PREDICTION OF CLINICAL DEPRESSION IN ADOLESCENTS USING FACIAL IMAGE ANALYSIS

Kuan Ee Brian Ooi, Lu-Shih Alex Low, Margaret Lech and Nicholas Allen[*]

School of Electrical and Computer Engineering, RMIT University, Melbourne 3001, Australia
[*]ORYGEN Research Centre and Department of Psychology, University of Melbourne, Melbourne 3010, Australia
{k.ooi, lushih.low}@student.rmit.edu.au, margaret.lech@rmit.edu.au, nba@unimelb.edu.au

ABSTRACT

This study investigates the possibility of using facial images to determine if a non-depressed person is likely to develop clinical depression within the next 1-2 years. The prediction method uses a typical classification approach of training and testing. Class models were determined using image data from adolescents who were either "at risk" or "not at risk" of depression. The risk factor was confirmed through 2 years of follow-up data collection. Two feature extraction approaches were compared; the eigenface (PCA) features and the fisherface (PCA+LDA) feature. The nearest neighbor (NN) classification was implemented using person independent and person dependent approaches. Best results were with the fisherface (PCA+LDA) method, providing prediction accuracy of 51% with the person independent approach and 61% when using a person dependent approach.

1. INTRODUCTION

Clinical depression has been recognized as a serious medical illness that affects both the health and well being of a person. If untreated, depression can lead to a range of negative effects from mild mood discomforts to suicides. In Australia, 2,101 suicide cases were reported in 2005 [11]. Depression appears at different stages of the life cycle, including adolescents. It has been reported that at least 14% of adolescents aged 4-12 years were diagnosed with depression or showed some signs of mental disorder. This figure rose to 19% for the 13-17 years age-group [14]. As such, an early diagnosis of depression in adolescents is pertinent. However, the ability to predict depression could provide for more holistic information to help reduce the adverse effects of the disorder or prevent it altogether.

Current conventional diagnostic methods used by psychologists involve subjective self-evaluation tests and interviews. For example, Roberts *et al.* [12] compared two types of scales for identifying depression among adolescents. Furthermore, in predicting depression in adolescents, recent psychological studies investigated the correlation of brain structure with depressive symptoms [17]. Different research fields have tested multiple methods to determine the causes of clinical depression and objective measures to detect or predict the risk of clinical depression. Among which are speech processing that analyses vocal parameters and facial recognition that examines facial parameters and expressions. Both speech and facial image have shown promising results in automatic detection of depression. In particular, current studies on speech analysis experimenting with various speech features achieved relatively high accuracy in the classification of clinical depression [2], [7], [9]. Moore *et al.* [9] experimented on prosodic, vocal tract and glottal features in the detection of depression. The results showed a highest accuracy of 96% when a combination of prosodic and glottal features was used. Low *et al.* [7] investigated five different feature categories, (Teager energy operator (TEO), cepstral, prosodic, spectral and glottal features) and found consistently higher results for the TEO feature compared to the other features and feature combinations. Also, the addition of TEO features to other features provided significant improvement of the classification results. Moreover, studies of the facial image analysis for automatic detection of depression are relatively limited in contrast to the number of studies conducted on speech; partially attributed to the fact that image analysis is generally computationally more demanding. Existing reports on facial image analysis show that, like speech, facial parameters and expressions also carry important distinctive information for depression diagnosis. Maddage, *et al.* [8] reported 85.5% of correct classification for depressed subjects using Gabor wavelet features extracted at facial landmarks while the active appearance modeling (AAM) method, representing facial images which was used by Cohn *et al.* [2], had an accuracy of 79% in identifying depressed and non-depressed participants.

This study focuses on predicting the currently non-depressed adolescents who are "at risk" of developing symptoms of clinical depression within the next 1-2 years, using facial image analysis. The prediction is based on classification into two classes: "at risk" (AR) and "not at risk" (NAR). Training data used to represent these two classes was confirmed through the follow up data collection conducted 2 years from the time when the database of video recordings was collected. To our knowledge this is the first study of its kind and it has a preliminary character aiming to determine if facial images captured 1-2 years before development of typical symptoms of depression can provide useful predictive information. The above hypothesis is tested using two effective yet relatively simple facial feature extraction methods representing the appearance-based approach used widely in facial analysis and recognition field. The first approach is the principal component analysis (PCA) method commonly used for image analysis and dimensionality reduction known as eigenface [16] in various face recognition and emotion detection tasks. The second approach is the fisherface method using linear discriminant analysis (LDA) combined with PCA, providing a class specific dimensionality reduction [1]. LDA forms a scatter that optimizes the classification process by maximizing the between-class variance and minimizing the within-class variance, whereas PCA maximizes the scatter of all projected samples. Although these two projection methods have been widely used in face recognition [1], [13], [16], it is yet to be explored in the area of depression detection.

2. IMAGE DATA

The depression prediction experiments were conducted on a clinical database of video recordings. Data was obtained as a result of research cooperation with the ORYGEN Youth Health Research Centre. It comprises video recordings of discussions conducted between parents and their adolescents (12-13 years of age) participating in two different types of family interactions: event-planning interaction (EPI) and problem-solving interaction (PSI). The video recordings were annotated based on the Living-In-Family-Environments (LIFE) coding system [4], [5] by psychologists. During each session, the family members were asked to conduct a discussion on an assigned topic reflecting the type of interaction (EPI or PSI) for a length of 20 minutes per interaction. These video recordings were made during the first (T_1) stage of data collection when a total of 191 (94 female & 97 male) adolescents diagnosed as non-depressed were recorded. After two years during the second (T_2) follow-up stage, it was determined using conventional diagnostic methods that some of these initially non-depressed adolescents developed symptoms of depression where 15 participants (6 male & 9 female) reported suffering from Major Depressive Disorder (MDD) and 3 participants (1 male & 2 female) having Other Mood Disorders (OMD).

In order to examine if facial images can provide vital predictive information before a typical medical diagnosis of depression can be made, datasets representing two groups of participants were selected from this database. The first group contained 15 control adolescents who were non-depressed in T_1 and did not show any symptoms of depression between T_1 and T_2; this group is referred to as "not at risk" (NAR). The second group contained 15 adolescents who were non-depressed in T_1 but developed the major depressive disorder (MDD) between T_1 and T_2; this group is referred to as "at risk" (AR). It is important to note that all of the images (for both groups; "at risk" and 'not at-risk") used in the prediction experiments represented adolescents that were diagnosed as non-depressed when these images were captured in T_1.

3. METHOD

Image sequences for each adolescent representing two separate video sessions, each of which lasting 20 minutes at a capture rate of 25 frames per second were extracted. This resulted in a total of approximately 30000 image frames per person and per video session. The raw image frames were then preprocessed using the SMQT face detector [10] to automatically detect frames representing the frontal face views and only these frames were selected for the subsequent face analysis. Each of these front-face-only frames was cropped to a smaller image size of 94 x 94 pixels resulting in a vector of dimension 8836. These image vectors were used to calculate two types of features, the eigenfaces and fisherfaces. Both types of features provided training and testing sets of feature vectors used in the classification process. The following sections discuss the computation of the eigenface (PCA) and fisherface (PCA+LDA) features and the basics of the nearest neighbour (NN) classifier.

3.1. Feature Extraction

3.1.1. Eigenfaces (PCA)

The principal component analysis (PCA) determines a set of eigenvectors calculated for a covariance matrix of an array of images. The eigenvectors represent basic images also called eigenfaces. A linear combination of eigenfaces can be used to approximate different types of human faces. PCA helps to reduce the feature dimensionality by mapping the original array of images into a lower-dimensionality space. The application of the eigenfaces (PCA) method followed steps described in [16].

Firstly a training array T (N x M) of face images was generated, where each of the M columns represented an N-dimensional image vector. A mean face vector of the training array was calculated and subtracted from the original array T and a scatter matrix S was obtained. The eigenvectors (eigenfaces) and the eigenvalues of the array S were calculated. In order to reduce the data dimensionality only the eigenfaces that represent the largest eigenvalues were kept. This way the original dimension for a single image representation was reduced from N=8836 dimensions to p=500. The training and the testing images were then projected onto the p-dimensional space which means that each image was represented as linear combinations of the p-dimensional eigenfaces. The weights obtained from these linear combinations were used as the training feature vectors stored in two arrays: $\boldsymbol{W}_{pAR}$ for the AR training class and $\boldsymbol{W}_{pNAR}$ for the NAR training class.

3.1.2. Fisherfaces (PCA)

The fisherface method [1] combines PCA with LDA which is used to find the vectors in the space which best discriminate among classes and the PCA is used as a pre-processing step for the LDA setup. The main concept of LDA is to maximize the between class scatter matrix S_B and to minimize the within class scatter matrix S_W which are defined by equation (1) for the former and equation (2) for the later.

$$S_B = \sum_{i=1}^{C} M_i (x_i - \mu)(x_i - \mu)^T \quad (1)$$

$$S_W = \sum_{i=1}^{C} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \quad (2)$$

C represents the number of classes where M_i is the number of training samples in class i, μ_i is the mean image of class X_i and x_k is the k-th image of class X_i .

The projection matrix $\boldsymbol{W}_{LDA}$ is a set of the eigenvectors with the largest eigenvalues of $S_W^{-1} S_B$. The fisherface method avoids S_W being singular using the PCA as a preprocessing step by projecting the image set to a lower dimensional space. Thus, by integrating the LDA with PCA, the optimal projection is defined by equation (3). In this experiment, two set of weights were generated: $\boldsymbol{W}_{optAR}$ for the AR training class and $\boldsymbol{W}_{optNAR}$ for the NAR training class.

$$W_{opt}^T = W_{LDA}^T W_{PCA}^T \quad (3)$$

3.2. Feature Classification

In this experiment, testing weights were obtained by projecting the testing images onto the lower dimensional space of the different training classes for classification. Here, four different testing weights were obtained for each individual testing image: 1. W_{tpAR} (projection on the subspace of W_{pAR}) 2. W_{tpNAR} (projection on the subspace of W_{pNAR}) 3. W_{toptAR} (projection on the subspace of W_{optAR}) 4. $W_{toptNAR}$ (projection on the subspace of W_{optNAR}).

From the extracted features, classification was performed using the NN approach [3] that searches for its nearest category which is found with the standard Euclidean squared distance measure (4) for classification decision.

$$d = \|x - y\|^2 \quad (4)$$

4. EXPERIMENTS AND RESULTS

4.1. Experimental setup

The experiments tested a two-class prediction problem in which the facial images were classified into two classes: "at risk" of depression (AR) and "not at risk" of depression (NAR). Two types of the predictive classification were tested: person independent and person dependent. Although, the gender dependency has been reported in previous depression detection studies [6], [7], [8], [9], a relatively small number of data representing different genders prohibited us from testing this otherwise important factor.

The assessment of the classification results was based on three parameters: specificity, sensitivity and accuracy typically used in a binary classification process. These parameters were defined as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \quad (5)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100\% \quad (6)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (7)$$

The positive class label was represented by the AR class and the negative label was represented by the NAR class. TP represented the number of true positive outcomes (the number of AR adolescents classified as AR), FP represented the number of false positive outcomes (the number of NAR adolescents classified as AR), TN represented the number of true negative outcomes (the number of NAR adolescents classified as NAR) and FN represented the number of false negative results (the number of AR adolescents classified as NAR).

In the person independent as well as the person dependent tests, approximately 50% of the dataset, 7 NAR adolescents and 8 AR adolescents were used for training and the remaining 50% were used for testing. The training and testing processes were repeated for 3 turns of cross validation, and the results were averaged over these three runs. This process was conducted independently for three types of family interaction combinations used during the recording sessions (EPI, PSI and combined EPI+PSI) and for two types of feature extraction/data compression methods (eigenface/PCA and fisherface/PCA+LDA). The system performance was determined using the parameters given in equations (5)-(7).

4.2. Person independent classification

In the case of a person independent prediction, all available images for the 15 AR and 15 NAR adolescents were put together (i.e. the data was not separated for each adolescent) and used to generate the class balanced training and testing sets for the classification process. Table 1 shows the performance results for the person independent classification using two different feature extraction methods and three different types of family interactions.

Training/Testing Features	Training/Testing Dataset	Classification Accuracy in %		
		Sensitivity	Specificity	Accuracy
Eigenface (PCA)	EPI session only	45	44	45
	PSI session only	47	54	50
	EPI+PSI sessions	46	50	48
Fisherface (PCA+LDA)	EPI session only	55	40	47
	PSI session only	53	50	51
	EPI+PSI sessions	55	45	50

Table 1. Person independent classification performance result

4.3. Person dependent classification

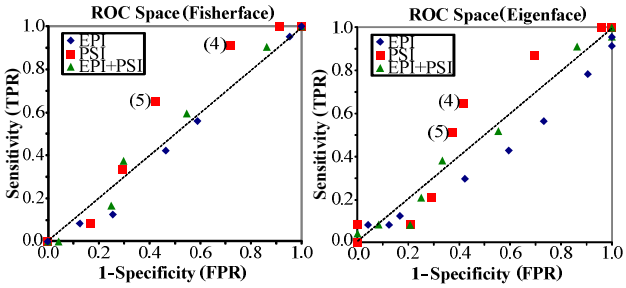
4.3.1. Selection of an optimal decision rule for the person dependent classification

From the medical prevention point of view, the predictive diagnostic method for depression should have a relatively large true positive rate (TPR) i.e. sensitivity, at the cost of lower values of false positive rate (FPR) i.e. 1-specificity. TPR to FPR ratio was controlled by the classification decision rule based on two thresholds. The first threshold ζ_{AR} specifying the percentage of images classified as AR above which the overall classification for a given AR adolescent was determined as AR, and the second threshold, ζ_{NAR} specifying the percentage of images classified as NAR above which the overall classification for a given NAR adolescent was determined as NAR. The classification results based on each of the 9 decision rules listed in Table 2 were used to calculate the parameters given in equations (5)-(7) respectively.

Rule	ζ_{AR}	ζ_{NAR}
1	10%	90%
2	20%	80%
3	30%	70%
4	40%	60%
5	50%	50%
6	60%	40%
7	70%	30%
8	80%	20%
9	90%	10%

Table 2. Classification decision rules

The first three sets of 9 points corresponding to the 3 interactions and the fisherface features were plotted in Fig. 1 (a), whereas the three sets of 9 points corresponding to the 3 interactions and the eigenface features are plotted in Fig. 1 (b). The ROC (Receiver Operating Characteristic) plots in Fig. 1 were used to determine which of the 9 classification decision rules provides the best TPR to FPR ratio. The point closest to the upper left corner of the ROC plain represents the best decision rule. The best combination giving approximately the highest TPR while having a reasonable un-skewed FPR was a balanced 50% criterion corresponding to classification decision rule number 5 in Table 2.



(a) Fisherface feature (b) Eigenface feature

Fig. 1. ROC space analysis

4.3.2. Results of the person dependent classification using an optimal classification decision rule

Based on the 50% criterion, Table 3 shows the results of person dependent classification accuracy. Two distinctive observations from the results show that the fisherface feature and the PSI session perform consistently better than the rest of the experiments. Others have shown that the LDA performed better than PCA in various cases of face recognition [1], [13]. In this experiment, the fisherface method performs slightly better than the eigenface

method. Also the PSI session was consistently giving the highest accuracy. This finding is consistent with other clinical depression experiments conducted by Low *et al.* [6] and has been supported by the fact that the PSI session involves conflictual interactions which are strong correlates of adolescents' depression in family environments where often negative emotions are expressed [15].

Training/Testing Features	Training/Testing Dataset	Classification Accuracy in %		
		Sensitivity	Specificity	Accuracy
Eigenface (PCA)	EPI session only	43	41	42
	PSI session only	51	63	57
	EPI+PSI sessions	38	67	52
Fisherface (PCA+LDA)	EPI session only	56	41	49
	PSI session only	65	58	61
	EPI+PSI sessions	60	45	52

Table 3. Person dependent classification performance result

5. CONCLUSIONS AND FUTURE WORK

Feasibility of using facial features in an automatic prediction of clinical depression in adolescents was investigated. Prediction task was performed using a classification approach based on two class models trained on a long-term observational data and representing adolescents that are "at risk" (AR) of developing depression within 1-2 years and adolescents that are "not at risk" (NAR) of depression. In the classification process, the eigenface (PCA) feature extraction and data reduction algorithm was compared with the fisherface (PCA+LDA) method. In both cases the nearest neighbour (NN) classifier was used. Image samples used in the experiments were recorded during two types of family interactions: PSI (problem solving interaction) and EPI (event planning interaction).

The highest prediction accuracy was achieved using the fisherfaces (PCA+LDA), providing 51% accuracy for the person independent classification during the PSI session and 61% when using the person dependent approach. These classification results are relatively low, and do not provide a definite answer to the question of whether it is possible to predict depression from facial images before a conventional diagnosis can be made. However, it is interesting that the PSI session provided consistently higher prediction results than the EPI session and the combination of PSI and EPI sessions. In the case of PSI, it was possible to obtain classification accuracy that was about 7%-11% above a pure chance level. This can be contributed to the fact that during the problem solving task a controversy is more likely to be elicited between speakers than during the event planning session. This in turn could lead to an increased effort to control the behavior and to present "nice" or "neutral" facial expressions during the recording sessions. The higher classification rates at the PSI session could therefore indicate that there is a difference in the degree to which the AR and NAR adolescents can control their facial expressions.

Further work needs to be done to determine the best type of features and classifiers for the prediction task. Different features such as the Gabor wavelet and the AAM should be tested, as well as more advanced classifiers such as the Gaussian mixture model (GMM) and the support vector machine (SVM). As both face and voice carry overlapping as well as complimentary information, further studies should also look at the possibility of integrating the facial image analysis with the speech signal analysis.

6. ACKNOWLEDGMENT

The authors would like to thank the ORYGEN Youth Health Research Centre, Australia for their support in this research.

7. REFERENCES

[1] P. N. Belhumeur, et al., "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions, Pattern Analysis and Machine Intelligence*, vol. 19, pp. 711-720, 1997.

[2] J. F. Cohn, et al., "Detecting depression from facial actions and vocal prosody," in *Affective Computing and Intelligent Interaction and Workshops Conf.*, pp. 1-7, 2009.

[3] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions*, *Information Theory,* vol. 13, pp. 21-27, 1967.

[4] H. Hops, et al., "Methodological issues in direct observation: Illustrations with the living in familial environments (LIFE) coding system," *Journal of Clinical Child Psychology*, vol. 24, pp. 193 - 203, 1995.

[5] N. Longoria, et. al., Living in Family Environments (LIFE) Coding. A Reference Manual for Coders. *Oregon Research Institute*, 2006.

[6] L. S. Low, et al., "Content based clinical depression detection in adolescents," in *Proc. European Signal Processing Conf.*, pp. 2362-2365, 2009.

[7] L. S. Low, et al., "Detection of Clinical Depression in Adolescents' Speech During Family Interactions," *IEEE Transactions, Biomedical Engineering*, vol. PP, pp. 1-1, 2010.

[8] N. C. Maddage, et al., "Video-based detection of the clinical depression in adolescents," in *Proc. IEEE Int. Conf. Engineering in Medicine and Biology Society*, pp. 3723-3726, 2009.

[9] E. Moore, et al., "Critical Analysis of the Impact of Glottal Features in the Classification of Clinical Depression in Speech," *IEEE Transactions, Biomedical Engineering*, vol. 55, pp. 96-107, 2008.

[10] M. Nilsson, et al., "Face Detection using Local SMQT Features and Split up Snow Classifier," in *Proc. IEEE Int. conf. Acoustics, Speech and Signal Processing*, pp. II-589-II-592, 2007.

[11] B. Pink, "Suicides," Australian Bureau of Statistics, Sydney: *Commonwealth of Australia*, 2007.

[12] R. E. Roberts, et al., "Screening for Adolescent Depression: A Comparison of Depression Scales," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 30, pp. 58-66, 1991.

[13] J. Ruiz-del-Solar and P. Navarrete, "Eigenspace-based face recognition: a comparative study of different approaches," *IEEE Transactions, Systems, Man, and Cybernetics*, Part C: Applications and Reviews, vol. 35, pp. 315-325, 2005.

[14] M. G. Sawyer, et al., "The Mental Health of Young People in Australia," M. H. a. S. P. Branch, Canberra: *Commonwealth of Australia*, 2000.

[15] L. Sheerber, H. Hops, B. Davis, "Family processes in adolescent depression," *Clin Child Fam Psychol Rev,* vol. 4(1), 19-35, 2001.

[16] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 586-591, 1991.

[17] M. B. H. Yap, et al., "Interaction of Parenting Experiences and Brain Structure in the Prediction of Depressive Symptoms in Adolescents," *Arch Gen Psychiatry*, vol. 65, pp. 1377-1385, December 1, 2008.