**LITERATURE SURVEY**

**TITLE:** “What reveals about depression level? The role of multimodal features at the level of interview questions,”

**ABSTRACT:** Early depression detection can enable timely intervention. Automatic depression detection has relied on features extracted from individual-level data, which may be too coarse to support effective detection. Existing detection models have largely overlooked interview questions commonly used in clinical depression assessment. This research proposes a two-layered multi-modal model for depression detection, which not only extracts features from responses at a level of individual interview questions, but also identifies [semantic categories](https://www.sciencedirect.com/topics/computer-science/semantic-category) of those questions. The evaluation results demonstrate that the proposed model outperforms the state-of-the-art methods for depression detection. The research findings have broad and cross-disciplinary implications.

**TITLE:** “Depression recognition based on dynamic facial and vocal expression features using partial least square regression,”

**ABSTRACT:** Depression is a typical mood disorder, and the persons who are often in this state face the risk in mental and even physical problems. In recent years, there has therefore been increasing attention in machine based depression analysis. In such a low mood, both the facial expression and voice of human beings appear different from the ones in normal states. This paper presents a novel method, which comprehensively models visual and vocal modalities, and automatically predicts the scale of depression. On one hand, Motion History Histogram (MHH) extracts the dynamics from corresponding video and audio data to represent characteristics of subtle changes in facial and vocal expression of depression. On the other hand, for each modality, the Partial Least Square (PLS) regression algorithm is applied to learn the relationship between the dynamic features and depression scales using training data, and then predict the depression scale for an unseen one. Predicted values of visual and vocal clues are further combined at decision level for final decision. The proposed approach is evaluated on the AVEC2013 dataset and experimental results clearly highlight its effectiveness and better performance than baseline results provided by the AVEC2013 challenge organiser.

**TITLE:** “Automated assessment of psychiatric disorders using speech: A systematic review,”

**ABSTRACT:** There are many barriers to accessing mental health assessments including cost and stigma. Even when individuals receive professional care, assessments are intermittent and may be limited partly due to the episodic nature of psychiatric symptoms. Therefore, machine-learning technology using speech samples obtained in the clinic or remotely could one day be a biomarker to improve diagnosis and treatment. To date, reviews have only focused on using acoustic features from speech to detect depression and schizophrenia. Here, we present the first systematic review of studies using speech for automated assessments across a broader range of psychiatric disorders.

**TITLE:** “Deep learning for depression recognition with audiovisual cues: A review,”

**ABSTRACT:** With the acceleration of the pace of work and life, people have to face more and more pressure, which increases the possibility of suffering from depression. However, many patients may fail to get a timely diagnosis due to the serious imbalance in the doctor-patient ratio in the world. Promisingly, physiological and psychological studies have indicated some differences in speech and facial expression between patients with depression and healthy individuals. Consequently, to improve current medical care, many scholars have used deep learning to extract a representation of depression cues in audio and video for automatic depression detection. To sort out and summarize these works, this review introduces the databases and describes objective markers for automatic depression estimation (ADE). Furthermore, we review the deep learning methods for automatic depression detection to extract the representation of depression from audio and video. Finally, this paper discusses challenges and promising directions related to automatic diagnosing of depression using deep learning technologies.

**TITLE:** “Artificial intelligent system for automatic depression level analysis through visual and vocal expressions,”

**ABSTRACT:** A human being's cognitive system can be simulated by artificial intelligent systems. Machines and robots equipped with cognitive capability can automatically recognize a humans mental state through their gestures and facial expressions. In this paper, an artificial intelligent system is proposed to monitor depression. It can predict the scales of Beck depression inventory II (BDI-II) from vocal and visual expressions. First, different visual features are extracted from facial expression images. Deep learning method is utilized to extract key visual features from the facial expression frames. Second, spectral low-level descriptors and mel-frequency cepstral coefficients features are extracted from short audio segments to capture the vocal expressions. Third, feature dynamic history histogram (FDHH) is proposed to capture the temporal movement on the feature space. Finally, these FDHH and audio features are fused using regression techniques for the prediction of the BDI-II scales. The proposed method has been tested on the public Audio/Visual Emotion Challenges 2014 dataset as it is tuned to be more focused on the study of depression. The results outperform all the other existing methods on the same dataset.

**TITLE:** “Detection of clinical depression in adolescents’ speech during family interactions,”

**ABSTRACT:** The properties of acoustic speech have previously been investigated as possible cues for depression in adults. However, these studies were restricted to small populations of patients and the speech recordings were made during patients' clinical interviews or fixed-text reading sessions. Symptoms of depression often first appear during adolescence at a time when the voice is changing, in both males and females, suggesting that specific studies of these phenomena in adolescent populations are warranted. This study investigated acoustic correlates of depression in a large sample of 139 adolescents (68 clinically depressed and 71 controls). Speech recordings were made during naturalistic interactions between adolescents and their parents. Prosodic, cepstral, spectral, and glottal features, as well as features derived from the Teager energy operator (TEO), were tested within a binary classification framework. Strong gender differences in classification accuracy were observed. The TEO-based features clearly outperformed all other features and feature combinations, providing classification accuracy ranging between 81%-87% for males and 72%-79% for females. Close, but slightly less accurate, results were obtained by combining glottal features with prosodic and spectral features (67%-69% for males and 70%-75% for females). These findings indicate the importance of nonlinear mechanisms associated with the glottal flow formation as cues for clinical depression.

**TITLE:** “Multichannel weighted speech classification system for prediction of major depression in adolescents,”

**ABSTRACT:** Early identification of adolescents at high imminent risk for clinical depression could significantly reduce the burden of the disease. This study demonstrated that acoustic speech analysis and classification can be used to determine early signs of major depression in adolescents, up to two years before they meet clinical diagnostic criteria for the full-blown disorder. Individual contributions of four different types of acoustic parameters [prosodic, glottal, Teager's energy operator (TEO), and spectral] to depression-related changes of speech characteristics were examined. A new computational methodology for the early prediction of depression in adolescents was developed and tested. The novel aspect of this methodology is in the introduction of multichannel classification with a weighted decision procedure. It was observed that single-channel classification was effective in predicting depression with a desirable specificity-to-sensitivity ratio and accuracy higher than chance level only when using glottal or prosodic features. The best prediction performance was achieved with the new multichannel method, which used four features (prosodic, glottal, TEO, and spectral). In the case of the person-based approach with two sets of weights, the new multichannel method provided a high accuracy level of 73% and the sensitivity-to-specificity ratio of 79%/67% for predicting future depression.

**TITLE:** “Detecting depression: A comparison between spontaneous and read speech,”

**ABSTRACT:** Major depressive disorders are mental disorders of high prevalence, leading to a high impact on individuals, their families, society and the economy. In order to assist clinicians to better diagnose depression, we investigate an objective diagnostic aid using affective sensing technology with a focus on acoustic features. In this paper, we hypothesise that (1) classifying the general characteristics of clinical depression using spontaneous speech will give better results than using read speech, (2) that there are some acoustic features that are robust and would give good classification results in both spontaneous and read, and (3) that a `thin-slicing' approach using smaller parts of the speech data will perform similarly if not better than using the whole speech data. By examining and comparing recognition results for acoustic features on a real-world clinical dataset of 30 depressed and 30 control subjects using SVM for classification and a leave-one-out cross-validation scheme, we found that spontaneous speech has more variability, which increases the recognition rate of depression. We also found that jitter, shimmer, energy and loudness feature groups are robust in characterising both read and spontaneous depressive speech. Remarkably, thin-slicing the read speech, using either the beginning of each sentence or the first few sentences performs better than using all reading task data.

**TITLE:** “A study of acoustic features for depression detection,”

**ABSTRACT:** Clinical depression can be considered as a soft biometric trait that can help to characterize an individual. This mood disorder can be involved in forensic psychological assessment, due to its relevance in different legal issues. The automatic detection of depressed speech has been object of research in the last years, resulting in different algorithmic approaches and acoustic features. Due to the use of different algorithms, databases and performance measures, deciding which ones are more suitable for this task is difficult. In this work, the performance of different acoustic features for depression detection was explored in a common framework. To do so, a depression estimation approach in which the audio data is segmented and projected into a total variability subspace was used, and these projected data was used to estimate the depression level by performing support vector regression. The data and evaluation metrics were the ones used in the audiovisual emotion challenge (AVEC 2013).

**TITLE:** “Depaudionet: an efficient deep model for audio based depression classification,”

**ABSTRACT:** This paper presents a novel and effective audio based method on depression classification. It focuses on two important issues, \emph{i.e.} data representation and sample imbalance, which are not well addressed in literature. For the former one, in contrast to traditional shallow hand-crafted features, we propose a deep model, namely DepAudioNet, to encode the depression related characteristics in the vocal channel, combining Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) to deliver a more comprehensive audio representation. For the latter one, we introduce a random sampling strategy in the model training phase to balance the positive and negative samples, which largely alleviates the bias caused by uneven sample distribution. Evaluations are carried out on the DAIC-WOZ dataset for the Depression Classification Sub-challenge (DCC) at the 2016 Audio-Visual Emotion Challenge (AVEC), and the experimental results achieved clearly demonstrate the effectiveness of the proposed approach.