ISSN No:-2456-2165

A Review of Depression Analysis using Facial Cues

Snehabharathi S K
Department of Electronics and Instrumentation Engineering
RVCE, Bengaluru, India

Dr. Prasanna Kumar. S. C
Professor, Department of Electronics and Instrumentation Engineering, RVCE, Bengaluru, India

Abstract— Depression is a common mental illness, which is affected by most people of the world. Most of people who are suffering from depression need treatment. By careful examination of Emotions, the early detection of depression is possible. This review presents an in-depth study of the various papers on depression analysis from emotions of facial images of patients. There are various methods used for facial recognition, feature extraction and classification of depression. There are various datasets used AVEC, Clinical Depression dataset from BlackDog Institute and others are used. 5 facial recognition and 5 feature extraction methods are studied. We found that the literature has primarily focused on viola jones method for face detection methods (54%) and deep learning methods for feature extraction (45%). Discussion on limitations of the methods conceived over the past year as well as future perspectives on various methods to improve performance are also provided.

Keywords:- *Depression Detection, Viola Jones Face Detection, Deep Learning.*

I. INTRODUCTION

A person is affected by depressive disorder, it hinders the normal functioning of the patient, and it is painful for both the person with the disorder and the care taker. Depression is a common issue but results in serious illness but with proper diagnosis and treatment condition of the patient will get better. Intensive research into this illness has resulted in the development of medications, psychotherapies and other methods to treat people with this disabling disorder. According to World Health Organization (WHO) depression is commonly worldwide mental disorder that affects more than 300 million people regardless of their ages[1]. The clinical evaluations depends on the depression screening instrument used like Beck Depression Inventory (BDI) [2], Hamilton Rating Scale for Depression(HRSD)[3],Quick Inventory of Depress Symptoms Self Report(QIDS-SR)[4]. In this review paper there is a discussion on various methods of depression analysis based on facial emotions. There are many papers from various publications like IEEE, Springer etc. These papers have used many datasets like are University of Pittsburgh depression dataset (Pitt) [5], a Black Dog Institute depression dataset (BlackDog) [6], and Audio/Visual Emotion Challenge depression dataset(AVEC)[7], Japanese Female Facial Expression(JAFFE)[8] Database etc. have been used. JAFFE contains only images of different emotions and doesn’t deal specifically with depression but the authors have correlated basic emotion for depression analysis.

II. LITERATURE SURVEY

The following section discusses about various depression detection techniques using facial cues. The process has 5 stages i.e. Preprocessing, Face detection, Feature extraction, Feature selection and Classification. Table 1 shows various methods used for these stages in various papers. N. C. Maddage et al.,[9] have used viola Jones face detector, Gabor wavelet feature extraction and classification were done using GMM. Both gender based and gender independent modelling is done. Accuracy obtained was 78.5%. They suggested that performance can be improved with larger dataset. J. F. Cohn et al.,[10] have developed own dataset of depression patients. They have used FACS model for face detection, Active appearance model for feature extraction and SVM classifier. Accuracy of 79% was obtained. They suggested that the use of multimodal techniques along with voice processing can improve performance. I. T. Meftah et al.,[11], have used Plutchik model to detect depression from emotions. They have used KNN classifier and classified based on number of successive negative days in the period of 25 days.

Steps	Methods
1.Pre-processing	Normalization, OpenFace, SG filter
2.Face detection	Viola jones face detector, Facial Action Coding System (FACS), Space-time Interest Points based on Histogram of Gradient (STIP-HOG), Eigen face recognition, Kanade-Tomasi Lucas (KLT) tracker
3.Feature Extraction	Gabor Wavelet features, Active Appearance Models (AAM), Convolutional Neural Networks (CNN), Stacked Denoising Autoencoder (SDAE), Histogram of Oriented Gradient (HOG)
4.Feature Selection	Principal Component Analysis (PCA), Min-Redundancy Max-Relevance (mRMR), Correlation based feature selection(CFS).
5.Classifier	Gaussian Mixture Models (GMM), Surface Vector Machine (SVM), K Nearest Neighbor (KNN), Deep Convolutional Neural Networks (DCNN), Random Forest (RF).

Table 1:- Stages and methods used

S. Alghowinem et al.,[12] have used AAM model for feature extraction, GMM and SVM for classification. They have used Blackdog depression dataset. Accuracy of 76.8% is obtained. They suggested fusing multimodal approach can provide higher performance. J. Joshi et al.,[13] have used Pittsburgh dataset, STIP-HOG method for feature extraction and SVM for classification. Accuracy of 91.7% is obtained. They suggested fusion scenario as part of specific histogram may contain overlapping information due to occlusion may improve results. S. Alghowinem et al.,[14] have used BlackDog dataset. They have used AAM features and GMM and SVM classifiers. Accuracy of 71.2% is obtained. S. Alghowinem et al.,[15] have located 74 points in the eye region and 126 statistical features from AAM are extracted. Gender based and gender independent classification was done using SVM classifier and overall accuracy of 75% was obtained. Y. Katyal et al.,[16] have used JAFFE dataset, viola jones and Eigen face recognition methods. SVM was used for Classification. Accuracy of 70% was obtained. S. Alghowinem et al.,[17] have used AVEC, Pittsburg and Blackdog depression datasets. They have used AAM features along with Head pose measurement. SVM classifier was used. Accuracy of 73.1% was obtained. They had overfitting problem and if overfitting is reduced then results could be improved. O. M. Alabdani et al.,[18] have used facial expression and body movement features extracted from ANN and SVM classifier was used. They suggested by using physiological features like heart beat can be used for analysis. A. Pampouchidou et al.,[19] have used AVEC dataset, KLT tracker for face detection, KNN for classification. Accuracy of 74.5% was obtained. They suggested to develop general person independent approach and combination of methods can improve accuracy. X. Li et al.,[20] have used correlation-based feature selection method, KNN, SVM, RF classifier. Accuracy of KNN was 81%, SVM was 76.4%, RF was 79.3%. They suggested adding new features, using different methods to increase accuracy and detection rates. S. Alghowinem et al.,[21] have used Blackdog dataset, extracted 185 statistical features from AAM and SVM classifier. Accuracy of 79% was obtained. They suggested use of large dataset and fully automated system can improve performance. H. Dibeklioğlu et al.,[22] have used AAM feature extraction, SDAE feature selection and GMM classification. The accuracy obtained was 72.9%. The limitation is that it cannot detect dynamic features. A. Pampouchidou et al.,[23] have AVEC dataset, openface for preprocessing missing data, extraction of statistical features and classification using different methods done for both gender based and gender independent classification. They suggested that by adding new features performance can be improved. Maarten Milders et al.,[24] have studied the depression patients and have found that depression may affect the detection of positive stimuli. Patients with depression detected fewer happy faces that matched healthy patients. Q. Wang et al.,[25] have used AVEC dataset, extracted facial features using AAM. They have extracted 49 statistical features and SVM classifier was used. They obtained accuracy of 78.85%. Gavrilescu, M et al.,[26] have used 3 layered neural networks with 16 personality factors. Facial features were extracted using FACS. Accuracy of 18% was obtained. They have given relation between Action Units and 16 personality factors. Pampouchidou, et al.,[27] have used AVEC dataset. They have used Gabor filter, DCNN and HOG for feature extraction. Motion representation is included by using motion histogram image. Accuracy of 87.4% was obtained. S. Al-gawwam et al.,[28] have used AVEC dataset, SG filter for preprocessing, eye landmarks and eye blink average duration of 150 to 300ms were extracted. Many classifiers like SVM etc were used. Accuracy of 92.95% was obtained.

III. DATASET DESCRIPTION

The datasets used in the literature reviewed in this study are AVEC[7], BlackDog [6] and University of Pittsburgh depression dataset (Pitt)[5]. For easier reference, Table 2 summarizes and compares the selected subsets of each dataset. This table gives specifications like language used, Male to female ratio, total time duration, hardware used, performance measure, sampling rate etc.

Dataset	**BlackDog**	**Pitt**	**AVEC**
Language	English (Australian)	English (American)	German
Classification	Severely Depressed /Healthy Control	Severe/Low depression	Severe/Low depression
Number of subjects per class	30	19	16
Males-Females	30-30	14-24	9-23
Procedure	open ended questions interview	HRSD clinical interview	human-computer interaction experiment (story telling)
Symptom severity measure	QIDS-SR	HRSD	BDI
Mean score (range)	19 (14-26)	Severe:22.4 (17-35) / Low:2.9 (1-7)	Severe:35.9 (30-45) / Low:0.6 (0-3)
Equivalent QIDS-SR Score [33]	19 (14-26)	Severe:17 (13-26) / Low:2 (1-5)	Severe:20 (16-22) / Low:1 (0-2)
Total Duration (minutes)	509	355.9	33.2
Average duration / subject (in min)	8.4 (± 4.4)	9.4 (± 4.3)	1.0 (± 0.8)
Hardware	1 camera,1 microphone	4 cameras, 2 microphones	1 web camera,1 microphone
Audio sampling rate	44100 Hz	48000 Hz	44100 Hz
Video sampling rate	30 fps	30 fps	30 fps

Table 2:- Summary of datasets[17]

IV. REVIEW OF METHODS

The various methods used for pre-processing, advantages and limitations are tabulated in Table 3. Normalized fiducial points: There are 49 facial fiducial points. It is normalized to remove noise and scaled for proper detection. Later the movements like head nods, turns and inclinations are smoothed prior analysis. OpenFace: The 2D facial landmarks are detected and aligned images are extracted. Only detected images are used for further processing. SG filter: It reduces the effect of irregular noise and keeps relevant signal information. SG filter is calculated as,

$$S_J^* = \frac{\sum_{i=-m}^{i=m} C_i S_j + i}{N} \quad (1)[26]$$

where S is the main signal, S* is the filtered signal, Ci is the constant for the ith smoothing, and N is the data samples number in the smoothing window which equals to 2m + 1, where m refers to the half-width of the smoothing window. Finally, j is the running index of the ordinate data in the original data table. It uses smoothing along with differentiation. The face detection algorithms along with their advantages and limitations are given in Table 4. Viola jones face detection: It uses Haar features along with AdaBoost learning algorithm. It detects visual features from large set and detects face regions. This program is directly used from OpenCV library FACS: It uses 17 Action units (AUs) which are used to detect facial features related to depression. The interval, mean duration ratio of the onset phase to total duration, and the ratio of onset to offset phase is computed. The STIP-HOG framework: It uses 2D Harris corner point detector, HOG is calculated for each frame. It captures even small change in a video. Large data subsets are created. Hence it takes more computation time. Eigen Facial Recognition: It minimizes variance within a class and maximizes variance between classes simultaneously. Eigenvectors are dependent on orthogonal linear transformation. To find features firstly, Organizing the data set into single matrix, mean calculation and subtract mean from each dimension to know the direction of maximum variance. KLT tracker: Firstly, face region is initialized and tracking video using KLT tracker. A pseudo-image of face is processed by Curvelet Transform. LBP descriptor is calculated to form feature vector. It preserves the motion information with short feature vector.

Various feature extraction methods are use Table 5 tabulates these methods. Gabor wavelet features extracted at different facial landmarks. It is represented by a feature vector. Training is done using the Gaussian mixture models (GMM). Gabor function (initial wavelet) is given by,

$$g_{\alpha,\xi,a,b}(x) = |a|^{-1/2} g_{\alpha,\xi}\left(\frac{x-b}{a}\right)$$

(2)[9]

for a ∈ R + (scale) and b ∈ R (shift). The energy localized around x = 0 as and wavelets are normalized. The discrete set of gabor wavelets forms a frame. Active Appearance Models(AAM) performs a gradient descent search to fit appearance and shape of the model. The shape s of an AAM is described by a 2D mesh, which is triangulated. The mesh vertex coordinates describe the shape. The shape can be expressed as a base shape s0 plus a linear combination of m shape vectors s.

Various methods are used for selection of features. They are PCA, CFS and mRMR methods. PCA is a statistical technique used to measure association of between the variables, direction of the data and its relative importance and allows to remove those eigenvectors which are not important. The principal components(T) of X is T=X.W. where W is a q*q weights matrix are the eigenvectors of XTX. Correlation based feature selection (CFS) identifies features subset which are more correlated with the class. Finally, 6 features including count of fixation, frequency, average of fixation duration, mean of pupil size mean and count of saccade etc. were selected for classifying movement of eye. The mRMR algorithm [11] was used for feature selection. mRMR is an incremental method for minimizing redundancy while selecting the most relevant features based on mutual information.

Pre-processing Methods	Advantages	Limitations
1.Normalization	It reduces tracking errors.	Data inconsistency.
2.OpenFace	Detects facial landmarks and resizes image.	It is not efficient for blurred images.
3.SG Filter	Reduces noise and stores data needed.	It requires predetermined filter values for better result.

Table 3:- Pre-processing methods

Face Detection Algorithms	Advantages	Limitations
1.Viola Jones detector	Features are invariant to pose and orientation change.	Hardship to locate features because of illumination, noise, occlusion and complex background.
2.FACS	Allows detailed analysis of facial expression events.	Limited to facial expression. Difficult to code the dynamic movements.
3.STIP-HOG	It identifies small changes.	The large data subsets makes it complex.
4.Eigen face recognition	Recognition is simple and efficient, raw data can be used.	Recognition rate decreases under varying pose and illumination. It is sensitive to scale.
5.KLT tracker	Preserving the motion data.	It fails to track if displace is large.

Table 4:- Face detection algorithms

Feature Selection Methods	Advantages	Limitations
1.Gabor Wavelet features	Better suited to spatial frequency tuning. multi-resolution and multi-orientation properties.	High dimension and high redundancy.
2.AAM	An effective means to separate identity and intra-class variation.	Model results rely on starting approximation.
3.CNN	Accuracy in image recognition problems.	High computational cost, slow and needs large training data.
4.SDAE	It transforms high-dimensional, noisy data to a lower dimensional, meaningful representation.	Overfitting problem.
5.HOG	It shows invariance to geometric and photometric changes.	It is variant to object orientation as it has small spatial regions.

Table 5:- Feature extraction methods

Classifier	Advantages	Limitations
1.GMM	It classifies static postures and non-temporal pattern recognition.	It fails when dimensionality of data is too high.
2.SVM	It works with unstructured data and gives better results.	It takes long time for training larger datasets.
3.KNN	Simple classifier.	It only uses the training data for classification.
4.DCNN	Higher performance. It can be adapted to new problems relatively easily	Requires a large amount of data. It is extremely computationally expensive to train.
5.RF	Extremely flexible and have very high accuracy even in case of missing data	Complexity

Table 6:- Classifiers Used

Let Sm−1 be the set of selected m − 1 features, then the mth feature can be selected from the set {F − Sm−1} as:

$$\max_{f_j \in F - S_{m-1}} \left[I(f_j, c) - \frac{1}{m-1} \sum_{f_i \in S_{m-1}} I(f_j, f_i) \right]$$

(3)[11]

where I is the mutual information function and c is a target class. F and S denote the original feature set, and the selected sub set of features, respectively.

There are 5 classifiers used. Table 6 tabulates these classifiers with their advantages and limitations. Gaussian Mixture Model (GMM) gives very high performances in classifying video content. Here K number of Gaussian densities cover the feature space with feature vectors. Each Gaussian density is called a component of the mixture. The statistical parameters like mean, variance and the weight associated Gaussian component are trained. Deep learning is tool which is self-learning. It identifies patterns in datasets. It can be designed to contain many intermediate layers for extraction of features. It is more efficient than other networks. It has the convolutional and pooling layers. In convolution layer there are weights connected to feature maps. This weighted sum is fed into the Rectified Linear Unit (ReLU). ReLU is simple and efficient and also it improves convergence while classification. It rectifies and avoids reducing gradient problem. These methods have been used in depression analysis using facial cues and provides good performance in classification of datasets. The activation function is m(h) = max(α*h, h) . Where h is the data sample and α is learning rate, which is adjusted to reduce error. Support Vector Machines(SVM) algorithm is supervised learning algorithm. It's a binary linear classifier which classifies samples into two classes after training by recognising the suitable hyperplane which has maximum distance between the classes K-Nearest Neighbor algorithm (KNN) classification is used to classify similar data based on Selected Euclidean distance from the test instance. The classification result is defined as linear combination of the emotional class. Random Forest gives more accurate prediction by building several decision trees and later by merging them. Let training set U = u1, ..., un with the responses V = v1, ..., vn. It randomly selects replacement for training set and fits trees to that sample location. Later the predictions for test sample is done by mean value of the predictions from all the regressed trees.

V. CONCLUSION

The Various methods used for Depression detection using facial cues. The Deep learning methods are more efficient and provides better results up to 90%. The Random Forest is the most robust and accurate classifier as it gives stable results even in case of missing data. Various other Deep learning methods can be explored for better results and cross-cultural datasets can be developed for better results.

REFERENCES

[1]. Depression and Other Common Mental Disorders: Global Health Estimates. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.
[2]. A. J. Rush, M. H. Trivedi, H. M. Ibrahim, T. J. Carmody, B. Arnow, D. N. Klein, J. C. Markowitz, P. T. Ninan, S. Kornstein, R. Manber et al., “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,” Biological psychiatry, vol. 54, no. 5, pp. 573–583, 2003.
[3]. M. Hamilton, “A rating scale for depression,” Journal of neurology, neurosurgery, and psychiatry, vol. 23, no. 1, p. 56, 1960.
[4]. A. T. Beck, R. A. Steer, R. Ball, and W. F. Ranieri, “Comparison of beck depression inventories-ia and-ii in psychiatric outpatients”, Journal of personality assessment, vol. 67, no. 3, pp. 588–597, 1996.
[5]. S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Breakspear, and G. Parker, “From Joyous to Clinically Depressed: Mood Detection Using Spontaneous Speech,” in Proc. FLAIRS-25, 2012, pp. 141–146.
[6]. Y. Yang, C. E. Fairbairn, and J. F. Cohn, “Detecting depression severity from intra- and interpersonal vocal prosody,” IEEE Trans. on Affective Computing, 2013.
[7]. M. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, S. Schnieder, R. Cowie, and M. Pantic, “AVEC 2013: the continuous audio/visual emotion and depression recognition challenge,” in Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge. ACM, 2013, pp. 3–10.
[8]. M J Lyons, J Budynek & S Akamatsu, "Automatic Classification of Single Facial Images", in IEEE Transactions on Patter Analysis and Machine Intelligence 21 (12): 1357-1362 (1999).
[9]. N. C. Maddage, R. Senaratne, L. A. Low, M. Lech and N. Allen, "Video-based detection of the clinical depression in adolescents," 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, 2009, pp. 3723-3726.
[10]. J. F. Cohn et al., "Detecting depression from facial actions and vocal prosody," 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, Amsterdam, 2009, pp. 1-7.
[11]. I. T. Meftah, N. Le Thanh and C. Ben Amar, "Detecting depression using multimodal approach of emotion recognition," 2012 IEEE International Conference on Complex Systems (ICCS), Agadir, 2012, pp. 1-6.
[12]. S. Alghowinem, R. Goecke, M. Wagner, G. Parkerx and M. Breakspear, "Head Pose and Movement Analysis as an Indicator of Depression," 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, 2013, pp. 283-288.
[13]. J. Joshi, A. Dhall, R. Goecke and J. F. Cohn, "Relative Body Parts Movement for Automatic Depression Analysis," 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, 2013, pp. 492-497.
[14]. S. Alghowinem, "From Joyous to Clinically Depressed: Mood Detection Using Multimodal Analysis of a Person's Appearance and Speech," 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, 2013, pp. 648-654.
[15]. S. Alghowinem, R. Goecke, M. Wagner, G. Parker and M. Breakspear, "Eye movement analysis for depression detection," 2013 IEEE International Conference on Image Processing, Melbourne, VIC, 2013, pp. 4220-4224.
[16]. Y. Katyal, S. V. Alur, S. Dwivedi and Menaka R, "EEG signal and video analysis based depression indication," 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, Ramanathapuram, 2014, pp. 1353-1360.
[17]. S. Alghowinem, R. Goecke, J. F. Cohn, M. Wagner, G. Parker and M. Breakspear, "Cross-cultural detection of depression from nonverbal behaviour," 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, 2015, pp. 1-8.
[18]. O. M. Alabdani, A. A. Aldahash and L. Y. AlKhalil, "A framework for depression dataset to build automatic diagnoses in clinically depressed Saudi patients," 2016 SAI Computing Conference (SAI), London, 2016, pp. 1186-1189.

[19]. A. Pampouchidou et al., "Video-based depression detection using local Curvelet binary patterns in pairwise orthogonal planes," 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, 2016, pp. 3835-3838.

[20]. X. Li, T. Cao, S. Sun, B. Hu and M. Ratcliffe, "Classification study on eye movement data: Towards a new approach in depression detection," 2016 IEEE Congress on Evolutionary Computation (CEC), Vancouver, BC, 2016, pp. 1227-1232.

[21]. S. Alghowinem et al., "Multimodal Depression Detection: Fusion Analysis of Paralinguistic, Head Pose and Eye Gaze Behaviors," in IEEE Transactions on Affective Computing, vol. 9, no. 4, pp. 478-490, 1 Oct.-Dec. 2018.[22] H. Dibeklioğlu, Z. Hammal and J. F. Cohn, "Dynamic Multimodal Measurement of Depression Severity Using Deep Autoencoding," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 2, pp. 525-536, March 2018.

[22]. Maarten Milders, Stephen Bell, Emily Boyd, Lewis Thomson, Ravindra Mutha, Steven Hay, Anitha Gopala,"Reduced detection of positive expressions in major depression",Psychiatry Research,Volume 240,2016,Pages 284-287, ISSN 0165-1781.

[23]. Q. Wang, H. Yang, Y. Yu, Facial Expression Video Analysis for Depression Detection in Chinese Patients, J. Vis. Commun. Image R. (2018).

[24]. Gavrilescu, M. & Vizireanu N.J," Predicting the Sixteen Personality Factors (16PF) of an individual by analyzing facial features" EURASIP Journal on Image and Video Processing (2017) 2017: 59. Springer International Publishing, Online ISSN: 1687-5281.

[25]. Pampouchidou, A., Pediaditis, M., Maridaki, A. et al. "Quantitative comparison of motion history image variants for video-based depression assessment" EURASIP Journal on Image and Video Processing (2017) 2017: 64. Springer International Publishing Online ISSN:1687-5281.

[26]. S. Al-gawwam and M. Benaissa, "Depression Detection From Eye Blink Features," 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Louisville, KY, USA, 2018, pp. 388-392.

[27]. S. Mantri, D. Patil, P. Agrawal and V. Wadhai, "Cumulative video analysis based smart framework for detection of depression disorders," 2015 International Conference on Pervasive Computing (ICPC), Pune, 2015, pp. 1-5.

[28]. A. Pampouchidou et al., "Facial geometry and speech analysis for depression detection," 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, 2017, pp. 1433-1436.

[29]. T. Yang, C. Wu, M. Su and C. Chang, "Detection of mood disorder using modulation spectrum of facial action unit profiles," 2016 International Conference on Orange Technologies (ICOT), Melbourne, VIC, 2016, pp. 5-8.

[30]. S. Song, L. Shen and M. Valstar, "Human Behaviour-Based Automatic Depression Analysis Using Hand-Crafted Statistics and Deep Learned Spectral Features," 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, 2018, pp. 158-165.

[31]. L. Yang, D. Jiang and H. Sahli, "Integrating Deep and Shallow Models for Multi-Modal Depression Analysis — Hybrid Architectures," in IEEE Transactions on Affective Computing.

[32]. [32] E. A. Stepanov et al., "Depression Severity Estimation from Multiple Modalities," 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom), Ostrava, 2018, pp. 1-6.