

## A computer vision based image processing system for depression detection among students for counseling

Namboodiri Sandhya Parameswaran, D.Venkataraman

Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India

---

### Article Info

#### Article history:

Received Apr 11, 2018

Revised Jun 12, 2018

Accepted Jan 7, 2019

---

#### Keywords:

Computer vision  
Depression detection  
Facial features  
Feature extraction  
Image processing

---

### ABSTRACT

Psychological problems in college students like depression, pessimism, eccentricity, anxiety etc. are caused principally due to the neglect of continuous monitoring of students' psychological well-being. Identification of depression at college level is desirable so that it can be controlled by giving better counseling at the starting stage itself. If a counselor identifies depression in a student in the initial stages itself, he can effectively help that student to overcome depression. But among large number of students, it becomes a difficult task for the counselor to keep track of the significant changes that occur in students as a result of depression. But advances in the Image-Processing field have led to the development of effective systems, which prove capable of detecting emotions from facial images, in a much simpler way. Thus, we need an automated system that captures facial images of students and analyze them, for effective detection of depression. In the proposed system, an attempt is being made to make use of the Image processing techniques, to study the frontal face features of college students and predict depression. This system will be trained with facial features of positive and negative facial emotions. To predict depression, a video of the student is captured, from which the face of the student is extracted. Then using Gabor filters, the facial features are extracted. Classification of these facial features is done using SVM classifier. The level of depression is identified by calculating the amount of negative emotions present in the entire video. Based on the level of depression, notification is send to the class advisor, department counselor or university counselor, indicating the student's disturbed mental state. The present system works with an accuracy of 64.38%. The paper concludes with the description of an extended architecture using other inputs like academic scores, social content, peer opinions and hostel activities to build a hybrid system for depression detection as future work.

Copyright © 2019 Institute of Advanced Engineering and Science.  
All rights reserved.

---

### Corresponding Author:

Namboodiri Sandhya Parameswaran,  
Department of Computer Science and Engineering,  
Amrita School of Engineering,  
Coimbatore, Amrita Vishwa Vidyapeetham, India.  
Email: cb.en.p2cvi16004@cb.students.amrita.edu

---

## 1. INTRODUCTION

In college students, depression is the result of the social change due to emergence of the internet, smart phones and different social media sites. Majority of students tend to conceal their psychological problems due to the social stigmas related to depression and also due to peer pressure. Some students remain totally unaware of their psychological problems and thus remain deprived of any help that may prove vital to their mental health. It becomes a difficult task for the counselor to keep track of the significant changes that occur in students as a result of depression in a large number of students. Thus we need and automated system

that captures images of students and analyze them for effective depression detection. Facial expressions are the most important form of non-verbal communications to express a persons' emotional or mental state. A large number of studies are currently undergoing on 'Facial feature analyses' for emotion recognition from images which effectively help in prediction of mental health condition of human beings. This study proposes an automated system that detects depression levels in students by analyzing frontal face images of college students. To predict depression, a video of the student is captured, from which the face of the student is extracted. Then using Gabor filters, the facial features are extracted. Classification of these facial features is done using SVM classifier. The level of depression is identified by calculating the amount of negative emotions present in the entire video.

A comparison of Manual FACS coding and Automated FACS coding for finding out Facial Expressions of depressed, showed high similarity in results of both the methods [1]. Highly depressed patients were found to exhibit low presence of smile (AU12) or sadness (AU 15). They showed the high presence of contempt (AU14) and disgust (AU10) along with smile. Figure 1 shows action units found to be present in depression videos [1]– (a) AU 10 – Disgust, (b) AU 12 – Happy, (c) AU 14 – Contempt, (d) AU 15 – Sad.



Figure 1. Action Units found to be present in depression videos [1]– (a) AU 10 – Disgust, (b) AU 12 – Happy, (c) AU 14 – Contempt, (d) AU 15 – Sad

The results pointed out that the most accurate action unit for depression detection was AU14 (action unit related to contempt). In [2], the identification of depression was done by analysing facial landmark points. The distances between them were found out using euclidean and city block distance methods. Here both video and audio features are extracted and then fused together and then classified. In [3] a cross database analysis of three main datasets – 'Black Dog Institute depression dataset (BlackDog), University of Pittsburgh depression dataset (Pitt), and Audio/Visual Emotion Challenge depression dataset (AVEC) has been done which analysis the three datasets individually as well as by combining them for detection of depression features. The dataset was generalized into eye activity data, head pose data, feature fusion data and hybrid data. Of all, the eye activity modality showed better performance. The results indicated that if variability in training data is more the testing results will be better. In [4], three different methods are discussed for emotion recognition. One is use of AU rather than AAM features for classification where AU 14, proved to be the most accurate AU for depression identification. The second method is by using the appearance features from the AAM for classification using SVM and the third is multimodal fusion of vocal and video features. This study claims that during clinical interviews of the depressed, the depression symptoms are communicated nonverbally and can be detected automatically. Another study for finding out depression from facial features has been done by measuring 'Multi-Scale Entropy' (MSE) over time period on the patient interview video. [5] MSE captures the variations that occur in the video across a single pixel. The videos of patients who had lower depression levels were highly expressive of their emotions and such videos showed high entropy levels, otherwise the entropy level was low.

In [6] patients were asked to wear devices to observe their heart-rate, sleep pattern, their reduction in social interaction, their GPS location to check if they are skipping work etc. for depression analysis. Data collection of depressed patients has also been done in [7] by indicating them film-clips to catch the outward appearances of feelings and furthermore by giving an assignment of perceiving negative and positive feelings from various facial pictures. In [8], for a video, the face region is first manually initialized and then KLT (Kanade-Tomasi-Lucas) tracker is utilized to extract curvature information from the picture. Video based approach indicated more precision as it sums up the face area all the more precisely. A technique for face recognition with the assistance of Gabor Wavelet has likewise been proposed. [9]. Here recognition of faces invariant to Pose and Orientation is done. The features extracted are classified with the help of SVM classifier. This framework claims to outperform other face recognition techniques. The work in [10] proposes an improved face recognition system which uses Stationary Wavelet Transform for feature extraction and Conservative Binary Particle Swarm Optimization for feature selection. The proposed method claims to give good performance under cluttered background and is much effective and robust to changes due to

illumination, occlusion, and expression. Utilization of landmark points [11] to compute the LBPH of facial feature reduces LBP histogram’s dimension, which is used for face detection. Too few landmark points result in loss of features. Therefore more landmark points need to be extracted to improve the true positive rate of the recognition process. In [12] the eye and eyebrow features are detected with 4 and 3 feature points for each eye and eyebrow respectively. One can also divide the eyebrow into three equal parts: the inner, the centre and the outer part as in [13]. Feature points can also be detected using different template matching techniques [14]. Facial expression recognition can also be done in two phases: manually locating fourteen points in face region and create a graph with edges [15] that connect such points and then training artificial neural networks to recognize the six basic emotions. The process of facial feature extraction can also be done using Artificial Neural Networks Multilayer Perceptron (MLP) with back-propagation algorithm training the ANN with a number of examples, called learning set [16] and then assigning weights to make the network capable of classifying facial expressions. Features of video and audio information are separated from the video utilizing a Movement History Histogram (MHH) which represents to the qualities of minute changes that occur in face and vocal appearances of the depressed [17]. Emotion recognition from faces can also be detected using Randon and Wavelet transforms. The Randon process projects the 2D image into Randon space and the DWT framework extracts the coefficients at the second level decomposition [18]. The fundamental facial features chosen are eye, nose and mouth locales that can be separated by applying Haar feature based Adaboost algorithm. This strategy diminishes the face preprocessing time for large databases. Facial Activity Units are additionally being recognized, where a combination of various facial activity units can form distinct complex facial expressions for better investigation [19]. On the off chance that the students' depressed feelings are mapped to their actions in classroom, their enthusiastic state can be seen if they are discouraged or not, and in light of this the instructor can help the student by giving careful consideration to that specific student as in [21]. In the event that diverse faces in a same scene demonstrate a similar positive or negative emotion, it would comprehend the entire circumstance of the scene, regardless of whether subjects in the scene are upbeat or in the case of something incorrectly is going on in the scene as in [24]. The work in [25] proposes a system that identifies depression in college students by finding out the presence of low level of happy features in frontal face videos of students. If the happy features are low in the video the student is predicted of having depression. In [26] the process of emotion recognition is done based on speech signal processing and emotion training recognition. The prosodic parameters from speech signals and the facial features from the video signals are extracted and classified parallelly. Both the classifier results are combined using ‘Bimodal’ integration for the final expression recognition result. A face recognition system which represents a face using Gabor-HOG features is proposed in [27]. The face image is filtered using a Gabor Filter bank. The Gabor magnitude images are obtained and the Histogram of Oriented Gradient is computed on these magnitude images. The results show that the fusion of both the methods outperforms the performance of both the processes when performed individually. A feature selection algorithm is proposed in [28], which uses 2D Gabor wavelet transformation to process only the eye and nose regions of face images which shows higher accuracy in detection of multi-pose and multi expression face. Table 1 shows analysis tabulated. Table 1 depicts the analysis of main five papers taken for reference wich include the depression features extracted in each paper, the limitations of each paper and the possible future work that can be underdaten for each particular research paper.

Table 1. Analysis Tabulated

Papers	Depression features extracted	Limitations	Future scope
“Social Risk.....”	Action Units	Interviews in general are less structured.	Depression related questionnaires may capture depressive facial expressions
"Cross-cultural ....”	Eye movement and Head pose movement	Training on specific datasets - prevents from generalizing to different observations.	More varied datasets can be created
"Discriminating clinical ...”	Unsupervised features- Multi Scale Entropy,Dynamical analysis, Observability features	Unsupervised features are used in an exploratory setting.	Features can be classified according to their discriminatory power
"Facial geometry...”	Facial landmarks’ (video) & Statistical descriptors (audio) are fused.	Non depressed Individuals not classified properly	Optimizing of features for detection of non depressive features.
"Video-based...”	Face region was manually initialized & then tracked with KLT	Reinitialized of face region required if the tracked points are below a threshold.	Can consider face as a whole for the entire video.

## 2. RESEARCH METHODOLOGY

In [1], the most accurate action unit one for depression detection depicted as AU14. Based on this theory, the current study proposes a system that will be trained with features of happy, neutral, contempt and disgust faces. Then in the testing phase, videos of college students will be collected while they are answering different questionnaires. The students' facial features will be extracted and classified by SVM classifier for depression detection. Depression detection will be done by overall presence of happy, neutral, contempt and disgust features throughout the video frames and student will be classified as having low, moderate or high depression. The architectural diagram of the proposed automated system can be modeled in the following way.

### 2.1. Proposed Architectural Diagram

Figure 2 shows architectural diagram for the proposed 'Depression Detection' system.

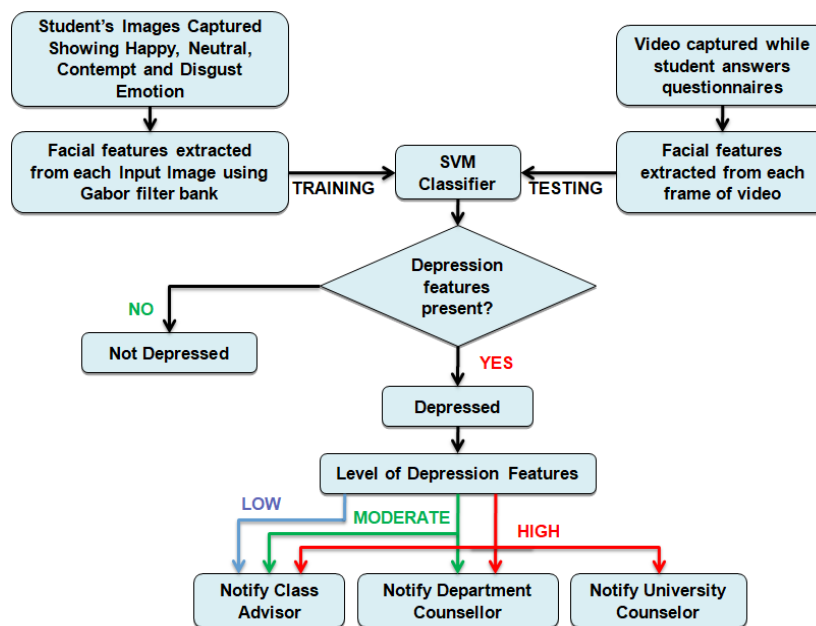


Figure 2. Architectural diagram for the proposed 'Depression Detection' system

## 2.2. Description of Proposed Architectural Diagram.

### 2.2.1. Training Dataset Creation

In addition to happy, contempt and disgust emotions, the emotion 'Neutral' face also implies lack of interest, or emotionless face which may be put forth by the depressed. The input is consequently a dataset of happy, neutral, contempt and disgust faces. For collecting the input dataset a GUI is created that captures images (for each 4 emotions) of the student as Figure 3 below:



Figure 3. Faces of student captured for training set

The training dataset created contains 40 images each of Happy, Neutral, Contempt and Disgust faces. Finally we have a total of 160 images in the input training dataset.

**2.2.2. Face Detection and Feature Extraction**

Once the training set is created, face from each image is detected using the Viola Jones Face Detection algorithm. This algorithm makes use of Haar features, which when convolved throughout the image, we get high output values only at those regions that match the pattern of the haar features and then using Adaboost algorithm and cascading classifiers, it detects a face as in Figure 4(b). Facial features from each face image are extracted using Gabor filters. A Gabor filter bank of 40 filters is created using 5 scales (2, 3, 3.5, 4 and 5) and 8 orientations (0, 23, 45, 68, 90, 113, 135 and 158.) as in [20]. The Gabor filter bank of 40 filters created is shown in Figure 4 (c). For a face detected, the Gabor features extracted are shown in Figure 4 (d).

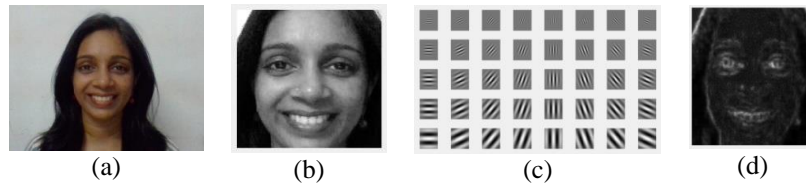


Figure 4. (a) Input image, (b) Face detected, (c) Gabor Filter Bank (d) Gabor Features

For every image a Gabor feature vector is formed as an 'n x 1' column vector as in Figure 5 (a). Feature vectors of all the input images found out and combined to form a feature vector of 'n x 160' feature set as in Figure 5 (b). The dimension of this feature set is very high and so Principal Component Analysis (PCA) is applied to this feature set for dimensionality reduction. Thus we get a '160 x 160' reduced dimension feature set after applying PCA as in Figure 5 (c). This 'Gabor Feature Set' is the input feature set for training. Classes are assigned to each feature vector. The Happy and the Neutral images are considered as positive class and hence assigned the value '+1' and the Contempt and Disgust images are considered as negative class and hence assigned the value '-1'. Finally we get a Gabor Feature Set for training with '160 x 161' dimension with the 161<sup>st</sup> column as the class value as in Figure 5 (d).

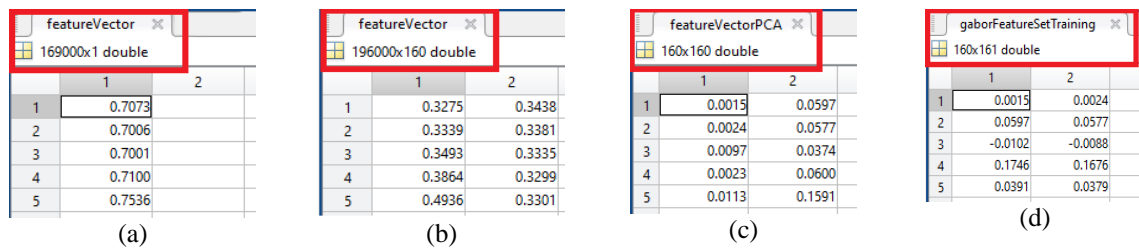


Figure 5. (a) Feature vector for one image; (b) Feature vector set for 160 images; (c) PCA applied feature set; (d) Feature set assigned with classes

**2.2.3. Dataset Creation for Testing**

For testing, a GUI is created, where the student is given a link to answer a simple online 'Depression Analysis Test' as shown in Figure 6(a). The system captures the frontal face video of the student, using the system webcam. This video is converted into frames and from each frame, the face is cropped and the Gabor features are extracted in the same way as in the training phase. The Gabor feature vector for all the frames are concatenated to form a test feature set. For a sample video of 160 frames the test feature set is as shown in Figure 6(b).

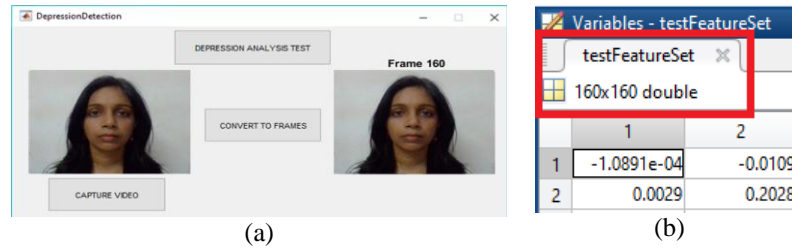


Figure 6. (a) GUI for capturing student's video for testing; (b) Test Feature Set for 160 frames

### 2.3. Classification with SVM

The input feature set is given to a Support Vector Machine classifier for training. The Support Vector Machine is a model that splits the two sets in the best possible way. This is the best split because it is the widest margin that separates the two groups. This line is called the hyperplane. The nearest points are called the Support Vectors.

$$w^T x + b = 0 \quad (\text{Equation of hyperplane}) \quad (1)$$

$$f(x) = \sum_i \alpha_i y_i (x_i^T x) + b \quad (\text{Equation of function}) \quad (2).$$

Where, the set  $\rightarrow X_i$  are the support vectors. Since  $w^T x + b = 0$  and  $c(w^T x + b) = 0$  define the same plane for positive support vectors:  $w^T x_+ + b = +1$  and for negative support vectors:  $w^T x_- + b = -1$ . Then the margin is given by:

$$\frac{w}{\|w\|} \cdot (x_+ - x_-) = \frac{w^T(x_+ - x_-)}{\|w\|} = \frac{2}{\|w\|} \quad (3)$$

To obtain the optimal hyperplane we need to maximize the margin  $\frac{2}{\|w\|}$  or we minimize the weight vector  $\frac{1}{2}(\mathbf{w} \cdot \mathbf{w})$ . Since it becomes a constrained optimization problem this problem can be converted to unconstrained optimization problem by using LaGrange multiple.

$$= L(w, b) = \frac{1}{2}(w \cdot w) - \sum \alpha_i \cdot y_i \cdot (w \cdot x_i) - \sum \alpha_i \cdot y_i \cdot b + \sum \alpha_i \quad (4)$$

Here 'w' has to be minimized and bias term 'b' has to be maximized. First, we take the derivative of the LaGrange with respect to 'b' to get:

$$\frac{\partial L}{\partial b} = \sum_{i=1}^m \alpha_i \cdot y_i = 0 \quad \text{Where, } m \text{ is the number of feature vector} \quad (5)$$

This is one of the constrains we have now. Then we take the derivative of LaGrange with respect to w to get:

$$\frac{\partial L}{\partial w} = \sum_{i=1}^m \alpha_i \cdot y_i \cdot x_i \quad \text{Where, 'm' is number of training samples} \quad (6)$$

When we substitute the above weight expression with the original expression of the LaGrange:

$$\sum_{i=1}^m \alpha_i - \frac{1}{2} \sum \alpha_i \cdot \alpha_j \cdot y_i \cdot y_j (x_i \cdot x_j) \quad (7)$$

Thus the Decision rule depends mainly only on the dot product of the unknown samples  $(x_i \cdot x_j)$ . Given a point 'z', the decision whether the point belongs to class 1 or class 2:

$$D(z) = \text{sign}(\sum_{i=1}^m \alpha_j \cdot y_j \cdot x_j \cdot z + b) \quad (8)$$

If the sign is positive then 'z' is classified to class '1' if negative 'z' is classified to class '-1'. The SVM classifier classifies the test data and gives the predicted classes. As in Figure 7, first image is classified as 1- positive image, image 2 - (-1) so negative image, image 3 is classified as 1- so positive image and so on all the 160 images get classified to get a 160 X 1 matrix of predicted classes.

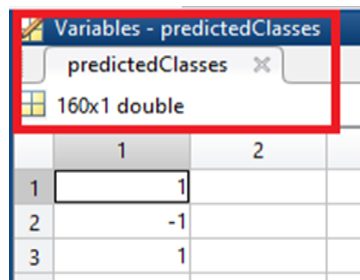


Figure 7. Predicted classes for the 160 test frames

**2.3.1. Depression Level Identification**

For identifying the level of deression from the video we need to find out the total amount of negative emotions in the video frames. The student’s emotion level may change within the time duration of the video. The video is therefore divided into three parts of equal time duration. As depicted in the Table 2:

- If all the three parts of the video have more of positive emotions – the student can be classified as having ‘No Depresseion’.
- If first two parts of the video show positive emotion and the third part show negative emotion, the video is classified as ‘Low Depression’ since only end part of the video is showing negative emotion.
- If two parts of the video is showing positive emotion, then the student may be suffering from ‘Mild Depression’, as most parts of the video show positive emotion.
- If out of the three parts, two parts of the video show negative emotions, then the student is mostly showing negative expressions so predicted as having ‘High Depression’.

Table 2. Depression Level Identification Table

Time Duration	First Part of Video	Middle Part of Video	Last Part of Video	Depression Level
<b>Features Present (Happy, Neutral – Positive class- ‘Positive’)</b>	Positive	Positive	Positive	No Depression
<b>Contempt and Disgust – Negative class – ‘Negative’)</b>	Positive	Positive	Negative	Low Depression
	Positive	Negative	Positive	Mild Depression
	Positive	Negative	Negative	High Depression
	Negative	Positive	Positive	Mild Depression
	Negative	Positive	Negative	High Depression
	Negative	Negative	Positive	High Depression
	Negative	Negative	Negative	High Depression

**3. EXPERIMENTAL RESULTS AND ANALYSIS**

Here videos five different students were taken for experimental analysis. For a single video, each frame of the video was analysed manually and, based on the emotion present they were assigned as having positive ‘+1’ or negative ‘-1’ emotion. These are thus the actual classes of the test video frames. The classifier predicted each frame to belong to either positive or negative class.

Table 3. Confusion Matrix for video I

Emotion	Negative (Actual)	Positive (Actual)	Total
<b>Negative (Predicted)</b>	65	16	81
<b>Positive (Predicted)</b>	41	38	79
<b>Total</b>	106	54	160

Table 3 represents the confusion matrix of actual and predicted classes for the test video frames. Overall 160 images of the test video where considered. 65 video frames where correctly classified as having negative emotion and the remaining 16 frames incorrectly classified as positive class. For the positive emotion frames, out of the 79, 38 were correctly classified as positive and the remaining 41 were wrongly classified. For this particular video the classifier worked with an accuracy of 64.38% as shown in Table 4

below which depicts the Performance metrics of the system for a sample video. Here Error percentage is quite low – 35.62%. Sensitivity which is the ability of the system to correctly classify as having negative emotion (true positive rate) is around 61.32% where as specificity- the ability of the system to correctly identify positive emotion (true negative rate) is 70.37%. Precision which depicts how close different samples are to each other is 80.25%.

Table 4. Performance metrics of the system for a sample video

Performance	Value (%)
Accuracy	64.38
Error	35.62
Sensitivity	61.32
Specificity	70.37
Precision	80.25
FalsePositiveRate	29.63
F1_score	69.52

As shown in Table 5, five videos were considered for testing. For all the five videos first 160 frames were considered for testing. The video was then divided into three equal parts. The sum of the positive and the negative emotion for each part was found out. If positive emotions were more compared to negative emotions then that part of the video was labeled as ‘Positive’. The ‘Actual Emotion State’ of the videos was found out by calculating the amount of the positive and the negative emotions present in the actual class. Similarly, the ‘Predicted Emotion State’ of the videos was found out by calculating the amount of the positive and the negative emotions present in the predicted class.

Table 5. Identifying Level of Depression

Vid eo	Actual -ve	Actual +ve	Actual Emotion State	Predi cted - ve	Predi cted +ve	Accura cy (%)	First Part Emotion	Second Part Emotion	Third Part Emotion	Predicted Emotion State
I	106	54	High Depression	81	79	64.38	Negative	Positive	Negative	High Depression
II	83	77	Mild Depression	72	88	51.88	Positive	Negative	Positive	Mild Depression
III	41	119	Not Depressed	84	76	55.63	Negative	Negative	Positive	High Depression
IV	88	72	Mild Depression	91	69	54.38	Positive	Negative	Negative	High Depression
V	101	59	High Depression	79	81	42.50	Positive	Positive	Negaive	Low Depression

Out of the five videos, Videos – I and II showed same Actual and Predicted emotional state. One of the videos, Video - IV with Mild Depression was predicted as ‘High Depression’. For the remaining two videos, Videos –II and V, the Actual and the Predicted Emotional state were contradicting. The system works with a maximum accuracy of 64.38%. If the student is predicted as having ‘Low Depression’, the ‘Class Advisor’ is sent a notification mail indication the mental state of the student. If the student has ‘Mild Depression’, the ‘Class Advisor’ and the ‘Department Counsellor’ are notified. If the student has got ‘High Depression’ along with the ‘Class Advisor’, the ‘Department Counsellor’ and the ‘University counsellor’ are also informed about the student’s disturbed mental state.

The analysis of this work depicts that, using algorithm proposed in the current study, the presence of depression features can be effectively found out even for a small duration of video. This process can in turn be applied for a video of any large duration and depression features can be identified effectively. This works proves that if the system is trained effectively with the images of depression features alone, the identification of depression in videos can be successfully done with video modality alone. Many of the previous works dealt with identification of all the basic six human emotions, but here only the identification of four main emotions - happy, contempt, disgust and neutral are considered which are mainly found in depressed as in [1]. This in turn reduces the training and testing overload and improves the classifier performance. In this work, the main focus was to find out depression in students, who are not formerly diagnosed with depression. This system does not make use of any standard emotion recognition databases for training. Instead it captures



the student face emotions itself for training the classifier. The testing video is taken captured at the same time, with the same camera under the same background conditions. This helps classifier to efficiently identify emotions from the video of the same person whose images were taken prior for training the classifier.

#### 4. CONCLUSION AND FUTURE WORK

This study was undertaken for finding out the level of depression in five different videos of college students. The presence of 'Happy', 'Neutral', - (positive emotion) and 'Contempt' and 'Disgust'-(Negative emotion) facial features, which are found prominent in depression videos were found out and analysed. The dataset for training and testing was captured separately and the facial features of the same were classified using a Support Vector Machine classifier. The amount of the positive and negative emotions in each video was analysed and the videos were predicted as videos with 'High Depression', 'Mild Depression' or 'Low Depression'. The classifier predicted the outcomes with a maximum accuracy of 64.38% accuracy.

The more the number of training samples, the more accurate will be the classifier prediction. The testing videos captured are of more than thousand frames, out of which only the first 160 frames were considered here for testing purpose. This process can be done for the entire video, by finding out the key frames of the video, by using a key frame extraction technique in the future work. The current study deals only with the recent videos of the student. However, for more accurate depression detection, the history of the student should also to be taken into consideration. Therefore, in the future work, more videos of the same student, taken at different time duration can be considered. This may help to analyse and compare the past and the present mental state of the student and provide more information to the process of depression level identification.

Depression detection from videos alone forms only a part of the whole process of identifying depression. Those students, who may be classified as not depressed, may be victims to depression in the future. For this reason, their other activities have to be continuously monitored. This includes the continuous monitoring of their academic activities, their extra curricular activities and also their social activities. Monitoring academic activities include monitoring the student's grades and attendance. Decreasing in grades or attendance may also be due to a student's extra curricular activities, like engaging in sports or arts. If a student's grades or attendance are poor and they are not active in other mediums like arts or sports also, then they may be at a high risk of falling into depression. Hence students' extra curricular activities have also to be continuously monitored for identification of depression. In addition to this, there should also be a way of monitoring a student's social media content because if the students' social media content show a negative attitude towards life, then such a student may be a victim of stress and depression. Furthermore, for students residing in hostels, input from the hostel authorities regarding the activities of a student within the hostel should also be considered for monitoring the students' day to day activities. If the student leaves the hostel premises to college, but in turn skips classes by indulging in other negative activities, then there is a risk of the student falling into a negative state of mindset which may eventually lead to depression. The future work to this study is to form an elaborate model of depression identification process, by taking all the above mentioned factors into consideration and combining it with the current work of identifying depression with images.

#### ACKNOWLEDGMENT

The authors would like to extend the heartfelt gratitude to the faculty-in-charge of Amrita-Cognizant Innovation Lab, Department of Computer science and Engineering, Amrita school of Engineering, Coimbatore for the support extended in carrying out this work.

#### REFERENCES

- [1] Girard, Jeffrey M., Jeffrey F. Cohn, Mohammad H. Mahoor, SeyedmohammadMavadati, and Dean P. Rosenwald. "Social risk and depression: Evidence from manual and automatic facial expression analysis." In Automatic Face and Gesture Recognition (FG), 10th IEEE International Conference and Workshops on, pp. 1-8. IEEE, 2013.
- [2] Pampouchidou, A., O. Simantiraki, C-M. Vazakopoulou, C. Chatzaki, M. Pediaditis, A. Maridaki, K. Marias et al. "Facial geometry and speech analysis for depression detection." In Engineering in Medicine and Biology Society (EMBC), 39th Annual International Conference of the IEEE, pp. 1433-1436. IEEE, 2017.
- [3] Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of clinical epidemiology*. 46(12):1417-32. 1993
- [4] Cohn, Jeffrey F., Tomas Simon Kruez, Iain Matthews, Ying Yang, Minh Hoai Nguyen, MargaraTejera Padilla, Feng Zhou, and Fernando De la Torre. "Detecting depression from facial actions and vocal prosody." In Affective Computing and Intelligent Interaction and Workshops. ACII 2009. 3rd International Conference on, pp. 1-7. IEEE, 2009.

- [5] Harati, Sahar, Andrea Crowell, Helen Mayberg, Jun Kong, and ShamimNemati. "Discriminating clinical phases of recovery from major depressive disorder using the dynamics of facial expression." In Engineering in Medicine and Biology Society (EMBC), 38th Annual International Conference of the, pp. 2254-2257. IEEE, 2016.
- [6] Tasnim, Mashrura, RifatShahriyar, NowshinNahar, and Hossain Mahmud. "Intelligent depression detection and support system: Statistical analysis, psychological review and design implication." In e-Health Networking, Applications and Services (Healthcom), 18th International Conference on, pp. 1-6. IEEE, 2016.
- [7] Pampouchidou, Anastasia, Kostas Marias, ManolisTsiknakis, P. Simos, Fan Yang, and FabriceMeriaudeau. "Designing a framework for assisting depression severity assessment from facial image analysis." In Signal and Image Processing Applications (ICSIPA), International Conference on, pp. 578-583. IEEE, 2015.
- [8] Maddage, Namunu C., RajindaSenaratne, Lu-Shih Alex Low, Margaret Lech, and Nicholas Allen. "Video-based detection of the clinical depression in adolescents." In Engineering in Medicine and Biology Society, (EMBC). Annual International Conference of the IEEE, pp. 3723-3726. IEEE, 2009.
- [9] Karthika R, Parameswaran L. *Study of Gabor wavelet for face recognition invariant to pose and orientation.* In Proceedings of the International Conference on Soft Computing Systems, pp. 501-509. Springer, New Delhi.2016.
- [10] Babu, S. Hitesh, Sachin A. Birajdhara, and Samarth Tambad. "Face Recognition using Entropy based Face Segregation as a Pre-processing Technique and Conservative BPSO based Feature Selection." Indian Conference on Computer Vision Graphics and Image Processing, pp. 46. ACM, 2014.
- [11] Xiang, Gao, Zhu Qiuyu, Wang Hui, and Chen Yan. "Face recognition based on LBPH and regression of Local Binary features." In Audio, Language and Image Processing (ICALIP), International Conference on, pp. 414-417. IEEE, 2016.
- [12] Moreira, Juliano L., Adriana Braun, and Soraia R. Musse. "Eyes and eyebrows detection for performance driven animation." In Graphics, Patterns and Images (SIBGRAPI), 23rd SIBGRAPI Conference on, pp. 17-24. IEEE, 2010.
- [13] Florea, Laura, and RalucaBoia. "Eyebrows localization for expression analysis." In Intelligent Computer Communication and Processing (ICCP), International Conference on, pp. 281-284. IEEE, 2011.
- [14] Phuong, Hoang Minh, Le Dung, Tony de Souza-Daw, Nguyen TienDzung, and ThangManh Hoang. "Extraction of human facial features based on Haar feature with Adaboost and image recognition techniques." In Communications and Electronics (ICCE), Fourth International Conference on, pp. 302-305. IEEE, 2012.
- [15] Tanchotsrinon, Chaiyasit, SuphakantPhimoltares, and SaranyaManeroj. "Facial expression recognition using graph-based features and artificial neural networks." In Imaging Systems and Techniques (IST), pp. 331-334. IEEE, 2011.
- [16] Owayjan, Michel, Roger Achkar, and MoussaSkandar. "Face Detection with Expression Recognition using Artificial Neural Networks." In Biomedical Engineering (MECBME), 3rd Middle East Conference on, pp. 115-119. IEEE, 2016.
- [17] Meng, Hongying, Di Huang, Heng Wang, Hongyu Yang, Mohammed AI-Shuraifi, and Yunhong Wang. "Depression recognition based on dynamic facial and vocal expression features using partial least square regression." In Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge, pp. 21-30. ACM. 2013.
- [18] Ali H, Sritharan V, Hariharan M, Zaaba SK, Elshaikh M. Feature extraction using Radon transform and Discrete Wavelet Transform for facial emotion recognition. *InRobotics and Manufacturing Automation (ROMA), 2016 2nd IEEE International Symposium*, pp. 1-5. IEEE. 2016.
- [19] Tian YI, Kanade T, Cohn JF. Recognizing action units for facial expression analysis. *IEEE Transactions on pattern analysis and machine intelligence*. 23(2):97-115. 2001
- [20] Haghighat M, Zonouz S, Abdel-Mottaleb M. CloudID: Trustworthy cloud-based and cross-enterprise biometric identification. *Expert Systems with Applications*. 42(21):7905-16. 2015.
- [21] Sahla KS, Kumar TS. Classroom Teaching Assessment Based on Student Emotions. In *The International Symposium on Intelligent Systems Technologies and Applications* pp. 475-486. Springer, Cham. 2016.
- [22] Nehru, Mangayarkarasi, and S. Padmavathi. "Illumination invariant faces detection using viola jones algorithm." In Advanced Computing and Communication Systems (ICACCS), 4th International Conference on, pp. 1-4. IEEE, 2017.
- [23] Vikram, K., and S. Padmavathi. "Facial parts detection using Viola Jones algorithm." In Advanced Computing and Communication Systems (ICACCS), 4th International Conference on, pp. 1-4. IEEE, 2017.
- [24] Athira, S., R. Manjusha, and LathaParameswaran. "Scene Understanding in Images." In *The International Symposium on Intelligent Systems Technologies and Applications*, pp. 261-271. Springer International Publishing, 2016.
- [25] Venkataraman, D., Parameswaran, N. S. "Extraction of Facial Features for Depression Detection among Students." *International Journal of Pure and Applied Mathematics*, International Conference on Advances in Computer Science, Engineering and Technology, pp. 455-462, 2018.
- [26] Wang, Yutai, Xinghai Yang, and Jing Zou. "Research of emotion recognition based on speech and facial expression." *Indonesian Journal of Electrical Engineering and Computer Science 11*, no. 1: 83-90. 2013.
- [27] Ouanan, Hamid, Mohammed Ouanan, and BrahimAksasse. "Gabor-HOG Features based Face Recognition Scheme." *Indonesian Journal of Electrical Engineering and Computer Science 15*, no. 2: 331-335. 2015.
- [28] Lin, Chuan, Xi Qin, Guo-liang Zhu, Jiang-hua Wei, and Cong Lin. "Face detection algorithm based on multi-orientation Gabor filters and feature fusion." *Indonesian Journal of Electrical Engineering and Computer Science 11*, no. 10: 5986-5994. 2013.