International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v10i10.5735
***Article Received**: 28 July 2022 **Revised**: 05 September 2022 **Accepted**: 12 September 2022*

Video Based Deep CNN Model for Depression Detection

Gyanendra Tiwary[1], Dr. Shivani Chauhan[2], Dr. Krishan Kumar Goyal[3]
[1]Department of Computer Science and Engineering
Bhagwant University
Ajmer, India
gyanendra.tiwary@gmail.com
[2]Department of Computer Science and Engineering
Bhagwant University
Ajmer, India shivanichauhanbit@gmail.com
[3]Department of computer applications,
Raja Balwant Singh Management Technical Campus ,Agra. India kkgoyal@gmail.com

***Abstract*—** Our face reflects our feelings towards anything and everything we see, smell, teste or feel through any of our senses. Hence multiple attempts have been made since last few decades towards understanding the facial expressions. Emotion detection has numerous applications since Safe Driving, Health Monitoring Systems, Marketing and Advertising etc. We propose an Automatic Depression Detection (ADD) system based on Facial Expression Recognition (FER).

We propose a model to optimize the FER system for understanding seven basic emotions (joy, sadness, fear, anger, surprise, disgust and neutral) and use it for detection of Depression Level in the subject. The proposed model will detect if a person is in depression and if so, up to what extent. Our model will be based on a Deep Convolution Neural Network (DCNN).

Keywords: Facial Expression Recognition (FER), *Machine Learning, Deep Convolution Neural Network (DCNN), , Neural Network, Emotion Recognition.*

I. INTRODUCTION

Our face is a portrait of our feelings, most of the times it is the first thing from which we can conclude one's emotion at the moment. We do it every day naturally, while talking to someone or showing some product, image or advertisement we observe the reaction of the person Infront of us by his/her face and we almost understand what he/she feels about it even before a single word said. Due to this property of human face in last few decades multiple attempts have been made to developed methods and algorithms for facial data processing and understanding. With the recent advancements in the field of DCNN models, it has become computationally easy to apply complex operations on images, which help drawing useful conclusions in realistic time bounds.

According to an updated (30th January, 2020) WHO [17], globally more than 264 million people of all ages are suffering from depression. Depression has become major reason for disability as per WHO. This study also says that 14% of the world's population have some or the other psychiatric problem. There are a range of problem from Major Depressive Disorder (MDD), Bipolar Disorder (BD), Post Traumatic Stress Disorder (PTSD), Generalized Anxiety Disorder (GAD) to Mild Mood Swings. In developing countries like India we rarely care about mental health, despite of the fact that Mental health is as important as physical heath and in most of the cases Psychiatric Health affects Physiological Heath and Work Efficiency severely. The problem is multifold, along with slow progression of diseases and ignorance of symptoms, in most of the cases, traditionally there are no pathological tests available, which can scientifically tell about presence or extent of mental illness. It is only when we see some physical problems, we decide to see a doctor. As per a survey by WHO, depression is major cause of suicide [18]. Looking at these studies, now people are getting aware and concerned about these issues. In india along with emergency Fire, Police and Ambulance services, Ministry of Health and Family Welfare along with some NGOs is running India Suicide and Stress Helpline numbers to assist someone in need. An ADD system can greatly help people in understanding and raising the alarm.

There is more than ever need of healthcare professionals all over the world. Due to the current pandemic of COVID '19 [19][20][21] the entire world is maintaining social distancing and people are not socializing. Work from home, increasing

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v10i10.5735
***Article Received**: 28 July 2022 **Revised**: 05 September 2022 **Accepted**: 12 September 2022*

screen time, increasing unemployment, less physical activities, changing schedules, home schooling, fear of getting infected etc are causing poor mental conditions and the cases of Major Depression Disorder (MDD) episodes are increasing [20]. An effective mobile app with ADD system can regularly assist people in checking their Mental Health and will reduce the dependance of Psychiatrist.

Human face has multiple muscles and by using them we may generate very complex expressions.

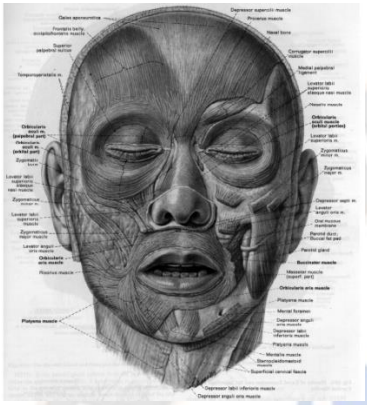

Fig 1. Human Facial Muscles [16]

Capturing any expression as an image or video we tend to understand the feeling/mood of the person. Our brain is very efficient to understand these expressions and easily analyses it but it is very difficult for computers to process on Images and/or Video data majorly due to high density and quantity of complex unmanaged data even for a small information.

After the evolution of Deep Neural Network and Convolution Neural Networks the processing of large number of images and video data has become possible even with ordinary computing capabilities. This shift has opened multiple ways towards making efficient image data processing systems and using it to solve complex problems such as Depression Detection, Making and adjusting Marketing strategies, Sentiment Analysis etc. Today using online resources such as Google Cloud, Tensor-Flow etc we may work on huge projects even at the convenience of our mobile phone.

II. RELATED WORK

Ekaterina et el [1] has proposed ML based algorithm for FER. Their major contribution is towards emotion recognition using facial image. The model is tuned and parameters were set for optimal functioning. Their model accuracy was 69% for public test sample and 59% for private samples.The emotion recognition accuracy has also been determined for both the groups. Happiness recognition accuracy was heighest (88%) and accuracy for sadness is lowest (57%)

Vizireanu et el [2] worked on depression detection and also classified stress scales using DASS (Depression Anxiety and Stress Scale). They have taken input as video and used FACS (facial Action Coding System). In this work, they have made three layer model. Active Appearance Model (AAM) is been used in first layer. In the second layer Action Unit (AU) along with multiclass SVM (Support Vector Machine) is being used. AU matrix has been made with the help of SVM. The final third layer is designed using FFNN (Feed Forward Neural Network) which is being used to analyze the pattern obtained in the form of matrix from second layer. This pattern is further fed into DASS levels to identify the level of stress. For depression they got 87.2% accuracy. But they got 90.2% accuracy for stress whereas 77.9% for anxiety. Since the average time to get the result is 64 secs, this model could be used for real time systems for detection of stress. Their model is also 93% accurate to differentiate a healty person from someone with MDD (Mild Dipressive Disorder) and PTSD (Post Traumatic Stress Disorder). Apart from this, they also claimed that the proposed model gives 85% accuracy for GAD (Generalized Anxiety Disorder).

Wang et el. [3] has proposed depression detection model based on reaction time (r-t) and eye movement (e-m) of the subject. They claimed that r-t and e-m features can represent the attention bias of the subject. Using these features, they differentiated a healthy and depressed person. They used SVM classifier and few normal control methods for their model. They completed their study with 86% accuracy.

In the work by Bhatia et al [4], they have analyzed the patient's head movement during their interviews with therapist. They established a relation or synchronization between patient and therapist head movements. While the patient is undergoing HRSD (Hamilton Rating Scale for Depression) questionnaire they have recorded the session. They majorly focused on patients of MDD. They have done these recordingsin a gap of 7 week's and collected 3 recordings for each pataints in 21 weeks time. From 2D videos, complete 3D head movements with reference horizontal (pitch) and vertical (yaw) axis were analyzed. By analyzing head angles Windowed Cross-Correlation patterns are recorded.A unique peak picking algorithm is being used to pick the abnormalities and HLM (Hierarchical Linear Model) is used for analyzing the changes in head movements. The results were showing the correlation between patient-therip[ist head movements. As per the authors it is the first time someone analysed the relationship between pateint's tharipist head movements with depression levels. The variabilioty within dyad was more than between dyad.

Audio Visual Emotion Challenge (AVEC) in it's 2019 competition posed the problem for detecting Depression using AI over cross cultural image dataset. Song at el. [5] proposed a comparison between multimedia processing methods to the Machine Learning methods for understanding the emotional health of humans based on Audio-Visual data. In this challenge The test sets are made available for

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v10i10.5735
***Article Received**: 28 July 2022 **Revised**: 05 September 2022 **Accepted**: 12 September 2022*

multimodal data processing to emotion and health research community. Here they have also compared the various aspects of real life health and emotion data. This paper presented all details of this competition such as Guidelines, Data Sets, The Baseline System on all three competition challenges:

i) SoMS (State of Mind System)- by listening positive and negative stories, what's the level of mood changes
ii) DDS (Depression Detection System)- by recordings of interviews taken by virtual AI driven agent, we have to detect depression level (PHQ-8 score)
iii) CES (Cross cultural Effect Sensing)- Here participants have to detect Arousal, Valance and liking extent on across cultural data set. For training they have been provided German and Hungarian data sets and testing has to be done on Chinese data sets solely.

For making baseline systems, they kept height level of transparency and realism. They have used open source softwares and used same number of epochs as allowed for any participant. They have also shared all scripts and outcomes in all three cases and the results shows that:

i) The best results were obtained when systems were trained on static scores and outcome taken on it's dynamic view. This means when the subject heared the story to before he/she heared any story. This behavior can be explained by Inertial Emotional Theory.
ii) Detecting depression was the most difficult and results were also not quite good.
iii) Detecting emotion in cross cultural environment is quite challenging task. In that also, detection using audio is far more difficult then detecting using video. The linguistic differences and different acoustics make this a ill posed problem. The data in the challenge has also been recorded in real time noisy environment, so understanding diffent expressions for diffenent emotions in different cultures are very difficult.

Hadid et al. [6] used Local as well as global 3D CNN for detecting depression. They have proposed a novel video based health monitoring system. To analyze the patterns in facial images and to detect depression they used deep neural network based architecture. They processed spatio temporal data separately. They combined a 2D CNN with a RNN (Recurrent Neural Network). Though combining these may detoriate the result further, but with recent convolutional 3D (C3D) networks helps to properly model them and improve the overall performance of the system. In this paper they claimed that these C3D networks are yet not been explored for depression detections and this is the first time they are proposing it in this work. They have used global (full face) along with local (eyes only) features and fused it with C3D network. The reason why they have selected eye area for local processing is because, eyes are more accurate to detect depression. Global average pooling is used to conclude spatio temporal features without fully connected layers to avoid over-fitting and find potential parameters. They have used AVEC 2014 dataset and their model outperforms most of the state of art depression detection systems.

Guo et al. [7] proposed video based depression detection model using spatio temporal features. They have used BDI-II (Beck Depression Inventory-II) for leveling the extent of depression. A 3D Convolutional NN applied at two different scales. Then a RNN is used to extract features from spatiotemporal data. The complete model can be viewed as RNN-C3D model which process global and local features to estimate depression levels.

Liu et al. [8] used local second order gradient for depression detection. Though physiological studies has established the fact that facial cues changes with the level of depression, but they mostly look at major changes. In this work they proposed a noble Local Second Order Gradient Cross Pattern (LSOGCP) model. They initially extract the features using higher order gradient and cross coding model on each frame. Then video representation using histogram is generated. At last a intra group classification and inter group regression is applied to calculate level of depression. They have used AVEC 2013 & 2014 datasets for their experimental analysis.

Huang et al. [9] proposed a method combining Computer Vision, Signal Processing and Sentiment Analysis. Depression causes significant changes in acoustic, linguistic and facial expressions which can be used as prominent biomarkers of depression. A multimodal model has been proposed to combine audio, video and text features to identify level of depression. The biomarkers vary as per gender. The audio and facial markers for males and females are quite different. In this work also they have concluded that socio cultural and gender variation plays an important role in depression detection using facial and audio cues.

Lul et al. [10] has also combined acoustic and face features for depression detection. They focused on acoustic rhythm and sparse facial data to improve the accuracy. In this work they mentioned the false positive results of other models and promoted multimodel systems to reduce that. They used Cepstrum method and spectrum subtraction to enhance speech frequency identification of depressed person. They extracted large variation rate by short term energy and Mel-frequency cepstral coefficients. Differential speeches are also analyzed for time and frequency domains. Moreover this paper has also uses orthogonal match persuit algorithm for

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v10i10.5735
***Article Received**: 28 July 2022 **Revised**: 05 September 2022 **Accepted**: 12 September 2022*

getting sparce facial data which later combined with facial and voice emotion features.The result shows 81.41% accuracy for depression detection. Noramally a professional doctor's efficiency goes about 47.3%, which can also be improved by just applying the algorithms proposed here to 71.54%. The authors claimed that their model can easily be incorporated with the existing hospital setup and will also reduce cost to effectively detect the state of depression. They further claimed that if only speech mode is turned on, another 6.76% efficiency can be increased. The process of depression detection is a inter disciplinary area of research and authors promoted that computer scientist should join hand with medical practitionars to make this domain more accurate with more methods. They also proposed to use this model for criminal investigation and interrogation, effective online education, safty and security services and also to entertainment industry as well.

Morency et al. [11] proposed behavioral descriptor based on audio-visual information for emotional disorder analysis. They proposed a model to detect Stress, Depression and PTSD. This model is focused on self assessment of these strongly connected disorders using a questionnaire. They used factor analysis for detecting generic disorders. They recommended to use automatic behavior descriptors which can quantify the disorder in objective manner, instead of manual annotation based methods. The objectivity of the model will help healthcare professionals to diagnose the disorder in a better manner. They used a dataset called DAIC (Distress Assessment Interview Corpus) contains interviews of paid participants with the clinical experts. They have also established correlation between the proposed behavioral descriptors with general distress measures. Their study also shows the source of these fidgeting behaviors They considered the non verbal emotional features captured using the self assessment questionnaire along with the audio-visual data and manual annotations. As they found a very close realation between all the disorders (stress, depression and PTSD), they focused on vertical gaze directional data, extent and duration of smile, voice monotonicity and quality and also the leg and hand fidgeting. The manual self adaptors and fidgets are combined with automatically extracted gaze, smile and voice features. In this work they have highlighted several statistical correlation between the distress level and non verbal features. They have jotted down the following four findings: 1) In the state of distress, subject shows a downward angle in gaze. Along with this, the face and eye gaze can also be identified automatically for a person in depression. 2) The intensity and duration of smiles can also be automatically analyzed for a person in stress. 3) The level of monotonicity and acoustic tension can also be automatically modeled for a person in stress. and 4) The hand movement and leg fidgets can be analyzed manually. Based on these features the level of stress and depression can accurately been analyzed.

Mehta et. al. [12] proposed a model for depression detection using vocal motor coordination. It has been observed that during distress the coordination between vocal source, prosodies and vocal tract is also changes. It has been established that these characteristics are also related to psychomotor retardation. Generally the pataint shows slugishness in vocal articulation and this eventually affects the voice production and pitch. In this work the authors picked these features during formant frequencies and delta mel cepstrum. These feature domains suggest the possible changes in vocal tract articulation frequencies. To collect this information multiple experiments with time variying different locations for microphones has to be conducted. They have used AVEC 2013 dataset for experimenting their model. They proposed a GMM (Gaussian Mixture Model) based multimode regression method to calculate the extent of depression and stress. The RMSE (Root Mean Square Error) they obtained is 7.42 and MAE (Mean Absolute Error) 5.75 on BDRS (Beck Depression Rating Scale).

III. PROPOSED METHODOLOGY

The objective of this research is to develop an ADD system with the help of DCNN. Through input image we will train a Deep Convolutional Neural Network model to identify 7 basic human facial expressions joy, sadness, fear, anger, surprise, disgust and neutral. The model will then take these classified images and use it to detect the depression level using Depression Anxiety Stress Scale (DASS).

Fig 2. Preprocessing

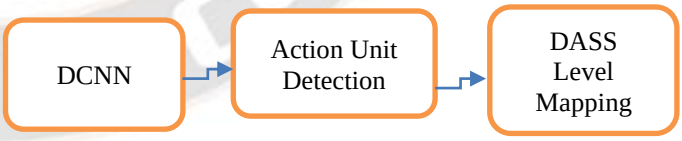

Fig 3. Depression Detection

Fig-2 shows the preprocessing steps. Here our intention is to filter and clean the input data to apply Facial Action Coding System (FACS) to identify basic emotions with help of Action Units (AU). We used viola Jones method to detect face and extract it from the background image. We used this image and extracted features like Eye, Mouth etc.

Now on this extracted feature, we applied Ekman FACS and detect the basic emotion. This classified image is then fed to a classic 3 layer DCNN and classified as per

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v10i10.5735
***Article Received**: 28 July 2022 **Revised**: 05 September 2022 **Accepted**: 12 September 2022*

DASS levels. In our work, we have used FER 2013 dataset. Total 7 emotions have been covered in this dataset.

Basic Emotion	Number of Images
Angry(0)	4594
Disgust(1)	548
Fear(2)	5122
Happy(3)	8990
Sad(4)	6078
Surprise(5)	4003
Neutral(6)	6199

Table 1. FER 2013 Data Set

IV. RESULTS

In this work we have taken input as video stream. Which is then fed into frame separator module. Then we extracted face from each frame using Vola-Jones method. This facial image is then used for feature extraction for FACS. The FACS features are fed into the next layer of DCNN. Here we have we have used max-pooling with sigmoid and softmax activation functions. Count of each class in output is shown in fig 4.

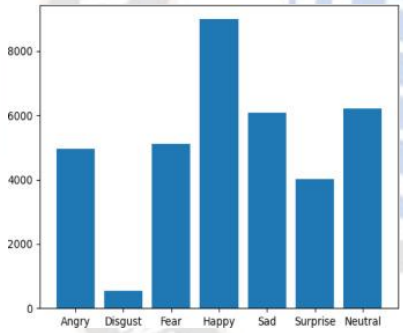

Fig 4. Output Class Data Count

Multiple epochs were run on the model and the accuracy curve can be seen in fig 5. We achived 82% accuracy after 15 epochs.

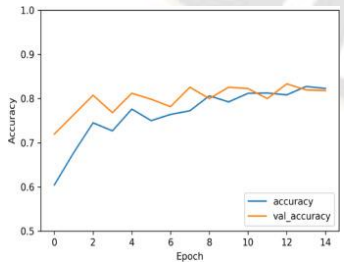

Fig 5. Model Accuracy V/S Epochs

VI. *Conclusion and Future Work*

In the present work, we have constructed an Automatic Depression Detection (ADD) model, which can be proved quite helpful to patients dealing with any kind of Depression or Anxiety. Using Vola-Jones algorithm for facial images detection has given us better result in terms of ill posed images. This method has shown quite good response for images taken in the wild (i.e. low illumination, non-frontal images, less contrast etc). Feeding these images to Ekman's FACS has given us good scores for various emotions. FACS has became a gold standard for any computer based classification for emotion detection. Various psychiatric researchers has proposed methods and questionnaire which can directly be used to classify the current status of the patient. DASS levels have helped us to do that in the current study. Using DASS levels we could easily classify among MDD, PTSD, BD etc. The work has shown a good result and we may further improve it by introducing Generative Adversarial Networks (GANs) This kind of model could be further developed and deployed as a mobile or web application to reach the mass.

REFERENCES

[1]. Ekaterina Ivanovaa and Georgii Borzunov: "10th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2019 (Tenth Annual Meeting of the BICA Society)" (Page No. 244–248) – 2019: Optimization of machine learning algorithm of emotion recognition

[2]. Mihai Gavrilescu * and Nicolae Vizireanu: Sensors 2019, 19, 3693; doi:10.3390/s19173693, 2019 "Predicting Depression, Anxiety, and Stress Levels from Videos Using the Facial Action Coding System"

[3]. Zeyu Pan, Huimin Ma, Lin Zhang, Yahui Wang IEEE ICIP 2019, 2019 DEPRESSION DETECTION BASED ON REACTION TIME AND EYE MOVEMENT

[4]. Shalini Bhatia, Roland Goecke, Zakia Hammal, Jeffrey F Cohn Proc Int Conf Autom Face Gesture Recognit. 2019 May ; 2019: . doi:10.1109/FG.2019.8756509. 2019 Automated Measurement of Head Movement Synchrony during Dyadic Depression Severity Interviews

[5]. Gudni Johannesson, & Nazzal Salem. (2022). Design Structure of Compound Semiconductor Devices and Its Applications. Acta Energetica, (02), 28–35. Retrieved from http://actaenergetica.org/index.php/journal/article/view/466

[6]. Fabien Ringeval, Björn Schuller, Michel Valstar, Nicholas Cummins, Roddy Cowie, Leili Tavab, Maximilian Schmitt, Sina Alisamir, Shahin Amiriparian, Eva-Maria Messner, Siyang Song, "Association for Computing Machinery. ACM ISBN 978-1-4503-5983-2/18/10.", 2019 "Audio/Visual Emotion Challenge 2019: State-of-Mind, Detecting Depression with AI, and Cross-Cultural Affect Recognition"

[7]. Wheidima Carneiro de Melo1, Eric Granger2 and Abdenour Hadid "978-1-7281-0089-0/19/$31.00 c 2019 IEEE", 2019 "Combining Global and Local Convolutional 3D Networks for Detecting Depression from Facial Expressions"

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v10i10.5735
***Article Received**: 28 July 2022 **Revised**: 05 September 2022 **Accepted**: 12 September 2022*

[8]. Mohamad Al Jazaery and Guodong Guo IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, 2020 "Video-Based Depression Level Analysis by Encoding Deep Spatiotemporal Features"

[9]. Rolf Bracke, & Nouby M. Ghazaly. (2022). An Exploratory Study of Sharing Research Energy Resource Data and Intellectual Property Law in Electrical Patients. Acta Energetica, (01), 01–07. Retrieved from http://actaenergetica.org/index.php/journal/article/view/459

[10]. Mingyue Niu, Jianhua Tao, Bin Liu 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), 2019 Local Second-Order Gradient Cross Pattern for Automatic Depression Detection

[11]. Aven Samareh, Yan Jin, Zhangyang Wang, Xiangyu Chang & Shuai Huang IISE Transactions on Healthcare Systems Engineering , 2018 "Detect Depression from Communication: How Computer Vision, Signal Processing, and Sentiment Analysis Join Forces"

[12]. Jian Zhao1 • Weiwen Su1 • Jian Jia2 • Chao Zhang1 • Tingting Lu1 Springer Science+Business Media, LLC, part of Springer Nature 2017, 2017 "Research on depression detection algorithm combine acoustic rhythm with sparse face recognition"

[13]. "Stefan Scherer a,∗, Giota Stratou a, Gale Lucas a, Marwa Mahmoud a,b, Jill Boberg a, Jonathan Gratch a, Albert (Skip) Rizzo a, Louis-Philippe Morency" Image and Vision Computing 32 (2014) 648–658, 2014 "Automatic audiovisual behavior descriptors for psychological disorder analysis"

[14]. James R. Williamson, Thomas F. Quatieri, Brian S. Helfer, Rachelle Horwitz, Bea Yu, Daryush D. Mehta, 2013 "Vocal Biomarkers of Depression Based on Motor Incoordination"

[15]. Prudhvi Raj Dachapally , 2019 Facial Emotion Detection Using Convolutional Neural Networks and Representational Autoencoder Units

[16]. Zhenhai Liu, Hanzi Wang(□), Yan Yan, and Guanjun Guo, 2015 "Effective Facial Expression Recognition via the Boosted Convolutional Neural Network"

[17]. PAUL VIOLA, MICHAEL J. JONES, 2003 Robust Real-Time Face Detection

[18]. P. Ekman, "Darwin, deception, and facial expression," Annals of the New York Academy of Sciences, vol. 1000, no. 1, pp. 205–221, 2003.

[19]. https://www.who.int/news-room/fact-sheets/detail/depression

[20]. https://apps.who.int/iris/bitstream/handle/10665/131056/9789241564779_eng.pdf

[21]. https://www.thelancet.com/journals/lanpsy/article/PIIS2215-0366(20)30482-X/fulltext

[22]. https://www.thelancet.com/journals/lanpsy/article/PIIS2215-0366(20)30491-0/fulltext

[23]. https://www.cdc.gov/mmwr/volumes/69/wr/mm6932a1.htm

[24]. https://psycnet.apa.org/fulltext/2020-63541-001.html