

Prediction of Depression using Convolutional Neural Networks from Audio Visual Expression

Bhagyashree Lambture¹, Amruta Aphale²

¹Assistant Professor, Department of Artificial Intelligence and Data Science,
AISSMS Institute of Information Technology, Pune, India

²Assistant Professor Department of Artificial Intelligence and Data Science,
AISSMS Institute of Information Technology, Pune, India

Corresponding Author: Bhagyashree Lambture (blambture@gmail.com),

Amruta Aphale (aphaleamruta@gmail.com)

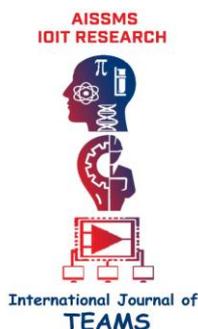
Article Information

Article history:

Received Nov 1, 2022

Revised Dec 2, 2022

Accepted Dec 16, 2022



ABSTRACT

Depression is a temperament issue that causes a persistent sentiment of bitterness and loss of interest. It influences how you feel, think and act and can prompt variety of emotional and physical issues. Especially when long-lasting and with moderate or severe intensity, depression may become a serious health condition. Depression can be recognized by the facial expressions from video and spoken expressions from audio. Convolutional Neural Networks (CNN) was initially created for the purpose of learning descriptive raw waveforms and deeply learnt features for visual expressions. Second, vocal data processing employs the MFCC approach. The most typical application for it is speaker detection for audio features.. To measure the depression scale the PHQ-9 questionnaires assessment is used to predict the presence and severity of depression. It may use to predict the depression scale to analyze the users/patients emotional or mental health.

KEYWORDS: Depression analysis, Deep learning, Visual expression, Vocal expression, PHQ etc.

1. INTRODUCTION

Depression is a perspective issue that raises an industrious opinion of ruckus. Additionally called significant burdensome issue or clinical sadness, it influences how you feel, think and carry on and can prompt an assortment of passionate and physical issues. With regards to nations, India is the most discouraged nation on the planet, as indicated by the World Health Organization, trailed by China and the USA. Comprehensively, in excess of 264 million individuals of any age experience the ill effects of sorrow. Depression is a main source of handicap worldwide and is a significant supporter of the general worldwide weight of illness. There are powerful mental and pharmacological medicines for moderate and serious wretchedness. Depression is unique in relation to normal state of mind variances and brief enthusiastic reactions to challenges in regular daily

existence. Particularly when dependable and with moderate or extreme power, gloom may turn into a genuine wellbeing condition. It can make the influenced individual endure enormously and work ineffectively busy working, in the home and at school. Depression may, in the best case scenario, lead to self-destruction. Every year, self-destruction causes the deaths of about 800,000. Self-destruction is the primary cause of death among those aged 15 to 29. The downturn rates taking off among kids as youthful as 12 and youthful grown-ups up to 25 and the rate among school age grown-ups (20-21) has expanded the most. Affecting an expected 300 million individuals, sadness is the most-well-known mental issue and by and large influences ladies more regularly than men.

Signs of depression vary from person to person, so it is important to monitor the symptoms in early stages because it affects person's daily lifestyle. Feeling

helpless, losing interest in daily activity, changing appetites or weight, excessive or unwilling sleeping, irrational behavior, losing energy, feeling self-hatred, fixation issues, and inexplicable throbs or torments are all negative reactions to depression. Depression can be cause because of budgetary challenges, joblessness, worry from work or instruction, loss of friends and family, youth physical or sexual maltreatment, relationship inconveniences, and inconsistent parental treatment and so forth.

A rapidly expanding area of research is automatic depression assessment based on facial expression. The current thorough audit of approaches in use, as described in more than sixty distributions over the last ten years, is focused on picture processing and machine learning calculations. There is no all-encompassing criteria that illustrate which highlight is typically helpful for treating depression. Deep learning has already been successfully used to combat depression. There are numerous variations of the deep learning system, including auto-encoders, convolutional neural networks, probabilistic models, and single-layer learning models. Convolutional neural networks have been widely applied in many approaches to achieve cutting-edge profitability. Be that as it may, the surface order situation has end up being genuinely effective. The CNN-based methodology having informational index with perceptible pictures and the best-distributed minuscule pictures. In this study, we appropriately examine how using CNN to evaluate facial and voice emotions can benefit the expectation of depression seriousness. The proposed research was done to forecast the level of depression based on the most recent input of vocal and visual data. The PHQ-9 is the 9-question depression size of Patient Health Questionnaire (PHQ). The PHQ is a self-controlled variant screening device that evaluates 12 mental and passionate wellbeing issues. The PHQ-9 inquiries depend on symptomatic models of depression and get some information about the patient's involvement with the most recent fourteen days.

Questions are about the degree of enthusiasm for getting things done, feeling down or discouraged, trouble with resting, vitality levels, dietary patterns, self-discernment, capacity to think and speed of working and musings of self-destruction. Reactions go from "0" (Not by any means) to "3" (almost consistently). The PHQ-9 is also the most widely used depression measure in the National Health Service of the United Kingdom, which requires providers to use a depression screening tool while treating depression.

Table- I: PHQ-9(Patient Health Questionnaire- Depression Severity Level)

Total Score	Depression Severity
1-4	Minimal Depression
5-9	Mild Depression
10-14	Moderate Depression
15-19	Moderately severe Depression
20-27	Major/Severe Depression

2.LITERATURE SURVEY

Early identification and treatment of depression is fundamental in advancing reduction, forestalling backslide, and lessening the passionate weight of the sickness. The majority of decisions made today are emotive, inconsistent among experts, and expensive for those who may be in desperate need of assistance. The convolutional neural network (CNN) and multipart intelligent preparation methods suggested in this paper offer a novel approach to dealing with robotized despondency location in discourse. 2568 speech samples from 77 non-discouraged and 30 discouraged people were used to evaluate the model. In test led, information were applied to remaining CNNs as spectrograms—pictures auto-produced from sound examples. The test results acquired utilizing diverse ResNet structures gave a promising gauge exactness arriving at 77%. The framework can be utilized freely, or as a component of a more perplexing, half and half, or multimodal arrangement. The primary focal points of this technique are its relative effortlessness, combined with its best in class exactness. [1]

Medicinal services twitter examination manages the wellbeing related tweets through wistful investigation by the patients themselves. The use of notion examination has developed colossally. It's another stage for association through patients, huge numbers of them taking an interest in dynamic for better after effects of human services treatment. Its application in social insurance can possibly examine and improve the soundness of a nation. Using Multinomial Nave Bayes and the Support Vector Regression (SVR) Algorithm as a classifier, we looked at tweets about people's general well-being for Depression and Anxiety from the combined tweets. [2]

An expanding number of individuals experiencing emotional wellness conditions resort to online assets (specific sites, internet based life, and so on.) to share their sentiments. Early depression location utilizing internet based life information through profound learning models can assist with changing life directions and spare lives. Be that as it may, the exactness of these models was not fulfilling because of this present reality imbalanced information disseminations.

A profound learning model (X-A-BiLSTM) for depression recognition in unbalanced online life

information was developed to address this problem. Two key components make up the X-ABiLSTM model: the XGBoost, which is used to reduce information unevenness, and the Attention-BiLSTM neural system, which enhances grouping limit. The Reddit Self-revealed Depression Diagnosis (RSDD) dataset was chosen, consisting of about 107,000 coordinated control clients and 9,000 clients who claimed to have been diagnosed with depression. Results show that, on the RSDD dataset, our methodology fundamentally outperforms the previous best in class models.[3]

Depression is a typical mental issue in more established individuals which is difficult particularly in low and center salary nations. In this investigation, exhibit a novel methodology for identification of depression utilizing clinical information acquired on-going Mysore investigations of Natal impacts on Maturing and Health (MYNAH), in south India where the individuals have experienced a complete evaluation for psychological capacity, emotional wellness and cardio metabolic messes. The proposed model created utilizing XGBoost helps in recognition of depression in MYNAH associate by utilizing highlight choice using the readily available data for more precise evaluations. [4]

Significant depressive issue is among the most well-known and destructive psychological well-being issues. A few profound learning models have been recommend for video-put together identification of misery based with respect to the outward appearances of subjects. To anticipate the downturn level, these designs are regularly demonstrated for relapse with Euclidean misfortune. Therefore, they don't use the information circulation, nor investigate the usual relationship between face images and depressive symptoms, and have restricted power to noisy and ambiguous marking. In order to accurately predict depression levels by circulatory learning, it proposes a deep learning design. It depends on another desire-fate task that enables evaluation of the concealed information appropriation over depressive levels, where anticipated circulation estimations are advanced to move toward the reality levels. The strategy can produce accurate depression levels that are much below the vulnerability mark. Broad tests AVEC2013, AVEC2014 datasets show that the proposed design speaks to a compelling methodology that can outflank cutting edge strategies. [5]

Depression is a main source of mental sick wellbeing, which has been found to expand danger of early demise. Also it is a significant reason for self-destructive ideation and prompts noteworthy impedance in everyday life. Feeling man-made brainpower is a field of continuous examination in feeling discovery, explicitly in the sector of text mining. The coming of web based media sources has brought about noteworthy client information being accessible for estimation examination of text and pictures. It expects to apply normal language handling

on Twitter channels for leading feeling examination concentrating on depression. Singular tweets are named nonpartisan or negative, in view of a curated word-rundown to distinguish depression propensities. During the time spent class forecast, support vector machine and Naïve-Bayes classifier have been utilized. The outcomes have been introduced utilizing the essential grouping measurements including F1-score, exactness and confusion matrix. [6]

During depression functioning of nervous system may changes, for example conduct of the vocal folds. Portraying these adjustments in an exact way from discourse signals is a non-inconsequential assignment, as this commonly includes dependable detachment of the voice source data from them. By abusing the capacities of CNNs to take in task-pertinent data from the information crude signs, exploring a few techniques to demonstrate voice source related data for depression recognition. In particular, we explore displaying of low pass sifted discourse signals, straight expectation leftover signs, homo-morphically separated voice source signals with no repetition Sifted signs to discover voice source-related information for the place of discouragement. The examinations show that sub segmental level displaying of straight forecast leftover signals or zero recurrence separated signs prompts frameworks better than the best in class low level descriptor based frameworks and deep learning based frameworks demonstrating the vocal plot framework information.[7]

Depression as a confusion has been an incredible worry in our general public and has been never-endingly an intriguing issue for scientists on the planet. Despite the vast quantity of research into how different temperaments, as well as tension, melancholy, and stress, bolstered activity logs compiled by unavoidable figuring gadgets like cell phones, predicting discouraged states of mind keeps on being an open inquiry. Depression investigation and self-destructive ideation discovery framework is designed for anticipating the self-destructive acts dependent on depression. Author gathered constant information from understudies and by forcing them to fill out surveys like PHQ-9 (Parent wellbeing poll) comprising of inquiries like what's your age? Or, on the other hand, are you a typical student/college student? Additionally, it was treated into substantial information with pertinent features like age, sex, school normalcy, and so forth. Then, based on its seriousness, it is prepared and classified using classification machine calculations into five stages of depression: minimal or none, gentle, moderate, respectably extreme, and extreme. XGBoost was used to achieve the highest precision, for instance 83.87%, in this dataset. Additionally, data was collected through tweets, and using arrangement calculations, it was determined whether the tweeter was discouraged or not. For instance, the Logistic Regression classifier provided an 86.45% accuracy rate for the same data.[8]

Depression is a common mental condition that has a big impact on people's daily lives and careers. Current depression location depends solely on the clinical meeting and organized poll, devouring a ton of clinical assets and gambling a scope of abstract predispositions. Probably accomplish an advantageous and target Depression identification framework, which can help clinicians in their determination of clinical discouragement. Developed one of the largest datasets for depression by proposing an exploratory worldview based on picture cognizance to record the members' response times and eye development. The author uses a typical classifier of Support Vector Machine to describe sad individuals after removing the comparative R-T (Reaction Time) highlights and E-M (Eye Movement) inclusions that can reflect the member's consideration propensity and ordinary controls. This technique accomplishes precision up to 86%, which beats the past related strategy. System gets exceptional order performance.[9]

Depression is seen as the biggest supporter of worldwide handicap and a significant purpose behind self-destruction. It affects the language use reflected in the composed content. The key target of this investigation is to analyze Reddit clients' presents on distinguish any elements that may uncover the downturn perspectives of pertinent online clients. For such reason, author utilizes the Natural Language Processing (NLP) procedures and AI ways to deal with train the information and assess the proficiency of proposed technique. Author recognizes a dictionary of terms that are more normal among discouraged records. The outcomes show that the proposed technique can altogether improve execution precision. The quality and adequacy of the consolidated highlights (LIWC + LDA + bigram) are most effectively exhibited with the Multilayer Perceptron (MLP) classifier bringing about the top execution for depression recognition. Better execution improvement can be accomplished by legitimate element determinations and their numerous element combinations.[10]

Audio depression acknowledgment technique dependent on convolution neural system and generative enmity organize model. As a matter of first importance, preprocess the informational collection, evacuate the drawn out quiet fragments in the informational index, and join the rest into another sound document. The highlights of the speech signal, such as Mel-scale Frequency Cepstral Coefficients (MFCCs), transient liveliness, and spectrum entropy, are then eliminated based on the computation of sound distinction standardization. The information base for creating the model is the separated lattice vector highlight information, which speaks to the unique qualities of the subject's own voice. The prior model is improved with the use of DR AudioNet, and the acknowledgment arrangement is completed by harmonizing the two adjacent portions with the current sound section. The results of the exploratory analyses

on the datasets Avid-Corpus and DAIC-WOZ demonstrate that the proposed method effectively reduces the depression acknowledgment error in comparison to other existing approaches, and the RMSE and MAE values obtained on the two datasets are superior to the analysis calculation by over 5%. [11]

3. METHDOLOGY

The proposed method is used to create a framework that can automatically predict the level of depression using the client's vocal and visual cues. In this experiment, the framework devices looked at the severity of depression using the deeply acquired characteristics. CNN have proven to be a powerful procedure in image recognition, video analysis and natural language processing. To assess the depression, audio recordings of user are preprocessed and audio segments of the participants are acquired. With the aid of the MFCC method, low level features are recovered from audio segments. CNN uses the raw visual images to directly develop its deep learning skills. For processing voice information, the MFCC technique is used. CNN and MFCC based recurrent neural network is trained to evaluate user is depressed or non-depressed. To quantify the depression scale the PHQ-9 polls fast appraisal is utilized to anticipate the nearness and seriousness of depression. The thought process of plan the depression identification framework is to foresee the depression scale to break down the clients/patient's feeling or mental health.

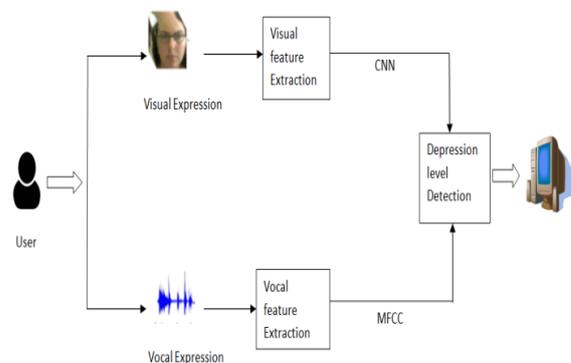


Fig. 1. System Architecture

Convolutional Neural Network for visual data processing: Convolutional neural network (CNN) is an uncommon engineering of artificial neural system, CNN utilizes a few highlights of the visual cortex. One of well-known employments of this design is picture characterization. The visual expression is gone through a progression of convolutional, nonlinear, pooling layers and completely associated layers, and afterward creates the yield.

Consistently, the convolution layer comes first. It has the image (a pixelated grid) embedded in it.

Imagine that the information grid is browsed beginning at the upper left corner of the image. The product then selects a smaller grid there, known as a channel (or neuron, or center). When it happens, the channel creates convolution, for instance moving the information picture along. The task of the channel is to replicate its attributes down to the first pixel values. After the filter has been applied to every position, a network is acquired. The system will comprise of a few convolutional systems blended in with nonlinear and pooling layers. At the point when the visual expression goes through one convolution layer, the yield of the primary layer turns into the contribution for the subsequent layer. Furthermore, this occurs with each further convolutional layer. After consummation of arrangement of convolutional, nonlinear and pooling layers, it is important to append a completely associated layer. This layer takes the yield data from convolutional networks.

MFCC for vocal data processing :The Mel Frequency Cepstral Coefficients (MFCC) is the most normally utilized sound highlights in speaker acknowledgment because of their strength in depicting the variety of low frequencies signal. In this work, the low-level highlights are characterized as the MFCC coefficients and they are removed from the preprocessed sound accounts. The discourse signal is first isolated into outlines. A cepstral highlight vector is then produced for each casing. The sound signs are preprocessed. Next, the low-level sound descriptors are extricated and standardized. MFCC highlights are fed to the deep neural system for depression identification. The named training set is extended through transfer learning and information increase where new sound fragments of the members speech just are produced. At last, the MFCC-based Recurrent Neural Network is prepared to distinguish depression/non-depression.

4. RESULTS AND DISCUSSION

Experts in brain science, software engineering, and associated controls are paying more and more attention to automatic depression detection. In order to identify depression, vocal and visual emotions are then used. The system is used to recognize the victim's depression severity level based PHQ-9 quick assessment. CNN is having better accuracy when it is dealing with big data. It helps to decrease the error rate and easily classifies the small object. The MFCC is the most ordinarily utilized sound highlights in speaker acknowledgment because of their vigor in portraying the variety of low frequencies signal.

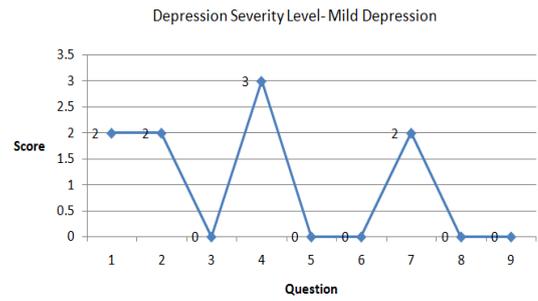


Fig. 2. Depression Severity Level (Mild Depression)

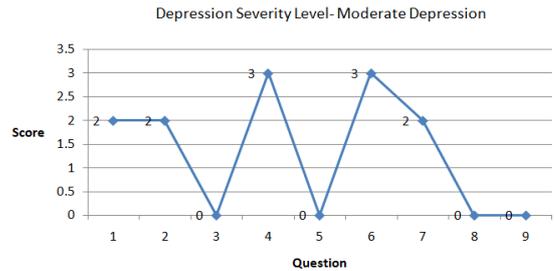


Fig. 3. Depression Severity Level (Moderate Depression)

As per fig. 2 and fig. 3 we get the following scores for Depression Severity Level:

Table- II: Result

Score Result	Depression Severity Level
9	Mild Depression
12	Moderate Depression

5. CONCLUSION

Behavior of compared to a non-depressed person, the behavior of a discouraged person exhibits relative shift in speech pattern (vocal expressions) and outward appearances (visual expressions)..CNN and MFCC based intermittent neural system is prepared to assess client/patient is depressed or non depressed and the PHQ-9 surveys speedy evaluation investigate the depression seriousness. Depression Detection System which is used to reduce the number of early deaths caused due to depression along with improvement of human safety. Automatic Detection System studies depression prediction based on visual information and vocal information. In our research attempt, the visual, vocal information and PHQ-9 depression assessment used to detect the severity level of depression. It may foresee the depression scale to analyze the users/patients emotional or mental health.

6. FUTURE WORK

For future work depression forecast is determine by utilizing text information alongside the accommodating solutions for conquer depression.

Conflict of Interest:There is no conflict of Interest in this work.

REFERENCES

- [1] Chlasta, Karol, Krzysztof Wołk, and Izabela Krejtz. "Automated speech-based screening of depression using deep convolutional neural networks." *Procedia Computer Science* Vol. 64, 2019, pp. 618-628.
- [2] Michael M. Tadesse, Hongfei Lin, Bo Xu, And Liang Yang, 2018, Detection Of Depression-Related Posts In Reddit Social Media Forum, *Ieee Access*, 2169-3536.
- [3] Zhiyong Wang, Longxi Chen, Lifeng Wang, Guangqiang Diao, 2019, Recognition Of Audio Depression Based On Convolutional Neural Network And Generative Antagonism Network Model, *IeeeAccess*.
- [4] P. Arora and P. Arora, "Mining Twitter Data for Depression Detection," 2019 International Conference on Signal Processing and Communication (ICSC), NOIDA, India, 2019, pp. 186-189, doi: 10.1109/ICSC45622.2019.8938353.
- [5] Q. Cong, Z. Feng, F. Li, Y. Xiang, G. Rao and C. Tao, "X-A-BiLSTM: a Deep Learning Approach for Depression Detection in Imbalanced Data," 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 2018, pp. 1624-1627, doi: 10.1109/BIBM.2018.8621230.
- [6] V. Arun, P. V., M. Krishna, A. B.V., P. S.K. and S. V., "A Boosted Machine Learning Approach For Detection of Depression," 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 2018, pp. 41-47, doi: 10.1109/SSCI.2018.8628945.
- [7] W. C. de Melo, E. Granger and A. Hadid, "Depression Detection Based on Deep Distribution Learning," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 4544-4548, doi: 10.1109/ICIP.2019.8803467.
- [8] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, 2017, pp. 858-862, doi: 10.1109/ISS1.2017.8389299.
- [9] S. P. Dubagunta, B. Vlasenko and M. Magimai.-Doss, "Learning Voice Source Related Information for Depression Detection," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 6525-6529, doi: 10.1109/ICASSP.2019.8683498.
- [10] S. Jain, S. P. Narayan, R. K. Dewang, U. Bhartiya, N. Meena and V. Kumar, "A Machine Learning based Depression Analysis and Suicidal Ideation Detection System using Questionnaires and Twitter," 2019 IEEE Students Conference on Engineering and Systems (SCES), Allahabad, India, 2019, pp. 1-6, doi: 10.1109/SCES46477.2019.8977211.
- [11] Z. Pan, H. Ma, L. Zhang and Y. Wang, "Depression Detection Based on Reaction Time and Eye Movement," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 2184-2188, doi: 10.1109/ICIP.2019.8803181..