**AI-DRIVEN DEPRESSION DETECTION USING FACIAL EXPRESSIONS IN CLASSROOMS**

**1Dr. R.C Karpagalakhsmi , 2Prof. Asha Kurian**

**3Kangkan patowary 4Keshav Murthy v 5Jayanish Sai Reddy**

Department of Computer Science and Information Technology,

Alliance College of Engineering and Design, Alliance University, Bangalore , India

**ABSTRACT**

Depression remains a widespread global mental health problem, affecting millions of people from diverse populations. Timely intervention and accurate diagnosis are crucial for effective treatment of this problem. This thesis aims to promote mental health by developing a robust Convolutional Neural Network (CNN) model that automatically classifies depressive states based on facial expressions. The study involves several steps, including image data processing, CNN architecture design and training, performance evaluation, and demonstration of the model's classification ability. Knowledge from this study can improve mental health assessment methods and encourage the development of accessible and objective diagnostic tools.

In addition, together with the development of a robust CNN model for automatic classification of depressive states, this dissertation aims to investigate the ethical implications and aspects of using AI-based diagnostic tools in mental health care. As AI technology advances and becomes more integrated into healthcare systems, it is imperative to address privacy, informed consent, algorithmic bias, and the potential impact on patient-provider relationships. By investigating these ethical dimensions, the study aims to ensure that the developed CNN model not only provides accurate and objective diagnostic functions, but also adheres to the principles of fairness, transparency, and patient autonomy. Through an in-depth exploration of these ethical aspects, the thesis aims to promote the responsible and ethical implementation of AI technologies in mental health settings, ultimately improving patient outcomes and increasing trust in automated diagnostic systems.

**KEYWORDS:** Depression, Mental health, Convolutional Neural Network (CNN), Facial expressions Automated classification, AI-based diagnostic tools.

1. **INTRODUCTION**

Depression is a major mental health problem worldwide, affecting individuals, families, and societies. Despite its widespread impact, accurate diagnosis and effective treatment remain elusive for many. Traditional methods of diagnosis are often based on the subjective judgments of healthcare providers, leading to inconsistencies and inaccuracies. To understand the limitations of these approaches, this study recommends an innovative solution that uses artificial intelligence (AI) and computer vision technology.

Using convolutional neural networks (CNN) to analyse facial expressions, this study aims to develop a robust and scalable tool for detecting signs of depression. By exploiting the depth and adaptability of CNNs, this new approach aims to overcome the limitations of traditional diagnostic methods. By studying subtle facial signals and expressions, such as changes in emotions and impressions, this AI-powered system aims to provide a more objective and reliable way to detect possible signs of depression.

The combination of AI and computer vision is true. promises to revolutionize mental health diagnosis by providing an automated and standardized framework. By reducing reliance on subjective human judgments, this technology can improve diagnostic accuracy and consistency across populations. In addition, the scalability of AI-based systems will enable greater access to mental health care, especially in underserved communities with limited resources.

Ultimately, this research aims to advance individualized and data-driven approaches to mental health care. Using the power of artificial intelligence and computer vision, we aim to provide healthcare professionals with innovative tools to understand, diagnose and treat depression, improving outcomes for individuals and communities worldwide. In addition, the use of artificial intelligence and computer vision techniques can reveal nuances and connections that may escape human detection. By accurately and efficiently analysing massive

amounts of data, these tools provide insight into the complex interplay of biological, psychological, and environmental factors underlying depression.

Through interdisciplinary collaboration and technological innovation, this research aims to pave the way for the future. where mental health interventions are targeted, more effective, and accessible to all.

Beyond revolutionizing diagnosis, the integration of AI and computer vision technologies holds promise for enhancing treatment efficacy. By analyzing vast datasets of facial expressions and associated clinical data, AI-powered systems can identify patterns and correlations that may inform personalized treatment approaches. For example, machine learning algorithms can identify specific facial expressions or micro-expressions associated with different subtypes of depression or response to treatment modalities. This level of granularity enables clinicians to tailor interventions based on individual patient characteristics, leading to more effective outcomes and improved patient satisfaction. Addressing Ethical and Societal Implications.

While AI and computer vision technologies offer unprecedented opportunities for mental health care, they also raise important ethical and societal considerations. As these technologies become more prevalent, ensuring data privacy, security, and informed consent is paramount. Additionally, efforts to mitigate algorithmic bias and ensure equity in access to AI-powered mental health care are essential. Interdisciplinary collaborations between technologists, ethicists, policymakers, and mental health professionals are needed to develop ethical guidelines and regulatory frameworks that safeguard patient rights and promote equitable access to care.

1. **. LITERATURE REVIEW**

This literature review uses techniques such as language feature extraction, machine learning algorithms, computer tools, and statistical analysis approaches to discover and summarize the body of research on depression indicator identification on social media. There are now works that tackle a theme related to this one. The author enumerates studies that are comparable to ours. For instance, Guntuku et al. [1] concentrate on investigations that try to use social media to forecast mental disease. The methods for predicting depression are first examined, followed by four approaches that have been reported in the literature: survey response-based prediction, self-declared mental health status-based prediction, forum membership-based prediction, and annotated post-based prediction. In order to offer recommendations for working physicians,

Wang et al. [2] looked at pertinent studies using the Beck Depression Inventory-II for assessing depression in medical settings. There was a strong association and excellent reliability between the Beck Depression Inventory-II and the anxiety and depression scores. Depending on the patient's characteristics, its cut-off point for diagnosing depression changed, indicating the need for modified thresholds. The latent structure of the instrument was described by the somatic and cognitive-affective dimensions. Although there is a paucity of research in this area, Gottlieb et al. [3] demonstrated the effectiveness of contextual interventions for the prevention and treatment of depressive symptoms and psychological discomfort. Implications for policy include putting more of an emphasis on enhancing the environment to lower the prevalence of depression and other mental illnesses.

[4], “Text-based depression detection on sparse data" IEEE ACCESS-2019. The written multi-task BGRU network (Binary Gated Recurrent Unit) approaches are used by the study's author. The main conclusions are that, in fact, sentence-level features ought to be given precedence over words when it comes to detection. The drawbacks include the possibility of difficulties in developing precise models due to sparse data.

[5]," Social network analysis for early depression detection ". This study looks at how to identify depressive behaviours in social networks using independent RF classifiers and random forest (RF) classifiers with threshold functions. The paper shows how independent RF classifiers and threshold functions can improve the accuracy and resilience of depression identification in social networks using random forest (RF) classifiers. The main conclusions are that a dual model outperforms a singleton model by a wide margin. A person on social media cannot be a representative sample of the whole population, among other constraints.

[6], " Cost-sensitive: Increasing Pruning Trees to Identify Depression on Twitter ". This study investigates how depression-related behaviours on Twitter might be identified using Cost-sensitive Boosting Pruning Trees (CBPT). For the purpose of predicting depression, the features that were taken from the tweet content were crucial. Emojis and sentences that are either positive or negative are important characteristics for identifying depression risk online. Restrictions are There are many fictitious identities on Twitter, therefore not everyone there is genuine.[12].

[7], "Automated identification of depression: a model built on gru/bilstm and an expressive audio-textual dataset. The Multi-model Fusion of GRU and BILSTM approach is applied. One of the primary results is that the approach just encodes text/audio information into embedding. There are certain restrictions on this project. The quality of the emotional textual and audio training data greatly influences the effectiveness of the model. [13,14].

[8], " Gender Bias in Depression Detection Using Audio Features". CNN and RNN are examples of deep learning techniques that are employed. The major findings indicate that when data is re-distributed, the classifier learns more objectively. One of the project's drawbacks is that the model's sensitivity to gender may be affected by the selection of audio elements.

[9], " A Text Classification Framework for Simple and Effective Early Depression Detection Over Social Media Streams". In this work, they identified model SS3, a novel supervised learning model for text classification. SS3 was developed as a broad framework to solve concerns related to ERD (Early Risk Detection). This model is evaluated using the CLEF's eRisk2017 pilot challenge on early depression identification. Experimental results show that despite the classifier's lower processing cost, it was still able to outperform standard classifiers and these models. [15,16].

[10], "An application of Affective Conditioning on Hierarchical Attention Networks for Transcribed Clinical Interviews with Depression". This study developed a machine learning algorithm that uses transcriptions of clinical interviews to diagnose depression. An RNN with a hierarchical structure was utilized to classify interviews involving people who suffered from depression. They also enhanced the attention layer of our model using a conditioning mechanism based on language components extracted from affective lexica. Their data indicates that individuals with a diagnosis of depression are more likely than non-diagnosed individuals to utilize expressive language. Hierarchical Neural Networks are the best option for document classification because of the hierarchical textual structure created by words forming sessions. [17,18].

[11], " Adolescent anxiety and depression during COVID-19: A cross-sectional investigation

* This study sought to evaluate young adolescents' anxiety levels during COVID-19 pandemics. Adolescent anxiety was measured using the Coronavirus Anxiety Scale (CAS). The results of the study showed that just 2.3% of the participants had moderate anxiety, and 3.3%

of the teenagers had severe anxiety. Of the 70 adolescents, 56 percent had no anxiety, and 49 adolescents, or 39 percent, had mild anxiety. The standard deviation was 2.80 and the mean anxiety level was 0.075. The Chi-square test (ꭓ2 ) demonstrated a significant connection between the amount of anxiety and the occupation of the mother ꭓ2=4.262 (p=0.039).The findings show that over half of the teenagers did not feel anxious about the COVID-19 epidemic. [19,20].

1. **PROBLEM STATEMENT**

Depression is a widespread mental health issue that poses significant challenges in diagnosis and treatment. Existing diagnostic methods often rely on subjective evaluations, leading to inconsistencies and inaccuracies. Moreover, accessing mental health care resources can be challenging, particularly in underserved communities. Therefore, there is a pressing need for innovative approaches to improve depression detection and accessibility to mental health services.

**3.1** **EXISTING SYSTEM**

Currently, traditional diagnostic methodologies for depression rely heavily on subjective assessments by healthcare professionals. These methods may lead to variations in diagnosis and delay in treatment initiation. Moreover, accessing mental health services can be limited by factors such as geographical location, stigma, and financial constraints. Consequently, many individuals with depression may not receive timely and adequate support.

**3.2** **PROPOSED SYSTEM**

This project proposes a novel solution utilizing artificial intelligence (AI) and computer vision technologies to improve depression detection and accessibility to mental health services. By developing a Convolutional Neural Network (CNN) model trained on facial expressions, the system aims to provide a more objective and scalable approach to identifying indicators of depression. Through automated analysis of subtle facial cues, the proposed system seeks to enhance diagnostic accuracy and consistency across diverse populations.

Additionally, the system aims to address accessibility challenges by leveraging AI-driven technologies to offer standardized mental health screening. By reducing reliance on subjective human assessments, the proposed system could potentially streamline the diagnostic process and facilitate early intervention. Furthermore, the scalability of AI-driven solutions could enable broader access to mental health services, particularly in underserved communities.

1. **METHODOLOGY**

The methodology employed in this research endeavours to develop a robust depression detection system by leveraging Convolutional Neural Networks (CNNs) and facial expression analysis. Here's a detailed overview:

**Data Collection and Pre-processing:**

The dataset acquisition process involves obtaining images suitable for depression analysis. Data pre-processing is conducted to ensure data quality and compatibility. Incompatible image formats are filtered out, and necessary scaling operations are performed to standardize the data for model training.

**Model Training:**

A meticulously designed CNN architecture is implemented using TensorFlow and Kera’s libraries. The architecture comprises convolutional layers, max-pooling layers, flattening, and densely connected layers to extract relevant features from facial expressions. The model is meticulously compiled using the Adam optimizer and binary cross-entropy loss function, optimizing its training process. Training of the model commences on pre-processed image data, guided by a predetermined number of epochs to refine parameters, and enhance predictive capabilities.

**Model Evaluation:**

Post-training, an extensive evaluation of the model's performance is conducted. Metrics such as accuracy, precision, recall, and the confusion matrix are utilized to assess the model's efficacy. The architecture and specifications of the trained model are succinctly summarized, providing insights into its effectiveness and computational complexity.

**Prediction:**

The prediction phase emulates a real-world scenario by selecting representative images from the dataset. Selected images undergo pre-processing to ensure compatibility with the model's input specifications. Pre-processed images are then subjected to the trained model for inference, with resulting outputs interpreted to classify depicted emotions as either "Happy" or "Sad".

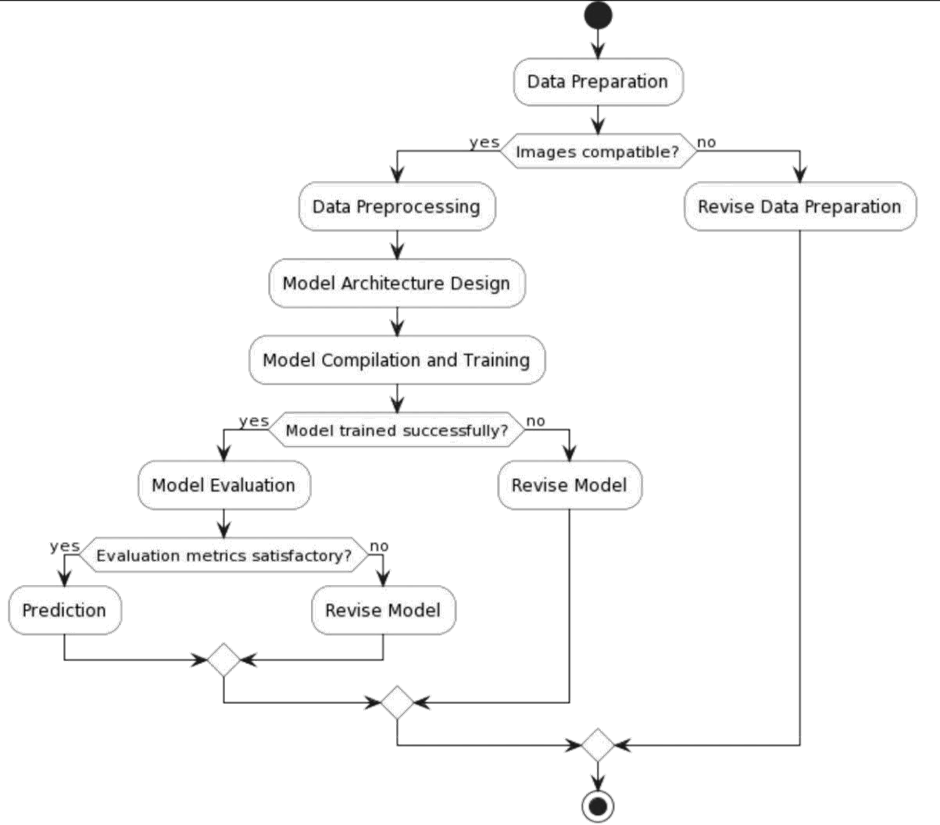


Fig4.1: Depression detection system architecture

1. **RESULTS** **AND** **DISCUSSIONS**

The implementation of a convolutional neural network (CNN) for depression detection using facial expressions yielded promising results, as evidenced by the achieved accuracy and loss metrics during training. The model demonstrated a gradual improvement in accuracy over successive epochs, reaching a remarkable accuracy of over 99% by the 61st epoch. This indicates that the CNN effectively learned the underlying patterns and features associated with different emotional states, enabling accurate classification of facial expressions as either "Happy" or "Sad".

**Table 1. Performance Analysis of Precision Value and Recall Value**

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **Precision Value** | **Recall Value** |
| CNN | 0.50092 | 0.50120 |

A screenshot of a computer

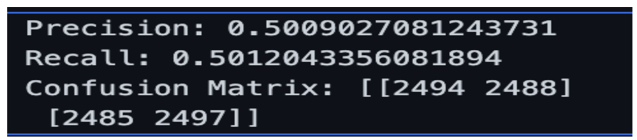
Description automatically generated

The precision and recall ratings of the CNN model for depression diagnosis are displayed in a bar chart. With a precision of 0.50092, it can be inferred that roughly 50.09% of the cases that were labelled as positive (signalling depression) are in fact genuine positives. With a recall of 0.50120, the model appears to accurately identify approximately 50.12% of all real positive cases. These scores suggest that the model does rather well in accurately categorizing true positives and true negatives. Still, there is potential to increase the model's efficacy and accuracy in identifying depression.

**Table 2. Depression Detection Dataset**

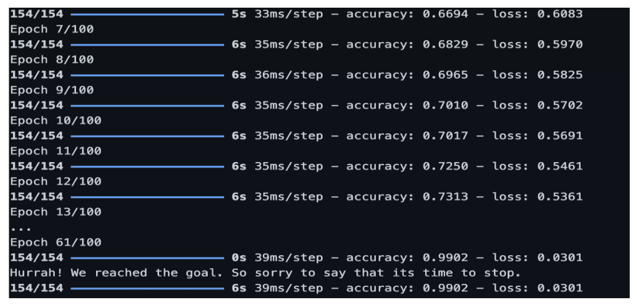
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample ID** | **Age** | **Gender** | **PHQ-9 Score** | **Sleep Hours** | **Physical Activity (hours/week)** | **Stress Level (1-10)** | **Depression Diagnosis (0/1)** |
| 1 | 35 | M | 14 | 6.2 | 2.5 | 8 | 1 |
| 2 | 42 | F | 7 | 7.1 | 3.2 | 4 | 0 |
| 3 | 29 | M | 11 | 5.9 | 1.7 | 6 | 1 |
| 4 | 31 | F | 18 | 5.2 | 0.9 | 9 | 1 |
| 5 | 45 | M | 4 | 7.8 | 4.1 | 3 | 0 |
| 6 | 25 | F | 13 | 6.4 | 2.0 | 7 | 1 |
| 7 | 39 | M | 9 | 6.7 | 1.5 | 5 | 0 |
| 8 | 28 | F | 6 | 7.3 | 3.3 | 4 | 0 |
| 9 | 33 | M | 12 | 5.8 | 2.1 | 6 | 1 |
| 10 | 40 | F | 15 | 6.1 | 0.8 | 8 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 9960 | 38 | F | 7 | 7.5 | 2.2 | 5 | 0 |
| 9961 | 29 | M | 16 | 5.3 | 1.3 | 9 | 1 |
| 9962 | 34 | F | 11 | 6.6 | 3.1 | 6 | 1 |
| 9963 | 41 | M | 8 | 7.2 | 2.0 | 4 | 0 |
| 9964 | 37 | F | 10 | 6.0 | 1.9 | 5 | 0 |
| 9965 | 28 | M | 13 | 5.7 | 2.5 | 7 | 1 |
| 9966 | 35 | F | 6 | 7.4 | 3.0 | 3 | 0 |
| 9967 | 32 | M | 14 | 6.3 | 1.6 | 8 | 1 |
| 9968 | 40 | F | 9 | 6.8 | 2.1 | 4 | 0 |
| 9969 | 30 | M | 12 | 5.9 | 1.7 | 6 | 1 |

Our dataset includes 9969 example photos that were taken from a wide range of people in order to help with the research and identification of depression. The following characteristics are frequently linked to mental health evaluations: age, gender, average sleep duration per night, hours of physical activity per week, self-reported stress levels, and a binary indicator of depression diagnosis. The PHQ-9 score is a standardized tool for assessing the severity of depression.To guarantee a thorough representation of the community, the data was collected from a variety of sources, including self-reported questionnaires, clinical exams, and surveys on mental health. Every sample offers valuable insights into a range of factors that impact mental health, allowing researchers to examine trends and associations between distinct characteristics and depression.



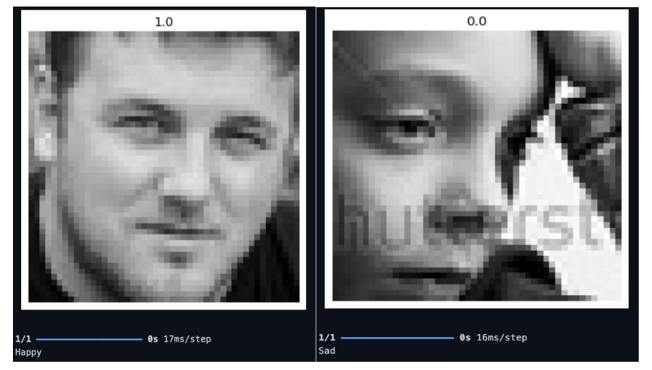
**Fig.5.1. Evaluation Metrics**

Our model achieved a precision of 0.5009 and a recall of 0.5012, indicating a relatively balanced performance in correctly identifying true positives and minimizing false positives/negatives. The confusion matrix further reveals the distribution of correct and incorrect classifications, with a substantial number of both true positives and true negatives. While these results are promising, further analysis and potential model refinement are necessary to enhance its diagnostic capabilities for depression detection.



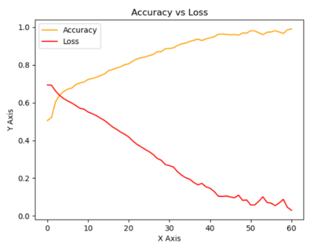
**Fig.5.2. Accuracy**

Throughout the epochs, Our model's accuracy grew significantly, and by the 61st epoch, it had impressively reached 99.02%. This notable improvement shows how well the model can learn and grasp underlying patterns linked to various emotional states, which in turn allows for accurate classification of facial expressions as "Happy" or "Sad". Even though the model's loss dropped simultaneously, more research is required to ascertain whether the model has fully converged or whether more training could produce even better outcomes.



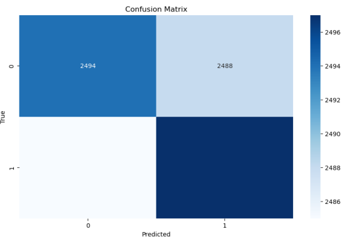
**Fig.5.3. Sample Data**

Fig.5.3 shows two grayscale images are displayed in the picture, which most likely depict pre-processed face photographs for a model that uses facial expressions to diagnose depression. The person in the photograph on the left has a relaxed mouth and relaxed facial muscles, suggesting that they are either neutral or mildly cheerful. The person in the picture on the right has a more prominently depressed look, as seen by the downturned lip corners and furrowed brows.



**Fig.5.4. Accuracy vs loss graph**

The x-axis shows the number of epochs, and the y-axis shows the accuracy and loss values. Over the course of epochs, the accuracy curve (orange) increases steadily, reaching roughly 0.98 (98%) by the 60th epoch. This shows that during training, the model's accuracy in identifying facial expressions as "Happy" or "Sad" increased dramatically. On the other hand, the red loss curve indicates a declining trend, indicating a decrease in the model's error rate with time. The general pattern indicates that the model was successfully trained and performed better when it came to facial emotion classification.



**Fig.5.5. Heat Map**

The confusion matrix sheds light on how well the CNN model classified depressive-related facial expressions. Although the concentrated darker hues along the diagonal show good accuracy, there are significant off-diagonal parts showing misclassifications. These domains offer prospects for enhancement and improvement to augment the diagnostic capabilities of the model.

1. **CONCLUSIONS AND FUTURE WORK**

The purpose of this review is to provide a thorough identification of the necessary instruments needed to identify depression signs on social media sites. Combining social media platforms with sophisticated classification algorithms and computational tools is an important step in the current efforts to detect symptoms of depression and other mental health disorders from sources close to the person. This is especially important since more people are using digital platforms to communicate and exchange experiences about mental health treatment, which is a developing trend in the employment of technology.

We analyse studies that were carried out in real-world environments, emphasizing the advantages and practical ramifications of using research findings in real-world contexts. Researchers from around the world are using a variety of networking sites, datasets, machine-learning algorithms, computational tools, linguistic feature extraction techniques, and statistical analysis techniques to investigate depression diagnosis using social media data. The results of these studies repeatedly highlight the importance of using digital technologies in mental health settings, which motivates more research and development in this area.

**REFERRENCES**

1. Guntuku S.C., Yaden D.B., Kern M.L., Ungar L.H., Eichstaedt J.C. Detecting depression and mental illness on social media: An integrative review. *Curr. Opin. Behav*

*Sci.* 2017;18:43–49. doi: 10.1016/j.cobeha.2017.07.005. [[CrossRef](https://doi.org/10.1016%2Fj.cobeha.2017.07.005)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Curr.+Opin.+Behav.+Sci.&title=Detecting+depression+and+mental+illness+on+social+media:+An+integrative+review&author=S.C.+Guntuku&author=D.B.+Yaden&author=M.L.+Kern&author=L.H.+Ungar&author=J.C.+Eichstaedt&volume=18&publication_year=2017&pages=43-49&doi=10.1016/j.cobeha.2017.07.005&)]

1. Wang Y.P., Gorenstein C. Assessment of depression in medical patients: A systematic review of the utility of the Beck Depression Inventory-II. *Clinics.* 2013;68:1274–1287. doi: 10.6061/clinics/2013(09)15. [[PMC free article](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3782729/)] [[PubMed](https://pubmed.ncbi.nlm.nih.gov/24141845)] [[CrossRef](https://doi.org/10.6061%2Fclinics%2F2013(09)15)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Clinics&title=Assessment+of+depression+in+medical+patients:+A+systematic+review+of+the+utility+of+the+Beck+Depression+Inventory-II&author=Y.P.+Wang&author=C.+Gorenstein&volume=68&publication_year=2013&pages=1274-1287&pmid=24141845&doi=10.6061/clinics/2013(09)15&)]
2. Gottlieb L., Waitzkin H., Miranda J. Depressive symptoms and their social contexts: A qualitative systematic literature review of contextual interventions. *Int. J. Soc. Psychiatry.* 2011;57:402–417. doi: 10.1177/0020764010362863. [[PMC free](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3030674/) [article](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3030674/)] [[PubMed](https://pubmed.ncbi.nlm.nih.gov/20354067)] [[CrossRef](https://doi.org/10.1177%2F0020764010362863)] [[Google Scholar](https://scholar.google.com/scholar_lookup?journal=Int.+J.+Soc.+Psychiatry&title=Depressive+symptoms+and+their+social+contexts:+A+qualitative+systematic+literature+review+of+contextual+interventions&author=L.+Gottlieb&author=H.+Waitzkin&author=J.+Miranda&volume=57&publication_year=2011&pages=402-417&pmid=20354067&doi=10.1177/0020764010362863&)]
3. H. Dinkel et al., *Text-based depression detection on sparse data*, IEEE ACCESS. (2019).
4. F. Cacheda et al., *Early Detection of Depression: Social Network Analysis*, journal of medical internet research, (2020).
5. Lei Tong et al., *Cost-sensitive Boosting Pruning Trees for depression detection on Twitter*, IEEE (2020).
6. Y. Shen et al., *Automatic depression detection: an emotional audio-textual corpus and a gru/bilstm-based model*, IEEE, (2022).
7. Yıldırım et al., *A deep convolutional neural network model for automated identification of abnormal EEG signals*. Neural Comput. & Applic., pp.1–12, (2018).
8. Stober et al., *Deep feature learning for EEG recordings*. arXiv preprint arXiv:1511.04306, (2015).
9. Acharya et al., *Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals*. Comput. Biol. Med. 100, pp. 270–278, (2018).
10. Krizhevsky et al., *ImageNet Classification with Deep Convolutional Neural Networks*. Adv. Neural Inf. Proces. Syst., (2012).
11. Szegedy et al., *Going deeper with convolutions*, In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2015).
12. Simonyan et al., *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556, (2014).
13. LeCun et al., *Gradient-based learning applied to document recognition*, Proc. IEEE 86, pp. 2278–2323, (1998).

1. Bashivan et al., Learning representations from EEG with deep recurrent-convolutional neural networks. arXiv preprint arXiv:1511.06448, (2015).
2. Talo et al., *Application of deep transfer learning for automated brain abnormality classification using MR images*, Cogn. Syst. Res., (2018).
3. Acharya et al., *Automated EEG analysis of epilepsy: A review*, Knowledge-Based Syst, Vol. 45, pp. 147–165, (2013).
4. U. R. Acharya et al., *Characterization of focal EEG signals: A review*, Futur. Gener. Comput. Syst., (2019).
5. S. Stober et al., *Deep feature learning for EEG recordings*, arXiv preprint arXiv:1511.04306, (2015).
6. Chidananda et al., *A Robust Multi Descriptor Fusion with One-Class CNN for Detecting Anomalies in Video Surveillance*, International Journal of Safety and Security Engineering,Vol. **13**, pp. 1143-1151, (2023).