DCNN-WiGRF: Depression Recognition through Deep Feature Extraction using Deep Learning and weight updated information gain Random Forest technique

[1]Ms. Pratiksha Meshram*, [2]Dr. Radhakrishna Rambola

[1]Department of Information Technology, SVKM's NMIMS,
MPSTME Shirpur campus, Maharashtra, India pratiksha.kaya@gmail.com,
[2]Associate professor, Department of computer Engineering, SVKM's NMIMS,
MPSTME Shirpur campus, Maharashtra, India Radhakrishna.Rambola@nmims.edu
*Corresponding Author Mail id: pratiksha.kaya@gmail.com

Cite this paper as: Pratiksha Meshram, Radhakrishna Rambola (2024) DCNN-WiGRF: Depression Recognition through Deep Feature Extraction using Deep Learning and weight updated information gain Random Forest technique *Frontiers in Health Informatics, 13* (3), 9162-9178

Abstract:

Depression is a general illness globally which causes severe implications. Depressive symptoms can predict earlier and it is considered as initial step in evaluation and treatment. There is a potential to generate computational intelligence model to detect the depression symptoms by using the depression datasets and machine learning advancements. Therefore, this research proposed DCNN-WiGRF algorithm to identify the depression disorder types such as normal, persistent disorder, bipolar disorder, and MDD. Initially the data is pre-processed and then augmented by using data augmentation process in which the rotation of image is performed. Further, the features are extracted by using Deep CNN algorithm from FER2013 dataset. Finally, the classification of the significant features is performed by using the Weight updated Informative Gain Random Forest algorithm based on queries and extracted features. Here the information gain is used because if the faster ability and it can exactly predict the depression from non-linear data. Thus, the depression prediction model can be adopted, which can contribute to classify the depression disorder types in order to provide earlier treatment.

Keywords: *Depression Disorders, Deep CNN, Informative Gain, Random Forest algorithm, Weight Updation*

1. Introduction

Depression disorder is considered as a common disease. According to WHO- World Health Organization, the depression disorder affected peoples are estimated as more than 300 million globally (Uddin et al., 2022). Depression may cause critical influence to human well-being and affect the performances at school, family, and workplaces which even resulted in self-injury. Younger stage depression is related with mood disorders and critical mental illness. Every year 0.8 million people die from suicide and it is considered as fourth major cause of death in adolescent age, stated by WHO. Incapability and disability are the major causes of depression. The depression occurrence in adult population is nearly 5 percent in cultures and 20 percent are in mild form like probable and partial depression(Depression, 2017; Kessler & Bromet, 2013). Among the adults, the most risk people with depression is the middle-age population. The depression occurrences are increasing and earlier intervention by the medical care can enhances the mental symptoms such as rumination and self-confidence absence and solves the somatic issues such as sleeping disorders and gastrointestinal problems in most of the cases(Weinberger et al., 2018).

Earlier prediction of depression symptoms followed by evaluation and treatment can improves the probabilities for restricting the symptoms and the disease, removes the negative results on health well-being, economic, personal and social life(Jones et al., 2018). The detection of depression symptoms is resource demanding and challenging. The present approaches are usually based on questionnaire surveys and clinical interviews from medical healthcare hospitals in which psychological assessment tables are used for mental disorder predictions. These questionnaires can diagnose psychological disorder roughly alone in case of depression. However recently Artificial intelligence approach employs certain computational techniques and linguistics in helps the machines to understand fundamental phenomena like emotions or sentiments from audio or texts or images or videos (Hantach et al., 2021).

In addition, deep neural network can used more in various research fields. It is considered as more robust compared with standard neural networks which possess two major disadvantages. The first disadvantage is considered as over fitting and computational time. Hence later the convolutional neural network has developed and shows higher power related with other approaches. CNN can extract the features across the training data which shows certain convolutional stacks for generating the various convolutional layers. CNN is majorly applied for the image and video assessment rather than decoding of temporal information.

Among various approaches, machine learning are widely used in evaluate the mental and physical states of human being from various datasets. Machine learning models are employed in performing important predictions. Hence the current research focuses on various machine learning models to predict the depression disorders from the reliable data source.

The major contribution of the study involves,

- To perform the pre-processing and data augmentation by randomly rotate the image dataset.
- To extract the significant depression features from depression dataset namely FER2013 using the Deep CNN network model for detect the important feature without any human supervision.
- To predict the depression disorder class namely normal, persistent disorder, bipolar dis-order and MDD by classify significant features using the Weight updated Informative Gain Random Forest algorithm based on the queries and extracted features.

1.2 Paper Organisation

The following section 2 describes the literature review of various depression types using the different artificial intelligence methods. Further section 3 describes the methodology of proposed DCNN-WiGRF method. The results and discussion is described in section 4. Finally, the study is concluded in section 5.

2. Related Works

The following section describes the related works of depression prediction using various artificial intelligence methods.

(He et al., 2021) suggested in considering the various features for identifying the depression using deep learning. More potential regression and representation patterns have been identified. The deep learning-based model can assist the doctors to assess the depressed patients also. In order to recognize the appearances from depressed patients, physiological signals, text, video, and audio from multi-modal depression and Chinese multimodal depressed patients progressed among various 570 cultures and races. (Priya et al., 2020) determines five severity levels with respect to stress, anxiety and depression. Five classification approaches such as Naïve Bayes, K-Nearest Neighbour, Random Forest, Decision tree, and Support vector machine have been considered. For the stress, anxiety, and depression scales, the significant variables identified as difficult to relax, scared without any reason and considered life as meaningless. While recognizing the psychological disorder these variables have been most significant.

(Shetty et al., 2020) focusing on better methods to improve the accuracy. CNN model, Naïve Bayes, Bernoulli Naïve Bayes, Support vector machine, and logistic regression, gradient boosting and random forest classifier have been analysed for better diagnosis. (Chekroud et al., 2016) improved the antidepressant treatment efficacy using various machine learning algorithms in order to assess the patients with symptoms. The variables have been identified for clinical remission prediction. (Islam et al., 2018) analysed temporal and emotional process and linguistics style using machine-learning techniques for classify the features. From more emotional features types, another approach used for paraphrase extraction of depression analysis (Meshram & Rambola).

(Andersson et al., 2021) assessed the postpartum depression predictions using demographic, psychometric and clinical data with the help of machine learning algorithms. From Sweden, the data has been collected with 4313 samples. Randomized trees algorithm ensured potential performance with 73 percent accuracy. However, for women without former mental health concerns, the accuracy obtained was 64 percent. Variables focused were anxiety and depression during pregnancy associated with personality and resilience. Higher postpartum depression risk issues in order to facilitate cost-effectiveness and follow-up. (Gao et al., 2018) reviewed different machine learning techniques utilized for prediction and classification of brain imaging, followed by major depression disorder prediction with respect to mood or control disorders. The study examined the treatment outcome detection for every patient. MDD biomarker identification associated with detailed overview assists the readers to identify the effective data mining technique.

(Hochman et al., 2021) analysed healthcare based clinical risk detection models focusing on postpartum depression risk prediction. It assists the postpartum women's clinical practice. In healthcare service, PPD risk prediction can serve as a warning sign for women before the depression symptoms alert identified. Secondary data analysis performed with respect to functional imaging, genetic data, and quantitative approach. (Bhakta & Sau, 2016) depression has detected as earlier from the socio-demographic and disease variables which assists the doctors in detecting the depression factor. Using GDS scale it has performed. For medical professionals the automatic predictive model development can provided the treatment accurate and faster. Five various learning procedures considered namely Decision table, Bayes net, Multilayer perceptron, and Logistic and SMO model. Better results have been obtained through Bayes Net. Optimization techniques have been required to avoid duplication confusion since depression is refer as multi-factorial illness. For more accurate depression prediction with feature selection, various nature inspired optimization algorithms have been utilized. The generalization of developed automatic prediction model has been assessed.

(Shah et al., 2020) analysed various techniques in order to predict depression from the text provided. The major focus is not on depression user classification but it can focus on to minimize the time in order to detect every user state. Word2Vec Embed with Meta features have been resulted in better output. The major drawback identified are the increased time in detection of depression. (Lin et al., 2018) examined the Taiwanese antidepressant outcomes using deep learning model. In evaluating the antidepressant treatment response with proper accuracy, deep MFNN framework has been developed. The proposed model results have been generalized in detecting the human disorders treatment which can be employed by developing the molecular prognostic and diagnostic tools in predicting the depression. The deep learning models evaluated using the independent greater sample sizes.

(Kumar & Chong, 2018) exploited the machine learning and correlation analysis for identifying the attributes showing certain impact in classifying the health status of patients. Correlation analysis has been attained which shows the depressive disorder symptoms with respect to emotional classification. For the bipolar patients, the results shows interesting correlations and the significant features have been extracted through physiological data. (Kessler et al., 2016) analysed the lifetime MDD disorder from 1056 respondents. The accuracy of developed machine learning model prediction has related with traditional logistic regression model. From the respondents, greater persistence chronicity exhibited as 40.8 to 55.8 percent severity indicator, suicide attempts recorded as 1.5 percent and frequent hospitalizations of 0.9 percent have been identified as machine learning predicted risks. From the patient self-reports MDD risk stratification models have been developed which improves the traditional methods in developing these models.

(Zandvakili et al., 2019) utilized new data-driven approach in evaluating the EEG coherence to defines the treatment in recognizing the clinical utility's candidate biomarkers. (Hatton et al., 2019) demonstrated the every individuals in developing the depression. Further machine learning based prediction model established in which most of the researches utilized the traditional statistical methods like logistic regression. In older individuals the predictive model's efficacy has not been assessed and thus the machine learning approach aimed at predicting the depressive symptoms compared with baseline of sub-threshold depression.

(Richter et al., 2020) employed cognitive behavioural tasks with respect to machine learning algorithms shows the several measures which characterize the subclinical psychiatric disorders. Higher level of depression or anxiety symptoms have been identified from anxious and depression individuals. (Yang et al., 2017) provided text, audio and video context related with AVEC2017 depression trial and the experimental outcomes outperformed the baseline results. Text descriptor-based paragraph vector embedding and HDR based 2D facial landmarks have been used. The multimodal hybrid context incorporating the DNN and DCNN model shows better performance. More evaluation has been required to boosting the performance of depression recognition.

(Zhou et al., 2018) focused on deep convolutional neural network fortified with global average pooling layer which are initially trained with facial expression information, allows in recognizing the prominent input image regions with respect to severity score of DAM - generated depression activation map. The overall recognition performance has been improved using the multi-region Depress Net approach. The method has been evaluated on two datasets and the results have been evaluated with respect to visual based depression identification. The learned deep model has expected to recognize the automated depression.

3. Proposed Methodology

In this research, the depression disorder is recognized using the deep feature extraction and machine learning techniques. Initially FER13 depression images dataset is pre-processed by performing resize the images. Followed by data augmentation process is performed for rotating the image. Further, the significant features are extracted from the augmented images using Deep CNN algorithm. Finally, Weight updated Informative Gain Random Forest algorithm is developed to classify the extracted features based on the queries provided. The overall flow of the proposed study is shown in below fig.1.

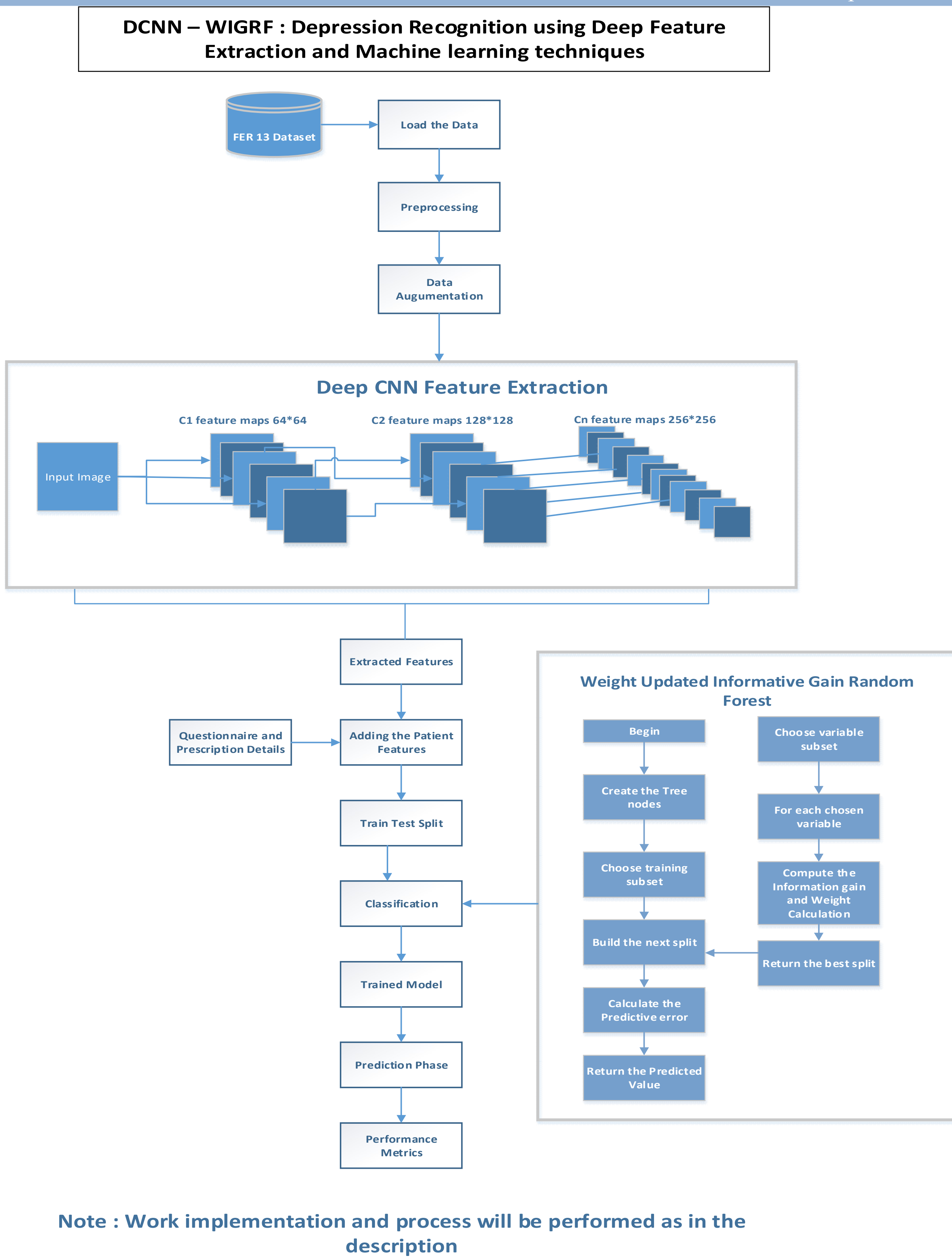


Fig.1. Proposed flow

3.1 Data Augmentation

The general data augmentation technique namely random rotation is used in this research. The image from dataset is rotated counter-clockwise or clockwise randomly with certain number of degrees with modifying the object position in frame.

3.2 Feature Extraction using Deep CNN

Convolutional Neural Network- CNN based deep learning approach which comprised with convolutional layers. The feature maps are extracting through convolutional layers from the image using various numbers of kernels. The pooling layers are comes followed by that which reduces these dimensions. Different pooling layers are also presented such as average pooling and max pooling. The network comprised with more layers like batch normalization layer, dense, dropout and flattern layers. The CNN architecture framework is shown below in fig.2. It is applied predominantly in widespread data like videos and images. Between the convolutional and fully connected layers, subsampling layers are presented. CNN can be applied automatically in datasets where relatively greater numbers of parameters and nodes are trained.

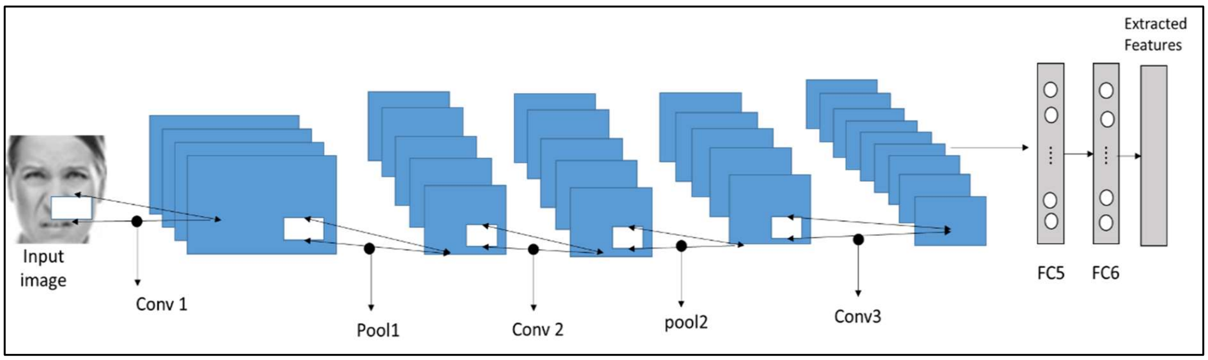


Fig.2. Convolutional Neural Network Architecture

Input layer

From the pre-processed data, the selected features are determined as convolutional neural networks input vector, which are termed as

$$X = [x_1, x_2, x_3 \ldots x_k] \quad (1)$$

From Eq. (1), no. of features per window is k after measurement. For speed up the model convergence, Min-Max Normalization formula is used in which the values in every dimension of data are transformed linearly and normalized to [0,1] range,

$$x = \frac{x-min}{max-min} \quad (2)$$

From Eq. (2), every column's minimum is min, whereas the maximum value is max.

Convolutional layer

The jth feature map output on first convolutional layer's ith unit, refer as,

$$x_i^{l,j} = \sigma\left(b_j + \sum_{a=1}^{m} w_a^j \, x_{i+a-1}^{l-1,j}\right) \quad (3)$$

From Eq. (3), for jth feature map, bias term is b, the kernel size is m, jth feature map weight is w_a , σ is the activation function, and the filter index is a. Here ReLu activation function has used and verified based on faster efficiency and training speed which is similar to the synaptic efficacy of human nervous system.

Max pooling layer

This type of pooling layer develops the combination statistics for the nearby output of $x_i^{l,j}$, can minimize the output's sensitivity and dimension which can attains preservation of scale-invariant feature. Maxpooling operation is utilized in this research. It can divides the convolutional layer output feature into multiple partitions, further identifying the maximum in every partition. The max-pooling layer output is,

$$x_i^{l,j} = max_{n=1}^{r}(x_{(i-1)*T_n}^{l-1,j}) \tag{4}$$

From Eq. (4), pooling size is n and pooling stride is T.

Training model

Deep CNN comprised with several pooling layers, fully connected layers and convolutional layers. After the pooling, activation function and convolution operations, the input data has mapped into the feature space of hidden layer. The fully connected layer combines finally the various local structural features which are learned from below layer for final classification performance. This research used only the pair of max-pooling and convolutional layer which are then flatten 2D data into 1D data, the neural network completed with fully connected layer. The soft-max classifier is utilized as model's top layer which can recognize four kind of transportation mode,

$$f(x) = \mathrm{argmax}_c\left(\frac{e^{x^{l-1}w_j}}{\sum_{n=1}^{N} e^{x^{l-1}w_n}}\right) \tag{5}$$

From Eq. (5), class label is c, index layer is I, no. of class is N, sample feature is x. With respect to Eq. (3) to (5), forward propagation has processed. From the input layer, information is propagated forward and moves through hidden layer and then to output layer which can attains the entire network output. an iteration of forward propagation yields network error value. The cost function of cross-entropy has adopted for measure the error value,

$$L(y) = -\frac{1}{n}\sum_x[ylna + (1-y)\ln(1-a)] \tag{6}$$

From Eq. (6), sample is x *, network predicted value is* a*, actual value is* y*, and no. of training samples is* n. The gradient descent algorithm has employed during the training process for parameters adjustment as bias b and weight w of layer by layer with respect to error backward-propagations. Specifically, the weight w_i is adjusted by informative gain and measures every parameter's adaptive learning rate. The procedure takes lesser memory and computationally effective. The above description is mentioned in following algorithm,

Algorithm 1

Input: $Depression\ Images$, $Output : Features$

Sliding Window Process

ef ← Extract Shadow Features

Normalize sf by equation (2)

regularization feature data, size = 64 Units − 128 Units

repeat:

Forward Propagation

cd ← Convolution2D(ef);

mp ← $\mathrm{Max}_{\mathrm{pooling(cd)}}$;

fc ← Fully_connected(mp);

class label ← relu(fc);

Backward Propagation:

conduct backward propagation with Adam;

Until wi convergences;

Use the trained network to predict the features

Batch Normalization

Batch normalization is a layer which permits every network layer to perform learning that is more independent. It is utilized to normalize the previous layer's output. in normalization, the activation function scales the input layer. Learning becomes so effective using batch normalization and it can be utilized as regularization to avoid model over fitting. The layer is included to sequential model for standardize the outputs or inputs. At several points, it can be used in between the layers. After the sequential model defining it has placed and after the pooling and convolutional layers.

Dropout layers

Dropout is considered as regularization approach which is used for over fitting prevention in model. Dropout are included to randomly switching certain neurons percentage in network. The outgoing and incoming connections are switched off when neurons are switched off. The model learning has enhanced further. After convolutional layers dropout layers are not in use and they used after network dense layers. It's always better to only switch off the 50 percent neurons.

Flattern layer

Flattern layer has converts the data into 1D array for next layer input. The convolutional layers output is flattern for creating single long feature vector. Further, it has connected to final classification model refer as fully connected layer.

3.3 Classification using Weight updated Informative Gain Random Forest algorithm

Here using the improvised random forest algorithm, the sample dataset FER13 in M- dimensional feature space of X and no. of good trees R are selected. From these R trees the no. of un-correlated good trees $Q-U-Corr$ are chosen. The algorithm to build an improvised random forest with X using $Q-U-Corr$ follows following steps,

Step 1: initial step is the data sampling which uses bagging method for generating the value of k in group of data subsets, $\{ BAGM_1, BAGM_2, \dots, BAGM_k \}$, using the random sample of dataset with replacement.

Step 2: the second step is the building of tree classifier. It uses each in-group of data subset $BAGM_i$ for building a tree and further provides the assessment value to tree. Continue all the steps until all trees are created and processed.

Step 3: the third step is tree ordering, which can sort all these k trees in descending order.

Step 4: For variable subset, calculate the Gini Index in information gain. For faster processing, generate the vectorized format of informative gain. The Gini index is used for to evaluate the attribute goodness to split in which the attributes with greatest Gini index value is selected as split in that node. Breiman who introduced Gini index as a measure of computing the data impurity(Nguyen et al., 2011). The following Eq. (7) shows the measurement of Gini index,

$$GI(K) = \frac{1}{2} - \left[1 - \sum_j p\,(r_f)^2\right] \quad (7)$$

From the Eq. (7), it shows that r_f is the relative frequency at K node. From this expression, it shows that the samples are on similar category with 0 impurity or else it shows positive value. The instances are classified with respect to classifications' majority votes from individual trees.

Informative gain is considered as entropy based algorithm determined by Shannon for feature selection(Shannon, 1948). It is used for entropy loss calculation. Entropy is the uncertainty measure in random variable and it is measures in bits which are describes the sample arbitrary collection' impurity. If there is CLC collection with P outcomes then the entropy of CLC is,

$$Entropy(CLC) = -\sum_{I\in P} r\,(I)\, log_2 r(I) \quad (8)$$

From Eq. (8), $r(I)$ is the proportion of $CLC \in I$ class. As discussed, entropy is the feature impurity measure. If there is 0 entropy value then all CLC instances belong to same class. For the A feature of sample set CLC, the informative gain CLC, A is expressed as,

$$Gain(CLC, A) = -Entropy(CLC) - \sum_{v\in A}\left(\frac{|CLC_v|}{CLC}\right). -Entropy(CLC) \quad (9)$$

From Eq. (9), CLC_v is considered as CLC subset in which A feature has v value. CLC_v is the no. of instances classified as CLC_v *and* CLC is no. of instances classified as CLC.

Step 5: Higher performance trees selection, choose the above R trees with higher AUC values.

Step 6: Building the Weight updated Informative Gain Random Forest algorithm: the correlations among R trees predicted probabilities are observed,

$$t = \begin{bmatrix} 1 & t(1,2) & t(1,3) \\ \vdots & \vdots & \vdots \\ t(t,1) & t(t,2) & 1 \end{bmatrix} * Gain(CLC, A) \quad (10)$$

From Eq. (10), observed t is utilized as an input in process of variable clustering and $Q-U-Corr$ correlated the clusters trees. In every cluster, the trees are sorted in ascending order. Select tree from every cluster with greater AUC value which provides $Q-U-Corr$ un-correlated higher performing trees and arrange the trees into Weight updated Informative Gain Random Forest.

The algorithm for Weight updated Informative Gain Random Forest is mentioned below,

Algorithm 2

1. Inputs: TC, NT, TR, NI
   [TC = Number of target classes (in our case 5)
   NT = Number of *R trees* required in decision making
   TR = Training dataset
   NI = Number of instances in TR]
2. For j = 1 to TC (*R trees* are obtained for each class)
3. *R trees* $_{\text{Init(POP)(Population Initialization)}}$
4. Init$_{\text{Weights(W)}}$(Weight Initialization as Wi $=\frac{1}{N}$for each instance i)
5. For k = 1 to NT (*R trees* evolved using Gini Random Forest)
   i. Fill POP. New set of *R trees*
   ii. Evolving GP string capable of recognizing Class TCj, keeping AUC as
   iii. fitness function

6. end For
7. end For
8. (NT $*$ TC) $R\ trees$ are acquired for every class TCj
9. Accumulative classification result of each class Cj is calculated by weighted sum of $t\ and\ output_t\ of\ the\ Images$

From the above algorithm, it shows that the informative gain can extends weight updation support over the instances, since can be not correctly classified in the before iteration. Further it can saves time for introducing the new Weight updated Informative Gain Random Forest, the whole new population is not generated rather updated weights over instance space seen in following iteration. Based on higher value the final test instances prediction is made from weighted sum of outputs.

4. Results and Discussion

The following section illustrates the proposed DCNN-WiGRF algorithm results for predicting the depression disorder types with respect to various performance metrics such as accuracy, precision, F1-score and Recall values.

4.1 Dataset Description

The depression dataset for this research is obtained from https://www.kaggle.com/msambare/fer2013. Here, the data comprised with pixel value as 48x48 grayscale face images of Normal, abnormal, persistent, Major Depressive Disorder- MDD, and bipolar disorder.

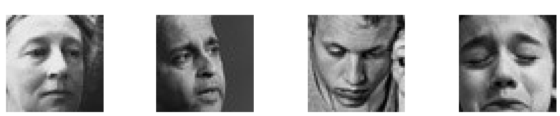
Table 1: Sample Depression disorder images

Normal

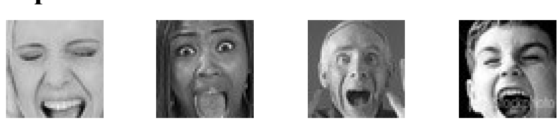
Abnormal

MDD

Persistent disorder



Bipolar Disorder



The face is nearly centred and resides in the space with similar amount in every image. The training set comprised with examples of 21010 and 2342 examples of test set. The sample images in the dataset are shown in below table 1. The following table-2 depicted the additional sample features which are added to images for every patient and it considered as queries which are to be enquired.

Table 2: Additional sample features added to the images for every patient- queries

Lack energy	**ofFeeling hopeless**	**Delusions**	**Mood swings**	**Trouble concentrating**	**Manic Episode**	**Over Energized**	**Over Happy**
0	0	1	0	1	1	0	0
1	1	0	0	0	0	1	0
1	0	0	1	1	0	1	1
0	0	0	0	1	0	1	0
0	1	0	0	1	0	1	1
1	1	1	1	0	1	1	1
1	0	0	0	0	0	1	1

4.2 Experimental analysis

The features which are extracted before sampling and after sampling are described in following table 3. Sampling is a random module with inbuilt function, returns specific items' length selected from sequence list. Here the abnormal values are increased after the sampling procedure and the rest are remaining same.

Table 3: Features- Before and after sampling

Sampling	**Normal**	**Abnormal**	**MDD**	**Persistent disorder**	**Bipolar Disorder**
Before Sampling	5245	436	3995	4830	4097
After Sampling	5245	5245	3995	4830	4097

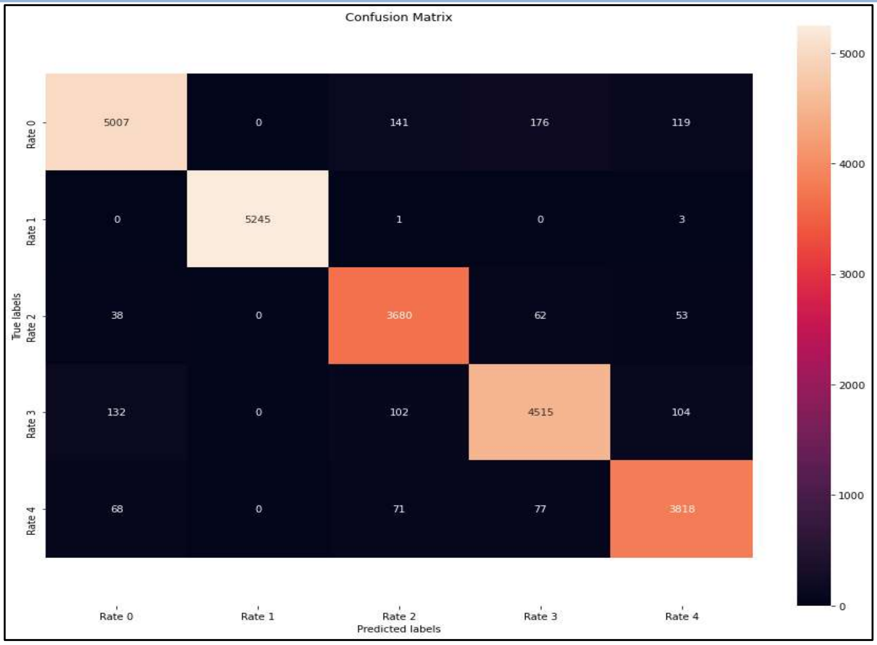
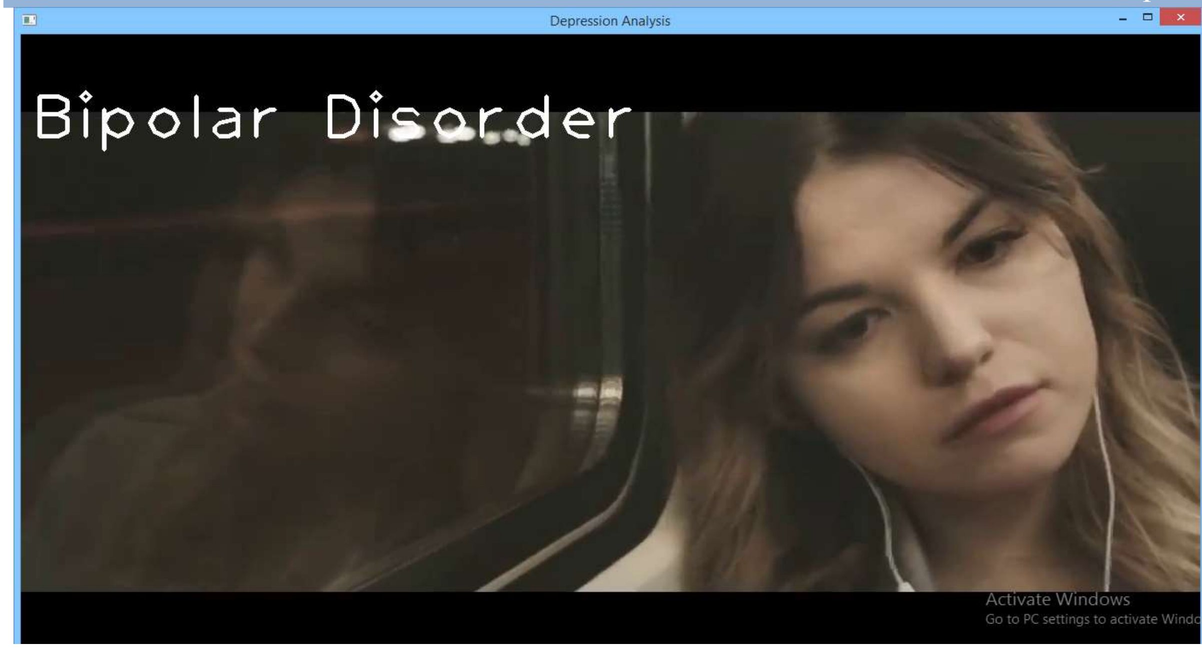


Fig.3. Confusion matrix

The confusion matrix is defined as table which are used to explain the classification model performance on set of testing data in which the true values are identified. The above fig.3 is the confusion matrix of proposed classifier with respect to test data. Followed by the predicted video results are shown in table 4.

Table 4: Predicted Depression types Results



The experimental results of before and after sampling the data is shown in following fig.4 with respect to various performance metrics such as accuracy, precision, recall and F1-score. Here after sampling the accuracy, precision, recall, and F1-score values obtained as 0.9516, 0.9521, 0.9518, and 0.9587 values respectively. However, before sampling the accuracy, precision, recall, and F1-score values obtained as 0.8745, 0.8698, 0.8612, and 0.8736 respectively.

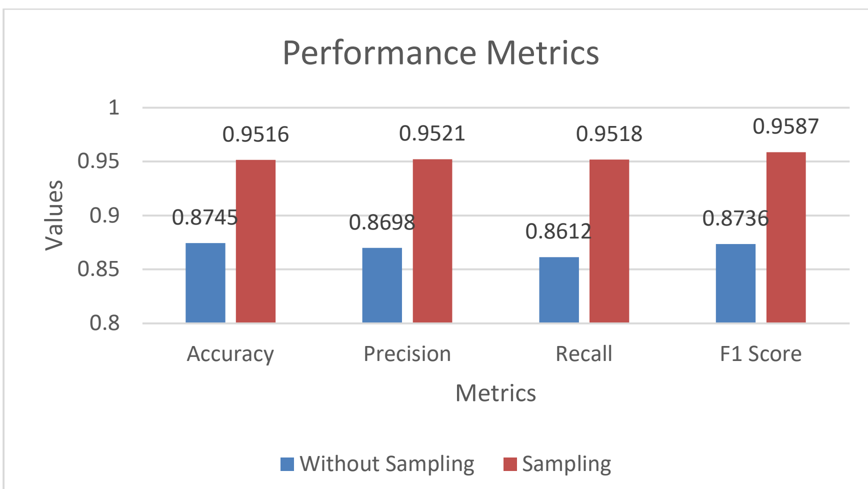


Fig.4. Experimental Results- Before and after sampling

4.3 Internal comparison

For every class of normal, abnormal, MDD, persistent disorder, and bipolar disorder, the proposed DCNN-WiGRF algorithm with respect to performance metrics precision, recall and F1-measure values' results are shown in fig.5. Here, in case of abnormal prediction all the metrics rates are high, followed by normal, persistent disorder, bipolar dis-order and MDD class values.

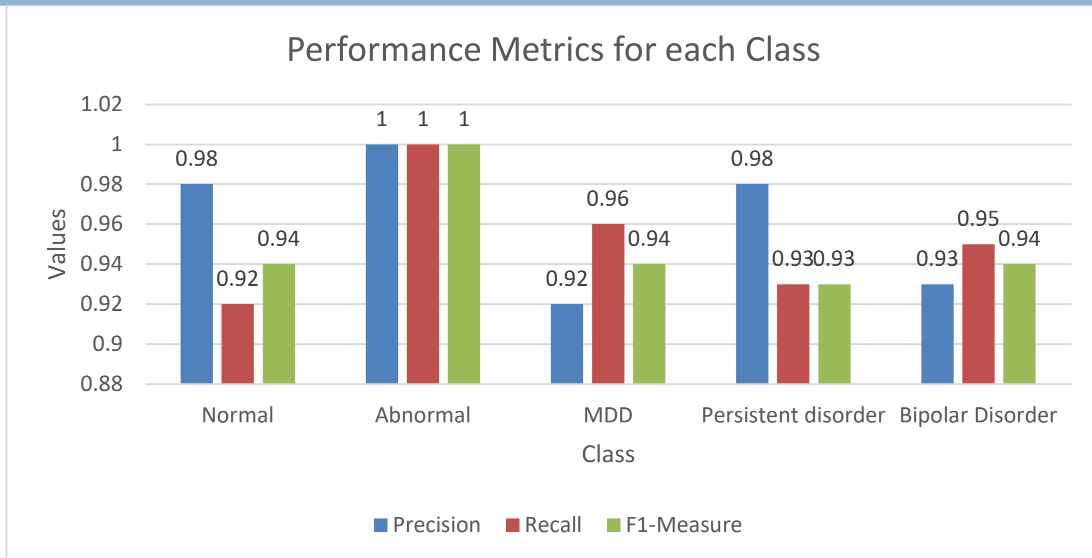


Fig.5. Depression class analysis results

Furthermore, the proposed DCNN-WiGRF algorithm results are compared with existing methods of Random Forest, SVM, K-Nearest Neighbours, Logistic regression, and Naïve Bayes algorithms shown in fig.6. The proposed method shows better results compared with the existing methods in terms of all metrics as accuracy as 98.9, precision as 98.2, recall as 98.5 and F1-score as 98.3 values. Here random forest, SVM, logistic regression shows lesser values compared with other algorithms.

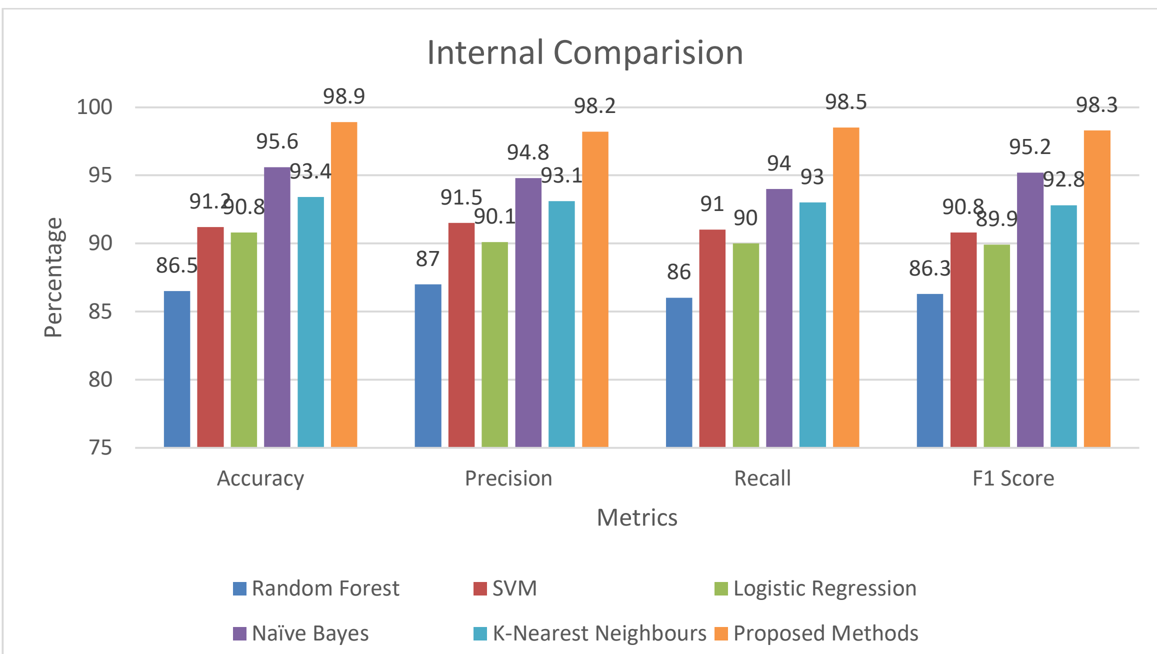


Fig.6. Internal comparison of proposed and existing methods

The following fig.7. Describes the proposed DCNN-WiGRF algorithm accuracy in depression disorder prediction as 98.9% which are compared with existing Support Vector Machine-SVM algorithm shows the average recognition accuracy as 63% value. The below fig.8. Clearly indicated the ROC curve of multiclass types of depression disorder.

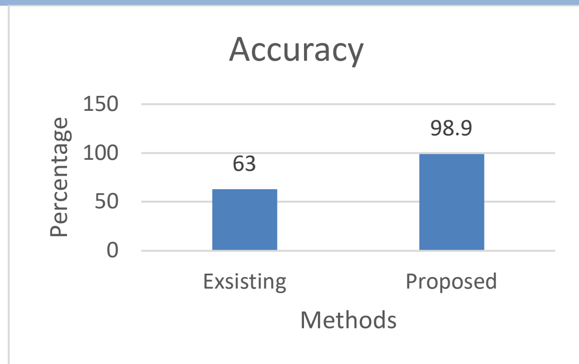


Fig.7. Comparative results- Accuracy

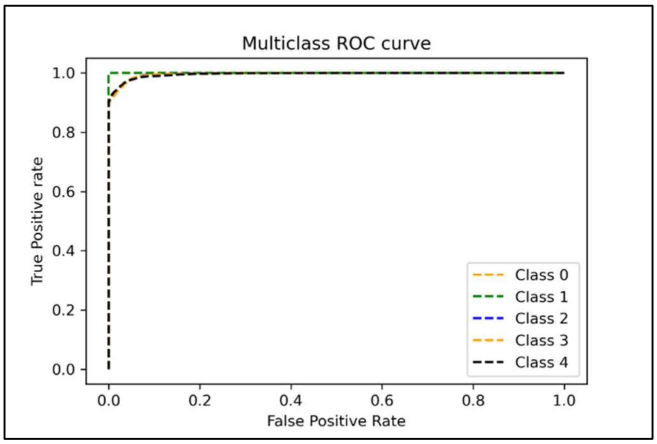


Fig.8. Multiclass ROC curve

5. Conclusion

To detect the depression disorder and its types automatically an effective computational intelligent approach is required. For that this research proposed DCNN-WiGRF algorithm for detecting the depression exactly from the FER2013 depression dataset. The proposed algorithm is effectively performed in extracting the significant features initially, further the proposed Weight updated Informative Gain Random Forest algorithm is classify the significant depressed features accurately. The proposed algorithm outperformed the existing algorithms such as Random Forest, SVM, K-Nearest Neighbours, Logistic regression, and Naïve Bayes algorithms in terms of accuracy, precision, recall and F1-score. Further when compared with existing studies the proposed algorithm shows better accuracy value as 98.9%. Thus, the depression prediction model can be adopted, which can contribute to classify the depression disorder types in order to provide earlier treatment.

Statements and Declarations

Acknowledgement

None

Conflict of Interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

All authors contributed to the study conception and design. All authors read and approved the final manuscript.

Data Availability

There is no Data Availability

References

Andersson, S., Bathula, D. R., Iliadis, S. I., Walter, M., & Skalkidou, A. (2021). Predicting women with depressive symptoms postpartum with machine learning methods. *Scientific reports*, *11*(1), 1-15.

Bhakta, I., & Sau, A. (2016). Prediction of depression among senior citizens using machine learning classifiers. *International Journal of Computer Applications*, *144*(7), 11-16.

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., . . . Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry*, *3*(3), 243-250.

Depression, W. (2017). Other common mental disorders: global health estimates. *Geneva: World Health Organization*, *24*.

Gao, S., Calhoun, V. D., & Sui, J. (2018). Machine learning in major depression: From classification to treatment outcome prediction. *CNS neuroscience & therapeutics*, *24*(11), 1037-1052.

Hantach, R., Lechuga, G., & Calvez, P. (2021). Key Information Recognition from Piping and Instrumentation Diagrams: Where We Are? International Conference on Document Analysis and Recognition,

Hatton, C. M., Paton, L. W., McMillan, D., Cussens, J., Gilbody, S., & Tiffin, P. A. (2019). Predicting persistent depressive symptoms in older adults: a machine learning approach to personalised mental healthcare. *Journal of affective disorders*, *246*, 857-860.

He, L., Chan, J. C.-W., & Wang, Z. (2021). Automatic depression recognition using CNN with attention mechanism from videos. *Neurocomputing*, *422*, 165-175.

Hochman, E., Feldman, B., Weizman, A., Krivoy, A., Gur, S., Barzilay, E., . . . Lawrence, G. (2021). Development and validation of a machine learning-based postpartum depression prediction model: A nationwide cohort study. *Depression and anxiety*, *38*(4), 400-411.

Islam, M., Kabir, M. A., Ahmed, A., Kamal, A. R. M., Wang, H., & Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. *Health information science and systems*, *6*(1), 1-12.

Jones, P. J., Mair, P., Riemann, B. C., Mugno, B. L., & McNally, R. J. (2018). A network perspective on comorbid depression in adolescents with obsessive-compulsive disorder. *Journal of anxiety disorders*, *53*, 1-8.

Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. *Annual review of public health*, *34*, 119-138.

Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Cai, T., . . . de Jonge, P. (2016). Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Molecular psychiatry*, *21*(10), 1366-1371.

Kumar, S., & Chong, I. (2018). Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states. *International journal of environmental research and public health*, *15*(12), 2907.

Lin, E., Kuo, P.-H., Liu, Y.-L., Yu, Y. W.-Y., Yang, A. C., & Tsai, S.-J. (2018). A deep learning approach for predicting antidepressant response in major depression using clinical and genetic biomarkers. *Frontiers in psychiatry*, *9*, 290.

Meshram, P., & Rambola, R. K. Diagnosis of depression level using multimodal approaches using deep learning techniques with multiple selective features. *Expert Systems*, e12933.

Nguyen, C., Nguyen, H., & Yong, W. (2011). Improving Classification Accuracy in Random Forest by Using Feature Impurity and Bayesian Probability. Proceedings on the International Conference on Artificial Intelligence (ICAI),

Priya, A., Garg, S., & Tigga, N. P. (2020). Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Computer Science*, *167*, 1258-1267.

Richter, T., Fishbain, B., Markus, A., Richter-Levin, G., & Okon-Singer, H. (2020). Using machine learning-based analysis for behavioral differentiation between anxiety and depression. *Scientific reports*, *10*(1), 1-12.

Shah, F. M., Ahmed, F., Joy, S. K. S., Ahmed, S., Sadek, S., Shil, R., & Kabir, M. H. (2020). Early depression detection from social network using deep learning techniques. 2020 IEEE Region 10 Symposium (TENSYMP),

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, *27*(3), 379-423.

Shetty, N. P., Muniyal, B., Anand, A., Kumar, S., & Prabhu, S. (2020). Predicting depression using deep learning and ensemble algorithms on raw twitter data. *International Journal of Electrical and Computer Engineering*, *10*(4), 3751.

Uddin, M. Z., Dysthe, K. K., Følstad, A., & Brandtzaeg, P. B. (2022). Deep learning for prediction of depressive symptoms in a large textual dataset. *Neural Computing and Applications*, *34*(1), 721-744.

Weinberger, A. H., Gbedemah, M., Martinez, A. M., Nash, D., Galea, S., & Goodwin, R. D. (2018). Trends in depression prevalence in the USA from 2005 to 2015: widening disparities in vulnerable groups. *Psychological medicine*, *48*(8), 1308-1315.

Yang, L., Jiang, D., Xia, X., Pei, E., Oveneke, M. C., & Sahli, H. (2017). Multimodal measurement of depression using deep learning models. Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge,

Zandvakili, A., Philip, N. S., Jones, S. R., Tyrka, A. R., Greenberg, B. D., & Carpenter, L. L. (2019). Use of machine learning in predicting clinical response to transcranial magnetic stimulation in comorbid posttraumatic stress disorder and major depression: a resting state electroencephalography study. *Journal of affective disorders*, *252*, 47-54.

Zhou, X., Jin, K., Shang, Y., & Guo, G. (2018). Visually interpretable representation learning for depression recognition from facial images. *IEEE Transactions on Affective Computing*, *11*(3), 542-552.