Enhancing Depression Prediction Accuracy Using Filter and Wrapper-Based Visual Feature Extraction

Suresh Mamidisetti [1,*] and A. Mallikarjuna Reddy [2]

[1] Department of Computer Science and Engineering, Anurag University, Hyderabad, Telangana, India
[2] Department of Artificial Intelligence (AI), Anurag University, Hyderabad, Telangana, India;
Email: mallikarjunreddycse@cvsr.ac.in (A.M.R.)
*Correspondence: sureshmamidisetti@gmail.com, 20eg305103@anurag.edu.in (S.M.)

***Abstract*—The pressing need for Artificial Intelligence (AI) applications in healthcare is evident, particularly in the context of depression prediction. Literature underscores the significance of visual cues as crucial indicators of depression. Primary objective of this work is to design a complete machine learning pipeline for more accurate depression prediction, which includes several stages like: data collection stage, feature extraction stage, feature selection stage, classification stage, and performance evaluation stage. Data collection involved video recording of participants (*n* = 219) while conducting emotion elicitation (triggering emotions by showing photos/videos) to depressed and non-depressed subjects. Then, numerous visual features like geometrical features and facial action unit features were extracted. Filter and Wrapper Feature Selection (FS) methods were used to extract the optimal feature set from high-dimensional visual features. In the Filter method, experiments are conducted using three strategies: quasi-constant strategy, mutual information gain, and linear discriminant analysis. In the wrapper method, experiments are conducted using three strategies: forward selection, backward elimination, and recursive feature elimination. Accuracy for the classification of non-depressed or depressed subjects was used as the performance metric. Obtained results with an accuracy of 85.6% show that the backward elimination approach (even though only ten features were selected) outperformed other experiments conducted in current work and also with the state-of-the-art methods. In addition to this, our method is also applied to publicly available benchmarking dataset to show its effectiveness on diverse dataset. These findings demonstrate the applicability of visual features using filter and wrapper feature selection method is reliable in depression prediction. Hence implications extend to a potential application in mental health assessment.**

***Keywords*—health care, artificial intelligence, depression detection, visual features, emotion elicitation, feature selection**

I. INTRODUCTION

Depression is a mental ailment indicated by the presence of constant sadness for at least two weeks, resulting in difficulty in performing daily activities. According to the World Health Organization, depression is the primary global cause of disability. In 2015, this mental disorder was estimated to affect over 320 million people. In the worst scenario, depression can lead up to the extent of suicide. Every year, more than 0.8 million individuals commit suicide, making it the second biggest cause of death [1]. The recent COVID-19 pandemic is another cause for a rapid surge in the incidence of depression, across the globe [2]. Although depression has adverse consequences, early diagnosis with the appropriate treatment can result in reducing its effects or even reversal can be possible [3]. With suitable counseling methods depression in the clients can be treated completely and they can be made to lead a better life, in future without being subjected to similar depressions.

Traditional methods for the diagnosis of depression include self-reports and clinician consultation, both of which are subjective assessments. Standard self-reports like Patient Health Questionnaire-9 (PHQ-9) [4] can be biased and depends upon how seriously an individual has reported its own mental status. Clinician consultation is dependent on availability and experience, and it might be [5] as well costly. To overcome these limitations, automatic depression diagnosis can be a promising solution for the quick and accurate diagnosis of depression by measuring the symptoms objectively [6].

Recently, tools/methods for automatic depression diagnosis based on artificial intelligence have drawn the attention of researchers. Methods based on Machine learning gained popularity to infer significant patterns from data of different modalities [7]. Data collected from various modalities like images, audio, text used in social media, smartphone usage, etc., are used to understand the variations in behavioral patterns between the depressed and non-depressed subjects [8–10]. Among all these modalities, visual manifestations are more prominent indicators of depression. Researchers have found that a high association exists between visual manifestations of a person and the presence of depression. Variations between the depressed and non-depressed subjects in

Manuscript received June 23, 2023; revised August 5, 2023; accepted August 30, 2023; published December 19, 2023.

For instance, depressed people exhibit fewer smiles, higher recurrent lip presses, higher mouth corners slanted downwards, etc. [11].

This study aimed to classify subjects into non-depressed and depressed using an end-to-end machine learning pipeline [12]. This pipeline includes several sequential stages such as data collection stage, feature extraction stage, feature selection stage, classification stage (using ML classifiers), prediction stage, performance evaluation stage. Some of the widely applied Machine Learning (ML) classifiers viz. Support Vector Machine (SVM), Random Forest (RF) is used in the present study to achieve high level of robustness and to improve accuracy in model predictions [13].

The major contributions of this work are listed below:

- Designing a complete machine learning pipeline solution to determine the subjects whether non-depressed or depressed.
- Creation of a visual data set using a well-known psychological experiment called emotion elicitation where facial emotions are triggered by presenting different kinds of happy/neutral/sad videos to the non-depressed and depressed subjects.
- Extraction of facial features such as geometric features, facial action unit features, and region unit features etc.
- Conducting a comparative study with two techniques of feature selection (namely wrapper method and filter method) to identify the subset of feature vectors for more accurate classification.
- Classification of non-depressed and depressed subjects with the individual feature vectors, reduced feature sets and combination of these two.

The remaining sections of this paper are arranged as follows: few related works are discussed in Section II. The dataset and the methodology adopted are described in Section III. Section IV presents the results obtained in the current study and related discussion is included. Finally, Section V presents the conclusion and limitations of the work as well as the scope for future potential.

II. Literature Review

Decades ago, researchers found that depressed people tend to exhibit fewer facial activities and reduced physical movements because of impaired cognitive function. This condition is called psychomotor retardation [14]. The clinician widely uses psychomotor retardation as a symptom for the diagnosis of depression. Researchers in computer vision, aim to investigate this behavioral manifestation through specific facial expressions like facial landmarks, facial action units, region units, etc.

Alghowinem *et al.* [15] utilized a clinically validated dataset of recorded interviews of depressed and healthy volunteers by the clinician. They computed several features like eye-activity related features, head pose, movements-related features. They investigated by performing the classification of depressed and non-depressed through individual feature vectors and also by fusing the different facial feature vectors. Their findings showed that combining the different facial features performed better than using individual feature sets in diagnosing depression.

Sumali *et al.* [16] used the recording of interviews between depressed and healthy volunteers with the clinician. These recordings were provided as input to detect 40 landmark coordinates. Several features, such as each landmark's speed, mouth region's area, etc., were computed to form feature vectors. These higher dimensional feature vectors were used to form a subset of feature sets using Least Absolute Shrinkage and Selection Operator (LASSO) feature selection procedure [17]. Then, the SVM classifier achieved the best performance in prediction of depressed and healthy class subjects.

Giannakakis *et al.* [18] have created a framework for identifying stress from facial clues. They recorded facial cues while participants were performing several tasks: self-introduction and paragraph reading, reaction when presented with neutral and unpleasant images, recalling negative stressed emotions from their past and watching neutral/adventure videos. Features extracted include mouth activity, eye-related events, head pose, etc. The feature selection strategies were employed to select robust features to classify participants stressed and neutral states.

Pediaditis *et al.* [19] used facial features with SVM classifier to predict depression and reported an accuracy of up to 73%. Facial expressions are one of the most commonly used cues to infer emotions, and SVM is a popular machine learning algorithm used for classification tasks. Nawaz *et al.* [20] used Electroencephalogram (EEG) signals with Principle Component Analysis (PCA) to predict depression and reported an accuracy of up to 78%. EEG is a non-invasive technique that measures electrical activity in the brain, and PCA is a dimensionality reduction technique that can help improve the performance of machine learning models.

Wang *et al.* [21] captured facial cues while participants are reading a neutral text. They were used to extract 36 facial key points. From these key points, and time series characteristics were obtained to investigate the variations in facial movements between healthy and depressed subjects. Their findings, based on statistical correlation coefficients show that facial prediction models are valid and reliable for the prediction of mental illness.

Byun *et al.* [22] used heart rate variability features with SVM classifier to predict depression and reported an accuracy of up to 74.4%. Heart rate variability refers to the variation in time intervals between successive heartbeats, and it is a widely studied physiological signal in emotion research. Hosseinifard *et al.* [23] used EEG signals with Logistic Regression (LR) classifier to predict depression and reported an accuracy of 83.3%. LR is another popular machine learning algorithm used for classification tasks, and it is often used as a baseline method for comparison with other machine learning models.

Wang *et al.* [24] collected the facial videos from depressed patients and healthy controls as well by showing them positive, neutral and negative images at China mental health hospital, then extracted facial features adopting person specific active appearance model [25] and found that statistical features of the movements of the mouth's corners, eyebrows, and eyes were effective in identifying the symptoms of depression using SVM classifier.

Tadalagi *et al.* [26] applied the LBP descriptor for extraction of features with illumination invariance, face detection using the Viola-Jones algorithm [27] and classification using the SVM to create a comprehensive model for increased accuracy and performance in diagnosing depression. Swati *et al.* [28] used the fisher discriminant ratio and univariate filter approach to select the best feature sets from facial regions, and proposed an Incremental Linear Discriminant Analysis ILDA-based wrapper feature selection method to get better results in video-based depression detection.

Existing research has following drawbacks that have affected the advancement of depression prediction methodologies. Firstly, a significant portion of these studies has heavily relied upon clinician-patient interactions as the primary means of data collection. Second, there has been a scarcity of investigations that delve into the extraction of facial features while simultaneously employing feature selection techniques. Third, the predominant exploration of deep learning methods has often been confined to datasets of small sizes which raised concerns about the applicability of these techniques in larger and more diverse datasets [29]. Lastly, the diversity inherent in demographic characteristics, such as age, culture, and backgrounds has posed a challenge to achieve consistent and uniform outcomes.

In light of these drawbacks, our present study takes a novel approach to address these challenges comprehensively. By creating a dataset that can avoid the necessity for clinician involvement, we pave the way for independent test instance generation without reliance on clinician support. Then instead of following the mainstream of multimodal methodologies, our study delves into the nuances of unimodal techniques with potential of applying a diverse array of feature selection techniques. Then, we leverage the efficacy of machine learning methods with the more modest dataset size. Additionally, our participant selection process meticulously balances gender representation and narrows the age range to 18 to 19, aligning our dataset more appropriately with the characteristics of this specific age group. Hence this work may be more suitable with this age group.

III. Materials and Methods

A. Dataset Description

The dataset consists of visual recordings of 219 volunteer participants (male-109, female-110 and their average age is 18.5) while conducting a task called emotion elicitation. This task is popular way to elicit emotions using photos and video clips in psychology studies [30]. During this task, each participant was presented with three types of video clips (positive/negative/neutral) carefully selected from popular scenes in films. This task is carried out for sparing 12 min for each participant. 54.8 h were spent in total throughout all the recordings. The experimental procedure with the type of clips, its description, and time duration is presented in Table I.

TABLE I. Experimental Procedure in Conducting Emotion Elicitation Task

Type	Description	Total time (min)
Blank display	Blank display in black color	1
Comedy video clips	The prank comedy videos which have received millions of views and positive comments were selected.	3
Blank display	Blank display in black color	1
Neutral video clip	Abstract shapes and a variety of color bars	3
Blank display	Blank display in black color	1
Tragedy video clip	A popular sad scene in which a widow mother prays god to save her life to take care of her children.	3

After this emotion eliciting activity, Patient Health Questionnaire-9 (PHQ-9) forms were also distributed to the participants. PHQ-9 is used to determine the psychological state of the participant. Each participant's mental state was classified using their PHQ-9 scores and given binary labels (depressed = 1; non-depressed = 0). The subjects in the study have been labeled as 101 depressed and 118 non-depressed in the presence of a paid psychologist. The following PHQ-9 score cut-offs have been utilized to examine the individual mental health, which were standardized by Ying [31].

- 0 to 4 score-the participant is categorized as a non-depressed.
- Any score greater than or equal to 5 is labeled as depressed.

B. Proposed Methodology

The current method follows the standard pipeline proposed in the book [32]. Fig. 1 presents the overview of the workflow of methodology in the present work. The video recordings created within the dataset served as the initial input for the pre-processing step. Subsequently, in feature extraction step, relevant features were extracted from the videos. These features were motivated by the features discussed in the related work and few others are hypothesized to distinguish between depressed and non-depressed. In feature normalization step, these features were normalized to ensure they operated within consistent ranges for subsequent analysis. In feature selection step, two methods are implemented. They are filter methods and wrapper methods. These techniques were employed to identify most relevant features within the dataset. Following the feature selection step, the subsequent steps involved the application of various classifiers. These classifiers were utilized to predict whether subject is depressed or non-depressed based on the selected features. Lastly, a performance evaluation step was conducted to analyze these methods. These steps are deeply presented in the following subsections.

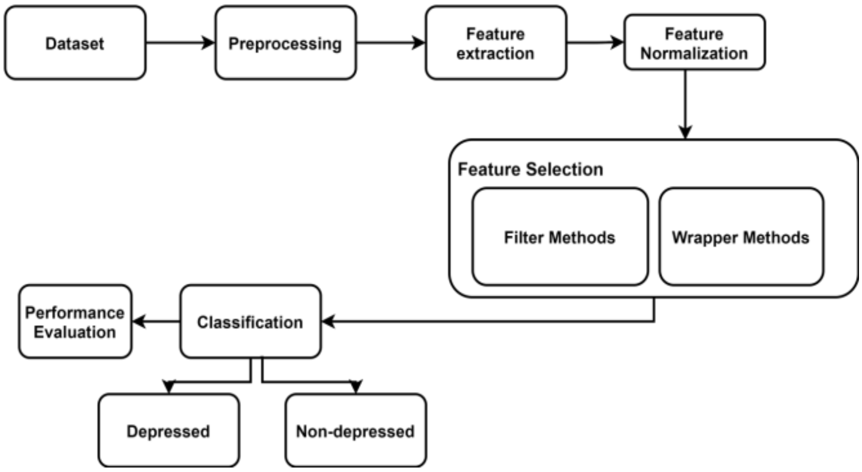

Figure 1. Overview of the workflow of proposed methodology in the present work.

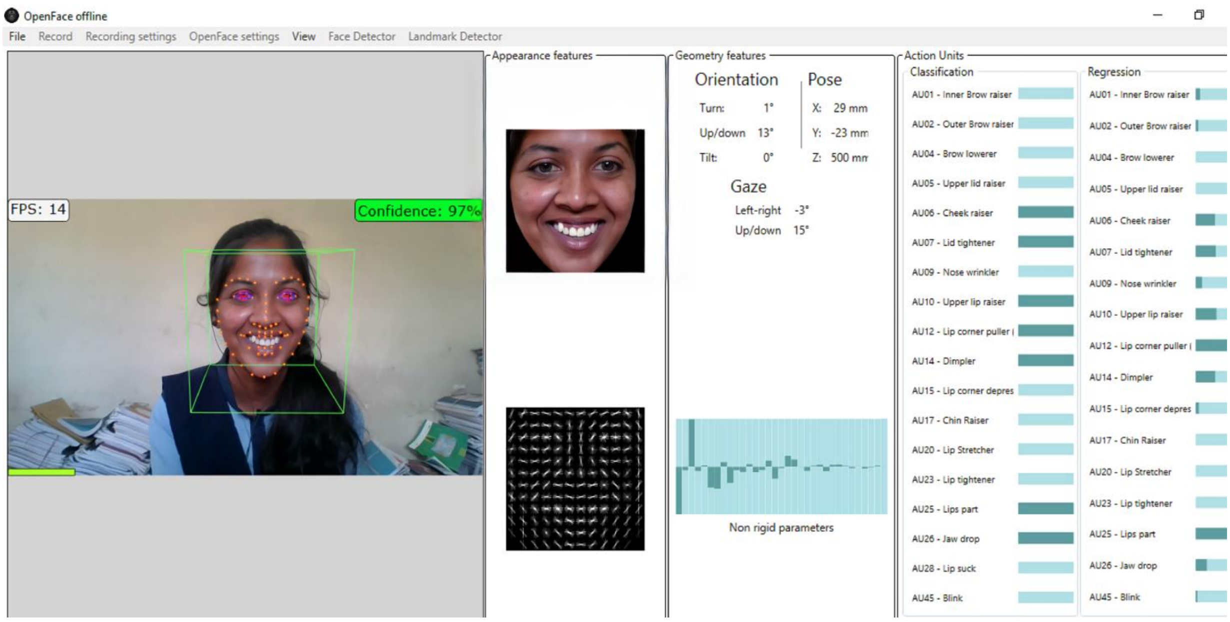

Figure 2. OpenFace toolkit Graphical User Interface for detecting head pose, eye gaze estimation, and facial landmarks (left column), identification of facial action units (right column).

1) Pre-processing

From the video recordings, portion of a clip where either the participant is invisible or someone else appeared in place of the participant, was removed manually. Then, OpenFace toolkit [33] was employed for conversion of the video files into numeric text (.csv) files. OpenFace is the state-of-the-art pre-processing application which generates several low-level features such as 68 facial landmarks including identification of facial action units, tracking of head poses, and estimation of eye gaze, etc., demonstrated in Fig. 2. In the present study, facial Action Units (AU) and facial landmarks are used.

2) Feature extraction

From the related work, it is obvious that depressed people have varied non-verbal behavior when compared with non-depressed people. The features that are extracted in the current study will quantify these variations such as lesser expression (happy/sad) occurrences, reduced zygomaticus activity (a muscle that draws the angle at the mouth to generate smile), veraguth fold (a fold between the eye brows) etc. considering these variations, a total of 279 features were computed, which are detailed in Table II.

The distance, displacement, and region unit's features were extracted using 68 facial landmark locations that are shown in Fig. 3. Facial Action Coding System (FACS) is the comprehensive system for depicting all visually observable facial movements. It divides face expressions into separate units of muscle activity known as Action Units (AUs) [34]. AU's and descriptions are shown in Fig. 4.

TABLE II. DETAILS OF THE FACIAL FEATURES

Type of features	Meaning	Statistical features	# of features
Distance features	Statistical features between pair of landmarks in each frame computed by Euclidean distance were labelled with dis0, dis1, dis2, dis3, dis4, dis5, dis6, and dis7 shown in Fig. 3	minimum, median, maximum, mean, kurtosis, skewness, standard deviation, mode, root mean square of dis0 to dis7	72
Displacement features	Statistical features of movements made by specific landmark in consecutive frames computed using Euclidean distance were labelled with dsp0, dsp1, dsp2, dsp3, dsp4 and dsp5 shown in Fig. 3	minimum, median, maximum, mean, kurtosis, skewness, standard deviation, mode, root mean square of dsp0 to dsp5	54
Region unit features	Statistical features of regions bounded by the left eye, right eye and mouth were computed by area of irregular polygon and labelled with R_1, R_2, and R_0 shown in Fig. 3	minimum, median, maximum, mean, kurtosis, skewness, standard deviation, mode, root mean square of R_0, R_1, and R_2	27
Action Unit (AU) features	To measure the facial muscles' movements, the facial action coding system is used. The occurrences of AU are marked as present (1) or absent (0) for Eighteen Action Units.	median, kurtosis, mean, mode, skewness, standard deviation, root mean square of 18 AUs.	126

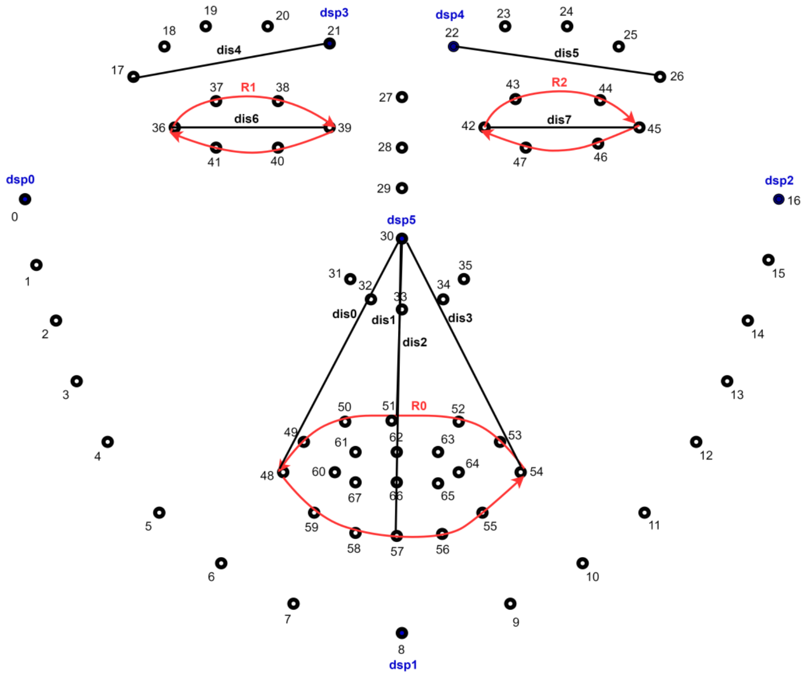

Figure 3. Facial features extraction using 68 landmark locations.

Figure 4. Visualization of few action units in Facial Action Coding System (FACS).

The displacements of six distinct landmark positions (labeled as dsp0, dsp1, dsp2, dsp3, dsp4, dsp5 shown in Fig. 3) using a single coordinate were obtained through Eq. (1):

$$Displacement(p, q) = \sqrt{(p_i - p_{i+1})^2 + (q_i - q_{i+1})^2} \quad (1)$$

where (p_i, q_i) are the coordinates of a landmark from frame i and (p_{i+1}, q_{i+1}) are the identical coordinates of a landmark from frame i+1, the values of i vary from 0 to n-1(the first frame is 0 and the last frame is n-1).

Eight distance measure metrics (labeled as dis0, dis1, dis2, dis3, dis4, dis5, dis6, dis7 shown in Fig. 3), between two pairs of coordinates were calculated using two coordinates through Eq. (2)

$$Distance(p, q) = \sqrt{(p_i - p_j)^2 + (q_i - q_j)^2} \quad (2)$$

where landmarks with two distinct coordinates in the same frame are represented by (p_i, q_i) and (p_j, q_j). For each frame, these distances were determined.

Three areas of the regions bounded by the left eye, right eye and mouth (labeled as R1, R2, R0 shown in Fig. 3), among multiple coordinates in that particular region were evaluated through Eq. (3)

$$RegionArea(p, q) = \frac{1}{2}\left|\sum(p_i q_{i+1} - p_{i+1} q_i)\right| \quad (3)$$

where collection of points in frame i is denoted as (p_i, q_i), (p_{i+1}, q_{i+1}) to (p_{n-1}, q_{n-1}), $i+1$ is expressed as 0 when $i=n-1$.

3) Feature normalization

Feature normalization is a standard technique to bring the values under different scales into a normalized standard form. The features that are extracted belong to different forms. For instance, action unit variations are seen in the scale of 0 to 1 but the region unit features values vary in different ranges depending upon the individual movements. Hence to standardize these values into the similar range and notations, we have performed min-max normalization [35].

4) Feature selection

In the present work, two kinds of feature selection techniques—Filter methods and Wrapper methods have been implemented. In filter methods, features are chosen on the basis of their scores according to the results of statistical tests for their correlation with the output variable. We used three filter-based approaches; (1) Quasi Constant approach; (2) Mutual Information gain; and (3) Linear Discriminant Analysis (LDA). In Features Selection, based on Wrapper method follows a greedy search approach which aims to find the best possible subset of features. We adopted three techniques in Wrapper method, they are: (1) Forward selection; (2) Backward feature elimination; and (3) Recursive feature elimination. The following sub-sections discuss in detail about these feature selection methods.

a) Quasi Constant Method (QCM)

The Quasi-Constant Method (QCM) is a filter method for feature selection that can be used to identify features that have little or no variation across a dataset. The basic idea of the QCM filter is to compute the variance of each feature across the dataset, and discard those features that have a variance below a threshold value. After experimentation we found 99% was helpful which led to improved accuracy in the case of our dataset.

Mathematically QCM filter can be represented as in Eq. (4):

$$z(j)= \begin{cases} 1, & if\ var\big(X(:, j)\big) > \ 99\%\ (Threshold) \\ 0, & otherwise \end{cases} \quad (4)$$

Here, the selected features can be represented as a binary vector z of length p, where $z(j) = 1$ if feature j is selected, and $z(j) = 0$ otherwise, $X(:, j)$ denotes the j^{th} column of the dataset.

b) Mutual information gain

Mutual information gain is a measure of the amount of information that one feature provides about another feature in a dataset. In the context of feature selection, mutual information gain can be used to quantify the relevance of each feature to the target variable, and to select a subset of features that are most informative for prediction.

Mathematically mutual information gain between two features *f_i* and *Y* can be defined as in Eq. (5):

$$I(f\_i;\ Y) = H(Y) - H(Y/f\_i) \quad (5)$$

where *H(Y)* is the entropy of *Y*, and *H(Y/f_i)* is the conditional entropy of *Y* given *f_i*.

The entropy of *Y* is defined as in Eq. (6):

$$H(Y) = -sum\left(p(y) \times log2\left(p(y)\right)\right), y\ in\ Y \quad (6)$$

where *p(y)* is the probability of *Y* taking on the value *y*.

The conditional entropy of *Y* given *f_i* is defined as in Eq. (7):

$$H(Y/f\_i) = -sum(p(f\_i, y) \times log2\ (p(f_i, y)/ p(f\_i))), (f\_i, y)\ in\ (f\_i, Y) \quad (7)$$

where *p*(*f_i*, *y*) is the joint probability of *f_i* and *Y* taking on the values (*f_i*, *y*), and *p*(*f_i*) is the marginal probability of *f_i*.

c) Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) seeks to identify a subset of features that maximally separate the classes in the dataset. LDA gives ranks to the features according to their discriminative power, and select a subset of features with the highest LDA gain. This subset of features can then be used for model training and prediction.

Let *sigma_k* and *mu_k* be the covariance matrix and mean vector of class *k*, respectively.

Then, the LDA gain of the i^{th} feature can be defined as the ratio of the between-class variance to the within-class variance, as shown in Eq. (8):

$$LDA(f\_i) = (mu\_k(i) - mu\_l(i)\hat{}2\ / sigma\_k(i)\hat{}2\ + sigma\_l(i)\hat{}2) \quad (8)$$

where *mu_k*(*i*) and *mu_l*(*i*) are the means of feature *i* in class *k* and class *l*, respectively, and *sigma_k*(*i*)^2 and *sigma_l*(*i*)^2 are the variances of feature *i* in class *k* and class *l*, respectively.

d) Forward selection

Forward selection involves starting with an empty set of features and iteratively adding features to the model, based on their individual performance.

The forward selection process is as follows:

- Start with an empty set of features S.
- Choose a ML classifier A, in our case Random Forest (RF), Support Vector Machine (SVM), Naive Bayes (NB), Logistic Regression (LR), and Decision Tree (DT).
- For each feature f in F:
  a. Add feature f to S.
  b. Train a model M using algorithm A and feature subset S.
  c. Evaluate the performance of M using a performance metric, in our case highest accuracy as parameter.

- Select the feature subset with the best performance and use it for model training and prediction.

e) Backward elimination

Backward elimination involves starting with the full set of features and iteratively removing the least important features, based on their individual performance.

For backward elimination the process is as follows:

- Start with the full set of features S = F.
- Choose a ML classifier A, in our case Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Naive Bayes (NB), and Decision Tree (DT).
- For each feature f in S:
    a. Remove feature f from S.
    b. Train a model M using algorithm A and feature subset S.
    c. Evaluate the performance of M using a performance metric, in our case highest accuracy as parameter.
- Select the feature subset with the best performance and use it for model training and prediction.

f) Recursive feature elimination

Recursive feature elimination involves iteratively eliminating the least important features and recalculating the model's performance, until the required number of features is obtained.

For recursive feature elimination the process is as follows:

- Start with the full set of features S = F.
- Choose a ML classifier A, in our case Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Naive Bayes (NB), and Decision Tree (DT).
- While S has more than k features:
    a. Train a model M using algorithm A and feature subset S.
    b. Evaluate the importance of each feature using a feature ranking method, in our case lasso feature selection is used.
    c. Remove the least important feature from S.
- Select the feature subset with k features and use it for model training and prediction.

5) Classification

Our primary aim is to classify whether a participant is depressed or non-depressed. The current problem is binary classification problem to identify an individual by the normalized optimal features selected from the feature selection techniques. We employed the following state-of-the-art classifiers viz. Random Forest (RF), Support Vector Machine (SVM), Naive Bayes (NB), Logistic Regression (LR), and Decision Tree (DT) for classification. The dataset is partitioned into training set (80%) and test set (20%) using random sampling without replacement. The ML models are built using the training set, whereas test data set is used for performance evaluation.

6) Performance evaluation

The performance metric called accuracy is adopted for evaluation of the current work. Level of correctness for the given model's performance is measured by accuracy performance metric. The reasons for selection of the accuracy performance metric are 1) it is extensively utilized in depression detection literature [36] and 2) it is best suited for the balanced dataset [37].

The performance metric accuracy is defined using Eq. (9). Where TP, TN, FP, and FN denotes quantity of true positives, true negatives, false positives, and false negatives, respectively.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{9}$$

Here TP: The participant who is depressed in reality is predicted to be depressed, TN: The participant who is non-depressed in reality is predicted to be non-depressed, FP: The participant who is non-depressed in reality is incorrectly predicted by the classifier as depressed also referred to as making a Type I mistake, FN: The participant who is depressed in reality is incorrectly predicted by the classifier as non-depressed (this is referred to as a "Type II error").

IV. Result and Discussion

The results of the experiment are described in the current section. The experiments are conducted on a 2.40 GHz Intel core i5 processor with Windows 10 operating system and 16 GB of RAM to evaluate the different model's performances. For the purpose of carrying out the current work, we employed the python programming language as well as libraries like sklearn, numpy, pandas, matplotlib, etc. All the experiments were carried out using the Jupiter python notebook.

The obtained dataset was processed in accordance with the following: (1) individual type of facial feature vectors; (2) concatenation of all the individual type of feature sets; (3) concatenation of the reduced feature vectors obtained after applying wrapper and filter feature selection methods from the individual feature sets. The results of the study are discussed in this section.

Table III lists the accuracies when individual types of facial feature vectors are considered. In this study, default parameters are used to prove that ML classifiers are capable to solve the current classification problem without the need of performing fine tuning. It presents an overview of individual type of facial features that are performed in the current study. It is evident that SVM classifier outperformed the others when using the individual facial features. SVM gave the accuracy of 62%, 64%, 62%, and 65% using distance, displacement, regional unit, and action unit features, respectively. SVM classifier performed better than, naive bayes, random forest, logistic regression and decision tree classifier while using different variant of facial features. From the table, it can also be seen that action unit features performed better over distance, displacement and regional unit features. Using action unit features, all the ML classifiers performed the best when compared over the distance, displacement and region unit features. Action unit features gave an accuracy of 65%, 60%, 63%, 60%, and 62% while using SVM, RF, NB, DT, and LR classifiers, respectively.

TABLE III. ACCURACIES FOR INDIVIDUAL FEATURES

Type of feature (# of features)	ML Classifier Used	Accuracy
Distance features (72)	SVM	62
	RF	56
	NB	60
	DT	58
	LR	60
Displacement features (54)	SVM	64
	RF	55
	NB	53
	DT	60
	LR	57
Regional Unit features (27)	SVM	62
	RF	55
	NB	58
	DT	53
	LR	49
Action Unit features (126)	SVM	65
	RF	60
	NB	63
	DT	60
	LR	62

All the individual type of features is fused to form a concatenated feature facial feature vector. Table IV lists the details of the results obtained when the combination of all the feature vectors is performed. It can be seen that SVM has performed better when the combination of the features has taken place. From this table, it is observed that such combination of features can lead to improvement of the model's performance. SVM gave an accuracy performance of 72% when others gave only 65%, 67%, 65%, and 64% while using RF, NB, DT, and LR classifiers, respectively.

TABLE IV. ACCURACIES WHEN ALL FEATURES ARE FUSED

ML classifierUsed	# of features	Accuracy
SVM	279	72
RF		65
NB		67
DT		65
LR		64

Now, further study has been performed to reduce the dimensionality of the feature vectors which will improve the model's accuracy. Table V lists the details of the current study when dimensionality reduction approaches were performed. Here, there were 279 initial features, and an SVM classifier was used. It can be noted that using wrapper method, resultant feature vector was considered with 10 features. From the conducted study it is hypothesized that ML classifiers are efficient for the classification of non-depressed and depressed subjects using only a minimal feature vector with size of 10.

From Table V, it is derived that SVM outperformed others when using backward elimination approach. The results obtained are optimistic because of the range that is considered. The threshold and the K values in the filter and wrapper methods were derived while model gives optimal results. Rigorous experimentation was conducted to decide the range of these values.

TABLE V. ACCURACIES OBTAINED WHEN APPLIED FEATURE SELECTION BY FILTER AND WRAPPER METHODS

	Variant of feature selection	# of features removed	# of features in resultant Feature vector	Tuned values range	Accuracy
Filter method	Quasi Constant Method	113	166	0.5–0.7	68
	Mutual information gain	174	105	0.5–0.7	78
	Linear Discriminant Analysis	159	120	0.5–0.7	68
Wrapper method	Forward selection	269	10	K=10–15	78
	Backward feature elimination	269	10	K=10–15	**85.6**
	Recursive feature elimination	264	15	K=15–20	79

An alternative method for evaluating the model's performance is using Receiver Operator Characteristic (ROC) curve is plotted in the present study [38]. Fig. 5 shows ROC curve of various classifiers that were employed. ROC curve is a graph plotted between false positive rate and true positive rate defined using Eq. (10) and Eq. (11). Sensitivity and 1-specificity are known as true positive rate and false positive rate respectively. In ROC curve, every data point refers to the sensitivity-specificity pair. From the Fig. 5, it is noted that among all classifiers SVM classifier leans towards the upper right corner. This implies that higher accuracy is achieved by the current depression detection model.

$$FPR = \frac{FP}{FP+TN} \quad (10)$$

$$TPR = \frac{TN}{TN+FP} \quad (11)$$

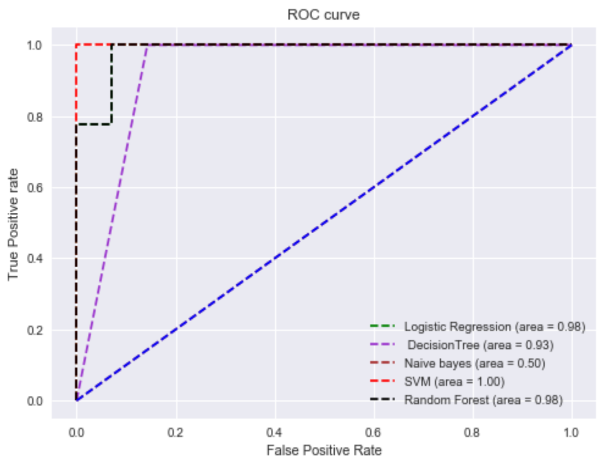

Figure 5. ROC curve of ML classifiers.

A. Comparison of the Proposed Method with the State-of-the-Art Methods

Table VI summarizes the methods and performance metrics used in different state-of-the-art works to predict depression. The proposed method in the table is backward elimination method using wrapper feature selection, and it achieved an accuracy of 85.6%. Backward elimination is a feature selection technique that starts with all features and iteratively removes the least significant ones until the desired level of performance is achieved.

Several findings can be observed form Table VI, They are as follows: our proposed work which has several features followed by feature selection methods gave better performance than using facial features followed by SVM classifier alone in the work [37]. These findings suggest that prevalent feature extraction using feature selection methods helps classifiers for classification of depressed and non-depressed subjects. In the work [20], feature transformation like PCA with the EEG signals were used. This can suggest that feature reduction technique may depend on the modality of the data like feature transformation gave better results using EEG modality and feature selection methods gave better results on visual modality.

Similar method to the current work is experimented in [22]. In this work, Heart Rate Variability (HRV) is used to extract 100 features and then apply recursion based feature elimination to input them to Support Vector Machine (SVM). The advantage over this work in the current method is inexpensive. HRV data needs equipment like fingerprint sensors, ECG machines etc. whereas our method does not need such clinical equipment's for data collection. In the study presented in [39], they used acoustic features and experimented with only few classifiers but lacks to include experiments with multiple classifiers. Our work analyses the performance over several classifiers using cost effective way of data collection.

TABLE VI. COMPARISONS OF THE PROPOSED METHOD WITH THE STATE-OF-THE-ART METHODS

Author and reference	**Method used**	**Performance metric**
Thati *et al.* [37]	Facial Features with SVM classifier	Accuracy up to 69%
Nawaz *et al.* [20]	EEG using Principle Component Analysis (PCA)	Accuracy up to 78%
Byun *et al.* [22]	Heart rate variability features using SVM classifier	Accuracy up to 74.4%
Janardhan *et al.* [39]	Acoustic features of speech using DES classifier	Accuracy of 77.3%
Our proposed method	Backward elimination method using wrapper feature selection	Accuracy of 85.6%

B. Performance of the Proposed Method with Benchmarking Dataset

We conducted tests to demonstrate the versatility of our suggested model in order to assess its broad suitability. Our method not only demonstrated efficacy within our particular dataset, but it also demonstrated potential for application across other datasets. Our methodology was applied to the Distress Analysis Interview Corpus-Wizard of Oz (DAIC-WOZ) [40], a publicly available benchmarking dataset used in the field of depression diagnosis. The DAIC-WOZ dataset includes audio and visual records of interviews between a real-life interviewee and a computer-generated interviewer acting as a supportive listener or a diagnostician. Participants participate in talks on a variety of themes, including personal experiences and emotional states, evoking authentic emotional displays. This well-established dataset is used by academics to develop and validate depression detection methods. Furthermore, its accessibility allows for method comparisons and cross-validation, accelerating research in automated depression detection technologies for real-world applications.

First, all the visual features in the current work are extracted on the DAIC-WOZ dataset and then best performer in the current method i.e., backward feature elimination is applied on these features. Table VII shows the details of the performance of the proposed backward feature elimination method on the DAIC-WOZ dataset. Our method yielded comparable performance parameters when applied on the benchmarking dataset. Among all the performance metrics recall gave a performance of 89.6% using the proposed technique. From Table VII, it can be inferred that the suggested method can be applied to diverse datasets with similar visual cues.

TABLE VII. PERFORMANCE METRICS VALUES OF PROPOSED METHOD ON A DAIC-WOZ DATASET

Performance metric	**Value obtained on DAIC-WOZ dataset (%)**
Accuracy	82.0
Precision	77.73
Recall	89.6
F1-Score	77.4

V. CONCLUSION

The integration of artificial intelligence methods into primary healthcare is an imperative need. Within this context, there arises a pressing requirement for studies that can effectively diagnose depression. Present study aims to design a complete machine learning pipeline for classification of depressed or non-depressed subjects. For this, we included several stages like: Data collection stage, feature extraction stage, feature selection stage, classification stage (using ML classifiers), and performance evaluation stage. In the data collection stage, videos of the participants were recorded while conducting a well-known experiment called emotion elicitation to analyze the visual differences between depressed and non-depressed. Then, numerous visual features like geometrical features, region units and facial action unit features were extracted. Then to obtain an optimal feature subset from the extracted high-dimensional visual features different Feature Selection (FS) techniques were employed. In the current study, we investigated using two FS techniques-Filter method and wrapper method. In the filter method, experiments are conducted using three approaches: quasi-constant approach, mutual information gain and linear discriminant analysis. In the wrapper method, experiments are conducted using three approaches: forward selection, backward elimination, and recursive feature elimination. We analyzed the performance of the classification model which classifies as depressed or non-depressed using accuracy as an evaluation metric. We further analyzed the performance of the model in different ways: with individual feature vectors, reduced feature sets and with their fusion. Among all the experiments conducted, the backward elimination approach (even though only ten features were selected and numbers of participants are only 219) outperformed others with an accuracy of 85.6%. Our method is compared with the state-of-the-methods in depression detection. In addition to this, our method is also applied to publicly available benchmarking dataset called as DAIC-WOZ dataset. Our method achieved comparable performance to demonstrate that our method works efficiently on diverse datasets with similar visual features. Our method outperformed others with significant margin. These findings demonstrate that visual features are predominant in depression classification of depressed and non-depressed subjects. The current work focused only on the visual features.

Our future work is aimed at the classification of depressed and non-depressed using multimodalities including speech and textual cues. We focus on using regression models to classify the levels of depression among the depressed patients like low, moderate, severe levels of depression. In future, we will also aim to longitudinal analysis that could focus on longitudinal data collection, allowing us to track the evolution of depression over time.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

The paper background work, conceptualization, methodology, dataset collection, implementation, preparing and editing original draft have been done by corresponding author Suresh Mamidisetti. Proof reading and supervision, have been done by A. Mallikarjuna Reddy; all authors had approved the final version.

Acknowledgment

This project was supported by students of Government Polytechnic, Hyderabad, India, and the authors would like to thank them for their help.

References

[1] WHO. (2017). Depression and other common mental disorders: Global health estimates. [Online]. Available: https://apps.who.int/iris/handle/10665/254610

[2] WHO. (2021). Mental health preparedness and response for the COVID-19 pandemic: Report by the director-general. [Online]. Available: https://apps.who.int/iris/handle/10665/359717

[3] S. Mamidisetti and M. R. A., "Multimodal depression detection using audio, visual and textual cues: A survey," *NeuroQuantology*, vol. 20, no. 4, pp. 325–336, 2022.

[4] K. Kroenke, R. L. Spitzer, and J. B. W. Williams, "The PHQ-9: Validity of a brief depression severity measure,"*Journal of General Internal Medicine*, vol. 16, no. 9, pp. 606–613, 2001.

[5] S. Guohou, Z. Lina, and Z. Dongsong, "What reveals about depression level? The role of multimodal features at the level of interview questions," *Information and Management*, vol. 57, no. 7, 2020.

[6] A. M. Reddy *et al*., "An efficient multilevel thresholding scheme for heart image segmentation using a hybrid generalized adversarial network," *Journal of Sensors*, vol. 2022, 2022.

[7] S. K. R. Mallikarjuna, A. Reddy, S. K. Reddy, and S. C. N. Kumar, "Leveraging bio-maximum inverse rank method for iris and palm recognition," *International Journal of Biometrics*, vol. 14, no. 3/4, pp. 421–438, 2022.

[8] M. Morales, S. Scherer, and R. Levitan, "A cross-modal review of indicators for depression detection systems," in *Proc. Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*, 2017, pp. 1–12.

[9] S. Mamidisetti and A. M. Reddy, "A stacking-based ensemble framework for automatic depression detection using audio signals," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 7, pp. 603–612, 2023.

[10] P. S. R. P. Thati, A. S. Dhadwal, and P. Kumar, "Multimodal depression detection: Using fusion strategies with smart phone usage and audio-visual behavior," *International Journal on Artificial Intelligence Tools*, vol. 32, no. 2, 2022.

[11] A. Pampouchidou, "Doctoral thesis automatic detection of visual cues associated to depression," *Hal Science*, vol. 11, 2018.

[12] S. Govathoti, A. M. Reddy, D. Kamidi, G. BalaKrishna, S. S. Padmanabhuni, and P. Gera, "Data augmentation techniques on chilly plants to classify healthy and bacterial blight disease leaves," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 6, pp. 131–139, 2022.

[13] V. NavyaSree, Y. Surarchitha, A. M. Reddy, B. D. Sree, A. Anuhya, and H. Jabeen, "Predicting the risk factor of kidney disease using meta classifiers," in *Proc. Second Mysore Sub Section International Conference (MysuruCon)*, 2022, pp. 1–6.

[14] J. S. Buyukdura, S. M. McClintock, and P. E. Croarkin, "Psychomotor retardation in depression: Biological underpinnings, measurement, and treatment," *Prog. Neuropsychopharmacol Biol. Psychiatry*, vol. 35, no. 2, pp. 395–409, 2011.

[15] S. Alghowinem *et al*., "Multimodal depression detection: Fusion analysis of paralinguistic, head pose and eye gaze behaviors," *IEEE Transactions on Affective Computing*, vol. 9, no. 4, pp. 478–490, 2018.

[16] B. Sumali, Y. Mitsukura, Y. Tazawa, and T. Kishimoto, "Facial landmark activity features for depression screening," in *Proc. 58th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, 2019, pp. 1376–1381.

[17] V. Fonti and E. Belitser, "Feature selection using lasso," *VU Amsterdam Research Paper in Business Analytics*, vol. 30, pp. 1–25, 2017.

[18] G. Giannakakis *et al*., "Stress and anxiety detection using facial cues from videos," *Biomedical Signal Processing and Control*, vol. 31, pp. 89–101, 2017.

[19] M. Pediaditis *et al*., "Extraction of facial features as indicators of stress and anxiety," in *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 3711–3714.

[20] R. Nawaz, K. H. Cheah, H. Nisar, and V. V. Yap, "Comparison of different feature extraction methods for EEG-based emotion recognition," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 910–926, 2020.

[21] X. Wang, Y. Wang, M. Zhou, B. Li, X. Liu, and T. Zhu, "Identifying psychological symptoms based on facial movements," *Frontiers in Psychiatry*, vol. 11, no. December, 2020.

[22] S. Byun *et al*., "Detection of major depressive disorder from linear and nonlinear heart rate variability features during mental task protocol," *Computers in Biology and Medicine*, vol. 112, no. August, 103381, 2019.

[23] B. Hosseinifard, M. H. Moradi, and R. Rostami, "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal," *Computer*

Methods and Programs in Biomedicine, vol. 109, no. 3, pp. 339–345, 2013.

[24] Q. Wang, H. Yang, and Y. Yu, "Facial expression video analysis for depression detection in Chinese patients," *Journal of Visual Communication and Image Representation*, vol. 57, pp. 228–233, 2018.

[25] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance and models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[26] M. Tadalagi and A. M. Joshi, "AutoDep: Automatic depression detection using facial expressions based on linear binary pattern descriptor," *Medical & Biological Engineering and Computing*, vol. 59, no. 6, pp. 1339–1354, 2021.

[27] S. Tikoo and N. Malik, "Detection of face using viola jones and recognition using back propagation neural network," *International Journal of Computer Science and Mobile Computing*, vol. 5, no. 5, pp. 288–295, 2017.

[28] S. Rathi, B. Kaur, and R. K. Agrawal, "Selection of relevant visual feature sets for enhanced depression detection using incremental linear discriminant analysis," *Multimedia Tools and Applications*, vol. 81, no. 13, pp. 17703–17727, May 2022.

[29] S. L. V. Papineni, S. Yarlagadda, H. Akkineni, and A. M. Reddy, "Big data analytics applying the fusion approach of multicriteria decision making with deep learning algorithms," *International Journal of Engineering Trends and Technology*, vol. 69, no. 1, pp. 24–28, 2021.

[30] J. J. Gross and R. W. Levenson, "Emotion elicitation using films," *Cognition and Emotion*, vol. 9, no. 1, pp. 87–108, 1995.

[31] Y. L. Zhang *et al*., "Validity and reliability of patient health questionnaire-9 and patient health questionnaire-2 to screen for depression among college students in China, "*Asia-Pacific Psychiatry*, vol. 5, no. 4, pp. 268–275, 2013.

[32] S. Agarwal, "Data mining: Data mining concepts and techniques," in *Proc. International Conference on Machine Intelligence and* Research *Advancement ICMIRA2013*, 2014, pp. 203–207.

[33] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. P. Morency, "OpenFace 2.0: Facial behavior analysis toolkit," in *Proc. 13th IEEE International Conference on Automatic Face & Gesture Recognition*, 2018, pp. 59–66.

[34] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003.

[35] T. Jayalakshmi and A. Santhakumaran, "Statistical normalization and back propagationfor classification," *International Journal of Computer Theory and Engineering*, vol. 3, no. 1, pp. 89–93, 2011.

[36] D. Ramalingam, V. Sharma, and P. Zar, "Study of depression analysis using machine learning techniques," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 7C2, pp. 187–191, 2019.

[37] R. P. Thati, A. S. Dhadwal, P. Kumar, and P. Sainaba, "A novel multi-modal depression detection approach based on mobile crowd sensing and task-based mechanisms," *Multimedia Tools and Applications*, vol. 82, pp. 4787–4820, 2022.

[38] A. Khan and S. Zubair, "An improved multi-modal based Machine learning approach for the prognosis of alzheimer's disease," *Journal of King Saud University Computer and Information Sciences*, vol. 34, no. 6, pp. 2688–2706, 2022.

[39] N. Janardhan and N. Kumaresh, "Improving depression prediction accuracy using fisher score-based feature selection and dynamic ensemble selection approach based on acoustic features of speech," *Traitement du Signal*, vol. 39, no. 1, pp. 87–107, 2022.

[40] J. Gratch *et al*., "The distress analysis interview corpus of human and computer interviews," in *Proc. Ninth International Conference on Language Resources and Evaluation LREC 2014*, Reykjavik, Iceland, 2014, pp. 3123–3128.