

Depression Detection by Facial Emotion Recognition Using Deep Neural Network

Amit Phadikar^{1*}, Himadri Mandal²

¹Dept. of Computer Science, Santal Bidroha Sardha Satabarshiki Mahavidyalaya, Goaltore, Paschim Medinipur, W.B., India

²Dept. of Electronics and Communication Engineering, Calcutta Institute of Technology, Uluberia, Howrah 711316, India

DOI: <https://doi.org/10.51583/IJLTEMAS.2025.1410000046>

Abstract: The task of facial emotion recognition is one of the widely used applications of image analysis as well as pattern recognition. In the biometric area of research, automatic face & facial expression recognition attracts researcher's interest. For classifying facial expressions into different categories, it is necessary to extract important facial features that contribute to identifying proper and particular expressions. In this paper, a depression detection system is proposed. This paper will give an overview of the Depression Detection System using Facial Emotion Recognition techniques and datasets. Depression Detection systems based on facial gesture enable real-time analysis, tagging, and inference of cognitive affective states from a video recording of the face. It is assumed that facial expressions are triggered for a time period when an emotion is experienced, and so depression detection can be achieved by detecting the facial expression related to it. Out of all the major 6 emotions present, depression plays a vital role. Depression is classified as a mood disorder. It may be described as feelings of sadness, anger, or loss that interfere with a person's everyday activities. Experimental results show that the scheme detects depression from real-time video capture from a camera with high accuracy.

Keywords: Emotion Recognition, Depression, Facial Emotions, Convolutional Neural Network, Deep Neural Network, Depression Level

I. Introduction

People experience depression in different ways. In certain cases, depression may lead to fatal cases. In order to avoid all of these, depression must be detected at the earliest, and the victim must be treated with appropriate remedies. The main objective of the project is to analyze the depression of a user using real-time video. Psychological problems in people, like depression, pessimism, eccentricity, anxiety etc., are caused principally due to the neglect of continuous monitoring of their psychological well-being. Identification of depression in the beginning is desirable so that it can be controlled by giving better counseling at the starting stage itself [1-3]. If a counselor identifies depression in the initial stages, he or she can effectively help that individual to overcome depression. But it becomes a difficult task for the counselor to keep track of the significant changes that occur in each & every individual as a result of depression. Thus, we need an automated system that captures facial images of persons and analyzes them for effective detection of depression. In the proposed system, an attempt is being made to make use of the image processing techniques to study the frontal face features of persons and predict depression [4]. This system will be trained with the facial features of positive and negative facial emotions. To predict depression, a video of the individual is captured, from which the face is extracted. The level of depression is identified by calculating the number of positive & negative emotions present in the entire video [5].

In deep learning, a Convolutional Neural Network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analyze visual imagery. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on the shared-weight architecture of the convolution kernels or filters that slide along input features and provide translation equivariant responses known as feature maps. Counter-intuitively, most convolutional neural networks are only equivariant, as opposed to invariant and translation. They have applications in image and video recognition, recommender systems, image classification, image segmentation, medical image analysis, natural language processing, brain-computer interfaces, and financial time series [6]. The CNNs are regularized versions of a multilayer perceptron. A multilayer perceptron usually means fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "full connectivity" of these networks makes them prone to overfitting data. Typical ways of regularization, or preventing overfitting, include: penalizing parameters during training (such as weight decay) or trimming connectivity (skipped connections, dropout, etc.) [7-8]. The CNNs take a different approach towards regularization: they take advantage of the hierarchical pattern in data and assemble patterns of increasing complexity using smaller and simpler patterns embossed in their filters. Therefore, on a scale of connectivity and complexity, CNNs are on the lower extreme. There are various techniques that can be kept in mind while building a deep neural network and are applicable in most of the computer vision problems.

A DNN (Deep Neural Network) is an artificial neural network that consists of more than three layers; it inherently fuses the process of feature extraction with classification into learning using FSVM and enables decision making [9-10]. The learning portion of creating models spawned the development of artificial neural networks. The ANNs utilize the hidden layer as a place to store and evaluate how significant one of the inputs is to the output. The hidden layer stores information regarding the input's importance, and it also makes associations between the importance of combinations of inputs. Deep neural networks, then, capitalize on the ANN component. They say, if that works so well at improving a model—because each node in the hidden layer makes both associations and grades the importance of the input to determine the output—then why not stack more and more of

these upon each other and benefit even more from the hidden layer? So, the deep net has multiple hidden layers. 'Deep' refers to a model's layers being multiple layers deep [11]. Cao et al. present a review on Deep learning-based depression recognition through facial expression in [12]. Kumar et al [13] proposed a scheme for early detection of depression through facial expression recognition and electroencephalogram-based artificial intelligence-assisted graphical user interface. The resulting AI assistant demonstrates high sensitivity, precision, and accuracy in the early detection of depression, establishing its potential as a reliable diagnostic tool. Authors argued that the application may be extended to clinicians, therapists, and hospitals for the identification of depression at its early stage. Li et al [14] includes a dual-scale convolution module, adaptive channel attention mechanism, and gradient class activation mapping technique for automatic diagnosis of depression. In which dual-scale convolution captures features of the facial region at different scales, and the adaptive channel attention captures the facial region with the most significant features. One well-known study trend is the combination of text, audio, and face features. Because they capture complementary signals, such as observable affect (sadness, dread) and related language cues of hopelessness or negativity, systems like these have been demonstrated to function better than single-modality models [15,16]. However, real-world or clinical implementation continues to prioritize ethical concerns like privacy, fairness, and explain ability [16,17].

In this paper, a depression detection system is proposed using a DNN. Depression detection systems based on facial gesture enable real-time analysis, tagging, and inference of cognitive affective states from a video recording of the face. Experimental result shows that the scheme detects depression from real-time video capture from a camera with high accuracy. The present depression classification model relies solely on facial expressions without considering other key indicators like speech, physiological cues, or behavioral data to make the scheme simple and easy to implement. The use of Convolutional Neural Networks (CNN) and Deep Neural Networks (DNN) enhances the model's accuracy and adaptability to real-time scenarios. The study also contributes meaningfully to the growing field of AI-based mental health diagnostics, showing potential for early, non-invasive depression screening.

The article is organized as follows: Section II presents the methodology of depression detection by facial emotion recognition using a neural network. The proposed implementation of the system is described in Section III. Section IV contains the results of experiments. In Section V, we present the concluding remarks.

II. Methodology Of Depression Detection By Facial Emotion Recognition Using Neural Network

The methodology of depression detection by facial emotion recognition using a Neural Network is shown in Figure 1. Figure 1 describes the broad description of how CNN can be used for diagnostic purposes for the improvement of social life. The system was designed with real-time functionality and ethical considerations. In order to facilitate smooth live camera interaction, the DNN is built to generate predictions at frame rates. In terms of ethics, the system prioritizes informed consent, data security, and privacy, especially as it deals with sensitive mental health information. Even while the system isn't meant to be a diagnostic tool, it can be a powerful tool to identify and monitor a young person's problems early on so that medical professionals can make better judgments. All things considered, the suggested system is a useful, effective, and morally conscious method of automatically identifying depression and facial expressions, and it has a lot of potential to grow into a multimodal affective computing application.

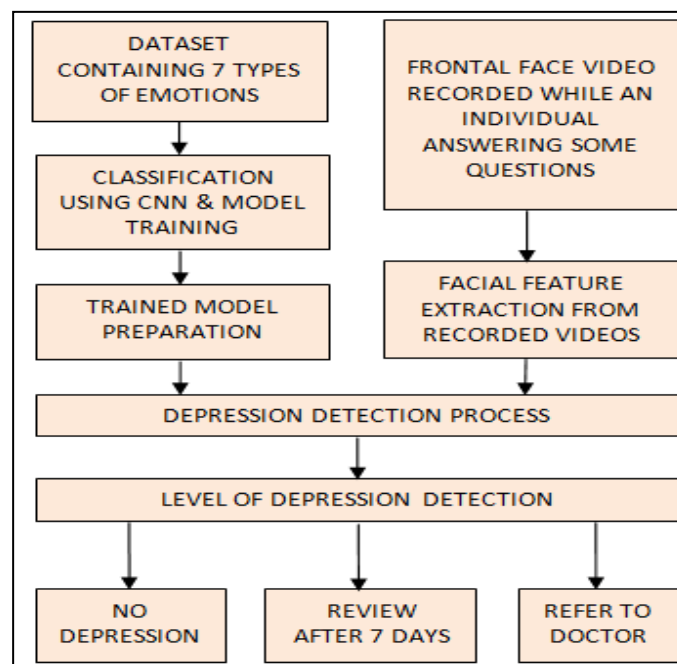


Fig. 1. Block Diagram of the System.

III. Proposed System & Description

The proposed implementation of the system is described below:

A. Facial Expression Dataset

For detecting depression, alone from images, it mainly depends on a clear and proper definition of a depressed face. There are many open-access facial expression datasets present on the internet. We have used a dataset for facial expression from Kaggle [18]. The dataset has 48×48-pixel grayscale images of faces. The training set consists of 28,709 examples with seven emotions (happy, sad, surprised, fear, angry, disgusted, and neutral).

B. Face Detection & Processing

The face circumference from the images was detected using the Deep Neural Network (DNN) Face detector. The DNN is one of the most robust face detection algorithms. It works well with occlusion, quick head movements, and can identify side faces as well. Moreover, it also gave the quickest fps (frames per second) among the others. Then, these detected rectangular facial expressions were clipped and recorded. Before that, the images were converted to grayscale images in order to avoid unnecessary density in the neural networks.

C. Convolution Neural Network Architecture

The CNN architecture is proposed mainly to cultivate the pixel values in the rectangular region containing facial expressions. The entire structure of CNN consists of facial expression data and includes the 3 stages, each of which has 2 convolution layers with 'relu' activation function followed by max-pooling layers, and 3 fully connected layers with 'relu' and softmax activation function. After all operations of convolution layers and max-pooling layers, each frame feeds to the fully connected layers, and the prediction of frames was processed with a classifier as seven different facial emotional states. Figure 2 shows the various steps of CNN.

D. Model Training

For training purposes, neural networks were implemented using Keras in Python. The model was trained for 50 epochs with a batch size of 128.

E. Real-time Video Capturing for Model Testing

The trained model was tested via real-time video capturing. A video of an individual was captured using a webcam & then it was converted into frames i.e., images. The face of that individual was detected from those images using the DNN face detector. After that, those detected frames were sent to the trained model in order to predict the emotion.

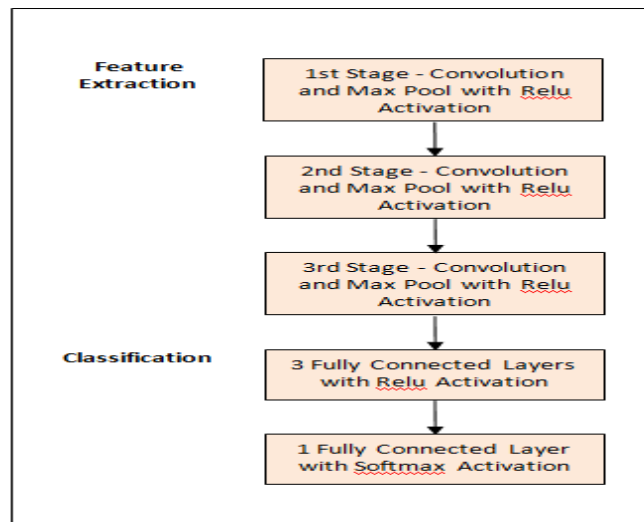


Fig. 2. Steps of CNN.

F. Depression Level Identification & Analysis

In this paper, the first 1200 frames for each video have been taken and divided into three (3) parts named First Part, Second Part, and Third Part. If all three parts of the video exhibit positive emotions, then the individual can be classified as having 'No Depression'. If the first two parts of the video show positive emotion, while the third part shows negative emotion, then the individual is classified as having 'Low Depression' since only the end part of the video is showing negative emotion. If two parts of the video are showing positive emotion, then the individual may be suffering from 'Mild Depression'. If out of the three parts, two parts or all three parts of the video show negative emotions, then the individual seems to be highly depressed, so it's classified as having 'High Depression'. Table 1 represents the depression level identification in detail.

Table I Depression Level Identification Table

Time Duration	First Part	Second Part	Third Part	Depression Level
Features Present	Positive	Positive	Positive	No Depression
Happy, Neutral – Positive class – ‘Positive’	Positive	Positive	Negative	Low
	Positive	Negative	Positive	Mild
Disgust, Sad – Negative class – ‘Negative’	Positive	Negative	Negative	High
	Negative	Positive	Positive	Mild
	Negative	Positive	Negative	High
	Negative	Negative	Positive	High
	Negative	Negative	Negative	High

Table II Output Table Of Predicted Depression Level

Video	First Part	Second Part	Third Part	Predicted Depression Level
Subject_1 (Video 1)	Positive	Positive	Positive	No Depression
Subject_2 (Video 2)	Positive	Positive	Positive	No Depression
Subject_3 (Video 3)	Negative	Negative	Negative	High Depression
Subject_4 (Video 4)	Positive	Positive	Positive	No Depression
Subject_5 (Video 5)	Negative	Positive	Positive	Mild Depression

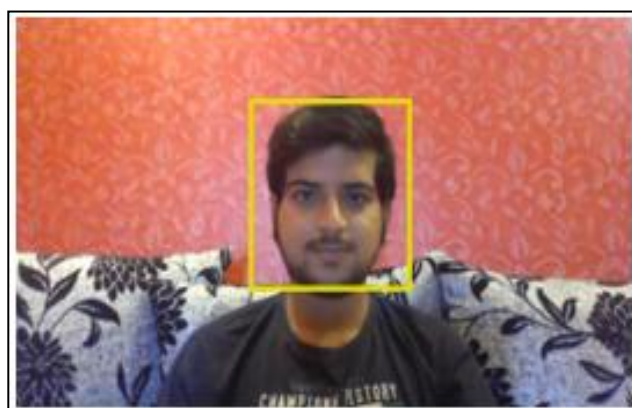


Fig. 3. Real-time Video Recording.

IV. Experimental Results

In this study, five (5) videos from five (5) different individuals have been taken for experimental purposes in order to identify the level of depression among them. Widely recognized assessment criteria, including accuracy, precision, recall, and F1-score, as well as confusion matrices for in-depth error analysis, were used to evaluate the system. Figure 3 shows one of the real-time video recordings of the process. Out of the five (5) videos captured from five (5) different persons, the 3rd and 5th persons have high and mild depression, respectively. If a person is diagnosed with ‘No Depression’, then a message will appear on the screen: “You don’t need to visit a doctor. Stay well.” If the person is diagnosed with ‘Low Depression’ or ‘Mild Depression’, then the system will show: “Come back review after seven (7) days.” and for ‘High Depression’, it will show that “Contact a doctor or any Psychiatrist immediately.” The system can also detect depression levels for short-duration or long-duration videos as well. In this

work, the main motto was to find out the depression of those individuals who were not previously diagnosed with depression. Table 1 shows the depression level identification table, whereas Table 2 shows the output table of predicted depression level.

Table Iii Confusion Matrix. Ld: Low Depression, Md: Mild Depression, Nd:No Depression, Hd: High Depression, Ta: Total Actual, Pca: Per-Class Accuracy

Actual \ Predicted	LD	MD	ND	HD	TA	PCA
Low Depression	16	2	1	1	20	80%
Mild Depression	2	15	1	2	20	75%
No Depression	1	1	18	0	20	90%
High Depression	1	2	0	17	20	85%

In Table 3, the confusion matrix shows the classification results. The model successfully identified “No Depression” (18), “High Depression” (17), “Mild Depression” (15) and “Low Depression” (16) with high Accuracy. Table 3 also shows the per class accuracy (e.g., No Depression 90%, and High Depression 85%). The Accuracy is the proportion of correctly classified samples and is defines as (Number of Correct Total Predictions/Total Predictions) =66/80=0.825. Table 4 shows the calculated Precision, Recall and F1-score values for various emotion classes. For any class i :

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \quad (1)$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (2)$$

$$F1_i = 2 \times \frac{Precision_i \times Recall_i}{Precision_i + Recall_i} \quad (3)$$

Table Iv Calculated Precision, Recall And F1-Score Values From Confusion Matrix (Table 3)

Emotion	Precision	Recall	F1-score
Low Depression	0.80	0.80	0.80
Mild Depression	0.75	0.75	0.75
No Depression	0.90	0.90	0.90
High Depression	0.85	0.85	0.85

V. Conclusion

A detailed study has been done on facial expressions related to depression, and a system has been proposed to detect the same. The system consists of two main modules: (1) a face detection module, implemented using DNN. For the testing phase, the videos of a total of five(5) different individuals answering depression detection questionnaires were taken. The video is then converted into frames, and from each frame, the face region is detected and saved to create a dataset of test faces. From the test sets, using the above proposed depression identification table, we were able to detect the depression level of an individual. (2) facial expression recognition, implemented using a CNN that mainly focuses on detecting facial expressions that can reflect depression of a person.

It can be further improved to have more decision taking capabilities and thus could be used for different applications. It aims to solve the problems faced by many individuals in their daily lives, which increases depression among them. The system also takes measures to ensure mental peace. The project mainly deals with emotion recognition and a depression analyzer. In the future trends, this can actually be implemented as a chatbot using Natural Language Processing. As for the current working, the desktop application is local. In the future, this application can be hosted on a website using an internet connection. The current application is basically a screening test before consulting a doctor. In the future days, a video consultation with a doctor or mental counselor can be arranged if the user is detected to be depressed. Moreover, the work can be extended by incorporating multimodal data such as voice analysis, body language, and text-based sentiment to improve detection accuracy.

Ethics Declarations

A. Ethical Approval

The submitted work is original and has not been published elsewhere in any form or language. This article contains no studies with human participants or animals performed by the author.

B. Data Availability Statement

The authors confirm that the data supporting the findings of this study are available within the article.

C. Authors Contributions

The authors confirm the responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

D. Funding

No funding was received for this article.

E. Competing Interests

The author declares no conflict of interest.

References

1. Girard JM, Cohn JF, Mahoor MH, Mavadati S, Rosenwald DP, (2013). Social Risk and Depression: Evidence from Manual and Automatic Facial Expression Analysis: Proc Int Conf Autom Face Gesture Recognit. pp. 1-8. doi: 10.1109/FG.2013.6553748.
2. Alghowinem, S., Goecke, R., Cohn, JF., Wagner, M., Parker G., and Breakspear, M (2015). Cross-Cultural Detection Of Depression From Nonverbal Behaviour: 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, pp. 1-8. doi: 10.1109/FG.2015.7163113.
3. Katikalapudi, R., Chellappan, S., Montgomery, F., Wunsch D., and Lutzen, K (2012). Associating Internet Usage with Depressive Behavior Among College Students: IEEE Technology and Society Magazine, vol. 31, no. 4, pp. 73-80. doi: 10.1109/MTS.2012.2225462.
4. Sahla, K.S., Senthil Kumar, T. (2016). Classroom Teaching Assessment Based on Student Emotions: In: Corchado Rodriguez, J., Mitra, S., Thampi, S., El-Alfy, ES. (eds) Intelligent Systems Technologies and Applications 2016. ISTA 2016. Advances in Intelligent Systems and Computing, vol 530. Springer, Cham. doi: 10.1007/978-3-319-47952-1_37
5. Venkataraman, D., and Parameswaran, NS (2018). Extraction of Facial Features for Depression Detection among Students: International Journal of Pure and Applied Mathematics, Volume 118, No. 7, pp. 455-463.
6. Parameswaran, NS., and Venkataraman, D., (2019). A Computer Vision Based Image Processing System for Depression Detection Among Students for Counseling: Indonesian Journal of Electrical Engineering and Computer Science, Vol. 14, No. 1, pp. 503-512. doi: 10.11591/ijeecs.v14.i1.pp503-512
7. Neha, S., Nivya, Shekar, PHC., Kumar, KS., Asha, VG (2020). Emotion Recognition and Depression Detection using Deep Learning: International Research Journal of Engineering and Technology (IRJET), Vol. 07 No. 08, pp. 3031-3036.
8. Alghowinem, S., Goecke, R., Wagner, M., Parker G., and Breakspear, M. (2013). Head Pose and Movement Analysis as an Indicator of Depression: Humaine Association Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, pp. 283-288, doi: 10.1109/ACII.2013.53.
9. Bouhabba, EM., Shafie AA., and Akmeliawati, R., (2011). Support Vector Machine for Face Emotion Detection on Real Time Basis: 4th International Conference on Mechatronics (ICOM), Kuala Lumpur, Malaysia, pp. 1-6, doi: 10.1109/ICOM.2011.5937159.
10. [10] Cohn JF. et al., (2009). Detecting Depression from Facial Actions and Vocal Prosody: 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, Amsterdam, Netherlands, pp. 1-7, doi: 10.1109/ACII.2009.5349358.
11. Tasnim, M., Shahriyar, R., Nahar N., and Mahmud, H., (2016). Intelligent Depression Detection and Support System: Statistical Analysis, Psychological Review and Design Implication: IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), Munich, Germany, pp. 1-6. doi: 10.1109/HealthCom.2016.7749494.
12. Cao,X., Zhai, L., Zhai,P., Li,F., He, T., He,L.(2025): Deep learning-based depression recognition through facial expression: A systematic review: Neurocomputing,vol. 627. doi: 10.1016/j.neucom.2025.129605.
13. Kumar, G., Das, T. & Singh, K., (2024). Early Detection of Depression Through Facial Expression Recognition and Electroencephalogram-Based Artificial Intelligence-Assisted Graphical User Interface: Neural Comput & Applic, vol. 36, pp. 6937–6954.
14. Li, M., Wang, Y., Yang, C., Lu Z., and Chen, J. (2024). Automatic Diagnosis of Depression Based on Facial Expression Information and Deep Convolutional Neural Network: IEEE Transactions on Computational Social Systems, vol. 11, no. 5, pp. 5728-5739., doi: 10.1109/TCSS.2024.3393247.
15. Phiri, D., Makowa, F., Amelia, V., Phiri, Y., Dlamini, L., Chung, M. (2025). Text-Based Depression Prediction on Social Media Using Machine Learning: Systematic Review and Meta-Analysis. J Med Internet Res, vol. 27, pp. 1-18. doi: 10.2196/59002.
16. Teferra, B., Rueda, A., Pang, H., Valenzano, R., Samavi, R., Krishnan S., Bhat V. (2024). Screening for Depression Using Natural Language Processing: Literature Review. Interact J Med Res, vol. 13, pp. 1-17. doi: 10.2196/55067.
17. Saha, D.K., Hossain, T., Safran, M. et al. (2024). Ensemble of Hybrid Model Based Technique for Early Detecting of Depression Based on SVM And Neural Networks. Sci Rep, vol. 14, pp.1-18. doi: 10.1038/s41598-024-77193-0.
18. Kaggle Data set: <https://www.kaggle.com/datasets>