresses negative or positive sentiments; 2) sentence-level techniques attempting to determine the polarity of sentiments in a single sentence; 3) aspect-based techniques devoted at recognizing sentiments associated with each subject in a document.

Sentiment analysis exploits Natural Language Processing (NLP) to mine texts according three main approaches [45]: a) knowledge-based approaches aimed to classify texts by affect categories based on the presence of unambiguous affect words; b) statistical methods which exploit machine-learning algorithms running on large databases of emotionally annotated text; c) hybrid approaches that combine knowledge-based and statistical techniques. Despite sentiment analysis has demonstrated good performances in predicting affect texts [46], only recently it has been utilized for discriminating either between depressed and healthy subjects or predicting depression severity degree through text features such as words-frequencies topic coverage range (e.g. word frequency within specific subject categories), and pronouns' distributions [41-42, 47]. In this context it has been shown that: a) depressed patients differentiate from controls in the frequency of use of first, second and third person singular pronouns [41] because of depressed subjects' tendency [48] to be self-focused; b) depressed subjects differentiate from controls in having poorer speech and being illogical [49]; c) depressed subjects differentiate from controls in using more negative than positive words [42].

A. Textual sentiment analysis instruments

A textual sentiment analysis instrument is developed to analyze text conversations mining them for emotions and moods. Textual sentiment analysis tools include dictionaries of words and sentences noted to have emotional, cognitive, and physical meanings, under the assumptions that written and spoken language reflects elements linked to emotions, feelings, and cognitive processes [42]. The ways these elements are mined differentiate the performances of each tool. Below, textual sentiment analysis tools employed to detect depressive states are discussed underlying their advantages and drawbacks.

ANEW: The Affective Norms for English Words (ANEW) was "*developed to provide a set of normative emotional ratings for a large number of words (nouns, verbs, adjectives and adverbs) in the English language. The goal is to develop a set of verbal materials that have been rated in terms of pleasure, arousal, and dominance and to provide standardized materials that are available to researchers in the study of emotion and attention*" [50, p.1]. Each word of the ANEW dictionary is evaluated by the Self-Assessment Manikin (SAM) scale, an affective rating system [51] that measures three affective dimensions: "valence/pleasure" (from positive to negative), "perceived arousal" (from high to low levels), and perceptions of "dominance / control" (from low to high levels), with a Likert scale ranging from 1 to 9. Any word rated greater than 5 is considered a positive word token, otherwise a negative one.

SenticNet: SenticNet is a publicly available semantic and affective resource for concept-level sentiment analysis, exploiting "*a knowledge base, of 200,000 commonsense* [words] *associated with* [their] *semantics, sentics, and polarity values... In particular, semantics define the* [explicit meaning] *of* [words], *sentics* [identify positive and negative associations words naturally carry with them] (*expressed in terms of pleasantness, attention, sensitivity, and aptitude) and polarity is* [a] *floating number between -1 and +1 (where -1 is extreme negativity and +1 is extreme positivity)"* (https://sentic.net/– last verified August 2020). SenticNet is particularly useful for detecting depressive states because it implements multiple degrees of word associations, making the measurement more accurate [52].

LIWC: Linguistic Inquiry Word Count (LIWC) [53] is a computerized word counting tool, working on Windows and Mac, focused on words linked to emotional and cognitive processes exploiting 4 main linguistic dimensions so-called *Linguistics, Psychological Processes, Relativity* and *Personal Contents*. The *Linguistics* dimension contains pronouns and verb tenses, and provides information about attentional allocation and temporal focus of attention. Dominant emotions and feelings (degree and valence), cognitive processes (that can reflect different cognitive mechanisms like reappraisal, or insecurity about the topic) and social relationships (reflecting how much the person is engaged in social relations) are linguistically described by the *Psychological Processes* dimension. The *Relativity* dimension covers linguistic concepts about attention towards external space, movements and time. The *Personal Contents* dimension comprises linguistic descriptions about interests towards jobs, pleasure activities, physical states and general topics like religion, money, and death. These linguistic dimensions are connected to depressive symptoms [1]. LIWC outputs a text file providing the percentage of words belonging to each dimension, even though each word can be located in more than one dimension. The tool is available in 17 different languages with freely available dictionaries enabling cross-language comparisons.

EmoLex: The NRC Emotion Lexicon (EmoLex) [54] is a list of English words associated with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). The EmoLex dictionary includes the most frequent English nouns, verbs, adjectives, and adverbs and each emotion is associated with a between 500 to 3.000 words' list, by a manual annotation through Amazon's Mechanical Turk service.

SÉANCE: SÉANCE is a tool is included in SALAT (The Suite of Linguistic Analysis Tools), an open-source toolkit that contains tools for automatic linguistic and word affect text analysis (www.linguisticanalysistools.org/tools.html – last verified August 2020). SÉANCE (Sentiment Analysis and Cognition Engine [55]) "*is a text analysis tool that includes 254 core indices and 20 component indices (e.g. Positive/negative adjectives, action, joy, fear/disgust, polarity nouns, polarity verbs, etc.) based on recent advances in sentiment analysis. In addition to the core indices, SÉANCE allows for a number of customized indices including filtering for particular parts of speech and controlling for instances of negation*". SÉANCE is free, easy to use and works on the most used operating systems (Windows, Linux, and Mac). It includes: 1) the Harvard IV-4 dictionary [57]; 2) the Lasswell dictionary [58]; 3) the ANEW database [50]; 4) the GALC database composed of lists of words pertaining to 36 specific emotions and two general emotional states (positive and negative, [59]); 5) the EmoLex database [54]; 6) the SenticNet database [52]; 7) an extension of WordNet [60]; 8) two large polarity lists [61] and 9) the rule-based sentiment analysis system VADER (Valence Aware Dictionary for Sentiment Reasoning [62]. Stasak and Epps [47] exploited SALAT and SÉANCE to predict depression.

IV. Multimodal Sentiment Analysis

Even though the detection and prediction of depressive states through behavioral features was already a sub-challenge of the international Audio/Visual Emotion Challenge (AVEC) in 2013, multimodal sentiment analysis was launched by AVEC 2017 in response to the need of more accurate solutions for affective computing [12]. The AVEC 2017 sub-challenge on depression asked "*participants to build a mathematical model to predict depression severity degrees*" exploiting concurrently audio, video, and text associated to DAIC interviews in the Wizard-of-Oz conditions. These interviews were conducted by an animated virtual agent controlled by a human experimenter [36]. The challenge provided audio features (among those fundamental frequency, MFCC coefficients, and formants frequency values) extracted through the COVAREP toolkit [63], video features (among those facial landmarks, action units, and head poses) obtained through the OpenFace toolkit [64], and text features (among those words' frequency) obtained through LIWC. The multimodal approach, particularly the topic modeling approach based on context-aware analyses of recordings improved significantly the results obtained from those obtained through context-unaware methods [65]. This trend was continued with the AVEC 2019 sub-challenge on Detecting Depression via Artificial Intelligent (AI) tools [66]. There, deep learning algorithms and convolutional neural networks were trained on the fusion of acoustic and linguistic contents of speech, plus facial and head poses features (always using the DAIC-WoZ (Wizard of Oz conditions) databases) to detect depressive states and predict their degree of severity. The results were slightly better than those obtained in 2017, but not outstandingly impressive suggesting further investigations both in structuring the computational models and identifying new invariant behavioral depressive features for their input [13].

V. Conclusion

Depression is the most widespread mental disorder and it has become compulsory to diagnose and prevent this pathology quickly and with increasingly accurate, reliable and objective methods. Textual analyses, considering linguistic attributes, such as grammar, syntax and vocabulary of depressed subjects, have been proved sufficiently different from controls to discriminate between them. In addition, behavioral features that include live signals obtained from speech production, facial expressions, head poses, and body movements have been considered and proved to successfully predict or detect depressive states. The natural next step to further increase the accuracy of detecting and predicting depressive states was to combine behavioral, textual and functional features, implementing "multimodal sentiment analysis". This methodology proved to significantly improve depression's prediction and detection accuracies obtained by using single modalities [12-13, 65]. These improvements were obtained on the DAIC-WoZ database, based on veterans' interviews. These subjects, in addition to depression, also suffer of post-traumatic stress disorders. They were interviewed by a virtual agent and this procedure fixed the interview context and discussion's topics. Subjects shared a war's experience too. All these commonalities among participants may have affected linguistic features, discussion's topics, and even prosodic features which in turn may have affected the detection and prediction accuracy of the proposed fusion procedures. There is a need to test this approach on different databases. These results call for further investigations, and further developments of automatic tools involved in the application of multimodal sentiment analysis, that extent to multiple multimodal databases (currently the most exploited is the DAIC in the WoZ condition), all age groups of users, and signals collected in real life scenarios. The EU community has funded the network research project MENHIR (Mental health monitoring through interactive conversations) aiming to "develop conversational technologies to promote mental health and assist people with mental ill health (mild depression and anxiety) manage their conditions" (https://menhir-project.eu/ – last verified August 2020). The project's partners are collecting interviews conducted by expert psychologists and counsellors. It appears that the language used by patients revolve on different topics, suggesting cross-cultural differences in the behavioral, prosodic, and linguistic displays of depression. Together with macroscopic signals visible and audible during interactional exchanges, also physiological signals such as skin conductance, heart rate, pupil dilation and saccadic eye movements [67-68] have been proved related to depression. Recently other features derived from functional activities, such as drawing and handwriting [69-70] have been used to detect negative moods as depression, anxiety and stress. These investigations exploited large non-clinical databases of 129 and 240 Italian subjects respectively assessed as depressed, anxious, and stressed through the Depression Anxiety and Stress Scale (DASS [71]). Online drawing and handwriting features (such as time keeping in-air or on paper the pen for completing specific drawing – i.e., pentagons drawing - and handwriting tasks – i.e., writing phonetically complete sentences) were used to discriminate depressed subjects from controls either through AI tools (a random forest algorithm [70]) or statistical analyses (ANOVA repeated measures [69]). Results were comparable to those obtained exploiting acoustic and linguistic features (in [70] depression was detected with an accuracy of 73% and a specificity of 99%). To our knowledge, the multimodal approach up to now has neglected these functional and non macroscopic features. Finally, it must be noted that depressed patients call for support 24 hours per day. There is not only a need to automate the prediction and detection of the disease and support clinicians with more reliable information on patients' daily feelings but also the need to automate the assistance of these patients with an empathic conversational assistant that coaches them in absence of a doctor, providing sufficient support to avoid desperate actions until a clinician reaches them. MENHIR aims to provide such continuous support through a conversational vocal assistant.

This short review on sentiment analysis is motivated by the aim to clarify outstanding issues on sentiment analysis tools for depression which have been, to our present knowledge neglected by literature, raises the interests of the CoginfoCom scientific community, and be a useful starting point for interested researchers.

AKNOWLEDGEMENTS

The research leading to these results has received funding from the EU H2020 research and innovation program under grant agreement N. 769872 (EMPATHIC) and N. 823907 (MENHIR), the project SIROBOTICS that received funding from Italian MIUR, PNR 2015-2020, D.D. 1735, 13/07/ 2017, and the project ANDROIDS funded by the program V:ALERE 2019 Università della Campania "Luigi Vanvitelli", D.R. 906 del 4/10/2019, prot. n. 157264,17/10/2019.

REFERENCES

[1] American Psychiatric Association, "Diagnostic and Statistical Manual of Mental Disorders 5th edition [DSM-5]," 2013.

[2] "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis", Lancet, vol. 388(10053), pp.1545–1602, 2016.

[3] World Health Organization, "Depression and Other Common Mental Disorders: Global Health Estimates," Geneva, Switzerland, 2017.

[4] M. Large, "Study on suicide risk assessment in mental illness underestimates inpatient suicide risk," BMJ, 352, i267, 2016.

[5] F. Scibelli et al., "How major depressive disorder affects the ability to decode multimodal dynamic emotional stimuli, Front. ICT,| https://doi.org/10.3389/fict.2016.00016, 2016.

[6] G. Purebl et al., "Joint action on mental health and well-being, depression, suicide prevention and e-health. Situation analysis and recommendations for action. https://ec.europa.eu/health/sites/health/files/mental_health/docs/2017_depression_suicide_ehealth_en.pdf, 2017

[7] K. van Zoonen et al., "Preventing the onset of major depressive disorder: A meta-analytic review of psychological interventions," International Journal of Epidemiology, vol. 43, pp. 318–329, 2014.

[8] J.R. Williamson et al., "Tracking depression severity from audio and video based on speech articulatory coordination," Computer Speech & Language, vol. 55, pp. 40-56, 2018.

[9] M. Valstar et al., "The continuous audio/visual emotion and depression recognition challenge," Proc. of the 3rd Audio/Visual Emotion Challenge, AVEC '13, Barcelona, Spain, 2013.

[10] M. Valstar et al., "3D Dimensional Affect and Depression Recognition Challenge," Proc. of the 4th Audio/Visual Emotion Challenge, AVEC '14, Orlando, FL, 2014.

[11] M. Valstar, et al., "Depression, mood, and emotion recognition workshop and challenge," Proc. of the 6th Audio/Visual Emotion Challenge, AVEC '16, Amsterdam, Netherlands, 2016.

[12] F. Ringeval et al., "Real-life depression, and affect recognition workshop and challenge", Proc. of the 7th Audio/Visual Emotion Challenge AVEC '17, Mountain View, CA, USA, 2017.

[13] F. Ringeval et al., "State-of-mind, detecting depression with AI, and cross-cultural affect recognition", Proc of the 9th Audio/Visual Emotion Challenge AVEC '19, Nice, France, 2019.

[14] J.C. Mundt, A.P. Vogel, D.E. Feltner, W.R. Lenderking, "Vocal acoustic biomarkers of depression severity and treatment response," Biol. Psych., vol. 72, pp. 580–587, 2012.

[15] N. Cummins et al., "A review of depression and suicide risk assessment using speech analysis," Speech Commun. 71, pp 10-49, 2015.

[16] G. Kiss, K Vicsi, "Investigation of cross-lingual depression prediction possibilities based on speech processing". Proc. of 8th IEEE CogInfocom 2017, Debrecen, HU, 2017.

[17] G. Kiss et al., "Detection possibilities of depression and Parkinson's disease based on the ratio of transient parts of the speech" Proc. of 9th IEEE CogInfocom 2018, Budapest, HU, 2018.

[18] M.J.H. Balsters, E.J. Krahmer, M.G.J. Swerts, A.J.J.M. Vingerhoets, "Verbal and nonverbal correlates for depression: a review." Curr. Psych. rev. 8, pp. 227–234, 2012.

[19] S. Scherer et al., "Automatic behavior descriptors for psychological disorder analysis," 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013. IEEE, Shanghai, pp. 1–8, 2013

[20] S. Alghowinem et al., "Head pose and movement analysis as an indicator of depression. Humaine Association Conference on Affective Computing and Intelligent Interaction. 2013, pp. 283–288.

[21] S. Poria, A. Mondal, P. Mukhopadhyay, "Evaluation of the intricacies of emotional facial expression of psychiatric patients using computational models," in Understanding Facial Expressions in Communication. Berlin, Germany: Springer, pp. 199–226, 2015.

[22] J.M. Girard et al., "Nonverbal social withdrawal in depression: evidence from manual and automatic analyses," Image Vis. Comput., vol. 32, pp. 641–647, 2013.

[23] J. Joshi, R. Goecke, G. Parker, M. Breakspear, "Can body expressions contribute to automatic depression analysis?" Proc. 10th IEEE Int. Conf. Workshops Autom. Face Gesture Recognition, pp. 1–7, 2013.

[24] P Lopez-Otero, L. Docio-Fernandez, C. Garcia-Mateo, "A study of acoustic features for depression detection". International Workshop on Biometrics and Forensics (IWBF) Valletta pp: 1-6, 2014.

[25] F. Honig et al., "Automatic modelling of depressed speech: Relevant features and relevance of gender," Proc. of Interspeech 2014, pp. 1248-1252, 2014.

[26] A. Esposito et al., "On the significance of speech pauses in depressive disorders: Results on read and spontaneous narratives" In Esposito A et al. (Eds) Recent Advances in Nonlinear Speech Processing, Smart Innovation, Systems and Technologies, vol. 48, 73-82, 2016.

[27] D. Bennabi, et al., "Psychomotor retardation in depression: a systematic review of diagnostic, pathophysiologic, and therapeutic implications," BioMed research international, https://doi.org/10.1155/2013/158746, 2013.

[28] B. Stasak, J. Epps, R. Goecke, "An investigation of linguistic stress and articulatory vowel characteristics for automatic depression classification," Computer Speech & Language, vol. 53, pp. 140-155, 2018.

[29] T. Taguchi et al., "Major depressive disorder discrimination using vocal acoustic features," Journal of Affective Disorders, vol. 225, pp. 214–220, 2018.

[30] S.B. Davis, P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Signal Process., vol. 28, pp. 357–366, 1980

[31] A. Mendiratta et al., "Automatic detection of depressive states from speech". In Esposito A et al. (Eds) Multidisciplinary Approaches to Neural Computing, Smart Innovation, Systems and Technologies 69, 301-314, Springer International Publishing, 2017

[32] S.S. Haykin, "Neural networks: a comprehensive foundation. Prentice Hall, 1999.

[33] M.N. Stolar, M. Lech, S.J. Stolar, N.B. Allen, "Detection of adolescent depression from speech using optimised spectral roll-off parameters," Biomed J Sci &Tech Res. Vol. 5(1), pp. 4350- 4359, 2018.

[34] J. Mundt et al., ,"Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology", J. Neuroling., vol. 20(1), pp. 50-64, 2007.

[35] F. Scibelli et al., "Depression speaks: automatic discrimination between depressed and non-depressed speakers based on nonverbal speech features". Proc. of IEEE ICASSP 2018, Calgary, AB, Canada, pp. 6842-6846, 2018.

[36] J. Gratch et al., "The Distress Analysis Interview Corpus of human and computer interviews" Proc. of LREC 2014, pp. 3123-3128.

[37] H. Dibekliğlu, Z. Hammal, J.F. Cohn, "Dynamic multimodal measurement of depression severity using deep autoencoding," IEEE J. Biomed. Health Informat., vol. 22(2), pp. 525-536, 2017

[38] M. Nasir et al., "Multimodal and multiresolution depression detection from speech and facial landmark features" Proc. AVEC 2016, Amsterdam, Netherlands, 2016.

[39] A. Pampouchidou, et al., "Depression assessment by fusing high and low level features from audio, video, and text," Proc. AVEC 2016, , Amsterdam, Netherlands, 2016.

[40] K. Collier et al., "Language evolution: syntax before phonology?" Proc. R. Soc. B, vol. 281(1788), http://dx.doi.org/10.1098/rspb.2014.0263, 2014.

[41] N. Shapira et al., "Using computerized text analysis to examine associations between linguistic features and clients' distress during psychotherapy" Journal of Counseling Psychology, published online ahead of print, Apr 30 2020;doi:10.1037/cou0000440, https://pubmed.ncbi.nlm.nih.gov/32352823/, 2020.

[42] K. B. Tølbøll, "Linguistic features in depression: a meta-analysis," Journal of Language Works, 4(2), pp. 39-59, 2019.

[43] B. Pang, L. Lee, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval, vol. 2(1-2), pp 1–135, 2008.

[44] A. Gandomi, M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," Int. J. of Information Management, vol. 35, pp. 137–144, 2015.

[45] E. Cambria, B. Schuller, Y. Xia, C. Havasi, "New avenues in opinion mining and sentiment analysis". IEEE Intelligent Systems, vol. 28(2), pp. 15–21, 2013.

[46] B. Ojamaa, P.K. Jokinen, K. Muischenk, "Sentiment analysis on conversational texts," Proc. of the 20th Nordic Conference of Computational Linguistics, NODALIDA 2015, Vilnius, Lithuania, pp. 233-237, 2015

[47] B. Stasak, J. Epps, "Elicitation design for acoustic depression classification: an investigation of articulation effort, linguistic complexity, and word affect," Proc. of InterSpeech 17, Stockholm – Sweden, 2017.

[48] R.E. Ingram, "Self-focused attention in clinical disorders: review and a conceptual model," Psychological Bulletin, vol. 107(2), pp. 156-176, 1990.

[49] C. Perlini et al., "Linguistic production and syntactic comprehension in schizophrenia and bipolar disorder," Acta Psychiatrica Scandinavica, vol. 126, pp. 363–376, 2012.

[50] M.M. Bradley, P.J. Lang, "Affective norms for English words (ANEW): Instruction manual and affective ratings," Technical report C-1, the center for research in psychophysiology, University of Florida, 1999.

[51] P.J. Lang, "Behavioral treatment and bio-behavioral assessment: Computer applications;" In J.B. Sidowski et al. (Eds.), *Technology in mental health care delivery systems*, pp. 119-137, Ablex Publishing, 1980.

[52] E. Cambria, D. Olsher, D. Rajagopal, "SenticNet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis" Proc. of AAAI 2014, pp. 1515–1521, 2014.

[53] J.W. Pennebaker, R.L. Boyd, K. Jordan, K. Blackburn, "The development and psychometric properties of LIWC2015". The University of Texas at Austin,Technical Report, pp. 1-25, 2015.

[54] S.M. Mohammad, P.D. Turney, "Crowdsourcing a word–emotion association lexicon," Computational Intelligence, vol. 29(3), pp. 436-465, 2013.

[55] S.A. Crossley, K. Kyle, D.S. McNamara, "Sentiment analysis and social cognition engine (SEANCE): An automatic tool for sentiment, social cognition, and social order analysis," Behavior Research Methods, vol. 49(3), pp. 803-821, 2017.

[56] Y.R. Tausczik, J.W. Pennebaker, "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods," Journal of Language and Social Psychology, vol. 29(1), pp. 24–54, 2010.

[57] http://textanalysis.info/pages/category-systems/general-category-systems.php – last verified August 2020

[58] H.D. Lasswell, J.Z. Namenwirth, "The Lasswell value dictionary," New Haven: Yale University Press, 1969.

[59] K.R. Scherer, "What are emotions? And how can they be measured?" Social Science Information, vol. 44, pp. 695–729, 2005.

[60] C. Fellbaum, "WordNet: An electronic lexical database," Cambridge: MIT Press, 1998.

[61] M. Hu, B. Liu, "Mining and summarizing customer reviews," In W. Kim & R. Kohavi (Eds.), Proc. of the tenth ACM SIGKDD international conference on knowledge discovery and data mining, pp. 168–177, Washington, DC, 2004.

[62] C.J. Hutto, E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text". In E. Adar & P. Resnick (Eds.), Proc. of the 8th AAAI conference on weblogs and social media, Palo Alto, pp. 216–225, 2014.

[63] G. Degottex et al., "COVAREP: A collaborative voice analysis repository for speech technologies," Proc. IEEE ICASSP 2014, pp. 960–964, 2014.

[64] T. Baltrůsaitis, P. Robinson, L.P. Morency, "Openface: an open source facial behavior analysis toolkit," Proc. of IEEE WACV 2016, pp. 1–10, 2016.

[65] Y. Gong, C. Poellabauer, "Topic Modeling Based Multi-modal Depression Detection", Proc. of AVEC'17, Mountain View, CA, USA, 2018.

[66] https://dl.acm.org/doi/proceedings/10.1145/3347320#sec5 – last verified August 2020

[67] A.Y. Kudinova et al., "Pupillary reactivity to negative stimuli prospectively predicts recurrence of major depressive disorder in women," Psychophysiology, vol. 53(12), pp. 1836-1842, 2016.

[68] M. Li et al., "Alleviated negative rather than positive attentional bias in patients with depression in remission: An eye-tracking study," J. Int. Med. Res. vol. 44, pp. 1072–1086, 2016.

[69] G. Cordasco et al., "Handwriting and Drawing Features for Detecting Negative Moods" In Esposito A et al. (Eds) Quantifying and Processing Biomedical and Behavioral Signals, Springer International Publishing, vol. 103, pp. 71-86, 2019

[70] L. Likforman-Sulem et al., "EMOTHAW: a novel database for emotional state recognition from handwriting and drawing. IEEE Trans. Hum.-Mach. Syst. 47, 273–284, 2017.

[71] P.F. Lovibond, S.H. Lovibond, "The structure of negative emotional states: comparison of the depression anxiety stress scales (DASS) with the beck depression and anxiety inventories" Behav. Res. Therapy 33(3), 335–343, 1995