How to Cite:
Rathod, V., Chohan, A., Nema, S., Nawghare, A., Devikar, P., & Agrawal, R. (2022). Improved remote mental health illness assessment and detection using facial emotion detection and speech emotion detection. *International Journal of Health Sciences*, *6*(S2), 9577–9590. https://doi.org/10.53730/ijhs.v6nS2.7508

Improved remote mental health illness assessment and detection using facial emotion detection and speech emotion detection

Vasundhara Rathod*
Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, India
*Corresponding author

Abhishek Chohan
Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, India

Sakshi Nema, Adhney Nawghare
Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, India

Prateeksha Devikar
Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, India

Rahul Agrawal
Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, India

***Abstract*---**The use of Artificial Intelligence in the healthcare sector has been recently observed. Projects are being made that integrate AI and therapeutic sessions[1]. A new area of study evolved where doctors, along with technicians, collaborate to create projects which will help give the old school therapy an advanced technical form. This study uses the original therapy techniques for mental health assessment and integrates it with machine learning models for facial emotion recognition and speech pattern recognition to get a better understanding of a patient's mental health condition and help them deal with it. This project assists the patient in the diagnosis of 11 different mental health conditions where the patient's emotional state is taken into consideration while diagnosis. Patients' social interactions are also being checked and analyzed regularly. This project calculates a verdict which declares whether the patient suffers from the diagnosed illness or whether they should retake the tests. In

Manuscript submitted: 27 March 2022, Manuscript revised: 18 April 2022, Accepted for publication: 9 May 2022

addition, the project maintains records of the patient's emotional and mental health journey.

***Keywords*---**emotion recognition speech, emotions through facial expressions, psychological disorders, remote diagnosis, Twitter sentiment analysis.

Introduction

Ever since the pandemic hit the world, followed by lockdowns, it has been observed that stress levels have increased in people. People losing their jobs, their habits, and change in everyday life has led to an increased amount of stress. Some of the reasons may be family problems, homesickness, loneliness, loss of motivation, etc. These factors may interfere with a person’s success in life, happiness, and satisfaction. People often tend to ignore the symptoms, which further exacerbates the situation. Thus, a continuous increase in their stress levels may result in psychological disorders. Such issues can be resolved by a simple approach i.e. consulting a counselor. But due to time and money constraints in counseling, most people choose not to treat themselves. Also concluded by[2], coronavirus has an indirect impact on the mental health of people due to lack of accurate remote mental health assessment systems and lack of proper treatments. As psychological diseases are as harmful as any physical disease, we came up with the solution to resolve this issue: an AI Counselling Assistant. This smart assistant will provide a questionnaire, and simultaneously study the facial expressions and the speech patterns of the individual answering the questions, analyze the person’s social media, and will deduce their mental health status. This assistant will recommend the psychiatrists or recommend a procedure to be followed and aid the patients in their recovery based on the results.

The main functional objectives of our application are as follows: The application helps in diagnosing the mental health of the patient by analyzing facial expressions, speech patterns, and interactive voice QNA based on the generalized test to get the probable diseases followed by projective tests to get the diseases suffering from and its severity. It determines the psychological health status of the patient by analyzing its social media activities. It helps in the creation of personalized virtual assistance for the user. It creates an easy-to-use and user-friendly environment. Personalized treatments based on the conclusion deduced from the counseling session are recommended to the user. Doctors are recommended based on the disease diagnosed from questionnaire, speech and facial models. It diagnoses various mental diseases and disorders like Depression, Dissociative identity disorder, Bipolar disorder, Anxiety, Schizophrenia and Early Psychosis, Obsessive Compulsive Disorder (OCD), Post-traumatic stress disorder (PTSD), Attention-deficit/hyperactivity disorder (ADHD), Eating Disorder, Suicidal Ideation, Postpartum depression test, Premenstrual dysphoric disorder test. It helps in creating a friendly interface between patients and psychiatrists.

Literature Review

One of the main challenges associated with remote mental health assessments and detection is fidelity while answering the questionnaire. Identifying emotion from facial expression is a tedious task in which even humans can't even tell the emotion exhibited from in the static image. Some of the earlier methods [3]involved recognition of muscle movements in static images to predict the emotion. There have been developments in using a CNN's for emotion classification and deep CNN's giving higher accuracy than the shallower models [4].

Training a Speech Emotion Recognition model is not an easy task. The factors like requiring large datasets containing many examples, correct extraction of features, and the classifier used hugely impact on the accuracy and loss of a model. Initially SER was performed using classifiers like Support Vector Machine, K-Nearest Neighbours, Bayes Classifier, etc. But the drawback of these classifiers was that they required big datasets in order to get a more accurate model. A model named Inception Net v3 along with TensorFlow was used in [5]Long Short-Term Memory Networks (LSTM) layers were used in the Convolutional Neural Network model in[6] giving a precision and recall of 88.01% and 88.86% respectively. In paper[7] DNN was combined with GMM, it worked fine even in the presence of noisy dataset. [7]Uses the Random Forest Classifier on SUSAS[8]dataset giving good accuracy, but the SUSAS dataset did not have samples for spontaneous speech. In paper [9], CNN along with deep Bi-LSTM beats the model consisting of only CNN layers when working on EMO-DB dataset giving an accuracy of 85.57%.

Tess[10], a psychotherapeutic device, uses NLP and indicates those expressions which are the cause of emotional distress of a person. Approximately 86% of people rated Tess bot with a positive rating. Chat bots like Wysa[11], Woebot[12]engage with users like a virtual psychotherapist. They help their users recognize the emotions they are experiencing and help develop skills as well as techniques to reduce anxiety. Wysa has a precision of 74.7% and 91% people rated wysa with a positive rating. Woebot has a precision of 98% in analysingself harm phrases.

There are psychotherapeutic devices like Avatar Project[13]which addresses auditory hallucinations. They are also used for schizophrenia treatment and improve medication adherence. It uses computer generated faces that interact with the patients using AI. Diagnostic accuracy with the avatar project is 85.7% while other text based has an accuracy of 71.4%. Special robots are also made for treatments. Some are made to look like animals, like Paro[14] looks like a seal, and helps dementia patients. The robot eBear[15]is integrated with Paro, to make a companion bot, which can be implemented in peoples' homes to act like health care assistants. They respond to speech and movement using dynamic dialog models. They help people to deal with loneliness while also reducing stress and anxiety, and improve mood and social connections.

Apart from the available technologies and virtual assistants, there are questionnaires available for different diseases, for example PHQ-9[16]for depression. Research on some questionnaires talks about accuracy (Table 1.1) while some talk about sensitivity and specificity (Table 1.2). Research says that for schizophrenia questionnaires, 64% patients, 77.2% experienced professionals and 70.2% inexperienced professionals found it effective.

Table 1.1: Accuracy of questionnaire

Disease Questionnaire	Accuracy(%)
Bipolar and anxiety	70.0
Obsessive-compulsive disorder	85.0
ADHD	89.5.0

Table 1.2: Sensitivity and specificity of Questionnaire

Disease Questionnaire	Sensitivity(%)	Specificity(%)
Depression	88.0	88.0
Suicidal	96.9	87.6
Postpartum	93.33	96.0

Material & Methods

Facial Emotion recognition

Dataset

The dataset is provided by Kaggle[17]which consists of 28709 images in form of a csv file and the corresponding emotions associated with the image as label ‘emotion’, with the following values: neutral, happy, sad, angry, fear, disgust, and surprise.

Data preprocessing

Images are taken from this csv file. Images in the csv were in the form grayscale in order to fit them into the model, they were resized and reshaped into a size of (48, 48, 3) image. These images were normalized and then batched into 28 batches each consisting of 1024 images. The emotion labels were converted into a one hot vector form, as it gives better accuracy as stated in [18] .

Network Architecture

Multi-layer convolutional networks[19]output a feature map so we have output which represent features associated with a particular emotion. Using a deep convolutional network[20]to classify emotions from the image into 7 basic emotions. The processed image is fed into the model as shape of (48,48,3). The image then underwent 3 blocks each block consisting of 2 conv2D layers, 1 max pooling layer and one dropout layer. The output obtained by the blocks was flattened out and fed to a group of 2 dense layers, the final one consisting of 7 units representing each emotion. The model outputs the emotions as a one hot vector.

Speech Emotion recognition

Dataset

When the user is answering the questionnaire using the voice-to-voice module, at that time their voices are recorded. This project uses Speech Recognition Model(SER) [21]to detect the emotions of the user while they answer.
RAVDESS[22]and TESS[23]datasets are used for our application. RAVDESS contains voices expressing different types of emotions from 24 different male and female actors. There are 1440 different sound files available. On the other hand, TESS has 2800 files. All together we have 4240 sound. Emotions expressed in this combined dataset: calm, neutral, happy, sad, angry, fear, disgust, and surprise.

Data preprocessing

The file from RAVDESS and TESS datasets is first converted into Pandas dataframes and then combined into a single dataframe. After that data augmentation is performed which helps the Speech Recognition Model to generalize better. Feature extraction helps in analyzing and finding relations with sound. In this step, sound is converted into a computer readable format so that operations can be performed on it. Various transformations are applied on our dataset to extract features, they are: Zero Crossing Rate[24], Chroma Shift[25], Mel Frequency Cepstral Coefficients(MFCC)[12], Root Mean Square(RMS)[26], Mel Spectrogram[27], Spectral Rolloff[28], Spectral Centroid[29]. After extracting features, they are stored in data frames.

Network Architecture

This project uses the Convolutional Neural Network (CNN) approach for Speech Emotion Recognition (SER).Operations were performed on sound files taken from RAVDESS [22] and TESS [23]. Mathematical data is obtained after data augmentation and feature extraction. This data is then passed for training. After a file containing the voice of a person is passed to this model, it detects emotion in the voice as an output.

Projective questionnaire

It is a heuristic questionnaire consisting of 21 questions which classify into 11 psychological diseases[30],[31][32][33],[34],[35],[36],[37], and [38]. It is the preliminary test conducted to streamline the most probable disease that the patient might have. Once the list of most probable diseases is obtained, the next step is to conduct disease specific tests for each disease. Each answer in the questionnaire has some weight (answers may contain weights for multiple diseases). Weight in each answer consists of a score (first character) followed by disease name (e.g. '2depression'). Once all the questions are asked, the score is calculated and according to the threshold value of each disease, the disease is listed down.

Following is the architecture of the SER using CNN [39]:

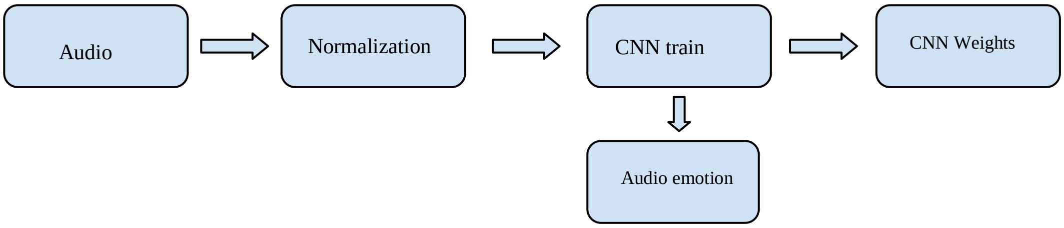

Fig 1:Architecture of SER using CNN

Emotion Score

Emotion Representation

Confining emotions into a one dimensional restricts the ability of the basic emotions to be combined and expressed into complex emotions. There are multiple models out there for multi-dimensional representation of emotion such as the big five personality theory also known as the five-factor model proposed in [40]which classifies emotion into extroversion, agreeableness, neuroticism, conscientiousness, and openness to new experiences. Morgan[41]proposed the multi-dimensional representation for emotions which are calm, happy, sad, and angry. Two of the proposed dimensions were Activation level and pleasantness level. But these representations proposed in [41]are limited to only four emotions. Mancini[42]provided a more diverse classification of emotions into valence and arousal, and we decided to proceed with those values. The emotion score used a multi-dimensional interpretation of emotions instead of one dimensional. The emotions were converted into pleasantness and activation level for emotions calm, happy, sad, and angry. The confidence percentage obtained from facial emotion recognition and speech recognition was multiplied with the Activation and pleasantness value range and then added with the base value of the range for each emotion accordingly. The final activation and pleasantness value represents the mean activation and pleasantness for all their exhibited emotions. Assigning a score to these values results in an average of pleasant emotions to be on a higher emotion score than the average of unpleasant emotions.

Table 2: Valence (pleasantness) and arousal (activation) from [42]

Emotion	Valence	Arousal
Happy	6.92	5.35
Neutral	4.24	2.79
Angry	4.15	4.42

Emotion	Valence	Arousal
Disgust	4.24	4.03
Sad	3.47	3.91
Fear	4.51	4.68

Verdict

Adding an extra layer of assurance to the mental illness classification, a verdict was taken. Because emotion prediction using solely facial or only spoken emotion recognition may occasionally result in expressed emotion being undiscovered, we combine the results from both, improving the likelihood of exhibiting emotion being identified. The basic emotions associated with a specific mental illness were taken into account and if the emotions exhibited by the user which were obtained from speech and facial emotion recognition did not match at least 60% of the emotions associated with that particular illness, the user was asked to take the test again. If the emotions matched with the associated emotions then the user was asked to consult a doctor and was provided with possible treatments.

Additional functions

Social media analysis

Users express their views, emotions and sentiment through social network sites like Twitter, Facebook etc. To get the analysis of their Twitter posts, an additional feature has been created in the application where the patient can get a review of sentiments of their Twitter posts. According to the research paper[43] the tweets and the twitter data are retrieved from Twitter social media with the help of Twitter API. Each message obtained is represented by a bag of words. After assigning individual scores to all the words, average polarity and subjectivity are calculated. After the sentiment analysis is done, the patients are shown output in the form of Tweets followed by their sentiment classifications and validations. The sentiment values are of three polarities- positive, negative, neutral.

Doctor recommendation and Patient history

The treatments and the recommendations are based on the diseases diagnosed like depression, anxiety, bipolar, ptsd, etc. and the level of severity that the patient is facing. If the severity of any disease is low or mild, then basic health recommendations related to the diseases are suggested. On the other hand, when the severity of the disease is greater, then a proper and strict treatment and schedule are suggested as some of them mentioned in [44]The best doctors according to their specialties are recommended for the patients. Patient's records consist of diseases and their score are maintained in order to track progress. The record is updated every time the patient takes the diagnosis test.

Result

Training of Speech Emotion Recognition

Below table shows the variation in accuracy when different classification methods and datasets were used.

Table 3: Accuracy of different datasets using classifiers MLP and CNN

Classifiers	Datasets Accuracy (%)				Combined Datasets Accuracy (%)	
	RAVDESS	CREMA-D	TESS	SAVEE	All 4	RAVDESS and TESS
MLP	59.0	40.6	98.9	68.1	57.9	84.8
CNN	60.09	49.46	98.81	67.22	60.3	87.96

Facial Emotion Recognition

Table 4: Evaluation metrics for facial recognition models

	Optimizer	Training Acc	Validation Acc
Alternate pooling and Convolutions	RMSprop	76.17	46.68
Alternate pooling and Convolutions	Adam	80.86	48.24
Three Block architecture	Adam	94.63	56.84
Three Block architecture	RMSprop	90.61	61.86

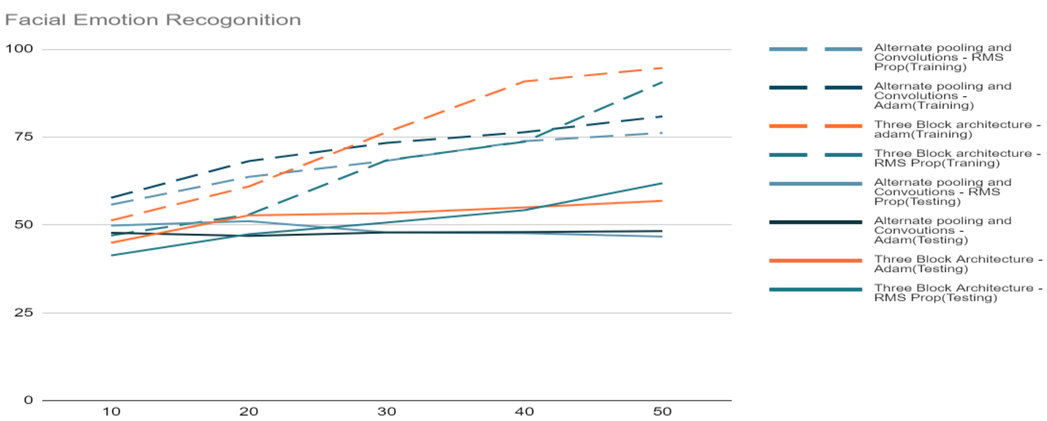

Fig 2: Evaluation metrics comparison between facial emotion recognition model and its variants

* Block architecture consists of 2 convolutions, 1 max pooling and 1 dropout layer in each block

Speech Emotion Recognition

For this model, training and testing logs are compared for different optimizers and block architectures. The optimizers used are RMSprop and Adam[45]. For the block architecture, each CNN block consists of three layers: Conv1D, MaxPooling and Dropout. Loss and accuracy are obtained for 3 CNN blocks first. After that the number of blocks was increased from 1 to 5 blocks.

Table 5: Evaluation metrics for speech recognition models

	Optimizers	No. of blocks	Loss	Accuracy (%)	Val_loss	Val_accuracy (%)
1	RMSprop	3	1.2090	58.71	1.0407	65.57
		4	1.1049	61.05	0.9499	68.24
		5	1.2074	57.44	1.0362	65.09
2	Adam	3	0.2902	89.06	0.4042	87.42
		4	0.1324	95.44	0.4708	87.96
		5	0.1606	94.30	0.5115	87.42

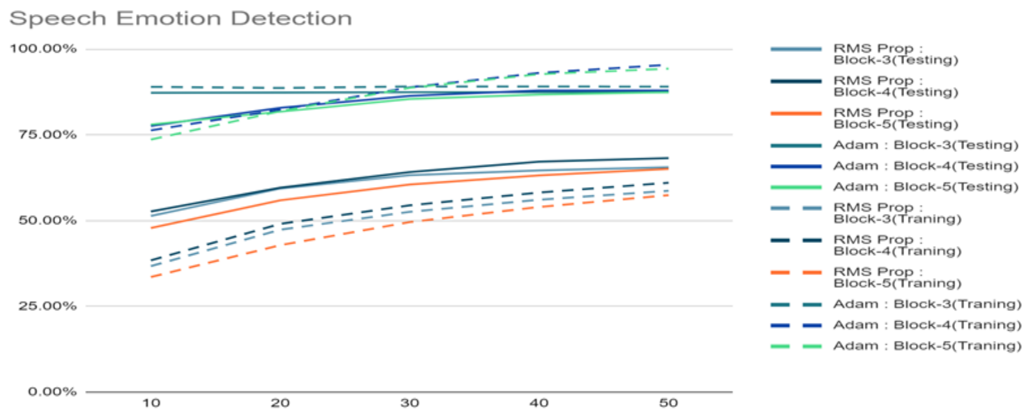

Fig 3: Evaluation metrics comparison between speech recognition model and its variants

Discussion

The Three block architecture implemented using RMSprop[45]as the optimizer provided the best results for accuracy in training and test set. For the optimizer

RMSprop, the accuracy and loss do not differ much when the number of CNN blocks is changed. Maximum accuracy achieved is 68.24% when the number of blocks is 4. Adam[45]optimizer, the combination of Momentum-based GD and RMSprop, is the best optimizer. Highest accuracy achieved is 87.96% for Adam optimizer for 4 CNN blocks. Combining emotions from both speech and face emotion detection models allows us to detect emotions that neither model can detect on its own. Thus, the models that provided the highest accuracy on both training and test sets among the several models that were trained and evaluated are chosen.

Threshold values for projective questionnaire

It is the value above which the disease is considered for the diagnosis. It is defined as the minimum percentage of maximum score of a disease in the given dataset. The threshold value is set to 60%.

Interpretation for questionnaire

According to the score and the threshold value, a decision is taken whether a disease will be considered for further diagnosis or not. If the obtained score percentage is greater than or equal to threshold value than the disease is considered otherwise it is not considered.

Threshold values for Facial Emotion Recognition

The activation value for each emotion proposed in [42]existed in a range, for eg. arousal value for 'happy' was in range 5-5.35 and similarly valence value for 'happy' existed in range 6-7 .The Emotional score assigned to a high valence score(above 6) and high activation score (above 5) was 100. For moderate valence and arousal score (5-6 and 4.5-5) score assigned was 75 and so on.

Training of Facial Emotion Recognition

In the finalized architecture block architecture is used with an optimizer as RMSprop[45]with a learning rate of 0.00001 and loss as categorical cross entropy. Model is trained for 70 epochs and attained accuracy of 92% on training set and 62% as test accuracy. While training the facial emotion recognition model two approaches are considered, first was using a shallower model with alternate convolutions, dropout layers which were pooled at end and a deeper model consisting of a block architecture where each block consisted of two convolution layers one max pooling and one dropout layer. In our training we found that block architecture achieved higher accuracy than the shallower model.

Training of Speech Recognition Model

The dataset is split into 70% training 30% testing set. A sequential model is initialized and different layers are added to it. The CNN model consists of Convolution layer (1D), Max Pooling layer, Activation layer, Dropout layer, dense layer. 4 blocks of Conv1D and MaxPooling1D are added along with 2 dropout layers. After that, one Flatten and two dense layers are added. Model is fit using

Reduce LR on Plateau which helps in reducing learning rate when a metric stops improving. 50 epochs are performed which results in a training accuracy of 95.44% and a testing accuracy of 87.96%. Clearly from data in table 3.1.1, CNN works better than Multi-Layer Perceptron (MLP). On one hand, 4 datasets, namely RAVDESS, TESS, CREMA-D[46]and SAVEE[47] are used. The problem with combining these datasets was that the number of files for some emotions, like calm, were very few, which in turn creates a less accurate model. On the other hand, when the same model was trained for only two datasets, RAVDESS and TESS, then accuracy was boosted. Hence, after some experimentation, it was found that SER works best with CNN classifier for two datasets.

Conclusion

Remote psychological illness detection does not exist as a combined application inculcating facial and speech emotion recognition models. Our initiative uses a multi-modal approach to improve the accuracy of remote mental disease identification. Also, the patient's emotional state is taken into consideration during diagnosis. It provides accurate scores and results considering all the models. Our application recommends treatment or doctors based on the severity of illness and stores their diagnosis records. The continual use of application also provides us with the degree of valence and arousal associated with an average person suffering from a particular mental health illness.

References

1. T. C. P. O. K. K. J. M. C. O. M. J. V. L. P. Rachel M. Satter, "Avatar-based simulation in the evaluation of diagnosis and management of mental health disorders in primary care," *Journal of Biomedical Informatics,* vol. 45, no. 6, pp. 1137-1150, 2012.
2. "Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis.," *Global Health,* vol. 57, no. 16, 2020.
3. G. L. M. F. C. I. F. a. J. M. M.S. Bartlett, "Automatic recognition of facial actions in spontaneous expressions.," *Journal of Multimedia,* 2006.
4. T. Rashid, ". Convolutional neural networks based method for improving facial expression recognition," *International symposium on intelligent systems technologies and applications,* pp. 73-84, 2016.
5. S. P. M. a. B. P. Roopa, "Speech emotion recognition using deep learning," *Int. J. Recent Technol. Eng.*
6. W. J. D. a. L. T. Lim, "Speech emotion recognition using convolutional and recurrent neural networks," *Asia-Pacific signal and information processing association annual summit and conference,* 2016.
7. S. S. I. I. Y. a. W. N. Hamsa, "Emotion recognition from speech using wavelet packet transform cochlear filter bank and random forest classifier," *IEEE,* pp. 96994-97006, 2020.
8. J. B.-G. S. S. R. a. P. B. Hansen, "Getting started with SUSAS: a speech under simulated and actual stress database," *In Eurospeech,* vol. 97, no. 4, pp. 1743-46, 1997.

9. Z. L. X. Z. Z. U. M. D. J. a. A. M. Peng, "Speech emotion recognition using 3d convolutions and attention-based sliding recurrent networks with auditory front-ends," *IEEE Access, 8,* pp. 16560-16572, 2020.
10. R. J. A. G. B. L. L. a. R. M. Fulmer, "Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial," *JMIR mental health,* vol. 5, no. 4, p. 64, 2018.
11. B. S. S. a. S. V. Inkster, "An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study," *JMIR mHealth and uHealth,* 2018.
12. C. S. S. a. Y. T. Ittichaichareon, "Speech recognition using MFCC," *International conference on computer graphics, simulation and modeling,* pp. 135-138, 2012.
13. R. C. T. O. P. K. K. M. J. O. C. J. M. a. P. V. Satter, "Avatar-based simulation in the evaluation of diagnosis and management of mental health disorders in primary care.," *Journal of Biomedical Informatics,* vol. 45, no. 6, pp. 1137-1150, 2012.
14. S. B. C. C. W. a. H. L. Šabanović, "PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia," *IEEE 13th international conference on rehabilitation robotics (ICORR) ,* pp. 1-6, 2013.
15. Kargarbideh, "Design and Development of the eBear: A Socially Assistive Robot for Elderly People with Depression," Doctoral dissertation, University of Denver, 2016.
16. K. S. R. a. W. J. Choo, "The PHQ-9," Journal of General Internal Medicine, vol. 16, no. 9, pp. 606-613, 2001.
17. P. a. C. A. Carrier, "Challenges in representation learning: Facial expression recognition challenge," Kaggle, 2021.
18. K. P. T. a. P. C. Potdar, "A comparative study of categorical variable encoding techniques for neural network classifiers.," International journal of computer applications, vol. 175, no. 4, pp. 7-9, 2017.
19. Y. K. K. a. F. C. LeCun, "Convolutional networks and applications in vision," Proceedings of 2010 IEEE international symposium on circuits and systems, pp. 253-256, 2010.
20. L. H. M. a. K. K. Ankile, "Deep Convolutional Neural Networks: A survey of the foundations, selected improvements, and some current applications," 2020.
21. D. a. D. L. Yu, " Automatic Speech Recognition," Springer London limited, 2016.
22. S. a. R. F. Livingstone, "RAVDESS Emotional speech audio," Kaggle, 2019.
23. E. J. Lok, "Toronto emotional speech set (tess)," Kaggle, 2019.
24. C. a. T. G. Panagiotakis, "A speech/music discriminator using RMS and zero-crossings.," 2002 11th European Signal Processing Conference. IEEE., pp. 1-4, 2002.
25. M. N. A. S. A. a. S. D. Kattel, "Chroma feature extraction," Conference: Chroma Feature Extraction using Fourier Transform, 2019.
26. C. a. T. G. Panagiotakis, "A speech/music discriminator using RMS and zero-crossings," 2002 11th European Signal Processing Conference, IEEE, 2002.
27. B. Thornton, "Audio recognition using mel spectrograms and convolution neural networks," 2019.

28. M. K. Z. a. V. D. Kos, "Acoustic classification and segmentation using modified spectral roll-off and variance-based features," Digital Signal Processing, vol. 23, no. 2, pp. 659-674, 2013.
29. M. F. M. B. E. P. D. a. T. A. Massar, "Fast computation of spectral centroids," Advances in Computational Mathematics, vol. 35, no. 1, pp. 83-97, 2011.
30. T. J. B. S. J. K. S. T. L. a. C. H. Emser, "Assessing ADHD symptoms in children and adults: evaluating the role of objective measures.," Behavioral and Brain Functions, vol. 14, no. 1, pp. 1-14, 2018.
31. P. Gillig, "Dissociative identity disorder," A controversial diagnosis. Psychiatry (Edgmont), vol. 6, no. 3, p. 24, 2009.
32. R. Hirschfeld, "The Mood Disorder Questionnaire: a simple, patient-rated screening instrument for bipolar disorder.," Primary care companion to the Journal of Clinical Psychiatry, vol. 4, no. 1, p. 9, 2002.
33. L. B. J. T. S. B. E. K. J. R. D. W. E. G. K. C. E. J. P. a. P. M. Horowitz, "Ask Suicide-Screening Questions (ASQ): a brief instrument for the pediatric emergency department. Archives of pediatrics & adolescent medicine,," vol. 166, no. 12, pp. 1170-1176.
34. C. T. J. G. D. a. B. S. Lancaster, "Posttraumatic stress disorder: overview of evidence-based assessment and treatment. Journal of clinical medicine," vol. 5, no. 11, p. 105, 2016.
35. E. Freeman, "Premenstrual syndrome and premenstrual dysphoric disorder: definitions and diagnosis.," Psychoneuroendocrinology, vol. 28, pp. 25-37, 2003.
36. B. R. P. J. a. M. J. Rapp, "Evidence-based assessment of obsessive–compulsive disorder. Journal of central nervous system disease," 2016.
37. H. M. U. J. A. Y. F. K. F. K. a. K. H. Takeuchi, "A questionnaire-based study of the views of schizophrenia patients and psychiatric healthcare professionals in Japan about the side effects of clozapine," Clinical Psychopharmacology and Neuroscience, vol. 14, no. 3, p. 286, 2016.
38. W. S. S. a. Y. M. Mahmud, "Postpartum depression: a survey of the incidence and associated risk factors among Malay women in Beris Kubor Besar, Bachok, Kelantan.," The Malaysian journal of medical sciences, vol. 9, no. 1, p. 41, 2002.
39. H. Murugan, " Speech emotion recognition using cnn.," International Journal of Psychosocial Rehabilitation, vol. 24, no. 6, 2020.
40. U. Barańczuk, "The five factor model of personality and emotion regulation: A meta-analysis. Personality and Individual Differences," vol. 139, pp. 217-227, 2019.
41. S. Morgan, "Categorical and dimensional ratings of emotional speech: Behavioral findings from the Morgan emotional speech set.," Journal of Speech, Language, and Hearing Research, vol. 62, no. 11, pp. 4015-4029, 2019.
42. G. B. R. A. S. A. F. a. T. E. Mancini, "Recognition of facial emotional expressions among Italian pre-adolescents, and their affective reactions," Frontiers in psychology, vol. 9, p. 1303, 2018.
43. R. R. a. A. P. Manguri, "Twitter sentiment analysis on worldwide COVID-19 outbreaks," Kurdistan Journal of Applied Research, pp. 54-65, 2020.
44. B. H. K. a. M. P. Burns, "Effective treatment for mental disorders in children and adolescents.," Clinical child and family psychology review, vol. 2, no. 4, pp. 199-254, 1999.

45. R. a. P. A. Poojary, "Comparative Study of Model Optimization Techniques in Fine-Tuned CNN Models.," International Conference on Electrical and Computing Technologies and Applications (ICECTA), pp. 1-4, 2019.
46. E. J. Lok, "Crowd Sourced Emotional Multimodal Actors Dataset (CREMA-D)," Kaggle, 2019.
47. T. Sunkaraneni, "SAVEE database," Kaggle, 2019.