

# Multimodal Spatiotemporal Representation for Automatic Depression Level Detection

Dubale Swapnil Arun<sup>1</sup>, Shelke Mayur Pandit<sup>2</sup>, Kapadane Vaibhav Shyam<sup>3</sup>,  
Savale Vishal Ashok<sup>4</sup>, Prof. C. S. Wagh<sup>5</sup>

Students, Department of Computer Engineering<sup>1,2,3,4</sup>

Guide, Department of Computer Engineering<sup>5</sup>

Navsahyadri Education Society Group of Institutions, Naigoan, Maharashtra, India

swapnildubale04@gmail.com<sup>1</sup>, shelkemayur4700@gmail.com<sup>2</sup>,

amvaibhav17@gmail.com<sup>3</sup>, vishalsavale33@gmail.com<sup>4</sup>

**Abstract:** *Humans being's cognitive system can be simulated with artificial intelligence systems. Machines and robots equipped with cognitive ability can automatically recognize a human's mental state through their gestures and facial expressions. An artificial intelligent system is proposed in this paper to monitor depression. In this world there are lot of people suffering from depression, and this impacting the economy very badly. Getting it checked and treated as soon as possible is important to cut costs and even lives can be saved too by early diagnosis of depression. It has been tried many times to figure out how depressed someone is using audio and visual features and the text analysis of conversational speech transcriptions. But it is hard to get the data and there are not lot of options for research. It is important to choose the right training data because there are not too many features that can be extracted from speech to detect depression level. Therefore the method described in this paper uses audio as well as video features to figure out how depressed someone is. This help in recognizable improvement in accuracy of detection of depression.*

**Keywords:** Depression, Face Expression, Audio Features, Machine Learning, Multimodal Depression detection.

## REFERENCES

- [1]. Niu M, Tao J, Liu B, et al. Automatic Depression Level Detection via lp-norm Pooling[C]//2019 Conference of the International Speech Communication Association (INTERSPEECH 2019). 2019: 4559-4563
- [2]. Zhou X, Jin K, Shang Y, et al. Visually Interpretable Representation Learning for Depression Recognition from Facial Images[J]. IEEE Transactions on Affective Computing, 2018.
- [3]. Lee J, Kim S, Kiim S, et al. Spatiotemporal Attention Based Deep Neural Networks for Emotion Recognition[C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018: 1513-1517.
- [4]. De Melo W C, Granger E, Hadid A, et al. Combining Global and Local Convolutional 3D Networks for Detecting Depression from Facial Expressions[C]//2019 IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), 2019: 1-8
- [5]. Al Jazaery M, Guo G. Video-based Depression Level Analysis by Encoding Deep Spatiotemporal Features[J]. IEEE Transactions on Affective Computing, 2018.