Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 32s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

User Responsive Automatic Method for Real-Time Depression Detection Using Deep Neural Network

Jiger Prakashchandra Acharya[1]
Research Scholar, Gujarat Technological University, Ahmedabad, Gujarat, India[1]
jigeracharya@gmail.com[1]

Dr.MilindS.Shah[2]
Professor, Electronics & Communication Dept., S.S.E.C-Bhavnagar, Gujarat, India[2]
mssvgec@gmail.com[2]

Cite this paper as: Jiger P. Acharya, Dr. Milind S. Shah, (2025) User Responsive Automatic Method for Real Time Depression Detection Using Deep Neural Network. *Journal of Neonatal Surgery*, 14 (32s), 4286-4297.

ABSTRACT

During last one decade the mental health is becoming a major concerning issue to society as number of various socio - economical, personnel and societal issues related to mental health are increased in exponential manner. Depression is mood disorders which result in severe disabling conditions which affect person's ability to cope with routine life and real-life challenges which are dynamic in nature. It may occur when person remain more than two weeks in negative state of mind continuously. Depression has observable behavioral symptoms related to affective and psychomotor domains which can be identified. Classical approaches majorly depend on person's behavioral analysis and family observations during clinical interviews, which are effective if it can be precisely defined and assessed but persons have tendency to conceal it so they are less effective while proposed method with Deep Learning techniques perform the said task with accuracy of 79% and open a new era for health care domain.

Keywords: *Mental health, Depression, SRI, PHQ-9, DMI-10, Visual behavioral analysis, Audio Video Emotion Challenge (AVEC), Visual features, DSM, Deep Neural Network (DNN) etc.*

1. INTRODUCTION

Depression will be leading cause of disability globally in all age groups which severely affects the global health index and social impact. It is an ongoing problem which may resolve with proper care but may reoccur via various risk factors. The medical community does not fully understand the causes of depression but mostly known various risk factors like environmental, psychological, social, financial factors, drug addictions change in genetic features, brain's neurotransmitter levels, and accidental events. Early detection plays a crucial role as by positive counseling, various physical and mental exercises like meditation and proper medication may resolve depression else situations become worst up to suicidal tendency. Human are expressive through verbal and nonverbal communication channels both channels are important to express feelings [1], [2]. Person can remain silent to suppress verbal communication and by fake expression or conceal expression can hide real scenario still researcher believes that facial expression is crucial input to present emotional status of person so psychologists try to map facial expressions to emotional states. When Person is feeling depression both verbal and nonverbal channel give strong indicators [3]. Number of researchers has contributed their work to identify depression which closely related with psychology, affective, computer and clinical domains so in real time it's a multi-challenging task. As domain start with Psychology, Computer Science and Clinical domain and last long to social domain the major key challenges are mentioned as under to map psychological domain to computer and validate it with behavioral/clinical predictions [4], [5].

2. METHODS

Human express their emotions through multiple channels which can be monitored for emotion analysis.[3],[4],[5].There are many approaches like Clinical, Visual, Verbal, Colour psychology, Self Response clinical inventory, EEG [6]etc. to determine the mental health of a person Major contribution of specific domain is as under.

Table 1: Authors contribution of domain related research

Sr	Authors	Type	Dataset
1	Becker et al. (1994)	Audio/Video	Dementia Bank Database Reddit Self-reported
2	Valstaret al. (2013)	Audio/Video	AVEC 2013
3	Pradhan et al. (2014)	Text data	SemEval-2014 Task 7
4	Lieberman et al. (2013)	Text data	Crisis Text Line
5	Valstaretal. (2014)	Audio/Video	AVEC 2014
6	Gratch J et al. (2014)	Audio/Video	The Distress Analysis Interview Corpus
7	Valstaretal. (2016)	Audio/Video	AVEC 2016
8	Videbech (2016)	Audio/Video /Report	The Danish Depression Database
9	Yates et al. (2017)	Text data	Depression Diagnosis (RSDD) dataset
10	Garcia-Ceja et al. (2018)	Audio/Video	Depression
11	Komal Anadkat et al. (2022)	Audio/Video/Social media text/EEG	FER2013, Ravdess, Kaggle's social media dataset, and EEG dataset

Visual as major channel for detection:

Face is the major channel of emotions indication. People reflect their feelings knowingly or unknowingly by their facial expression so face is considered as prime part of body for expression so based on this fact Ekman [1] categorize standard expression of six fundamental feelings along with Head orientation, posture and movement majorly use to recognize depression [7], [8].Girard et al [9] explore strong co-relation between visual direction and depression.

Table 2: List of major Visual Features for Behaviour Detection

Features	Association with Facial Features
1	Eyelid opening /squinting movement
2	Outward appearance event with inconstancy with force
3	Head appearance with direction and deployment
4	Face movement
5	Pupil enlargement/inclination
6	Iris development
7	Grin force and term
8	Eye stare
9	Dismissal articulation/negative event
10	Mouth corner appearance

In Colour psychology-based approach reveals how various colours impact person's mood and so behaviors. It enlightens colour impact on person's emotional response based on age and cultural background etc. [10]. Major evidence in this emerging area is based on the feedback result of anecdotal, but domain expert has made a few important conclusions that how colour theory correlate with behaviors. In 2020 study that surveyed the emotional associations of 4598 persons from 30 countries and conclude that people commonly associate certain colours with specific emotions as below:

Table 3: Colour association with Emotion and Mood

Colour	Percentage	Emotions	Mood Category
Black	51%	Sad	negative
White	43%	Relief	Positive
Red	68%	Love	Positive
Blue	35%	Relief	Positive
Green	39%	Contentment	negative
Yellow	52%	Joy	Positive
Purple	25%	Pleasant	Positive
Brown	36%	disgust	negative
Orange	44%	Joy	Positive
Pink	50%	Love	Positive

Many contributors used fusion of above in specific group called multi-modal data in which persons reflect their feelings through multiple channels [11], [12]. Which are more accurate and reliable than focusing only on a single modal like audio, video, text, or EEG etc. [6].

Real-time depression detection methodology:

This research develops a depression detection system which incorporates multiple data as inputs from image, video (stored and live) [ ]and response provided by user in quiz assessment. as shown in Fig. 1. Data acquisition, data processing, system construction, and model performance evaluation are described as follows

Model take input data from various channels like camera for static and video images input, Stored image and/or video upload options for testing is also available and third input is of quiz in which user response are process and based on the outcomes of all channels person's negative mood and its severity is identify.

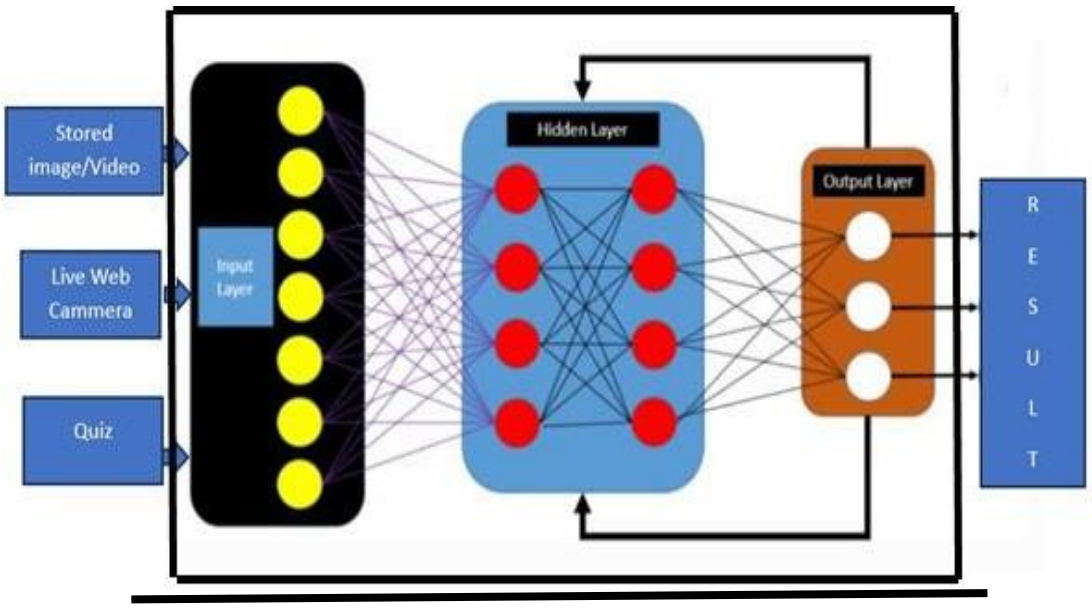

Figure 1: Depression detection Deep neural network model

Figure 2: Prototypic Facial Expression

Figure 3: Testing Facial Expression intra-class variability

Figure 4: Testing Facial Expression inter-class variability

As shown in above figure face images are classified broadly into positive and negative groups. Happy is positive bias images where Sad, Fear, Angry, Surprise and Disgust are negative biased images. Neutral has no biasing as per its name.

Data acquisition:

Depression detection model obtained data from three sources which are static/dynamic images capture from camera both live and stored and quiz responses.

Data channel 1: Input image from Camera in real time video

Image and/or Video from camera is captured and continuously images are identified in categorical expression and positive and negative expression is determined. As shown in above figure face images are classified broadly into positive and negative

groups. Happy is positive bias images where Sad, Fear, Angry, Surprise and Disgust are negative biased images. Neutral has no biasing as per its name.

Data channel 2: Stored image and video

Stored Image and/or Video from camera is captured and continuously images are identified in categorical expression and positive and negative expression is determined.

Data channel 3: Quiz response

First Quiz has unique general questionnaires associate with physical and mental health based on the answer selected by user it categorizes person's mood at Neutral, Happy and Sad identified from quiz response based on past 2 or 3 days experience.

Table 4: Self Response Inventory

Sr	Question (Select option based on past 2-3 days experience)	Outcome	Yes	No	Not Decided
1	Have you taken enough sleep?	Rest			
	Evaluates your physical well-being and rest. Adequate sleep is essential for health and productivity.				
2	Have you taken enough healthy food (in bf / lunch /dinner)?	Food			
	Assesses whether you're meeting your nutritional needs for energy and overall health.				
3	Have you done any exercise/yoga/workout?	Exercise			
	Reflects your physical activity levels, which are important for maintaining health, fitness, and mental well-being.				
4	Are you enjoying your daily routine?	Enjoyment			
	Indicates satisfaction and contentment with your daily life and habits.				
5	Can you select your clothes by yourself?	Decision			
	Measures independence and autonomy in decision-making and self-care.				
6	Can you prepare / suggest foods? (For bf / lunch /dinner)?	Selection/Creativity			
	Evaluates self-sufficiency in meal planning and preparation.				
7	Have you interact with your family members to share your thoughts or feelings?	Sharing feelings			
	Assesses emotional connectivity and communication with loved ones.				
8	Are you comfortable to communicate with your neighbours/colleuges/or friends?	Communication			
	Measures your social comfort and ability to engage with people in various settings.				
9	Do you spare some time to do your liking things?	Hobby /liking			
	Reflects your balance between responsibilities and personal enjoyment or hobbies.				
10	Can you read newspaper/books/articles easily?	Concentration			
	Evaluates cognitive function, literacy, and engagement with information.				
11	Can you watch TV/mobile/Movie easily?	Concentration			
	Measures ease of leisure activity and mental relaxation.				
12	Are you satisfied with your individual achievement/contribution to family/Society?	Self role			

	Reflects self-assessment of your accomplishments and your sense of purpose and fulfillment of your efforts and knowledge/Skills and attitude.				
13	Do you have interest in things which appeal you a lot previously?	Active Participation			
	Examines personal interests and desires related to romantic/pleasure connections.				
14	Do you feel your life going smoothly?	Satisfaction			
	Reflects your perception of life's trajectory and personal well-being.				
15	Do you feel tomorrow is better than present?	Optimistic			
	Indicates optimism and hope for the better future.				
16	Do you feel you are unique and no one can replace yourself?	Self-esteem			
	Reflects self-worth and a sense of individuality.				
17	Do you feel stress of your study/work /responsibilities?	Stress			
	Whether you experience feelings of pressure, anxiety, or overwhelm due to your academic, professional, or personal duties.				
18	Do you feel nothing going on track and after a lots of efforts from your side u can't do anything?	Out of control situation			
	if you ever feel like, despite putting in a lot of hard work and effort, things still aren't progressing as they should. It reflects a sense of frustration or helplessness,				
19	Do you feel frustrated or aloneness?	Helplessness			
	Reflects feelings of frustration, helplessness, or burnout. The outcome of this question could reveal several underlying emotional or psychological states, such as: Frustration or Disappointment, Helplessness or Powerlessness, Burnout, Lack of Progress or Direction, Self-Doubt.				
20	Selection of Colour base on your mood from these: Black/Yellow/Red/None	Colour			
	Mood association with colour.				
21	Have you come across with negative thoughts /suicidal thoughts?	Negative			
	Negativity				

If person having negative response to more than 7 questions of above it suggest negative mood of the person .Depending on the outcome of mood response from all of three channels are negative traditional Depression detection test PHQ-9[14] is suggested and based on the scores obtained from responses given in the quiz score is converted from numeric to depression outcome. The scores are converted in three classes as follows:

Table 5: Classification of S.R.I. Score

Sr	Score range	Level	Depression	Actions
1	0-4	Level 0	No Depression	Stay healthy physically and mentally
2	5-8 points	Level 1	Slight Depression	Repeat test, Take care of your mental health
3	9-14 points	Level 2	Moderate Depression	Repeat test ,still result is same advisable to consult medical professional
4	more than 15 points	Level 3	High Depression	Immediately consult medical professional

Model Architecture:

Deep Learning model of CNN (Convolution Neural Network) is used for creating, training and analyze expression from static image and video channels. Major components of the architecture are as below:

1. Convolution Layers: Multiple Convolution layers to extract features from images using multiple filters (32,64,128,256) of kernel sizes (3 x 3) which extract features from images.
2. Rectified Linear Unit (ReLU): function used for the introduction.
3. MaxPooling layers: perform reduction of spatial dimension reduction so computational complexity also reduced.
4. Dropout: this layer drops randomly some nodes in training phase so model does not have tendency of Over fitting.
5. Flatten Layer : Converts the 2D feature maps into a 1D vector for input to dense layers.

6 .Dense Layers: These are fully connected layers for seven expression classifications

Above layers include Image Pre-processing, Data Augmentation, data generation, training and Validation.

3. RESULTS

Table 6: Model performance evaluation

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	2.2673	0.2043	2.0278	0.2493
2	1.8677	0.2697	1.6808	0.3757
3	1.6180	0.3736	1.6879	0.3589
4	1.4435	0.4470	1.2770	0.5007
5	1.3103	0.4982	1.2653	0.5237
6	1.2351	0.5255	1.2923	0.5056
7	1.2070	0.5357	1.2488	0.5189
8	1.1475	0.5628	1.0932	0.5803
9	1.1403	0.5671	1.1497	0.5628
10	1.0962	0.5850	1.0529	0.5992
11	1.0821	0.5887	1.0385	0.6027
12	1.0651	0.5986	1.0442	0.5936
13	1.0482	0.6021	0.9976	0.6180
14	1.0344	0.6097	1.0922	0.6041
15	1.0152	0.6157	1.1200	0.5887
16	0.9960	0.6218	0.9868	0.6250
17	0.9973	0.6255	1.0060	0.6271
18	0.9561	0.6398	0.9771	0.6411
19	0.9605	0.6380	1.0074	0.6082
20	0.9468	0.6449	1.0069	0.6341
21	0.9391	0.6445	0.9674	0.6487
22	0.9247	0.6490	0.9729	0.6522

23	0.8960	0.6622	0.9405	0.6508
24	0.9082	0.6603	0.9881	0.6187
25	0.8918	0.6643	0.9616	0.6341
26	0.8988	0.6594	0.9764	0.6529
27	0.8824	0.6713	0.9893	0.6243
28	0.8660	0.6786	0.9972	0.6397
29	0.8636	0.6774	0.9671	0.6508
30	0.8598	0.6847	0.9640	0.6459
31	0.8403	0.6897	1.0170	0.6501
32	0.8273	0.6951	0.9798	0.6543
33	0.8183	0.6951	0.9737	0.6278
34	0.8195	0.6944	0.9706	0.6655
35	0.8011	0.6966	1.0212	0.6404
36	0.8174	0.6997	0.9421	0.6397
37	0.7934	0.7029	1.0191	0.6404
38	0.7851	0.7029	1.0593	0.6397
39	0.7841	0.7100	0.9531	0.6494
40	0.7779	0.7063	0.9565	0.6522
41	0.7669	0.7110	0.9169	0.6745
42	0.7594	0.7183	0.9593	0.6536
43	0.7406	0.7243	0.9267	0.6767
44	0.7310	0.7318	0.9420	0.6725
45	0.7191	0.7329	0.9537	0.6613

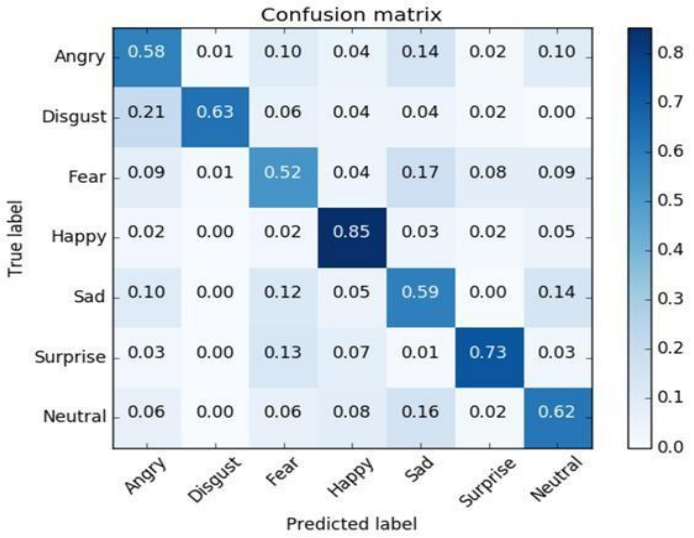

Figure 5: Confusion Matrix

Database:

The FER2013is open dataset (Facial Expression Recognition 2013) [13] which consist of facial images along with their emotion categories of person. The database have facial images of size 48×48 pixels of monochrome images with seven different categories of emotions Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. The dataset having 28709 images in the training set and 3589 images in testing set.

Confusion matrix of model represents:

True label (rows): The actual emotions present in the dataset (e.g., Angry, Disgust, etc.).

Predicted label (columns): The emotionspredicted by the classification model.

Each cell contains the proportion of samples predicted for a given emotion compared to the true label.

Key observations:

1. Diagonal values0.58 for angry, 0.85 for Happy represent correct predictions. Higher values along the diagonal indicate better performance for that specific emotion.The model predicts "Happy" correctly with 85%and “Neutral" correctly 62% of the time so reaming emotions are negative or converge to negative emotions.
2. Off-diagonal values: represent misclassifications where the model incorrectly predicts a different emotion.eg. "Angry" is misclassified as "Fear" 10% of the time; "Neutral" is sometimes predicted as "Sad" (16%) etc.
3. Best and worst performance: The model performs best for "Happy" (85% accuracy on true labels) and struggles with "Fear" (52%) and confuses it with "Sad"(17%) and "Neutral" (9%).
4. Heat map intensity: Darker blue cells represent higher values, indicating stronger agreement.

To calculate precision, recall, and F1-score for each expression calculated with standard formulas. Each value is derived directly from the matrix:

1. TP: Diagonal value for each class (e.g., 0.58 for Angry).
2. FP: Sum of the column excluding the diagonal value.
3. FN: Sum of the row excluding the diagonal value.

Table 7: Model performance evaluation

Class	Precision	Recall	F1-Score
Angry	0.5321	0.5859	0.5577
Disgust	0.9692	0.6300	0.7636
Fear	0.5149	0.5200	0.5174
Happy	0.7265	0.8586	0.7870
Sad	0.5175	0.5900	0.5514
Surprise	0.8202	0.7300	0.7725
Neutral	0.6019	0.6200	0.6108

Observations:

1. Best Precision: Disgust (96.92%), meaning the model rarely predicts Disgust incorrectly.
2. Best Recall: Happy (85.86%), meaning the model captures most Happy instances accurately.
3. Best F1-Score: Happy (78.70%), showing a balance between precision and recall.

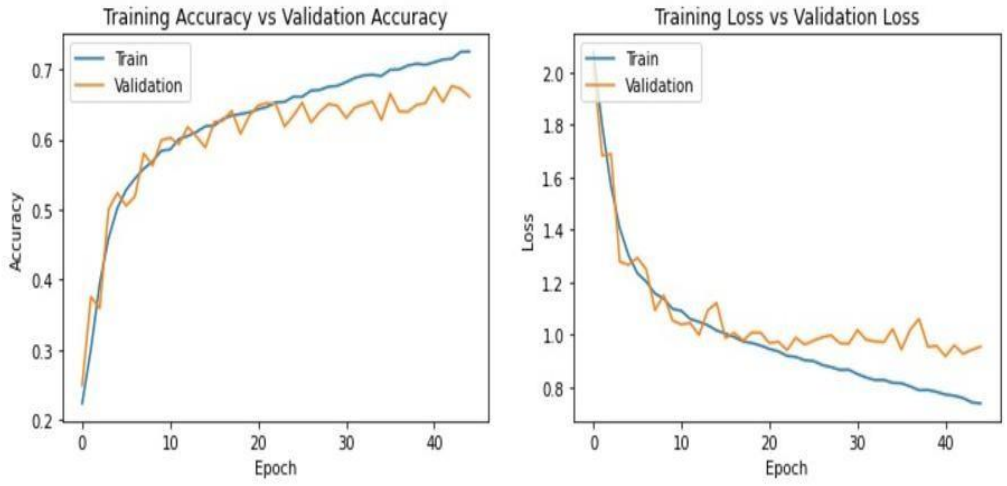

Figure 6: Accuracy vs Epoch & Loss vs Epoch on the Confusion Matrix

- **Table 8:Comparison of Proposed Method with Existing Approaches**

Ref No.	Year	Modality						fusion	Dataset	Accuracy
		Image	Audio	Text	EEG	SRI	Colour			
[16]	2015	√	√					Decision level	Fer+tfd	47.67
[17]	2017	√	√					Decision level	Recola	76.00
[18]	2018	√	√	√				Decision level	Cmu-mosei	49.10
[19]	2018	√	√	√				Feature level	Iemocap	77.60
[20]	2020				√			Feature level	Deap	89.53
[21]	2020	√	√	√				Hybrid	Iemocap	73.98
[22]	2021	√	√		√			FL,DL,HL	Ravdess, Eeg	78.75
[23]	2022	√	√	√	√			Weighted decision level	Combined customized dataset	93.00
[23]	2022	√	√	√	√			Weighted decision level	Real-world collected dataset	67.00
Proposed Approach	2025	√				√	√	Weighted decision level	Real-world collected dataset	79.00

4. DISCUSSION

Depression is among leading cause of disability in all age group globally and affects health care system severely. Millions of peoples affected and likely to be affected persons require psychological and / or clinical treatments. It more affected to woman as before and after pregnancy their hormones and body get changed. Teenagers are more emotional vulnerable and more connected to internet, social media, virtual world which make situations worst. If depression like mental health issues can be detected in initial stage it can cure with little care else situation may become worst.

This research is dedicated to such pre diagnostic self-automated tool which can produced reliable result by performing fusion of available classical methods. This research is a novel approach which analyzes classical Self Response Inventory (SRI) and Visual features of person without intervention of other person so one can be more secure and reliable result can be produced if tool is operated in neutral mode and dual approaches give significant result which is really boon for person who required clinical and /or psychological treatments.

There is future further research directions can be viewed which could be performed in near future for quality improvement of this automated tool. The model could focus on improving predictions for "Fear" "Sad" and "Neutral" where misclassifications are relatively high. To optimize the model's performance, especially for emotions like Fear, Sad, and Neutral, where misclassifications are significant, here are some detailed recommendations: First, if the participants' number could be increased performance of model will also improve. Second, quality of camera plays a major role and high-quality images contains visual features can retrieve specific information which further strengthen the visual behavioral analysis. Third, fusion of more modalities approaches like verbal features, social media text, EEG etc. could be improve the outcome.

Declaration

- We have not taken any financial support for this research paper.
- We have not any conflicting issues.

We have not published this work anywhere else.

REFERENCES

[1] P. Ekman, Handbook of Cognition and Emotion, JohnWiley& Sons Ltd, 2005, pp. 45-60.

[2] H. Ellgring, R. J. Eiser, K. R. Scherer, "Non-verbal Communication in Depression," Cambridge University Press, 2007.

[3] American Psychiatric Association (5th edition, 2013). Diagnostic and statistical manual of mental disorders. Washington: DC.

[4] Gupta R, Malandrakis N, Xiao B, Guha T, Van Segbroeck M, Black MP,PotamianosA,Narayanan SS (2014). Multimodal prediction of affective dimensions and depression inhuman-computer interactions. In International Audio/Visual Emotion Challenge and Workshop.

[5] A. Pampouchidou, P.G. Simos, K. Marias, F. Meriaudeau, F. Yang, M. Pediaditis, M. Tsiknakis, "Automatic Assessment of Depression Based on Visual Cues: A Systematic Review," IEEE Transactions on Affective Computing, vol. 10, no. 4, pp. 445-470, 2019.

[6] Komal Anadkat & Hiteishi M. Diwanji "Unimodal to Multimodal Emotion Recognition: A systematic Review" Test Engineering and Management "ISSN: 0193 4120, Volume 83 PP- 6423-6427 August 2020 (SCOPUS Indexed).

[7] Alghowinem S, Goecke R, Wagner M, Parker G, Breakspear M (2013). Head pose and movement analysis as an indicator of depression. In Humaine Association Conference on Affective Computing and Intelligent Interaction,283–288.

[8] Alghowinem S, Goecke R, Wagner M, Parker G, Breakspear M (2013). Eye movement analysis for depression detection. In IEEE International Conference on Image Processing,4220–4224.

[9] Jeffrey M Girard, Jeffrey F Cohn, Mohammad H Mahoor, S Mohammad Mavadati, ZakiaHammal, and Dean P Rosenwald, "Nonverbal social withdrawal in depression: Evidence from manual and automatic analyses," Image and vision computing, pp. 641-647, 2014.

[10] Internet resource: https://www.verywellmind.com/colour-psychology-2795824

[11] Morales, M., Scherer, S., & Levitan, R, "A Cross-modal Review of Indicators for Depression Detection Systems," Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology –- From Linguistic Signal to Clinical Reality, 2017.

[12] Y. Zhou, S. Scherer, D. Devault, J. Gratch, G. Stratou, L.-P. Morency, and J. Cassell, "Multimodal Prediction of Psychological Disorders: Learning Verbal and Nonverbal Commonalities in Adjacency Pairs,"

17thWorkshop on the Semantics and Pragmatics of Dialogue. SEMDIAL, pp. 160-169, 2013.

[13] R.Verma,"FER2013,"Kaggle,26-May-2018.[Online].Available: https://www.kaggle.com/datasets/deadskull7/fer2013.

[14] Pampouchidou, Anastasia, Kostas Marias, ManolisTsiknakis, P.Simos, Fan Yang, and Fabrice Meriaudeau (2015). Designing a framework for assisting depression severity assessment from facial image analysis. In Signal and Image Processing Applications(ICSIPA), International Conference on, pp. 578-583, IEEE.

[15] Patient Health Questionnaire-9,Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (1999). Patient Health Questionnaire-9 (PHQ-9) [Database record]. APA PsycTests.

[16] S. E. Kahou, X. Bouthillier, P. Lamblin, C. Gulcehre, V. Michalski, K. Konda, S. Jean, P. Froumenty, Y. Dauphin, N. Boulanger-Lewandowski, R. Chandias Ferrari, M. Mirza, D. Warde-Farley, A. Courville, P. Vincent, R. Memisevic, C. Pal, and Y. Bengio, "Emonets: Multimodal deep learning approaches for emotion recognition in video," Journal on Multimodal User Interfaces, vol. 10, no. 2, pp. 99–111, 2015.

[17] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, "End to-End Multimodal Emotion Recognition Using Deep Neural Networks," IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 8, pp. 1301–1309, Dec. 2017, doi:10.1109/jstsp.2017.2764438.

[18] S. Sahay, S. H. Kumar, R. Xia, J. Huang, and L. Nachman, "Multimodal relational tensor network for sentiment and emotion classification," Proceedings of Grand Challenge and Workshop on Human Multimodal Language (Challenge-HML), 2018.

[19] D. Hazarika, S. Poria, A. Zadeh, E. Cambria, L.-P. Morency, and R. Zimmermann, "Conversational memory network for emotion recognition in dyadic dialogue videos," Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 2018.

[20] M. M. Hassan, Md. G. R. Alam, Md. Z. Uddin, S. Huda, A. Almogren, and G. Fortino, "Human emotion recognition using deep belief network architecture," Information Fusion, vol. 10.1016/j.inffus.2018.10.009. 51, pp. 10–18, Nov. 2019, doi: 21. S. Siriwardhana, A. Reis, R. Weerasekera, and S. Nanayakkara, "Jointly fine- tuning 'bert-like' self supervised models to improve multimodal speech emotion recognition," Interspeech 2020, 2020.

[21] D. Priyasad, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Attention driven fusion for multi-modal emotion recognition," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020.

[22] J. Njoku, Angela C. Caliwag, W. Lim, S. Kim, H.-J. Hwang, and J.-W. Jeong, "Deep Learning Based Data Fusion Methods for Multimodal Emotion Recognition," The Journal of Korean Institute of Communications and Information Sciences, vol. 47, no. 1, pp. 79–87, Jan. 2022, doi: 10.7840/kics.2022.47.1.79.

[23] Komal D. Anadkat,Dr. Hiteishi M. Diwanji & Dr. Shahid Modasiya "Effect Of Preprocessing In Human Emotion Analysis Using Social Media Status Dataset" Reliability: Theory and Application "ISSN 1932-2321, Volume 17 Issue 1 PP-67 March 2022 (SCOPUS Indexed).

[24] Komal D. Anadkat & Dr. Hiteishi M. Diwanji "Effectof Activation Function in Speech Emotion Recognition On The Ravdess Dataset" Reliability: Theory and Application "ISSN 1932-2321, Volume 16 Issue 3 PP-63 September 2021 (SCOPUS Indexed).

[25] Internet resources like: doc player.net ,journalofbigdata., springeropen.com, deepai.org, arxiv.org ,pubmed.ncbi.nlm.nih.gov