STM JOURNALS

Current Trends in
Signal Processing

https://journals.stmjournals.com/ctsp

ISSN: 2277-6176 (Online)
ISSN: 2321-4252 (Print)
Volume 14, Issue 1, 2024
January—April
DOI (Journal): 10.37591/CTSP

Review CTSP

Face Emotion Recognition to Detect Depression

Kalpesh Patil[1], Jayesh Umap[2], Yash Rathor[3], P.P. Gaikwad[4,*]

Abstract

In the current competitive world, one of the most familiar and grave mental illnesses we encounter in humans is depression, also called as major depression or major depressive disorder. It makes one feel depressed and disinterested all the time, which has a bad impact on one's thoughts and behavior. This affects not only the victims but also people associated with them, such as family, friends, and society. If not treated properly, it can end up with adverse actions such as suicide or hurting others. Therefore, it is very essential to detect such people and provide them with the necessary required treatment. With advancements in artificial intelligence and deep learning, facial emotion recognition has always been in focus. Through the facial expression, we can detect the emotions of a person. Convolutional neural network method will be used to recognize facial expressions in real time from picture frames in video data. The video input will be taken through an external camera module fitted on the ESP32 chip. Image pre-processing will be done for better feature (expression, emotions) extraction. These processed images will then be sent into a deep learning model, which predicts whether the facial expressions indicate depression. Depending on whether depression is detected or not, a message will be sent to that particular person and also to their well-wishers.

Keywords: Emotion detection, convolutional neural network (CNN), deep learning, tensor flow, depression, emotion recognition model (ERM)

INTRODUCTION

Depression is a mental health condition that affects a person psychologically. Being very common. according to the World Health Organization (WHO) around 330 million of people on earth are currently suffering from depression [1]. Generally, depression does not occur from a single event but from a bunch of events, unless the single event is very dreadful. Good news is that the depression can be treated. Sooner the person suffering from depression gets treatment, more are the chances of recovery. The initial line of therapy for depression is psychological. In cases of moderate to severe depression, they might be taken in addition to antidepressant drugs. For moderate depression, antidepressant medicines are not necessary. Cognitive behavioral therapies can impart new ways of thinking, adjusting, or interacting with people. They may include both professional talk therapy and lay therapists under supervision. Talk therapy can be conducted in person or virtually. Self-help guides, apps, and websites all provide psychological therapy. There are various kinds of depression [2], and the severity of their symptoms varies from mild to severe. Major depression, melancholia, dysthymia, and psychotic depression are a few kinds of depression. It has a negative impact on people's emotions, thoughts, and behaviors. Depression makes it harder to keep up with daily tasks and interferes with regular relationships and employment. There are various feelings that are caused by depression, some of which are as follows- sad, irritated, guilty, frustrated and disappointed. Negative thoughts such as "I have failed", "It's my fault", "There is nothing good in

***Author for Correspondence**
P.P. Gaikwad
E-mail: patilkv745@gmail.com

[1-3]Student, Department of Electronics & Communication Engineering, Gaytri Nivas, Pune, Maharashtra, India
[4]Assistant Professor, Department of Electronics & Communication Engineering, Gaytri Nivas, Pune, Maharashtra, India

Received Date: April 01, 2024
Accepted Date: May 23, 2024
Published Date: May 30, 2024

Citation: Kalpesh Patil, Jayesh Umap, Yash Rathor, P.P. Gaikwad. Face Emotion Recognition to Detect Depression. Current Trends in Signal Processing. 2024; 14(1): 1–9p.

my life", "Life is not worth living" are induced by depression. Depression has various symptoms, which can be mainly classified into two classes — behavioral and physical. Behavioral symptoms include withdrawing from close people (friends and family), stopping enjoyment, not going out, becoming unproductive at work, and attraction to alcohol and drugs. Physical symptoms include being tired at all times, sleeping problems, feeling sick almost all the time, constant headache, and loss or change in appetite. People with depression are at increased risk of suicide. Research also shows strong relationships between depression and physical health, including cardiovascular disease, cancer, diabetes, and respiratory diseases. For these reasons, it is very important to detect depression at an early stage for proper treatment and preventing harmful measures.

Because of this, we have developed a machine learning model that uses the pictures obtained from the films to identify facial expressions. An external camera module is put on the ESP32 chip in order to record videos. The ESP32 CAM WiFi Module is used in numerous internet of things (IoT) applications [3]. With a footprint of just 40 × 27 mm and a deep sleep current of up to 6 mA. Bluetooth with OV2640 camera module 2MP for face recognition has an extremely competitive small-size camera module that can work independently as a minimum system. It is suitable for IoT uses such as smart home gadgets, industrial wireless control, and wireless monitoring. Using a DIP package, this module may be installed straight into the backplane to enable quick product manufacturing. It offers consumers a very reliable connection mechanism, making it ideal for use in a range of IoT hardware terminals. ESP incorporates a seven-stage pipeline design, two high-performance WiFi, traditional Bluetooth, 32-bit LX6 CPUs, and BLE Beacon. This device's main frequency adjustment range is 80 MHz to 240 MHz. In addition, it has on-chip, Hall, and temperature sensors. Images are processed using convolutional neural networks (CNNs). CNN is a deep learning algorithm having multiple layers and it specializes in image processing. The initial images are pre-processed by applying Gaussian blur for better feature extraction. These processed images are then sent to machine learning models for classifying whether the person is depressed or not depending upon the expressions.

A full discussion of previous work on facial expression detection may be found in the literature review. To solve these machine learning difficulties, they mostly use deep learning. A kind of machine learning called "deep learning" is based on the principle of "learning by example." This approach to artificial intelligence (AI) trains machines to handle data in a manner that is modelled after the human brain. One such deep learning system is CNN, which processes images with exceptional accuracy and whose description is provided in full below.

The paper consists of following contents: Literature Review, where we have discussed research done on the topic till now. Open Issues showcases a detailed description of drawbacks in the currently used technologies. In Proposed Methodology we have provided a clear explanation of our model's approach. Algorithm gives a detailed explanation regarding the algorithm used in the project. Flowchart depicts the flow of processes in the model. Acknowledgments as a vote of thanks and lastly the Conclusion presenting the closing statements followed by references.

LITERATURE REVIEW

There has been moderate research on facial expression recognition for depression detection as it has useful implementation in medical field. These works had various aspects in focus but a common goal was to improve efficiency of the depression detection using facial expression recognition. It mainly depends on factors such as lips, eyes, and wrinkles for facial expression recognition to detect depression. Different works focused on producing models using various methods for detecting depression are discussed further.

Nazira et al. [3] proposed a comprehensive framework for depression detection, integrating advanced deep learning techniques with traditional image processing methods. By combining CNN with OpenCV, the model achieves robust performance in accurately identifying indicators of depression

from image data. Moreover, OpenCV, in conjunction with Haar cascade classifiers, is utilized for face detection, a crucial aspect of depression diagnosis. Haar cascade classifiers are trained with a large dataset of facial images to detect various facial organs, enabling the identification of human faces within images. The accuracy of their detection system is found to be 81%.

Zhu [4] proposed a comprehensive methodology for depression detection by leveraging state-of-the-art dynamic encoding techniques to capture both short-term and long-term human behavioral information. The approach involves frame-level representation, multi-scale short-term facial temporal representation, multi-scale long-term facial temporal representation, and depression severity recognition. Facial action units (AUs), head positions, and gaze directions are concatenated from OpenFace 2.0 detections to create a matrix representation for every video frame in the frame-level representation. This matrix captures dynamic face traits that are essential for the investigation of depression. In the research, dynamic facial imaging algorithms are extended to store short-term facial dynamics, therefore addressing the difficulty of summarizing depression severity from single frames. To acquire multi-scale short-term facial behavior information, many time-windows are used. Feature vectors are smoothed using the ranking function, and temporal order parameter vectors are obtained using RankSVM.

Kantharia et al. [5] discussed in their paper about the two different types of input photos that are usually utilized in facial behavior recognition: learning images, which are used to train classifiers, and testing images, which are used to assess the learnt classifier. Frequently used sources for these photos include face expression picture databases such as MMI and JAFFE. The first phase is face detection, which entails identifying the facial region in the input image. Even with the success of many face identification techniques, problems with posture variation, occlusion of features, and imaging circumstances still exist. Four types of face detection approaches may be distinguished appearance-based, knowledge-based, feature-invariant, and template-matching techniques. Face detection is followed by feature extraction, which creates separable and classifiable vectors by extracting pertinent information from the face picture. Processing the frontal face in its whole or breaking it up into subsections for information gathering can accomplish this.

de Melo et al. [6] implemented the musculoskeletal disorders architecture. The fundamental building element of the MSN architecture consists of many three-dimensional convolutional filters with different spatial widths and temporal depths. Identity shortcut connections are employed inside the block to help train very deep networks while minimizing overfitting. This is achieved by employing residual connections. The resulting network captures extensive spatiotemporal information characterizing depressive behaviors. The output of the basic building block is defined by a function that learns the residual mapping using rectified linear unit (ReLU) activation functions and batch normalization. The first convolutional layer, which down samples the input spatially by a factor of two, is followed by successive convolutional layers with different spatial widths and depths. The outputs are concatenated to increase the depth of the final features. Within the lingering link, a convolutional layer with a $1 \times 1 \times 3$ kernel size and batch normalization is added to guarantee compatibility for the sum operation with the input.

OPEN ISSUES

In existing systems [4, 7], the web-based dataset can include built-in biases, such as inconsistent depression categorization or biases related to demographics. Results may be distorted due to biases in the dataset that impact the model's capacity to generalize. Normalizing data is important, but not every situation will benefit from the particular normalization method that was used. Depending on the features of the dataset and the network design, certain normalization techniques could be more appropriate. The use of AI systems for mental health diagnosis presents ethical questions about permission, privacy, and the abuse of personal information. Strong ethical standards, resolving any prejudices, and fairness concerns are important factors that require significant thought.

For the system discussed in reference [8], the MSN architecture could not be interpretable, like many deep learning models, which makes it difficult to comprehend how the model makes its predictions. In scenarios where interpretability is essential, such diagnosing mental health issues, the model's opaque structure might impede patients' and physicians' acceptance and confidence. A major factor in the MSN architecture's efficacy is the availability and quality of training data. Although face dynamics are included in the research, it is unclear how the model may be applied to different demographics and cultural situations. Predictions may become skewed or erroneous as a result of biases in the training data, particularly if the dataset does not fairly reflect the target population.

Regarding reference [5], using dynamic picture algorithms and spectral techniques, the method extracts dynamic representations and concatenates different face action units (AUs), head positions, and gaze directions. It is a sophisticated feature engineering process. Implementing and fine-tuning such complicated feature engineering procedures may add complexity, raise computing overhead, and call for domain knowledge. There is a dependence on the precision and dependability of automated recognition algorithms, such OpenFace 2.0, for the identification of facial AUs, head positions, and gaze orientations. The efficacy of depression recognition may be impacted by biased or noisy feature representations resulting from automated detection method inaccuracies or biases.

PROPOSED SYSTEM / METHODOLOGY

The proposed system overcomes all the limitations of existing system along with that it works efficiently and it is easy to use. So, in the proposed system we are going to develop a user interface and will have more focus on recognizing the facial expression for detecting depression. Convolutional neural networks, a few machine learning and deep learning algorithms, and a variety of Python modules for various operations will all be used in the development of this.

Data Collection and Pre-processing

We will be using AffectNet [9] a facial expression dataset which contains around 1 million pictures gathered from the internet through searches conducted in six different languages on three major search engines using 1250 emotion-related keywords We will be using this dataset for training, testing, and validating our facial expression detection model. The details of the dataset are provided in Table 1.

A region of interest (ROI) will be defined within each frame and visually identified by a blue boundary square. The most important thing to do after gathering all the photos is to add a Gaussian blur filter on them. Following the collection of all the photos, we will process them using a Gaussian blur filter. This is a crucial step in extracting different characteristics and getting the dataset ready for our upcoming processes.

Table 1. Dataset for training, testing, and validating our facial expression detection model.

Emotion	Total Images	Training Images	Testing Images
Happy	134915	90393	44522
Sad	25959	18171	7788
Surprise	14590	11380	3210
Fear	6878	5227	1651
Disgust	4303	3098	1205
Anger	25382	19797	5585
Neutral	75374	60299	15075
Contempt	4250	3060	1190

Feature Extraction

In order to make feature extraction easier, we will use a Gaussian Blur filter and threshold operation in our subsequent work after capturing frames with OpenCV [10]. In order to minimize noise and highlight important visual components in the collected frames, this step is crucial. The processed image will then be used as input for our CNN model which consists of various convolutional layers for better features and characteristics identification. Features such as facial action units (FAUs) focusing on facial muscle movements, intensity of expression, facial landmarks (nose, eyes, contours, mouth, lips), gaze analysis and skin texture (pallor or dark circles) will be extracted. These features then will be used as criteria for determining depression in our emotion recognition model (ERM).

Model Building

We will be creating an ERM, which will be trained using the images captured in the data collection phase for detecting depression. This tensor flow–based model will be equipped with the ability to classify whether a person is depressed or not, mainly using Haar cascade classifier [11]. Testing of the model will also be done using the dataset created in the first step. The model will also take time as a crucial element into consideration as a person suffering from depression is sad and in distress for a long period of time, rather than someone who is only unhappy for a short period of time. Other Python libraries are also used in the building of this model such as Keras, Matplotlib, Seaborn, and Numpy.

Real Time Emotion Recognition

The ESP32 camera module and the ERM will work in unison to provide the user to detect whether the person whose facial expression are being processed through the camera, is suffering from depression or not. Here the images captured from camera are sent to the trained and tested model which then detects the features in the images for depression and give the results accordingly.

Customized Intervention and Support

Depending upon the result from the model, customized intervention and support will be provided. For example, if a person is detected with depression, then the victim and their well-wishers will be notified/alerted through messages regarding the mental condition of the victim. Along with that, the victim will also be provided with the emergency helplines provided for depression. On the other hand, if a person is not suffering from depression, then they will be notified about this accordingly. The flowchart of the prosed system is shown in Figure 1.

ALGORITHM USED

A regularized kind of feed forward neural network, a CNN [12] learns feature engineering on its own through filter optimization. It is a particular kind of artificial neural network designed to process pixel input and recognize images. It is ideal for identifying and recognizing objects because of which we are implementing it in our model for detecting facial expression. It consists of multiple layers, which are as follows:

- *Input Layer:* The raw images are given as input to the CNN through the input layer. Here, the user's facial expressions are given as input to this layer.
- *Hidden Layers:* In order to learn aspects unique to the data, these layers carry out operations that modify the data. Convolutional, pooling, and fully connected (FC) layers are among them frequently. In our scenario, these layers are used to process the face expression photos.
- *Convolution Layer:* It consists of a convolutional kernel (matrix of numbers used in image convolution) which is used for extracting features from the input images. Performance of the layer is based upon the number of kernels. Here, different features like smile, eyes, wrinkles, edges, contour and textures will be extracted.
- *Activation Function:* Each layer consists of an activation function which is used to map input nodes to output nodes and also in normalizing the output. Now there are various types of activation functions but usually ReLU is used in CNN [13]. ReLU is a non-linear function which outputs the input directly if it is positive, otherwise it gives output as zero.

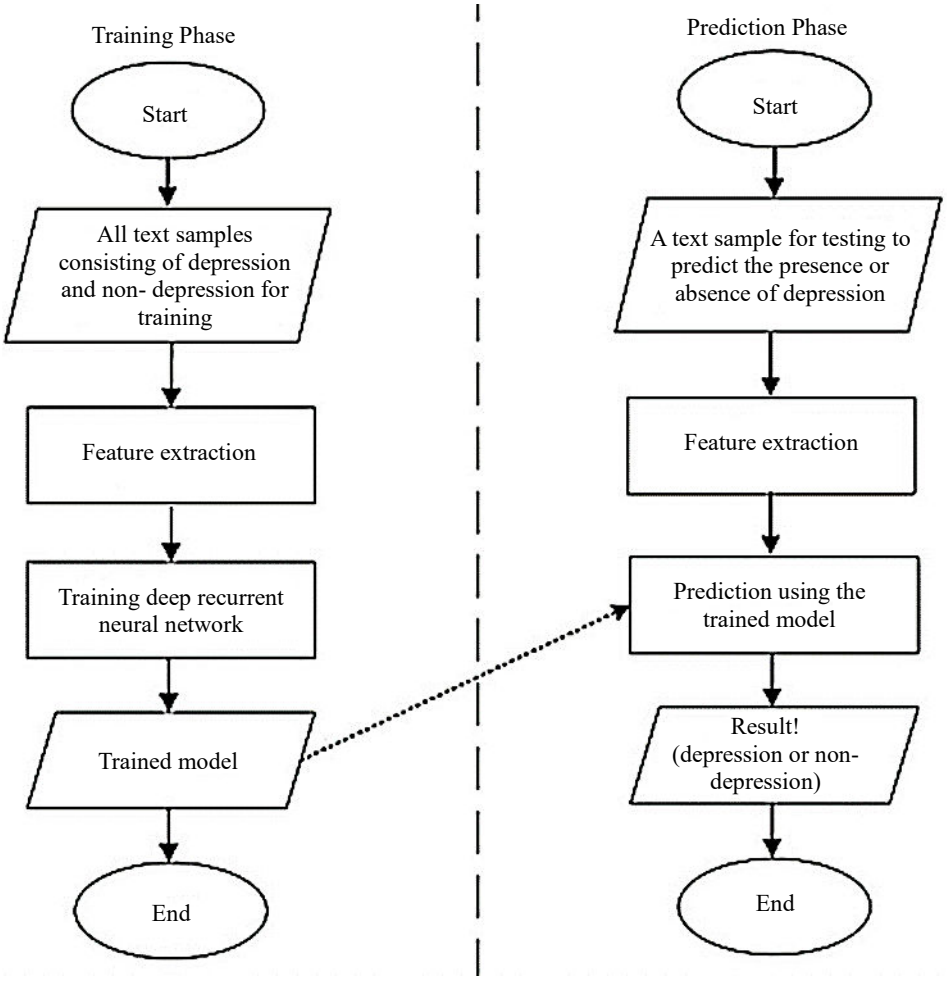

Figure 1. Flowchart for training and prediction phase of the model.

- *Pooling Layer:* They are primarily used to shrink the spatial size of the representation, which lowers overfitting (information loss) by reducing the number of parameters and calculations in the network.
- *Fully Connected Layer:* Based on the attributes that were retrieved from the earlier layers, images are classified in this layer. The photos will be divided into many classes in this instance.
- *Dropout Layer:* This layer is used to prevent over fitting as it nullifies the contribution of some neurons towards the next layer.
- *Output Layer:* This layer produces the desired final predictions and mainly uses sigmoid has an activation function. Here, the model will predict the desired label for the image given as input. CNN for facial expression recognition system is shown in Figure 2.

REQUIREMENTS

Software Requirements

- Core programming language: Python 3.8 or above.
- Frameworks: OpenCV, Tensor Flow, Keras, Seaborn, Haar cascade classifier, etc.
- IDE: Visual Studio Code [14], Jupyter Notebook, Pycharm.
- Message alert platform: Telegram

Hardware Requirements

- Computer/Laptop
- ESP32 external camera module

- 64-bit Windows 10 environment
- Minimum 6 GB RAM and 256 GB SSD
- Internet connectivity

RESULTS

The CNN analysis of various face emotions is shown in Figure 3 and confusion matrix of these expressions in presented in Table 2.

Evaluation metrics such as precision, accuracy and recall. Model precision is calculated to be 87.1%, followed by accuracy of 81.5% and recall of about 88.6%. To compute accuracy, precision, and recall, we take the values of the true positives, true negatives, false positives, and false negatives from the confusion matrix.

Table 2. Confusion matrix of precision, accuracy, and recall.

Metrics	Model Performance
Precision	0.871
Accuracy	0.815
Recall	0.886

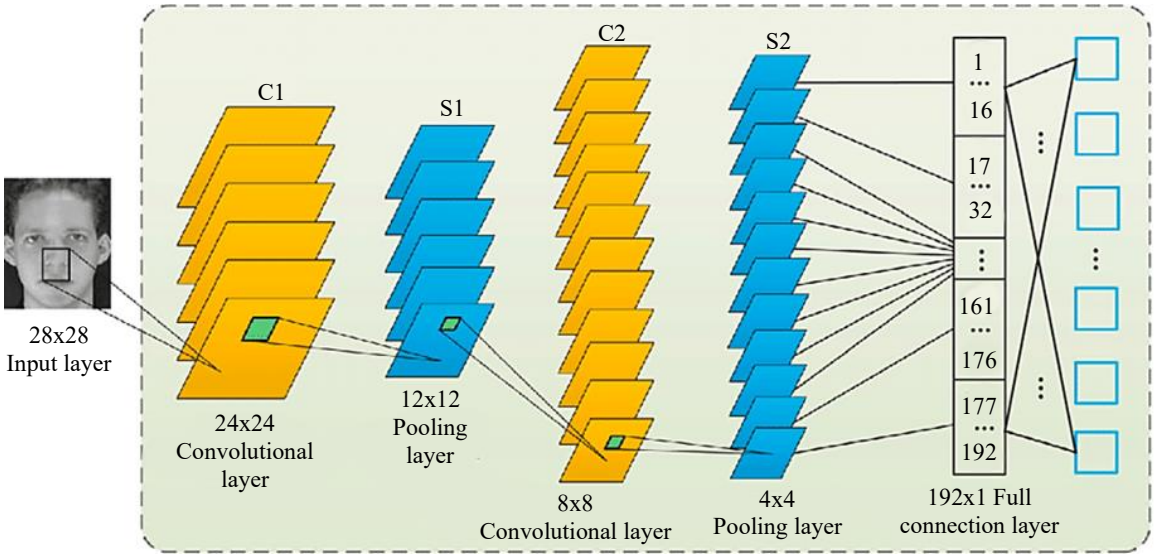

Figure 2. Convolutional neural network (CNN) for facial expression recognition.

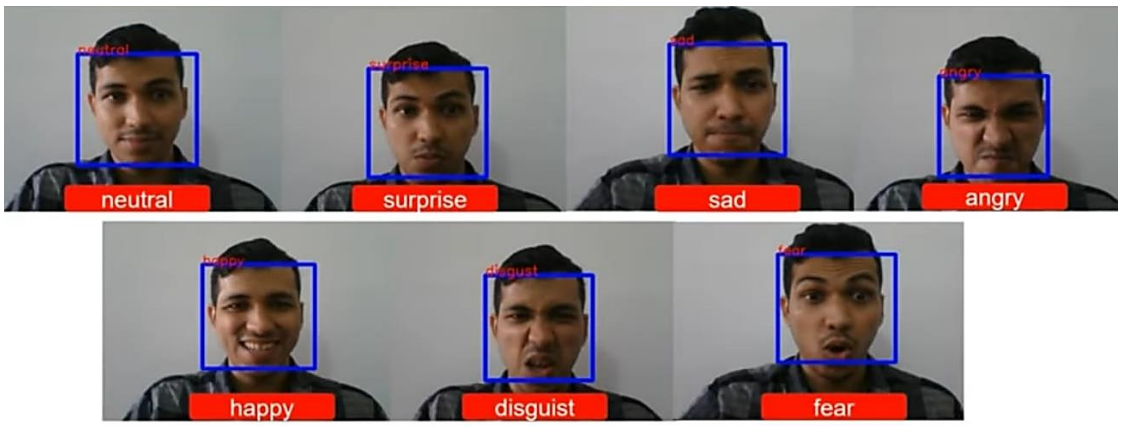

Figure 3. Types of facial emotions.

CONCLUSION

The proposed model will definitely play a crucial role in recognition and detection process of depression, which will be beneficial for early and better treatment. Facilitating a real-time detection, interpretation and then notifying the user regarding their mental state through messages, our model will be a stand-out overcoming various drawbacks of other models. Our model uses a CNN, which offers cutting-edge precision in image processing. It also makes use of various Python libraries for its other functionalities. Through this paper we have given a detailed overview of how, a real-time facial expression recognition model for depression detection can be created having high efficiency and accuracy.

Acknowledgments

We would like to express our deepest appreciation to Principal Dr. S. D. Lokhande for providing all the facilities necessary for completing this research. This research would not have been possible without the support from our Head of Department Dr. M. B. Mali. We are extremely grateful to our internal guide Mrs. P. P. Gaikwad who generously provided knowledge and expertise during the research. We are deeply indebted to our internal reviewers Mr. G. N. Gaikwad and Mrs. J. A. Desai for their valuable guidance and patience. A special thanks to all the staff members of our department teaching as well as non-teaching for their moral support. We would also like to acknowledge all our fellow classmates for their advice and suggestions. Lastly, we would like to thank our families who believed in us and always kept our spirits soaring high.

REFERENCES

1. Gillette H. Healthline Media. What percentage of people have depression? [Online]. April 24, 2024. Available at https://www.healthline.com/health/baby-blues-vs-postpartum-depression#faq
2. BetterHealth Channel. Depression explained. [Online]. 2024. Available at https://www.better health.vic.gov.au/health/conditionsandtreatments/depression
3. Nazira FA, Das SR, Shanto SA, Mridha MF. Depression detection using convolutional neural networks. In: 2021 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON), Dhaka, Bangladesh, December 3–4, 2021. pp. 9–13. doi: 10.1109/SPICSCON54707.2021.9885517.
4. Zhu Y. Automatic depression recognition using multi-scale facial behavior dynamics. In: 2022 IEEE 2nd International Conference on Information Communication and Software Engineering (ICICSE), Chongqing, China, March 18–20, 2022. pp. 88–93. doi: 10.1109/ICICSE55337.2022. 9828962.
5. Kantharia KJ, Prajapati GI. Facial behavior recognition using soft computing techniques: a survey. 2015 Fifth International Conference on Advanced Computing & Communication Technologies, February 21–22, Rohtak, Haryana, India, 2015. pp. 30–34. doi: 10.1109/ACCT.2015.132.
6. de Melo WC, Granger E, Hadid A. A deep multiscale spatiotemporal network for assessing depression from facial dynamics. IEEE Trans Affect Comput. 2022; 13 (3): 1581–1592. doi: 10.1109/TAFFC.2020.3021755.
7. Mollahosseini A, Hasani B, Mahoor MH. AffectNet: A database for facial expression, valence, and arousal computing in the wild. IEEE Trans Affect Comput. 2019; 10: 18–31.
8. Kabir MM, Ohi AQ, Rahman MS, Mridha MF. An evolution of CNN object classifiers on low-resolution images. In: 2020 IEEE 17th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET), Charlotte, NC, USA, December 14–16, 2020. pp. 209–213. doi: 10.1109/HONET50430.2020.9322661.
9. Krause FC, Linardatos E, Fresco DM, Moore MT. Facial emotion recognition in major depressive disorder: a meta-analytic review. J Affect Disord. 2021; 293: 320–328.
10. Mittal A. Haar Cascades. A general representation of training a Haar classifier. [Online]. December 20, 2020. Available at https://medium.com/analytics-vidhya/haar-cascades-explained-38210e579 70d

11. Anderson IM, Shippen C, Juhasz G, Chase D, Thomas E, Downey D, Toth ZG, Lloyd-Williams K, Elliott R, Deakin JW. State-dependent alteration in face emotion recognition in depression. Br J Psychiatry. 2011; 198 (4): 302–308.
12. Brownlee J. A Gentle Introduction to the Rectified Linear Unit (ReLU). [Online]. August 20, 2020. Available at https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/
13. Akhapkin RV, Volel BA, Shishorin RM, Ustyuzhanin DV, Petelin DS. Recognition of facial emotion expressions in patients with depressive disorders: a prospective, observational study. Neurol Ther. 2021; 10: 225–234.
14. Visual Studio Code. 2024. Setting up Visual Studio Code. [Online]. Available at https://code.visualstudio.com/docs/setup/setup-overview

