A REVIEW PAPER FOR PSYCHE - "A MENTAL HEALTH DETECTION SYSTEM"

1st Riya Magotra
Department of Computer Engineering
K.J. Somaiya Institute of Engineering and Information Technology, Sion
Mumbai, India

2nd Ritika Raina
Department of Computer Engineering
K.J. Somaiya Institute of Engineering and Information Technology, Sion
Mumbai, India

3rd Pallavi Patil
Department of Computer Engineering
K.J. Somaiya Institute of Engineering and Information Technology, Sion
Mumbai, India

***Abstract*— Stress is a term used to describe emotional distress. Our own mental health as well as that of those around us may be impacted. While anxiety is a common, potentially terrifying reaction to stress, it can also result in panic attacks. The most recent figures show that millions of people around the world experience one or more mental disorders1. Early diagnosis of mental illness can help with both disease progression and treatment in general.Effective feature detection is made possible by convolutional neural networks. A convolutional neural network is particularly good at identifying images because of the way that simple features that are discovered early in the network (like lines) may be combined to create more complex features (like the straightforward curve lines that make up a left eye). Audio datasets are used to train and assess classification algorithms like CNN. After using acoustic feature extraction as pre-processing, CNN is employed to accurately identify the audio based on emotions. This enables us to predict whether the person is anxious or not.This study reviews machine learning-based mental disease detection system developments. Early detection and management can improve patient outcomes for mental illness, a global issue. Speech, text, and physiological data train detection systems, according to the paper. Deep learning and support vector machines are also reviewed. The paper discusses data bias and interpretability issues in mental disease detection systems. Finally, these systems may improve mental health care results and save costs.**

Keywords*—*Mental Health, Sentiment Analysis, Deep Learning Algorithms, Convolutional Neural Network, Natural Language Processing

I. Introduction

A software programme or tool known as a mental disease detection system employs machine learning algorithms and other methods to recognise and classify mental health issues in people. In order to identify possible indicators of mental illness, the system can be configured to analyze data from a variety of sources, including medical records, behavioral assessments, and self-reported symptoms. Such a system's objective is to offer early intervention and stop the worsening of mental health disorders, improving patient outcomes. Recent technological developments have made it feasible to create more advanced techniques for detecting mental disease, which can increase the precision and effectiveness of diagnosis.

However, these tools should never be used in place of qualified clinical judgement or as a replacement for specialist mental health care.

The main causes of elements that affect mental health are probably a person's way of life, including stress at work, financial hardship, family problems, marital problems, violence, and environmental conditions. These circumstances may be a factor in mental health conditions like depression, worry, stress, and other psychological conditions that affect quality of life[10].

The effects of melancholy on people's psychology are profound. Depression will consequently impair a person's ability to concentrate, learn, and work efficiently, which will have a significant negative impact on the individuals' quality of life. Five of the top ten most common diseases in the world that render individuals disabled or incapable are mental illnesses, with depression coming in at number one. This is sufficient evidence that depression can seriously harm civilization.

Currently, a questionnaire poll is the primary method for diagnosing depression. Mental health is a major part of an individual's health. Users who take part in psychological surveys receive survey questionnaires from Healthcare institutions or psychological assessment organizations[2].

Each year, nearly 13% of children, 46% of adolescents, and 19% of adults battle with mental illness worldwide. Therefore, early diagnosis of depression is essential to save the lives of depressed people by preventing it from progressing to a severe and irreversible condition. Usually, the patient's behavior exhibits the signs of depression. As a result, physicians use questionnaires and talking therapy sessions as screening tools to gauge the severity of depression.

However, the effectiveness of the psychologist or psychiatrist will determine how well the talking session goes. Furthermore, due to the stigma associated with mental illness, depressed people are less likely to seek treatment. As a consequence, a sizable proportion of those who are depressed do not receive the best treatment and adequate recovery time. Therefore, developing appropriate and effective methods for diagnosing depression is a developing area of study, and recent advancements in instrument or sensor technology open up new avenues for doing so[3].

Different facial traits, such as eyes, eyebrows, and nose, can be extracted using feature extraction with the use of the CNN model, which are used to categorise face features, and gender and feelings are predicted. This is related to feature extraction in the current study because characteristics would be used to classify people and predict their emotional states[4].

II. Model architecture

Figure 1 depicts the model's suggested organisational structure for the system. Both the categorization of a person's gender and the classification of their facial expressions are used in this model. All of the models are educated using the same underlying data. Then, characteristics are extracted by utilising the gender classification model and the face emotion model. Then, the relevant features, that is, emotion features and gender features, are concatenated and transferred to a CNN-based classifier model, which is where the associated inferences are formed. In the work that has been detailed, a convolutional neural network, also known as a CNN, has been used in combination with emotion categorization in order to identify a wide range of emotions, some of which include happy, anger, sadness, neutrality, surprise, and fear, amongst others. The CNN model that is used to distinguish different emotions is constructed using input photographs from the dataset that have the dimensions 48*48*1. The CNN model contains a first layer that is comprised of 16 filters, with each filter having a size of [7*7]. Filters are a template for feature analysis that can be used to compare a given template to different parts of an image. This comparison may be accomplished by using filters. [4]

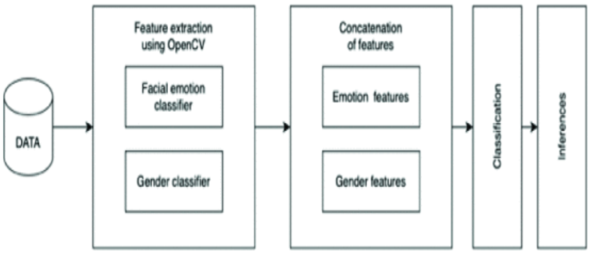


Fig. 1. Proposed System Architecture

The model was constructed by the use of the binary entropy loss function. In order to construct our model, we came up with a total of 80 epochs.

Both of our models made use of the OpenCV computer vision library. Python and Keras were the programming languages of choice for us when it came to putting both of these models into action. Image classification and machine learning, as well as the identification and categorization of various objects, are accomplished with its assistance. OpenCV is a free and open-source set of programming functions with the purpose of resolving issues that arise in computer vision.

CNNs, also known as convolution neural networks, are a kind of deep learning model that can classify pictures into a variety of categories. The CNN model makes use of a technique that is composed of many layers; the first layer is responsible for identifying edges, bright spots, and dark regions in pictures (a 2D array of pixels) that are fed into the model. In the second layer of analysis, the CNN model is able to recognise features of the face such as the eyes, the nose, and the mouth. The third layer, which is the last stage, of the CNN model is where complete facial features are recognised.

During the process of convolution, an input picture is used to generate a stack of images that have been filtered. In pooling, the picture stack is reduced in size by using a sliding window. The data is normalised using the ReLu technique, in which all negative values are changed to zeroes, also known as ReLu units, and the stack of pictures is changed into a rectified linear unit layer (ReLu), which does not include any negative values in the array. As the layers are stacked on top of one another, the output of one layer will ultimately become the input of the next one. These layers are repeated, and the photographs are continually filtered, in order to form one layer that is thoroughly linked throughout its whole. In order to gather voting weights in entirely connected layers and features in convoluted layers, the backpropagation approach is utilised. When the CNN model has finished processing an image, it generates a final output. The amount of incorrectness in the output, also known as error, reveals how accurately the calculated features and weights were determined. A method known as gradient descent is one that may be used to lessen the amount of error that is present in the calculated weights and features. The model will then identify the right response and output it to the user.

The most cutting-edge image classifiers that are now accessible make an effort to automatically identify key aspects of a picture in order to classify it without the assistance of a person. When it comes to this particular aspect, the efficacy of convolutional neural networks is superior to that of their forebears. The pre-processing of data in CNN may be done in a manner that is noticeably simpler, and many of its activities do not call for the involvement of a person. The CNN model is also more accurate than models that came before it, particularly early versions. Within the context of this model, the output dimensions are generated on the basis of the class labels. "Man" and "Woman" are the class labels used in our case when gender detection is being performed.

The names of the classes of emotions that may be detected by emotion detection are as follows: joy, sorrow, anger, fear, surprise, and neutrality. In the not-too-distant future, a number of different optimisation algorithms will be capable of being used to construct the model's total loss function. It is possible to use massive pre-trained models in order to improve the accuracy of the model. In addition, a sparse dataset may be used to train a model whilst taking into consideration each and every minute characteristic, including the elevation of the brows, nose, eyes, and lips. The dataset may also be trained using the user's apparel in line with the clothing conventions that are prevalent in a variety of geographical locations. Keeping this in mind allows for significant improvements to be made to the model's level of accuracy[4].

A. *Convolutional Neural Network(CNN)*

Following a series of layers that are used for convolutional and pooling operations, the model comprises fully linked layers where classification takes place. A two-dimensional matrix of EEG signal data serves as the model's input. This matrix is organized such that each row represents a distinct electrode and each column serves as a time sample. The model will provide a binary classification as its output, which will indicate whether or not the patient is likely to suffer from a mental disease.

The following describes the architecture of the model:

The input layer is responsible for receiving the two-dimensional matrix of EEG signal data.

Convolutional layers: The input data are processed via numerous layers of one-dimensional convolutional filters, with each filter in each layer identifying a separate feature. Following the completion of each convolutional layer, the output is sent into an activation function known as a rectified linear unit, or ReLU.

Max pooling layers: Max pooling layers are used to downsample the output of the convolutional layers, hence lowering the total number of parameters utilised by the model and ensuring a certain degree of translation invariance.

The output of the most recent max pooling layer is flattened into a 1D vector before the next layer is flattened.

Dense layers: There are numerous completely linked layers, and they are responsible for transforming the flattened output into a feature space with fewer dimensions. A sigmoid activation function is given the output of the last dense layer that was constructed.

The output layer makes use of a sigmoid activation function in order to create a binary classification that indicates whether or not the patient is likely to have a mental illness.

During the training process, the model is improved by using Adam optimisation and binary cross-entropy loss optimisation. The model is trained on a large dataset consisting of EEG signals that have been annotated with information on mental health. After the model has been trained, it is possible to use it to determine, based on the EEG signal data of a new patient, whether or not that patient is likely to have a mental disorder.

It is important to note that this is only one example of a CNN model architecture for the detection of mental disease. There may be changes based on the particular sort of physiological data that is being analysed as well as the particular mental illness that is being targeted.

Convolutional neural networks are classified as deep neural networks if they have three or more buried layers. Using the histogram of directed gradient approaches and scale-invariant feature transform, CNN can extract more data. Fig. 1 illustrates how CNN[5] can be used to extract traits from photos and learn from them.

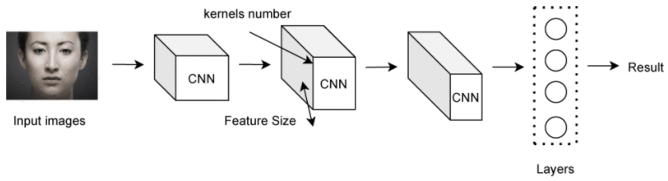


Figure 2: Depression picture classification

Figure 2 shows the results of a depression image classification project that made use of a convolutional neural network and an enriched dataset. Each square illustrates a different result of the convolution. Measurements are taken of both the breadth and the height. The breadth of each cube is equal to the number of kernels it contains. Each and every convolution should strive to achieve an activation layer in addition to group normalisation. Following the completion of the last convolution, it is subsequently turned into a linear collection of nodes.Each of the six convolutional layers in our proposed architecture for the CNN uses a ReLu activation function. This architecture is based on the sequential model. Because each layer makes use of a variety of convolution kernels, it is possible for that layer to produce a variety of feature maps. In the architecture that we recommend, the size of the kernel is (3,3). The scene for the photograph is black and white. The following is a mathematical description of the convolution technique:

$$y_j^l = f\left(\sum_{i\epsilon M_j} y_i^{l-1} * k_{ij}^l + c_j^l\right)$$

Here, y_i^{l-1} = the outcome of the characteristic map for the preceding layer. y_j^l is the jth convolution layer's ith channel output. Utilize the M_j input feature maps subset to determine the activation function u_l^j, f (). Convolutional kernels are denoted by k_{ij}^l and corresponding offsets by c_l^j.

The process of individually training each layer model using batch normalization has been completed. Normalize the output of the layer that came before it and cut down on the amount of overfitting for a more subtle regularization effect. In the configuration that we have given, the batch size is 128. The group normalization of the CNN layer has to be defined. As a consequence of this, the designs of deep neural networks become more stable, and convergence takes place more rapidly. One possible formulation of the equation for batch normalization is as follows:

$$Y_i'^{(m)} = \frac{y_i^{(m)} - \mu_C^{(m)}}{\sqrt{\sigma_C^{(m)^2}}}$$

Here, $y_i^{'(m)}$ represents the mth hidden unit standardized value, $\mu_c^{(m)}$ represents the mean value, $\sigma_c^{(m)^2}$ represents the mth hidden unit variance, and C represents the specific data batch.

The calculation is carried out with the help of the ReLU, which stands for "rectified linear activation function." The processing demands of the neural network are kept from expanding at an exponential rate thanks to ReLU's assistance. Zero is what is sent back by ReLU whenever it is given a non-positive number. If the value being inputted is positive, ReLU will output the same value as the input.

The definition of ReLU function is as follows:

$$\text{ReLU}=\max(0,y)$$

We are using layers that were generated by MaxPooling2D. According to our understanding, the capacity of the pool is (2). The input is downsampled across all of its spatial dimensions by choosing the greatest possible value for each input channel while it is being passed through an input window. After that, the Dropout layer is added to the model in order to prevent it from getting overfit. During the training phase, the dropout layer randomly set all of the input units to zero while still making use of the frequency rate. The capabilities of the network is restricted by the dropout layer.

The following formula may be used to determine the output of the neural network with dropout:

$$m[pk] = \sum_{r} n(p)pk$$

In this case, dropout chance n(p) is n where p = neuron output.

The thick layer is composed of a neural network that has a great number of connections. The neuron in the layer below a certain neuron is the source of information that is received by that neuron. The result of the last convolutional layer is what is fed into the thick layer that comes before it. The output of the convolutional layer is used by the dense layer, which then uses that information to categorise the images. The equation below indicated a layer that was fully linked across its entirety:

$$e(y) = Activation(u^{n}y + d)$$

Here, the dense layer bias value is given by d and the dense layer weight vector is defined by the expression u = [u 1, u 2, u 3, u 4,..., u n].

B. *Open CV*

In order to diagnose depression, we made use of the depression model.h5 and OpenCV in conjunction with the Haar cascade classifier. Face recognition in OpenCV is accomplished with the help of the Haar Cascade Classifier. A significant number of photographs are used in the training process for the Haar Cascade Classifier. The process flow of the Haar Cascade Classifier is shown in Fig.3[5].

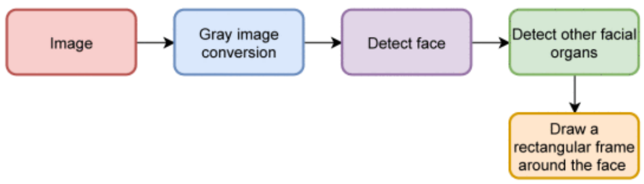


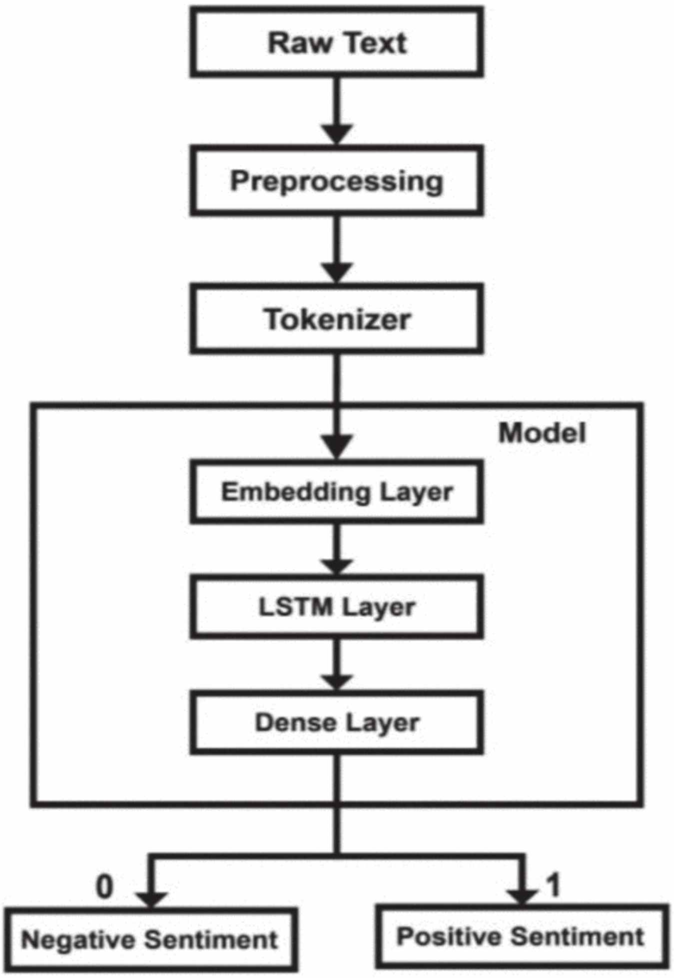
Fig.3. Haar Cascade Classifier's workflow

Essentially, the Haar Cascade Classifier turns the image into a grayscale one. The Haar Cascade classifier recognises additional facial features and confirms that the image comprises a human face. On the face organ, doodle a rectangle frame.

C. Sentiment Analysis

The first step of sentiment analysis is preprocessing, then tokenizing, and the last part is prediction. During preprocessing, the user's raw text will be converted to lowercase and any characters other than spaces and alphabetic characters will be removed. After that, an integer-token tokenizer is used to the preprocessed text in order to transform it into tokens. Through a method called as tokenization, the text that is meant to be read is sectioned off into its component parts. During the stage when predictions are being made, the tokens are sent to the embedding layer of the model. The embedding layer is responsible for converting the integer-tokens into real-valued vectors so that the LSTM layer can analyse the sequence. The conclusion is reached by using a dense layer with softmax serving as the activation function. The values 0 and 1 respectively represent negative and positive states of mind. Figure 4 illustrates the layout of the building[6].

Fig. 4. Architecture for sentiment analysis



D. Facial Expression

For the purpose of identifying facial expressions, a CNN model composed of 13 layers and separated into 4 blocks is used. The last of the first three blocks is called max-pooling, and it comes after two convolution layers in the first three blocks.

A dropout layer is inserted in the space between each block in order to avoid overfitting. Because of this, the network becomes less dense as the training progresses.

The ultimate block is composed of two dense layers that are entirely linked to one another, with the first dense layer having an output shape of (None, 128) and the second dense layer also having an output shape of (None, 128). (None, 7).

The top layer is comprised of a 7-way softmax classifier. This network, which accepts as input a 48 x 48 grayscale picture, generates the probability distribution of the seven classes ('angry,' 'disgust,' 'fear,' 'happy,''sad,''surprise,' and 'neutral'). The seven classes are 'angry,' 'disgust,' 'fear,' and 'happy,''sad,''surprise,' and 'neutral'.

In comparison to the conventional CNN Model, this model is able to provide more accurate predictions as the number of layers grows. Figure 5 illustrates the layout of the building[6].

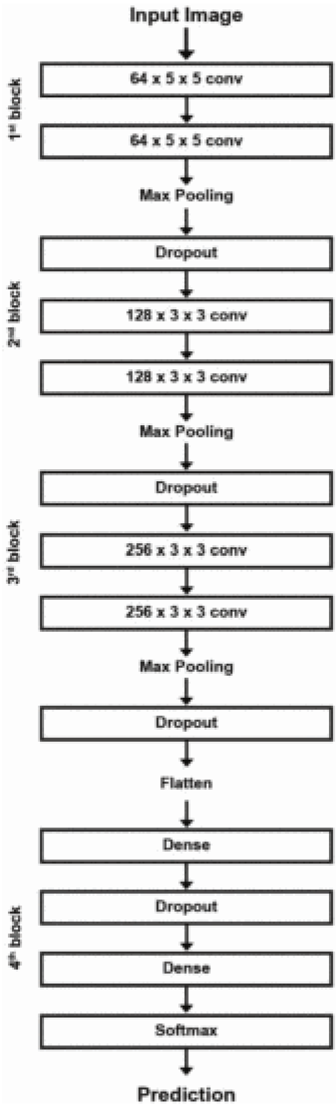


Fig.5. Architecture for facial expression

III. Experimental Results And Conclusion

A. Merging of Models

The 2D-CNN models that were trained on the face dataset, the eye-only dataset, and the mouth-only dataset are integrated in order to improve the overall classification performance.

For the purpose of putting the models through their paces, video clips of the test participants are recorded, and then the relevant video frames are extracted. After the video frames have been extracted, the MTCNN algorithm is used in order to recover the subject's face, eyes, and mouth regions. This step follows the extraction of the video frames. Following the incorporation of the extracted ROIs into the relevant trained models, the probabilities of classification are then calculated. The following is the outcome of computing the final class probabilities for a particular video clip by utilising the average probability for each ROI.

$$\Pr(face + eye - only + mouth - only) = e_{avg} + 0.5|e_{avg} - f_{avg}| + 0.5|m_{avg} - f_{avg}|.$$

B. Evaluation

We consider the following metrics to measure the performance of the proposed method which are given as:

$$Recall = \frac{TP}{TP + FN},$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN},$$

$$Precision = \frac{TP}{TP + FP},$$

$$F - score = \frac{2 \times Recall \times Precision}{Recall + Precision},$$

IV CONCLUSION

In conclusion, the diagnosis of mental disease is a difficult effort that requires the examination of a variety of data sources, such as physiological and behavioral data. This is a complicated and difficult undertaking. Growing interest in the development of automated systems for the identification and diagnosis of mental diseases has coincided with the development of methods for machine learning and deep learning. These systems have the potential to enable early identification and intervention, which might lead to better treatment results and an improved quality of life for persons who are afflicted with mental illness.

Finally, this study aids in identifying a person's mental condition. In order to determine the results of the project, many methods and algorithms were applied. It also illustrates previous research on the same subject and how it was solved, in addition to the project's original method. It provides a thorough overview of the solution and architecture that are suggested. Also, every module utilized in the Psychological Screening test is described in the article. This document aids in our comprehension of the project's overall characteristics, findings, and implications.[11] The Convolution Neural Network with Sequential model is used in this instance to detect facial expressions, while the LSTM model is utilized for sentiment analysis. The goal of the suggested methodology is to improve results from both written and facial responses. The accuracy rates for facial expressions and sentiment analysis are 80% and 73.45%, respectively. By utilizing other datasets or networks, the proposed models' accuracy can be improved.Other factors for analyzing depression include vocal expression, face microexpression, fluctuation in eye color, and behavioral changes in addition to facial and textual inputs. In order to increase accuracy, these traits can be applied in the future with the use of neural networks. The accuracy of forecasting a person's depression condition can be increased by using video frames in addition to text.[6]

ACKNOWLEDGMENT

Riya Magotra,Ritika Raina, contributed equally for this paper and were guided by Prof. Pallavi Patil, K.J. Somaiya Institute of Engineering and Information Technology.

REFERENCES

[1] R. Jadhav, V. Chellwani, S. Deshmukh and H. Sachdev, "Mental Disorder Detection : Bipolar Disorder Scrutinization Using Machine Learning," 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2019, pp. 304-308, doi: 10.1109/CONFLUENCE.2019.8776913.

[2] Y. Ding, X. Chen, Q. Fu and S. Zhong, "A Depression Recognition Method for College Students Using Deep Integrated Support Vector Algorithm," in IEEE Access, vol. 8, pp. 75616-75629, 2020, doi: 10.1109/ACCESS.2020.2987523.

[3] A. Seal, R. Bajpai, J. Agnihotri, A. Yazidi, E. Herrera-Viedma and O. Krejcar, "DeprNet: A Deep Convolutional Neural Network Framework for Detecting Depression Using EEG," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-13, 2021, Art no. 2505413, doi: 10.1109/TIM.2021.3053999.

[4] B. Divya Kannan and N. Ilamurugu, "SADGF: Surveillance based Anxiety Detection using Gender-based Facial Emotion Recognition," 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022, pp. 966-969, doi: 10.1109/ICACRS55517.2022.10029253.

[5] F. A. Nazira, S. R. Das, S. A. Shanto and M. F. Mridha, "Depression Detection Using Convolutional Neural Networks," 2021 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON), Dhaka, Bangladesh, 2021, pp. 9-13, doi: 10.1109/SPICSCON54707.2021.9885517.

[6] M. Mohan, A. K. Abhinav, A. Ashok, A. V. Akhil and P. R. Achinth, "Depression Detection using Facial Expression and Sentiment Analysis," 2021 Asian Conference on Innovation in Technology (ASIANCON), PUNE, India, 2021, pp. 1-6, doi: 10.1109/ASIANCON51346.2021.9544819.

[7] R. Acharya and S. P. Dash, "Automatic Depression Detection Based on Merged Convolutional Neural Networks using Facial Features," 2022 IEEE International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, 2022, pp. 1-5, doi: 10.1109/SPCOM55316.2022.9840812.

[8] Abdi, H. and L.J. Williams, Principal component analysis. WIREs Computational Statistics, 2010. 2(4): p. 433-459

[9] I. Syarif, N. Ningtias and T. Badriyah, "Study on Mental Disorder Detection via Social Media Mining," 2019 4th International Conference on Computing, Communications and Security (ICCCS), Rome, Italy, 2019, pp. 1-6, doi: 10.1109/CCCS.2019.8888096.

[10] R. A. Rahman, K. Omar, S. A. Mohd Noah, M. S. N. M. Danuri and M. A. Al-Garadi, "Application of Machine Learning Methods in Mental Health Detection: A Systematic Review," in IEEE Access, vol. 8, pp. 183952-183964, 2020, doi: 10.1109/ACCESS.2020.3029154.

[11] R. Jegan, S. Mathuranjani and P. Sherly, "Mental Stress Detection and Classification using SVM Classifier: A Pilot Study," 2022 6th International Conference on Devices, Circuits and Systems (ICDCS), Coimbatore, India, 2022, pp. 139-143, doi: 10.1109/ICDCS54290.2022.9780795.

[12] S. Hurayn, A. M. Shetty, S. V. Deshpande, V. Gupta, U. Ananthanagu and A. Professor, "Analysis of Mental Illness using Twitter Data," 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS), Bengaluru, India, 2021, pp. 1-4, doi: 10.1109/FABS52071.2021.9702565.

[13] A. Prakash, K. Agarwal, S. Shekhar, T. Mutreja and P. S. Chakraborty, "An Ensemble Learning Approach for the Detection of Depression and Mental Illness over Twitter Data," 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2021, pp. 565-570.

[14] M. P. Jain, S. Sribash Dasmohapatra and S. Correia, "Mental Health State Detection Using Open CV and Sentimental Analysis," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 465-470, doi: 10.1109/ICISS49785.2020.9315984.

[15] E. Saravia, C. -H. Chang, R. J. De Lorenzo and Y. -S. Chen, "MIDAS: Mental illness detection and analysis via social media," 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, USA, 2016, pp. 1418-1421, doi: 10.1109/ASONAM.2016.7752434.

[16] A. Wongkoblap, M. A. Vadillo and V. Curcin, "Detecting and Treating Mental Illness on Social Networks," 2017 IEEE International Conference on Healthcare Informatics (ICHI), Park City, UT, USA, 2017, pp. 330-330, doi: 10.1109/ICHI.2017.24.

[17] R. Kanaan, B. Haidar and R. Kilany, "Detecting Mental Disorders through Social Media Content," 2021 IEEE 3rd International Multidisciplinary Conference on Engineering Technology (IMCET), Beirut, Lebanon, 2021, pp. 23-28, doi: 10.1109/IMCET53404.2021.9665555.