A Novel Multimodal Method for Depression Identification

Rahul Singhal[1], Shruti Srivatsan[2], Priyabrata Panda[3]

[1]Department of Computer Science and Engineering, Jaypee Institute of Information Technology, Noida, India
[2]Department of Mathematics, University of Waterloo, Ontario, Canada
[3]Department of Electrical Engineering, Maulana Abul Kalam Azad University of Technology, Kolkata, India

E-mail: [1]rahulsinghal1904@gmail.com , [2]shrutisrisvce@gmail.com, [3]priyabratapanda20000@gmail.com

Abstract

Depression is one of the most prominent mental health issues, characterized by a depressed low mood and an absence of enthusiasm in activities. In terms of early detection, accurate diagnosis, and effective treatment, doctors face a serious challenge from depression, which is a serious global health issue. For patients with this mental disease to receive prompt medical attention and improve their general well-being, early identification is essential. For the purpose of detecting various psychological illnesses including depression, anxiety, and post-traumatic stress disorder, medical audio consultations along with survey responses have been used. A depressed individual displays a range of subtle signs that may be more easily identified by combining the results of multiple modalities. Multimodality involves extracting maximum information from data by using multiple modes, so that the deep learning model can be trained efficiently to give better results. Given that each modality functions differently, combining various modalities is not easy, and each origin of a modality takes on a different form. It is clear from the literature that is currently significant in the area that, combining the modalities yields positive outcomes. A trustworthy approach to identify depression is thus urgently needed because it continues to be a problem for many individuals in today's society. In this work, textual and audio features are incorporated related to the identification of depression, and a novel multimodal approach using an optimized Bi-directional Long Short -Term Memory model that recognizes premature depression is suggested for medical intervention before it develops further.

Keywords: Multimodality, Depression, Deep Learning

DOI: https://doi.org/10.36548/jtcsst.2022.4.001
Received: 07.09.2022, received in revised form: 06.10.2022, accepted: 21.10.2022, published: 01.11.2022

1. Introduction

One of the most common mental disorders prevalent in the world is Depression. It acts as a considerable contributor to the universal burden of diseases leading to self-harm. About 264 million people of various ages are said to undergo depression worldwide [1]. Unipolar depressive disorder, denoted by persistent sadness, is estimated to be one of the main causes of the problem by 2030 [2]. Almost 800 thousand people are said to commit suicide every year, according to a World Health Organization report. It is also considered to be the prevalent reason for death among adolescents aged between 15 and 19 [3]. Reports suggest that for every suicide, the attempts are much larger [4].

To date, doctors have recognised about six types of depression [5]. Some of those are as follows:

1. Major depression - It is the most common type of depression found in individuals, where they tend to lose interest in activities, and a dark mood prevails over the mind. Doctors often prescribe psychotherapy and medication as a treatment.

2. Bipolar disorder - Sometimes also known as manic depression, the individual goes through variations in the mood - ranging from extreme “high” level to depressive “low” level of thoughts.

3. Seasonal Affective Disorder (SAD) - SAD is a unique type of depression where the individual gets into a state of depression with the days getting shorter and the person receiving less sunlight. This often happens during the winter months. Light therapy and antidepressants can help.

4. Persistent Depressive Disorder - If the individual is suffering from depression for two years or more, it’s persistent depressive disorder, formerly known as “dysthymia”. Psychotherapy, medication or a combination of both can be used for treatment.

5. Perinatal depression - This can happen to women either during pregnancy or in the first 12 months after delivery. It affects one in seven women, having devastating effects on the woman and her family.

6. Premenstrual Dysphoric Disorder (PMDD) - PMDD is a severe form of premenstrual syndrome which occurs in women during their periods. Oral contraceptives and antidepressants can be used for treatment.

Depression can be detected by observing how the person behaves, in terms of their withdrawal from society or excess fatigue and loss of interest in things they enjoyed earlier. The high cost of the clinical diagnosis makes seeking help cumbersome for undiagnosed depressed people, especially during the early stages. According to telemedicine, an intelligent diagnosis system serves as a solution for diagnosing at an earlier stage. Lack of funding, a shortage of qualified medical professionals, and the stigma attached to mental illness are all obstacles for providing good care. People with depression are frequently misdiagnosed in nations of all income levels, while those who do not have the disease are over-diagnosed and given antidepressants far too frequently. Identification of mental disorders using computational models has harnessed significant attention in recent years. These systems are devised using Machine Learning (ML) and Deep Learning models which can aid the patient in evaluating to form a self-diagnosis [6].

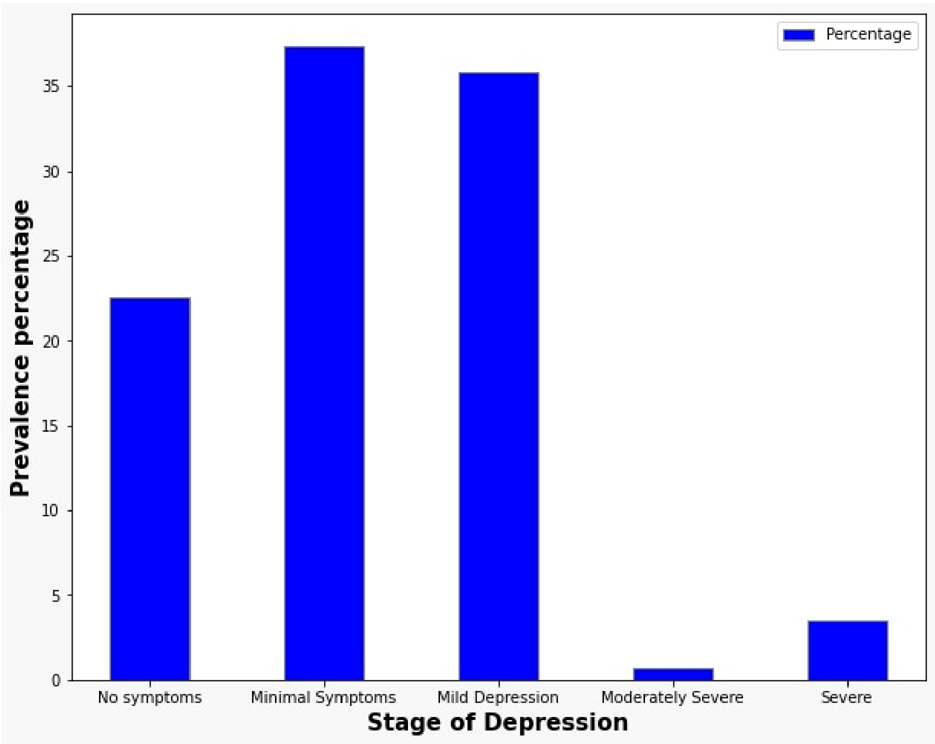

Figure 1. Presence of Depression among working professionals [6]

In a recent study, the existence of depression among IT professionals was found to be high as can be seen in Figure 1. There is an association between depression and factors like the usage of alcohol, problem substance usage, variety of roles played and varied shifts in the corporate workplace. As a result, sensitization about the significance of mental health must be included in the IT industry employee healthcare measures to abstain from the onset of depression. Early recognition of depression can be done by having programs such as employee assistance programs, peer intervention groups, and nominating occupational psychologists in companies [7]. During the pandemic, one can observe a steady increase in people with mental illnesses. It is said that one in five individuals suffer from depression, especially during the pandemic due to a constant fear of the virus and the imposition of lockdown. There has been a 25% increase in the prevalence of depression among individuals in the first year of the COVID pandemic as can be observed in Figure 2.

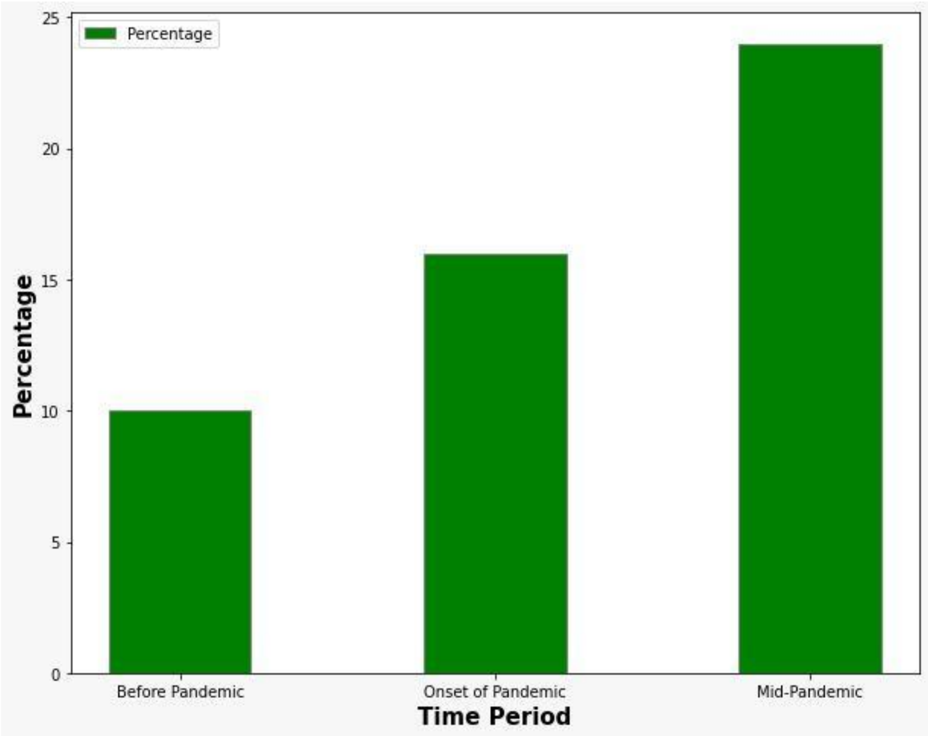

Figure 2. Impact of the Pandemic on Mental Health [8]

The future sections discuss the literature review, the multimodal approach proposed in this work along with the experimental results, conclusion and future work.

2. Literature Review

An inventive approach was developed to recognise mild versions of depression using Electroencephalography (EEG) by Xiaowei L et al. [9]. Deep recurrent Convolutional Neural Networks have been used in combination with three-channel images to classify mental load, resulting in 80% accuracy. Structural Magnetic Resonance Imaging (sMRI) data has been used for the classification of depression by modelling different machine learning models. A combination of Support Vector Machine (SVM) and IG-Random Tree yielded 85% accuracy [10]. Social media data has been mined and used for analysing the onset of depression in [11], obtaining an accuracy of 76% and precision of 86%.

A two-stage multimodal approach has been devised based on the time interval between posts on Instagram. The approach yielded F1-Score of 83%, serving as an early depression detector [12]. In [13], the authors discussed the idea of utilising fMRI and MDD to classify depression in teenagers and young adults. Facial stimuli were used in social cognition. The existing fMRI studies were examined in relation to five functional areas of functioning. Patel et al., reviewed and compared ML models used to identify depression [14].

A review emphasising studies was done on sMRI-based depression identification [15]. Twenty-seven papers on the effectiveness of MDD therapy were examined by Cohen et al. [16] and most of these studies employed Support Vector Machine and Logistic Regression for categorization. AUC was reported to be 0.84 where EEG signals performed better than MRI prediction. Binnewies et al., illustrated the connection between lifestyle, brain anatomy, and depression. No long-term alterations were visible in their work [17]. Regression using relevance vectors was used by Wang et al., to assess the potential of MRI with different modalities and there was compelling evidence to suggest late-onset depression [18]. The authors of [19] depicted that for empirical diagnosis, low-dimensional depiction created using partial least squares regression was advantageous. In comparison to classification methods that tried to make diagnoses purely based on functional connectivity, depression diagnosis based on estimated clinical ratings fared better.

Uddin et al., [20] used a deep model of RNN based on the Long Short -Term Memory (LSTM) approach to simulate two varied emotional states: depressive and non-depressive. DBN, CNN, and logistic regression were the other models implemented. The suggested method outperformed conventional methods on the two datasets giving an accuracy of 98% and 99% when one-hot and TF-IDF features were combined with RNN. CNN-LSTM, ST-CNN, 3D CNN, and 3D VGG networks were used by Mousavian et al. [21]. They constructed functional communication methods, one utilising the Smith Atlas as well as the other using cluster ICA. In order to achieve better accuracy with deep learning models, a new weighted activation function was presented in [22] which decreases the hidden layers. To assess the severity of depression, the weighted activation function was used and evaluated on the depression dataset. Soleiman et al. [23] used two models of CNN+LSTMsemi-contextual text processing and NLP approaches. Depression was detected by collecting the person's tenseness and breathiness throughout the same interview. The speech quality model scored 0.76, whereas the analysis model for text achieved the highest results with an F1-score of 0.8 for depressed persons. In [24], the authors introduced a three-layer LSTM autoencoder which was a successful late fusion method based on a straightforward feed-forward neural network and named it an adaptive nonlinear judge classifier.

The existing solutions were mostly focused on using one modality to detect depression or building a complex model. This work resolves these problems by taking both text and audio modalities and reusing the models from each of them by concatenating them. Thus, the performance is observed to be better than other solutions.

3. Proposed Methodology

3.1 Dataset Description

The dataset is the Distress Analysis Interview Corpus (DAIC) which was developed to support the detection of neurological problems like depression and anxiety [25]. The database consists of 189 sessions of conducted interviews ranging from 7 to 33 minutes, with people identified as sick and healthy based on one or many disorders. It consists of 59 depressed and 130 non-depressed individuals. The Wizard-of-Oz set (DAIC-WOZ) has a computer bot as an interviewer, which holds interviews in a comfortable manner. This way, the subject was able to open up about their issues making the interviewer test the indicators of various disorders [26]. Ellie, the animated digital interviewer, was managed by 2 human examiners, called wizards in another cabin, where the person was left alone in front of a huge computer monitor. Ellie interacted with the subject and assessed their condition by capturing verbal and non-verbal indicators in charge of the assessment of distress disorders, like depression [27].

3.2 Workflow

This work proposes a multimodal approach consisting of textual and audio data. Firstly, the data were loaded from the dataset. Later, each modality was processed with separate bidirectional LSTMs and concatenated as shown in Figure 3. Few of the performance metrics were recorded for the concatenated model, comparing the results with each modality. One can observe enhanced performance on combining the modalities. The work has been simulated using Python 3 in Jupyter Notebook. The workflow is depicted in the form of a flowchart in Figure 3.

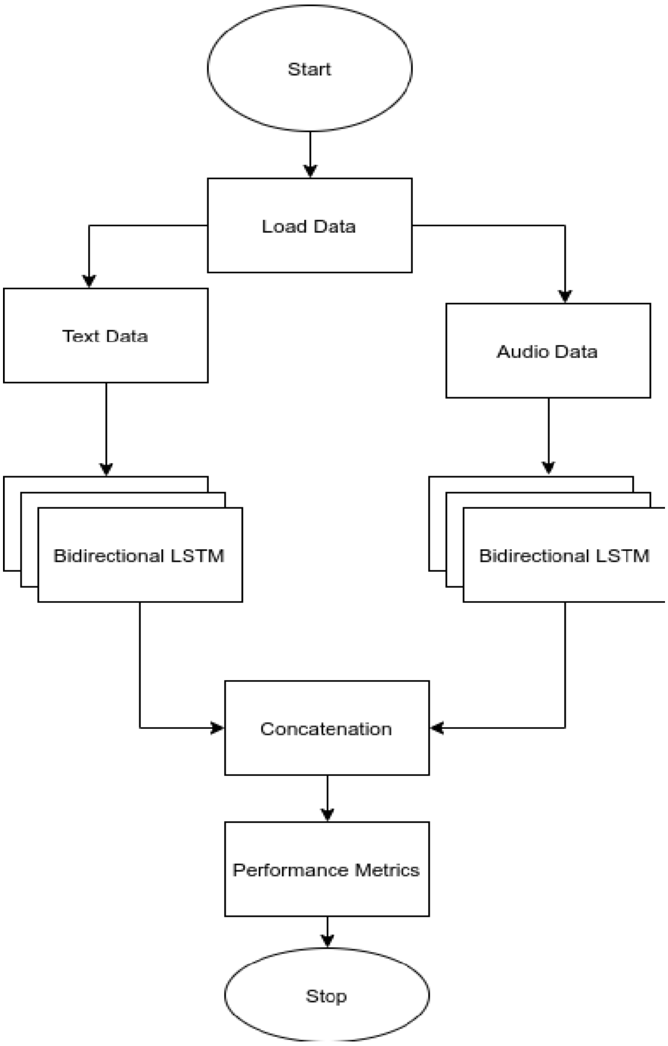

Figure 3. Flowchart of the Proposed Work

3.3 Data Preprocessing and Feature Engineering

Using time stamps to merge the audio and text modalities, data preprocessing was used to discover the time-independent interconnections between the two. To comprehend the context between words, the forced orientation was carried out utilising sentence-level granularity. Weight vectors were applied to the extracted features to manage how much data was passed to the following layers. In order to obtain the most meaningful information possible from each time step, the features were combined and delivered to the Bi-LSTM, which incorporates a gating mechanism to regulate the amount impact different modalities put on the final result.

3.4 Model Implementation

Bidirectional Long-Short Term Memory (BiLSTM) neural network consisting of 3 layers has been used to model the temporal changes of the interviews. They are useful in modelling sequential data. Despite consisting of different topologies, their weights were fixed and the outputs from both modalities were concatenated, forming a multimodal approach. The proposed feedforward network was composed of 3 layers with 128 hidden nodes and a ‘tanh’ activation function. The training examples were padded in the smaller set of text with the number matching with the larger set of audios, carried out by mapping examples that appeared in the same window of the interview. This was done since audio and text inputs for every branch of LSTM had different numbers of examples.

4. Experimental Results

The results of the various modalities along with their concatenation are tabulated below:

Table 1. Results of the Proposed Work

Model	Accuracy	Error Rate
Audio BiLSTM	0.62	5.11
Text BiLSTM	0.729	5.90
Multimodal approach	0.743	5.14

The multimodal approach has efficiently combined both the modalities to yield much better performance when compared to audio or text modality separately. Some of the hyperparameters explored in the BiLSTM model are tweaking the values of the learning rate, the number of epochs, the optimizer technique, the loss function and the activation function for the dense layer.

5. Conclusion and Future Work

One of the most widespread mental illnesses worldwide is depression and its early detection helps the subject to recover easily. This work focuses on a multimodal approach combining audio and text modalities to predict the onset of the psychological disorder. Bidirectional LSTM models with three layers have been used for both modalities and their concatenated approach yielded an accuracy of 74.3% which is better when compared to each of the modalities separately. The scope of this study can be extended further by combining with other larger datasets and also including other modalities. In the future, advanced deep learning models like Residual Perceptron Network and Fusion approaches can be experimented with.

References

[1] World Health Organization. (n.d.). Depression. World Health Organization. Retrieved October 15, 2022, from https://www.who.int/news-room/fact-sheets/detail/depression

[2] Mathers, C.D., & Lončar, D. (2006). Projections of Global Mortality and Burden of Disease from 2002 to 2030. PLoS Medicine, 3.

[3] World Health Organization. (n.d.). Suicide. World Health Organization. Retrieved October 15, 2022, from https://www.who.int/news-room/fact-sheets/detail/suicide

[4] World Health Organization. (n.d.). Preventing suicide: A global imperative. World Health Organization. Retrieved October 15, 2022, from https://www.who.int/publications/i/item/9789241564779

[5] Harvard Health. (2020, October 13). Six common depression types. https://www.health.harvard.edu/mind-and-mood/six-common-depression-types

[6] Pacis, D.M., Subido, E.D., & Bugtai, N.T. (2018). Trends in telemedicine utilizing artificial intelligence.

[7] Gandhi, P.A., & Kishore, J. (2020). Prevalence of depression and the associated factors among the software professionals in Delhi: A cross-sectional study. Indian Journal of Public Health, 64, 413 - 416.

[8] Li, X., La, R., Wang, Y., Hu, B., & Zhang, X. (2020). A Deep Learning Approach for Mild Depression Recognition Based on Functional Connectivity Using Electroencephalography. Frontiers in Neuroscience, 14.

[9] World Health Organization. (n.d.). Covid-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide. World Health Organization. Retrieved October 19, 2022, from https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide

[10] Kipli, K., Kouzani, A.Z., & Hamid, I.R. (2013). Investigating machine learning techniques for detection of depression using structural MRI volumetric features. International Journal of Bioscience, Biochemistry and Bioinformatics, 3, 444-448.

[11] Chatterjee, R., Gupta, R.K., & Gupta, B. (2021). Depression Detection from Social Media Posts Using Multinomial Naive Theorem. IOP Conference Series: Materials Science and Engineering, 1022.

[12] Chiu, C., Lane, H., Koh, J., & Chen, A.L. (2020). Multimodal depression detection on instagram considering time interval of posts. Journal of Intelligent Information Systems, 56, 25-47.

[13] Kerestes, R., Davey, C.G., Stephanou, K., Whittle, S., & Harrison, B.J. (2014). Functional brain imaging studies of youth depression: A systematic review☆. NeuroImage : Clinical, 4, 209 - 231.

[14] Patel, M.J., Khalaf, A.M., & Aizenstein, H.J. (2016). Studying depression using imaging and machine learning methods. NeuroImage : Clinical, 10, 115 - 123.

[15] Kipli, K., Kouzani, A.Z., & Williams, L.J. (2013). Towards automated detection of depression from brain structural magnetic resonance images. Neuroradiology, 55, 567-584.

[16] Cohen, S.E., Zantvoord, J.B., Wezenberg, B.N., Bockting, C.L., & van Wingen, G.A. (2021). Magnetic resonance imaging for individual prediction of treatment response in major depressive disorder: a systematic review and meta-analysis. Translational Psychiatry, 11.

[17] Binnewies, J., Nawijn, L., Tol, M.V., Wee, N.J., Veltman, D.J., & Penninx, B. (2021). Associations between depression, lifestyle and brain structure: A longitudinal MRI study. NeuroImage, 231.

[18] Wang, Z., Yuan, Y., Jiang, Y., You, J., & Zhang, Z. (2021). Identification of specific neural circuit underlying the key cognitive deficit of remitted late-onset depression: A multi-modal MRI and machine learning study. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 108.

[19] Yoshida, K., Shimizu, Y., Yoshimoto, J., Takamura, M., Okada, G., Okamoto, Y., Yamawaki, S., & Doya, K. (2017). Prediction of clinical depression scores and detection of changes in whole-brain using resting-state functional MRI data with partial least squares regression. PLoS ONE, 12.

[20] Uddin, M.Z., Dysthe, K.K., Følstad, A., & Brandtzaeg, P.B. (2022). Deep learning for prediction of depressive symptoms in a large textual dataset. Neural Computing and Applications, 34, 721-744.

[21] Mousavian, M., Chen, J., & Greening, S.G. (2020). Depression Detection Using Atlas from fMRI Images. 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), 1348-1353.

[22] Sanyal, H., Shukla, S., & Agrawal, R. (2021). Study of Depression Detection using Deep Learning. 2021 IEEE International Conference on Consumer Electronics (ICCE), 1-5.

[23] Solieman, H., & Pustozerov, E.A. (2021). The Detection of Depression Using Multimodal Models Based on Text and Voice Quality Features. 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), 1843-1848.

[24] Ceccarelli, F., & Mahmoud, M.M. (2022). Multimodal temporal machine learning for Bipolar Disorder and Depression Recognition. Pattern Anal. Appl., 25, 493-504.

[25] Gratch, J., Artstein, R., Lucas, G.M., Stratou, G., Scherer, S., Nazarian, A., Wood, R., Boberg, J., DeVault, D., Marsella, S., Traum, D.R., Rizzo, A.A., & Morency, L. (2014). The Distress Analysis Interview Corpus of human and computer interviews. LREC.

[26] Mathers, C.D., & Lončar, D. (2006). Projections of Global Mortality and Burden of Disease from 2002 to 2030. PLoS Medicine, 3.

[27] Morency, L. P., Stratou, G., DeVault, D., Hartholt, A., Lhommet, M., Lucas, G., ... & Rizzo, A. (2015, March). SimSensei demonstration: a perceptive virtual human interviewer for healthcare applications. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 29, No. 1).