



Visual Depression Diagnosis From Face Based on Various Classification Algorithms

Sana A. Nasser ^{a*}, Ivan A. Hashim ^b, Wissam H. Ali ^c

^a University of Technology, Department of Electrical Engineering, Baghdad, Iraq,
316333@student.uotechnology.edu.iq

^b University of Technology, Department of Electrical Engineering, Baghdad, Iraq,
30095@uotechnology.edu.iq

^c University of Technology, Department of Electrical Engineering, Baghdad, Iraq,
30088@uotechnology.edu.iq

Submitted: 20/05/2020

Accepted: 11/07/2020

Published: 25/11/2020

KEYWORDS

Visual depression diagnosis, Depression detection, Facial action coding system, KNN, SVM, PCA, LDA

ABSTRACT

Most psychologists believe that facial behavior through depression differs from facial behavior in the absence of depression, so facial behavior can be utilized as a dependable indicator for spotting depression. Visual depression diagnosis system (VDD) establishes dependents on expressions of the face that are expense-effective and movable. At this work, the VDD system is designed according to the Facial Action Coding System (FACS) to extract features of the face. The key concept of the Facial Action Coding System (FACS) to explain the whole face behavior utilizing Action Units (AUs), every AU is linked to the motion of unique or maybe further face muscles. Six AUs have utilized as depression features; those action units are AUs 4, 5, 6, 7, 10, and 12. The datasets that employed to evaluate the performance of the proposed system are gathered for 125 participants (30 males, 95 females); many of them are among 17-60 years of age. At the final step of the current system, four kinds of classification techniques were applied separately; those classifiers algorithms are KNN, SVM, PCA, and LDA. The outcomes of the simulation indicate that the best outcomes are achieved utilizing the KNN and LDA classifiers, where the success rate is 85%. New classification methods in the VDD system are the key contributions of this research, gather real databases that can utilize to compute the performance of every other VDD system based on face emotions, and choose appropriate features of the face.

How to cite this article: Sana A. Nasser, Ivan A. Hashim, Wissam H. Ali, "Visual Depression Diagnosis From Face Based on Various Classification Algorithms," Engineering and Technology Journal, Vol. 38, Part A, No. 11, pp. 1717-1729, 2020.

DOI: <https://doi.org/10.30684/etj.v38i11A.1714>

This is an open access article under the CC BY 4.0 license <http://creativecommons.org/licenses/by/4.0>

1. INTRODUCTION

Reliable Visual Depression Diagnosis system (VDD) is becoming an extraordinary method to reduce suicides around the world as well as other social issues, particularly after the development of computer vision.

Depressive is a mood disorders that weaken mental health and impact the daily living skills of the casualty, their relationships, and the healthcare system. In the latest survey, the World Health Organization (WHO) expected about 350 million persons globe-wide are influenced of depression [1]. Furthermore, depression is now the fourth-largest reason of illness in the world and is estimated to be the major cause in 2020 (current year) [1]. Additionally, The WHO estimates depression will become the primary reason for suicide in the subsequent 15 years. According to the American Academy of Sociology, about two-thirds of people who commit suicide are depressed at the time of their death, and the risk of suicide is about 20 times higher for people with major depression [32]; as shown in figure 1 which was built using the data from Twenge's work [31], and it demonstrates an investigation of extensive data for the rates of major depressive episodes (MDE) among 100k people in the United States of America for the years 2009-2017. Suicide-related outcomes (suicidal ideas, plans and attempts, and deaths by suicide) increased among young-adults ages 18 to 25 between 2009 and 2017, with smaller and less consistent increases among adults ages 26 and over. A successful cure for depressive disorders was obtained in most cases [2]. Furthermore, a shortage of funding and proficient health suppliers is a significant barrier to the successful detection of depression [1]. Also, present valuation approaches to identifying depression rely on reports and doctors' reviews through the patient through the interview; such tests can vary based on the expertise of the clinician and the procedures used to diagnose them while allowing such inadequate instruments to provide visible symptoms that are good indicators for depression. Therefore, limited knowledge of healthcare suppliers might affect the detection of depression. By a growing incidence of depression, current advancements in computer sight and processing of signals may play a crucial part in helping to overcome these barriers [3]. It must be pointed out that sadness is not depression because everybody has a variety of emotions over days and weeks, typically based on events and circumstances. When disappointed or suffering a loss, the person usually feels sad. Usually these feelings ebb and flow. They respond to input and changes. By contrast, depression tends to feel heavy and constant. People who are depressed are less likely to be acclaim, cheered, or consoled. Those who get well from depression sometimes welcome the probability of experiencing natural sadness again, to have a "bad day," as opposed to a heavyweight on their minds and spirit every day.

Most depression detection systems in the preceding work utilized ineffective classification procedures and the datasets that were gathered in agreed light conditions, and restricted environment. Last but not least, most of the earlier systems were based on a system that cannot operate with an invisible database.

Therefore, the key objectives of this research are to construct a system to diagnosis depression with a minimum cost and high accuracy, in addition to assist specialists in detecting depression before becoming suicidal. This system works with a dataset to be collected in an unregulated setting and with no physical interaction with the human body.

The remainder of the paper is as follows: Section 2 explains the mechanism for VDD system. Section 3 describes the details of the collected dataset. Steps that utilized to build the proposed VDD system are presented in section 4. Section 5 summarizes the performance comparison between VDD system classifiers. Section 6 explains in comparison the performance of the proposed system with previous work in details. Section 7 offers some valuable conclusion derived from this paper.

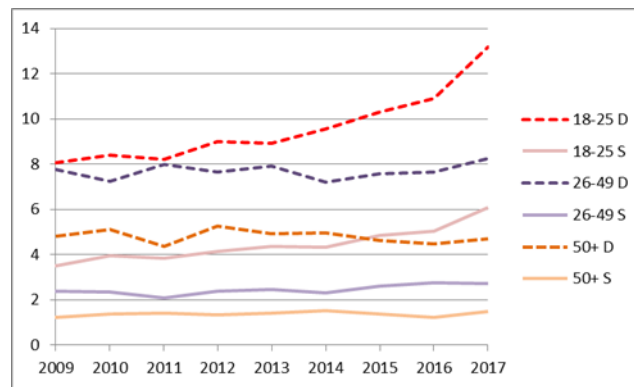


Figure 1: Statistics for depression and its association to suicide in the USA from 2009 to 2017 rated per 100,000 people; where D for depressed and S for suicide.

2. MECHANISM-BASED VISUAL DEPRESSION DIAGNOSIS SYSTEM

A VDD system is essential to detect depressive people without a clinician. Anyone can be diagnosed with depression based on this automated facial recognition system that has studied the signs of depression. Constructing a computer vision scheme for depression recognition built on the expressions of the face can be split for three main steps [4]. The first step is to capture and pre-processes the video; that means preparing the took videos for the next step called the features extraction stage. That means, this stage includes recording video for participant's face through an interview, edit the captured video; recognition of the face, and finding the significant facial landmarks on the images of the face that are required to extract the features. Next, the feature extraction stage is responsible for obtaining the depression features from a series of face pictures. This step utilizes to obtain the face signs which acts for the possible signs of depression. As a final step, the decision-maker is an appropriate classifier utilized to give the subject's status. The block figure of the VDD system utilized video databases as seen in Figure 2; those steps are described in the subsequent sectors.

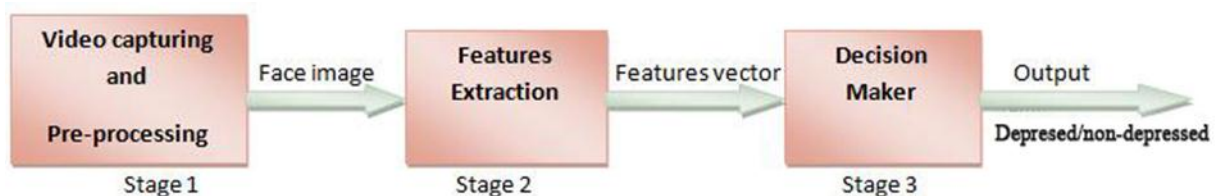


Figure 2: General Block Diagram of Visual Depression Diagnosis Systems

I. Capturing of the Video and Pre-processing

The digital camera is utilized to register the subject's facial expressions through an interview when the interviewer inquires a list of questions; the face of the subject should be on the front of a camera to record all the changes in facial expression. The editing of the video contains just a period once the participant is trying to think before answering the question. In order to provide this video recorded for the following step, it is necessary to detect the face of the subject and to locate those facial landmarks; it is essential to track the parts required to take out the features. The Viola-Jones face detector is the greatest common face detector to date [5]. This can process images especially rapidly, recognize a face with various skin colors, and identify the face with eyeglasses. This face tracker has been utilized in various facial expression detection approaches. The landmark of the face was tracked by relying on the bounding box for the face extracted from the face originator. The Constrained Local Neural Field (CLNF) method utilized in the current work as obtainable as in [6]. This procedure has the ability to diagnosis and path facial landmarks at weak light situations; although in the existence of an obstruction, plus may detect landmarks from an unknown database; the full form for that sample could be represented utilizing 4 different kinds of factors $\rho = [s, R, q, t]$,

these are s : is the scaling factor; $R2D$ for rotation of the object, t for term of the translation, and q for non-rigid shape term. The position of i feature indicated as x_i , which is arranged during those 4 parameters, is depending in on a Point Distribution Model (PDM) which defined in the equation bellow:

$$x = s.R2D.(\bar{x}_i + \phi_i q) + t \quad (1)$$

Where $\bar{x}_i = [\bar{x}_i, \bar{y}_i, \bar{z}_i]^T$: The average value for i feature, where ϕ_i : a $3 \times m$ principal element matrix.

II. Extraction of the Features

Extraction of the features contains extracting several facial features to choose it as a dependable indicator for identifying depression. Several methods are utilized to take out the feature from the face, the best common and automated techniques which can achieve the features from an invisible database based on FACS, those approaches can be found in [7][4][8]. FACS describes facial behavior using AUs that represent the movement of the muscles of the face. This FACS recognizes expressions of the face that depends on the detection of the AUs. The AU numbers determine the expression of the face based on FACS. A dependable and strong technique for diagnosing the AUs has utilized 2 kinds of features; geometry and appearance feature; as offered in [9].

III. Decision Maker

Classification of every VDD system tends to categorize the derived characteristics into two groups; namely depressive, and non-depressive. The features extracted are the AUs for the system; that depends on FACS. Four various kinds of classification were proposed to perform the current research; Support Vector Machine (SVM); K-Nearest Neighborhood (KNN), Principal Component Analysis (PCA), and Linear Discriminant Analysis (LDA).

K- Nearest Neighbor (KNN) algorithm is one of the most common methods, which used for classification and pattern recognition problems. It's a basic classifier and having better accuracy when contrasted to further classification methods. To realize the working of the current algorithm, let us consider two class problems. Every class most likely has a similar sample. In the case of an unidentified test model, the KNN classifier makes decisions depends on the closest distance among the unidentified input model and every nearest model per class applying popular distance computation approaches, like Euclidean and Manhattan. Then the unknown sample will be assigned to the class of the sample that given the closest distance, as shown in Figure 3. Eq. (2) can be given for the Euclidean distance calculation technique [10].

$$D = \sqrt{\sum_{i=1}^N (P_i - Q_i)^2} \quad (2)$$

Where: $P_i = (p_1, p_2, \dots, p_n)$ stands for the unknown test sample, $Q_i = (q_1, q_2, \dots, q_n)$ stands for a sample in the database trained, and N : refers to the size of the feature.

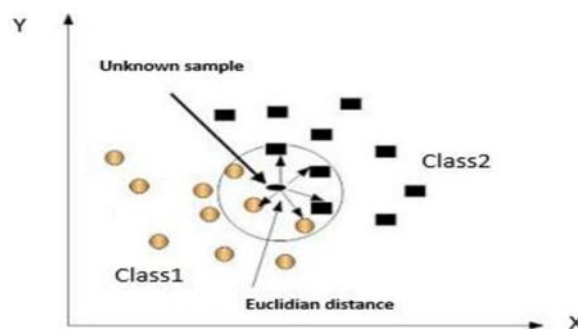


Figure 3: An example of using the KNN algorithm with two classes [11]

The SVM can define as learning algorithms for regression prediction problems and pattern recognition. The concept beyond SVM is to discover the optimal margin among the training samples and to hold the margin separately. A significant first step in the linear SVM input data training was to define the maximal hyperplane vector [12]. If there is a collection of training pattern :

$(x_1, y_1) \dots (x_i, y_i)$, where x_i : is the pattern of the input; y_i :The corresponding output of the target is $\in \{1, -1\}$. This array is linearly separable if the current vector w and scalar b are present.

$$w * x_i + b \geq 1 \quad \text{if } y_i = +1 \tag{3}$$

$$w * x_i + b \leq -1 \quad \text{if } y_i = -1 \tag{4}$$

The equation defined the best hyperplane, and the segregation is:

$$w * x_i + b = 0 \tag{5}$$

Equation (5) is the better boundary choice, as indicated in Figure 4.

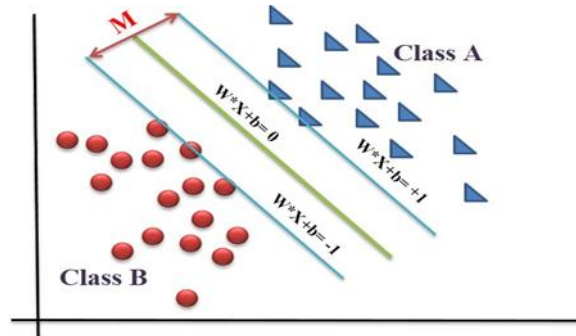


Figure 4: Optimal separating hyperplane between two classes [13]

PCA (principal component analysis) is the most ancient and best-renowned technique of multivariate data analyses. In generic terms, PCA applies a vector space and converts to reduce the dimensionality of huge datasets. By implementing mathematical projection, the original data, that might have involved numerous variables, can frequently be explained in just little variables (e.g., the major components). The main idea of the principal component assessment is to minimize the dimensionality of a set of data in which there a huge number of interlinked variables while remaining as much variation as possible in the data. The above reduction is obtained by converting to a new group of variables, the main components which are not correlated but which are ordered in such a way that the first little retain many of the variation present in all the initial variables. The calculation of the principal components reduces the problem of an eigenvalue-eigenvector for the positive-semi-specified symmetric matrix. Eigenvectors and eigenvalues are measures applied to quantify the direction magnitude and the direction of the variation taken by each axis. Eigenvector defines the angle and direction of the axis during the data space, and it quantifies the magnitude of the difference of the data on the axis as shown in Figure 5. A is an X_n matrix; λ is the eigenvalue, and X : is the eigenvector, the feature combinations are equivalent to the number of dimensions of the dataset.

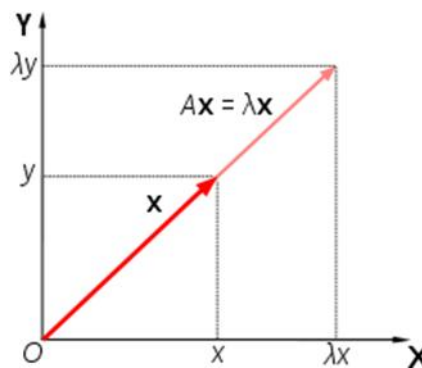


Figure 5: Eigenvectors and eigenvalues [14]

PCA researches for the better demonstration of the data in a smaller-squares sense [15]. This is achieved by analyzing the matrix of information covariance in its vectors and selecting the greatest consideration of them to profile a projection matrix. The coaching data in this situation is a matrix of X from column vectors, symbolize X_i , including the selected signal statistics. Suppose that there was a total of ρ these arithmetical profiles. N is the number of classes (e.g., modulation types) denoted in the training group. Similarly, let n be the number of training profiles in each class, so that $\rho = nN$

Describe the mean profile as:

$$M = \frac{1}{\rho} \sum_{i=1}^{\rho} X_i \tag{5}$$

The first step is to delete the mean profile in each of the ρ profiles, as it is common to all of them and therefore does not contain any useful classification information:

$$\bar{X}_i = X_i - m \tag{6}$$

These new training profiles are centered around the new training matrix. \bar{X} Defines the matrix of covariance as:

$$C = \bar{X}\bar{X}^T \tag{7}$$

In addition to resolving it to its eigenvectors w_i . The eigenvectors conforming to the best k eigenvalues form the dropping matrix W . In this implementation, all necessary individual vectors have been considered and included in the projection matrix. W columns describe the feature space in which the statistical profiles are to be estimated. The first project the oriented training matrix into this space, to decide the centroids for each class:

$$T = w^T \bar{x} = [w^T \bar{x}_1 | w^T \bar{x}_2 | \dots | w^T \bar{x}_p] \tag{8}$$

The centroid for the class l will be

$$g_l = \frac{1}{n} \sum_{i=1}^n w^T \bar{x}_i \tag{9}$$

To employ the classifier, calculate the projection of a test profile t in the feature area,

$$t' = w^T (t - m) \tag{10}$$

and then select the centroid having the least Euclidean space to the point t' [16].

Linear discriminant analysis (LDA); or discriminant job analysis, is a generalization regarding Fisher's linear discriminant, a method applied in statistics, style recognition, and machine learning to locate a linear combination of features that separate 2 or more classes from objects. The resulting collection may be utilized as a linear classifier for dimensionality decreasing before final classification, as shown in Figure 6 [17].

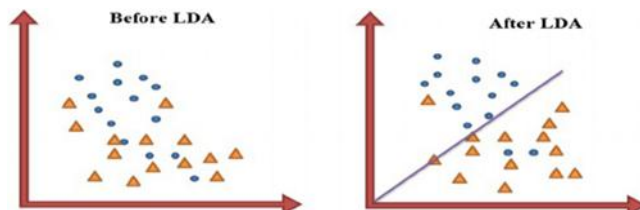


Figure 6: LDA classification [17]

LDA research a diverse projection matrix W in order to maximize the segregation among classes [15].

To clarify this more, it is needed to explain the class scatter matrix S_B and between the within-class scatter matrix S_w , described as:

$$S_w = \sum_{i=1}^N S_i \quad (11)$$

S_i Which is the covariance matrix of class i , and:

$$S_B = \sum_{i=1}^N n(m_i - m)(m_i - m)^T \quad (12)$$

Where m_i is the average profile of class i and m .

If those equivalent matrices were computed in the feature area shaped through projecting the profiles utilizing the projection matrix W , the results would be $W^T S_w W$ and, $W^T S_B W$ respectively. The issue for LDA is to find the matrix W that will maximize the ratio of the determinants of the two matrices as bellow,

$$\frac{|W^T S_B W|}{|W^T S_w W|} \quad (13)$$

From [18], Setting the matrix W in the expression above is equal to solving the Eigen problem.

$$S_B W_i = \lambda_i S_w W_i \quad (14)$$

where W_i is the number of features in the i_{th} column for $i = 1$ as with PCA, columns are organized in order to decrease their eigenvalues and an arbitrary number is retained.

3. COLLECTED DATASETS

There are only limited datasets utilized in the existing computer Vision researches on depression diagnosis that include, in particular, peoples' facial expressions when they are depressed. None of these databases have been made publicly available. So, it is essential to gather the dataset when designing the depression diagnosis system to check the system efficiency and the strength of the procedures utilized to construct the system. Databases are gathered from 125 persons (30 males, 95 females) among the ages of 17-60 years. Before the interview was done, each participant was given a questionnaire (Depression Rating Mechanism) called the Beck Depression Inventory (BDI) [19]. As a self-report inventory, which is one of the most widely used psychometric tests for measuring the degree of the severity of depression. Although the BDI was initially designed to be managed by qualified interviewers, it is most often a self-administrator. It usually takes 5-10 minutes to answer the questions' content, and the score is calculated by summing points to each one of the 21's questions. The result of the score calculated as 0 - 9 are considered no depression, 10 - 15 simple depression, 16 - 23 moderate depression, 24 - 36 severe depression, and over 37 extreme depression. The summary of the 125 participants' check-in datasets are shown below:

TABLE I: The result of the BDI scores for 125 samples

No depression	Simple to moderate	Serve to Extreme
33.6 %	50.4 %	16%

In addition, during the interview there were about 17 relevant questions asked for subjects. The questions are asked and repeated, whereas the camera (Cannon) captured the facial behavior through all sessions.

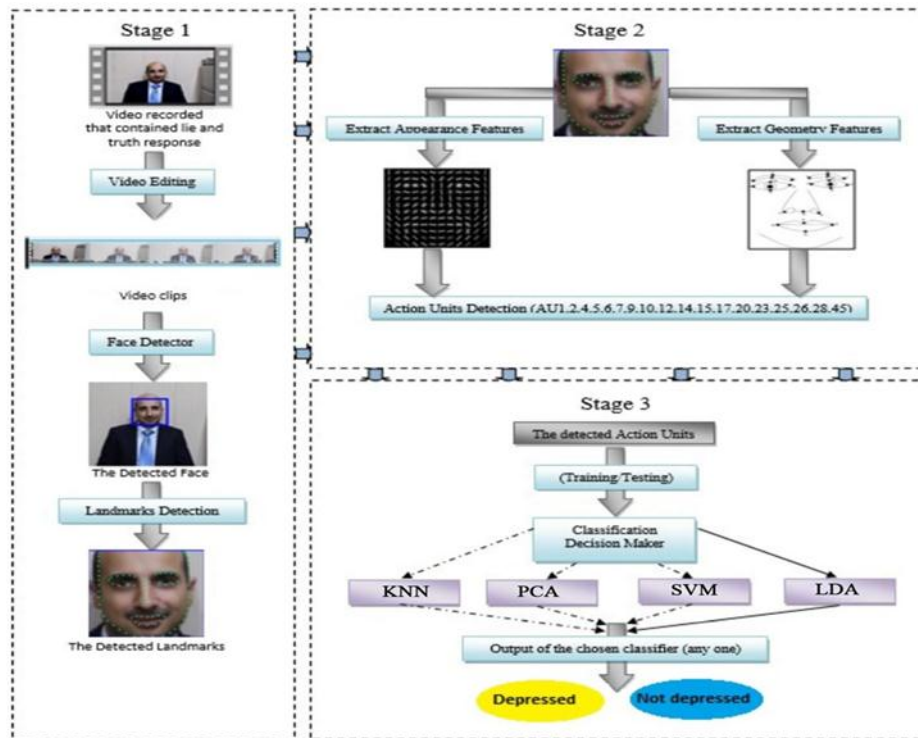
Databases are gathered under usual light conditions, backgrounds are extremely available, and there were no constraints on the subject's head location as seen in the study, Figure 7. This means, subject turns his/her head without any constraints. The gathered dataset is satisfied with normal facial expressions instead of acting or contriving. This was one of the difficulties faced in this work during recording video. This was one of the difficulties faced in this work during recording video as it is difficult to show the expressions from the face of depression person. It is difficult to be visible to a depressed person, and one of the symptoms of depression is a lack of contact with others like the rest of normal people. The database is collected from several educational institutions, which are; Al-Mamoun University College, Al-Karrada private high school for girls, Al-Sayab high school, and Gori kindergarten and nursery.



Figure 7: The interview session

4. THE PROPOSED VISUAL DEPRESSION DIAGNOSIS SYSTEM

In order to build the suggested visual depression diagnosis system, three steps are utilized. The first step in the design of the device is the video Capture and Pre-Processing, which is accountable for filming the video, facial realization, and finding multiple points that interesting in the subject face to arrange the video for the following level. The next step is Taking Out of The Feature. This step is about the diagnosed face; in order to obtain the main features of the AU described in this research on the basis of 2 types of features, geometry and appearance obtained from the face. The final step is the Decision-Making, it's applied to detect the depression participant depends on AUs extracted from the earlier step. Wherever several classification methods are proposed to distinguish between depressed and non-depressed, these classifiers are KNN, SVM, PCA, and LDA. Figure 8 shows the designing stages of the suggested system.



The stage's layout of the suggested visual depression diagnosis system

1. Capturing of the Video and Pre-Processing

Video is filmed with a video camera CANON 1300D. The result format of the video is the MOV extension at the resolution of 1920x1080 pixels per frame. In the next step, it's needed to editing the videos then take out features from it. Editing of the video has been done in order to recognize and acquire just the appropriate sections of these videos. The outcomes of editing are sets of video clips (every single video clip has from 0.35 to 1.6-second duration approximately). Each video includes the period of the subjects' thinks before answering the questions. The above editing is needed to remove the undesirable sections out of the original videos and separate the response to depression

from the response, not to depression. Psychoanalysis, depression can be detected through the moment earlier the answer to the question, which is a moment of think to reply.

After edit the gathered datasets, 280 video-clips gained, fifty percent of the video clips refers to the depression response, whereas fifty percent refers to the non-depression response. The mean duration to each single video clip is one second, the participant has around 2-10 video clips to both depression and non-depression reactions. The participant's face was identified as a pre-processing stage by applying the Viola-Jones facial detection. This scheme depends on the grayscale of the image to select the component pixels through using highlighted squares, a collection of Haar-like squares in the AdaBoost procedure. And it considered Four Haar-like highlighted squares. Separate Haar highlights are estimated to include whole data estimated to show the face of the region. AdaBoost has 31 layers of cascades with a threshold of 3, as described in VJA. Figure 8 demonstrates the recognition outcomes of the VJA of a subject, as established in that figure, the face went during the 31 layers of AdaBoost, at which point the skin's cheek was applied to the cascaded AdaBoost. Finally, the common limit border is chosen to use the superior AdaBoost layer to detect the face area. The landmark of the face was identified based on the faces bounding box acquired from the detector of the face; facial landmark is necessary for facial following-up, and for extracting geometry features. The Constraint Local Neural Field (CLNF) utilized for landmarks of the face is set the position of sixty-eight landmarks as seen in the Figure 9 below.

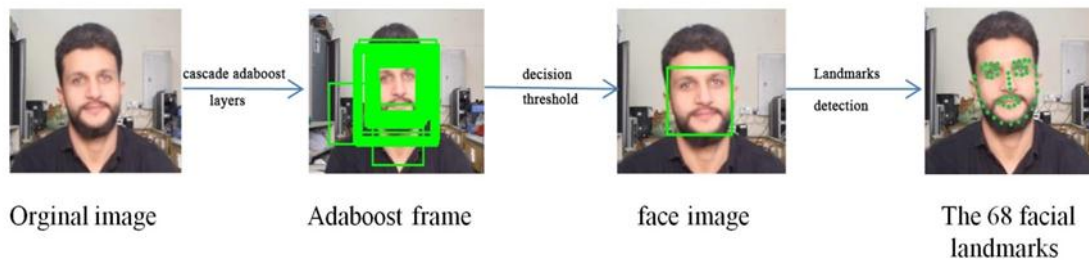


Figure 8: Diagnosis of the face and facial landmark recognition

II. Feature Extraction

The Suggested system identifies 18 facial AUs (AU 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28, and 45) some of these AUs don't affect the procedure of detecting depression. Based on the result of the proposed AUs detection system using collected datasets, show that some AUs not affected during a depressive and non-depressive response, roughly remain on the same status in all interviews. Only six AUs have an important effect during the participants' interviews; these are AUs 4, 5, 6, 7, 10, and 12. Table 2summarizes these six AUs that used in this work as potential markers for depressor finding.

TABLE II: The AUs that Effective in VDD system including its name and area

Action Unit	FACS Name	Facial Region
AU4	Brow lowerer	
AU5	Upper lid raiser	
AU6	Check raiser	
AU7	Lid Tightener	
AU10	Upper Lip Raiser	
AU12	Lip Corner Puller	

The description of the AU detection system is shown in Figure 10. The parameters of the non-rigid shape and the position of the landmarks defined the geometry feature. Before the extraction appearance feature, it needs to transform the face image into a common reference frame (it is

necessary for removing individual variances between faces due to scaling and rotation) and removing the non-facial information from the face image. Alignment and masking perform this task. Face image after removing all non-facial information was coded based on Histograms of Oriented Gradient (HOG) for extracting appearance features.

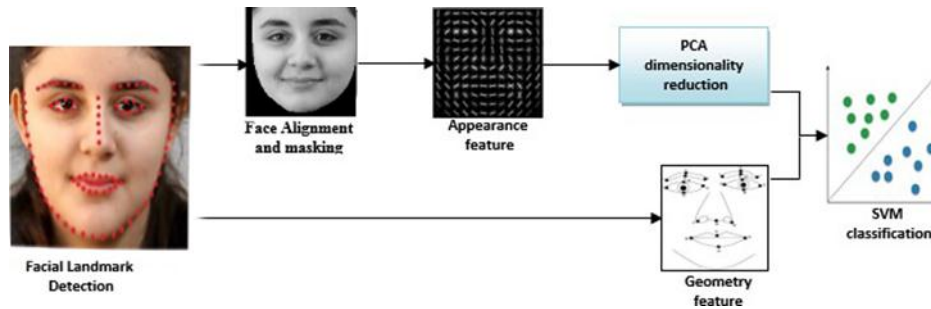


Figure 9: Flow diagram of the system for detecting Action Units

Open source CLNF algorithm is utilized for facial landmark detection and tracking, followed by feature extraction appearance features and geometry features. Masking and alignment are being used to align the face shape to the standard frame and remove non-facial details from the face picture. This is achieved by applying a transforming symmetry between the detected position of a landmark and the front representation of a face picture.

III. Decision Maker

There were four various kinds of classifiers proposed for the decision-making step: KNN, SVM, PCA, and LDA that utilized to distinguish depressed from non-depressed. The detection rate varies from one classifier to another, and it depends on the training database. The testing and training datasets were selected randomly from the gathered database. The gathered dataset is 280 video clips (140 video clips for depression response and 140 video clips for non-depression response). This dataset is divided into testing and training sets (50 percent training and 50 percent testing), 70 video clips are chosen for depression response (after feature extraction and detection AUs) randomly as mentioned above for train the classifier on depressive response and 70 video clips are randomly chosen for training the classifier on non-depressive response. The reminded video clips (70 for depression and 70 for non-depression response) used for the testing phase.

5. THE PROPOSED CLASSIFIERS PERFORMANCE COMPARISON

The performance for the suggested classifiers KNN, SVM, PCA, and LDA are compared based on the diagnosis rate of the result obtained when utilizing collected datasets, as displayed in Table 3, which shows the recognition rate of spot depressed when utilizing collected databases. KNN and LDA classifiers were the top and equal diagnosis accuracy results, which are 85%, based on collected datasets. PCA has accuracy equal to 83.58%, and SVM has an 82.14% success rate when using the same datasets. Both PCA and SVM detection accuracy results were lower than KNN and LDA, which had an equal performance rate of successful results. The above outcomes are obtained when the suggested system was doing test utilizing 50 percent of the gathered dataset. In contrast, the other 50 percent of the dataset is used to train classifiers on depressed and non-depressed facial features.

TABLE III: Comparison of the performance for the suggested visual depression diagnosis classifiers

Suggested Classifiers	Diagnosis Accuracy
KNN	85%
SVM	82.14%
PCA	83.57%
LDA	85%

6. COMPARISON OF SUGGESTED VDD SYSTEM'S PERFORMANCE WITH PREVIOUS WORKS

To evaluate the reliability of the suggested VDD system, the greatest outcomes of the proposed systems should be confronted with earlier works of literature; the best outcome of the suggested work was established on KNN and LDA. It is difficult to compare the results in the same environment; the reason for this difficulty is that the databases used by previous work are not openly available and different from the collected datasets used in the proposed VDD system. As instances of these variances are light conditions, the psychosomatic state of the subjects, and the existence of obstruction. Table 4 shows the comparison between the proposed system and the earlier works based on recognition techniques, the number of subjects, features details and accuracy; this table demonstrates various recognition techniques applied for recognition of the two classes in the preceding works. The efficiency of the suggested system depends on the number of subjects and recognition accuracy and also represents the two principal challenges of the proposed system.

TABLE IV: Comparison of the suggested VDD system's performance with related works

Reference No.	Year	Depression Recognition Techniques	Features Details	No. of participants	Accuracy
[20]	2009	Viola-jones and LMM	Eyes, eyebrows, nose, and mouth	8	75.6%
[21]	2014	SVM	The effect of anger, contempt and action unit AU04,07 AND 09	39	64.1%
[22]	2014	SVM	Full face	50	82%
[23]	2015	SVM	Eyes, full face	32	68.8%
[24]	2016	HMM	Six expression	26	65.38%
[25]	2016	VIOLA-JONES AND KLT	Smile	7	N/A
[26]	2017	SVM and LIBSVM AND HRSD	Mouth bend and wrinkles	18	85%
[27]	2017	SDAE	Head pose and facial landmarks	57	72.59%
[28]	2017	SVM	The gaze of eyes, head pose, and smile	18	90.82%
[29]	2018	Viola-jones	Happy, contempt, sadness and disgust faces (AU 10,12,14)	30	N/A
[30]	2019	LBPAND SVM	Full-face Ten features (eyes, nose, mouth, and cheeks)	N/A	68%
Proposed System	2020	KNN	Facial expressions based on six AUs	125	85%

As demonstrated in Table 4, the suggested VDD system outperforms to related works in terms of the dataset that collected and accuracy achieved.

7. CONCLUSION

The main objectives of the suggested system are to identify and incorporate the specific movement of the facial muscle as reliable indicators for the diagnosis of depression in an unregulated environment, and to establish a reliable device that assists clinicians in the diagnosis of depression. The results demonstrate the theory of the ability to use expressions of the face. That six Action Units (AUs 4, 5, 6, 7, 10 and 12) have a powerful effect of depression diagnosis dependable on the dataset gathered, the omitted action units (AUs 1, 2, 9, 14, 15, 17, 20, 23, 25, 26, 28 and 45) were not effective in the detection accuracy. The suggested VDD system is a self-regulating device that really can recognize any video clip as an input; however, it is not part of the trained data for this device.

References

- [1] "Depression." [Online]. Available: https://www.who.int/health-topics/depression#tab=tab_1. [Accessed: 01-Feb-2020].
- [2] T. P. Blackburn, "Depressive disorders: Treatment failures and poor prognosis over the last 50 years," *Pharmacol. Res. Perspect.*, vol. 7, no. 3, pp. 1–20, 2019.
- [3] S. Al-Gawwam and M. Benaissa, "Depression Detection from Eye Blink Features," in 2018 IEEE International Symposium on Signal Processing and Information Technology, ISSPIT, 2018, pp. 388–392.
- [4] J. F. Cohn et al., "Detecting depression from facial actions and vocal prosody," in Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009, 2009, pp. 1–7.
- [5] P. VIOLA and M. J. JONES, "Robust Real-Time Face Detection," *Int. J. Eng. Technol.*, vol. 7, no. 2, pp. 29–32, 2004.
- [6] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Constrained Local Neural Fields for robust facial landmark detection in the wild," in In Proceedings of the IEEE international conference on computer vision workshops, 2013, pp. 354–361.
- [7] E. S. Mikhailova, T. V. Vladimirova, A. F. Iznak, E. J. Tsusulkovskaya, and N. V. Sushko, "Abnormal recognition of facial expression of emotions in depressed patients with major depression disorder and schizotypal personality disorder," *Biol. Psychiatry*, vol. 40, no. 8, pp. 697–705, Oct. 1996.
- [8] Q. Wang, H. Yang, and Y. Yu, "Facial expression video analysis for depression detection in Chinese patients," *J. Vis. Commun. Image Represent.* vol. 57, no. November, pp. 228–233, 2018.
- [9] T. Baltrušaitis, M. Mahmoud, and P. Robinson, "Cross-dataset learning and person-specific normalization for automatic Action Unit detection," in 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG, 2015, vol. 2015-Janua, pp. 1–6.
- [10] İ. Babaoğlu, "Diagnosis of Coronary Artery Disease Using Artificial Bee Colony and K-Nearest Neighbor Algorithms," *Int. J. Comput. Commun. Eng.*, vol. 2, no. 1, pp. 56–59, 2013.
- [11] I. Nurwauziyah, U. D. S, I. G. B. Putra, and M. I. Firdaus, "Satellite Image Classification using Decision Tree , SVM and k-Nearest Neighbor," no. July, 2018.
- [12] B. Schölkopf, "SVMs - A practical consequence of learning theory," *IEEE Intell. Syst. Their Appl.*, vol. 13, no. 4, pp. 18–21, Jul. 1998.
- [13] A. Hulaj, A. Shehu, and X. Bajrami, "Support Vector Machine for the Classification of Images Captured by WMSN," in Proceedings - 2017 International Conference on Control, Artificial Intelligence, Robotics and Optimization, ICCAIRO, 2017, pp. 283–287.
- [14] Q. Wang, Q. Gao, X. Gao, and F. Nie, "Angle principal component analysis," *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 7, no. 5, pp. 2936–2942, 2017.
- [15] B. J. Frey, "Pattern Classification," in Graphical Models for Machine Learning and Digital Communication, 2018, p. 654.
- [16] D. Acquisition et al., "Naval Postgraduate," vol. 298, no. June, pp. 405–405, 2010.
- [17] K. Torkkola, "Linear Discriminant Analysis in Document Classification," in International Conference Data Mining Workshop on Text Mining, 2001, no. October, pp. 1–10.
- [18] T. Qin, T. Y. Liu, X. D. Zhang, D. S. Wang, and H. Li, "Global ranking using Continuous Conditional Random Fields," in Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference, 2009, pp. 1281–1288.
- [19] G. Jackson-Koku, "Beck depression inventory," *Occup. Med. (Chic. Ill.)*, vol. 66, no. 2, pp. 174–175, 2016.
- [20] N. C. Maddage, R. Senaratne, L. S. A. Low, M. Lech, and N. Allen, "Video-based detection of the clinical depression in adolescents," in in Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC, 2015, pp. 3723–3726.
- [21] G. Stratou, S. Scherer, J. Gratch, and L. Morency, "Automatic nonverbal behavior indicators of depression and PTSD : the effect of gender," *J. Multimodal User Interfaces*, vol. 9, 2014.

- [22] M. Senoussaoui, M. Sarria-paja, J. F. Santos, and T. H. Falk, "Model Fusion for Multimodal Depression Classification and Level Detection," Proc. 4th Int. Work. Audio/Visual Emot. Challenge, 2014, pp. 57-63.
- [23] S. Alghowinem, R. Goecke, J. F. Cohn, M. Wagner, G. Parker, and M. Breakspear, "Cross-cultural detection of depression from nonverbal behavior," in 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG ,2015, pp. 1-8.
- [24] T. H. Yang, C. H. Wu, K. Y. Huang, and M. H. Su, "Coupled HMM-based multimodal fusion for mood disorder detection through elicited audio-visual signals," J. Ambient Intell. Humaniz. Comput., vol. 8, no. 6, pp. 895-906, 2017.
- [25] S. Harati, A. Crowell, H. Mayberg, J. Kong, and S. Nemati, "Discriminating clinical phases of recovery from major depressive disorder using the dynamics of facial expression," in Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS,2016, pp. 2254-2257.
- [26] J. M. Girard, J. F. Cohn, M. H. Mahoor, S. Mavadati, and D. P. Rosenwald, "Social risk and depression: Evidence from manual and automatic facial expression analysis," 2013 10th IEEE Int. Conf. Work. Autom. Face Gesture Recognition, FG, pp. 1-8, and 2013.
- [27] H. Dibeklioglu, Z. Hammal, and J. F. Cohn, "Dynamic Multimodal Measurement of Depression Severity Using Deep Autoencoding," IEEE J. Biomed. Heal. Informatics, vol. 22, no. 2, pp. 525-536, 2018.
- [28] A. Jan, H. Meng, Y. F. A. Gaus, F. Zhang, and S. Turabzadeh, "Automatic Depression Scale Prediction using Facial Expression Dynamics and Regression Categories and Subject Descriptors," Proc. 4th Int. Work. Audio/Visual Emot. Chall., pp. 73-80, 2014.
- [29] D. Venkataraman, "Extraction of Facial Features for Depression Detection among Students," Int. J. Pure Appl. Math., vol. 118, no. 7, pp. 455-463, 2018.
- [30] B. G. Dadiz, C. R. Ruiz, T. Manila, and Q. Manila, "Detecting Depression in Videos using Uniformed Local Binary Pattern on Facial Features," pp. 413-422, 2019.
- [31] J. M. Twenge, A. B. Cooper, T. E. Joiner, M. E. Duffy, and S. G. Binau, "Age, Period, and Cohort Trends in Mood Disorder Indicators and Suicide-Related Outcomes in a Nationally Representative Dataset, 2005-2017," J. Abnorm. Psychol., vol. 128, no. 3, pp. 185-199, Apr. 2019.
- [32] "Understand the News." [Online]. Available: <https://www.vox.com/>. [Accessed: 09-Feb-2