Enhanced Depression Detection from Facial Cues Using Univariate Feature Selection Techniques

Swati Rathi[1]([✉]) [iD], Baljeet Kaur[1,2] [iD], and R. K. Agrawal[1] [iD]

[1] School of Computer and Systems Sciences, JNU, Delhi, India
swatirathi362@gmail.com, baljeetkaur26@hotmail.com, rkajnu@gmail.com
[2] Hansraj College, University of Delhi, Delhi, India

Abstract. Timely detection of depression and the accurate assessment of its severity are the two major challenges that face the medical community. To assist the clinicians, various objective measures are being explored by researchers. In literature, features extracted from the images or videos, are found relevant for detection of depression. Various feature extraction methods are suggested in literature. However, the high dimensionality of the features so obtained provide an overfitted learning model. This is handled in this work with the help of three popular univariate filter feature selection methods, which identify the reduced size of relevant subset of features. The combinations of univariate techniques with well-known classification and regression techniques are investigated. The performance of classification and regression techniques improved with the use of feature selection methods. Moreover, the proposed model has outperformed most of the video-based existing methods for identifying depression and determining its level of severity.

Keywords: Classification · Depression · Motion History Image · Regression · Univariate feature selection · Visual features

1 Introduction

Depression is a psychological disorder attributed to the presence of low mood and disinclination towards routine activities for a period generally longer than two weeks. It negatively impacts a persons well-being and is known to increase the risk of suicidal tendencies [1]. Of the 300 million people affected by depression globally, 57 million people (18%) belong to India [18]. Depression may be self-assessed using Patient Health Questionnaire (PHQ) [13] or may be determined through clinical interviews, which are based on Hamilton Depression Rating Scale [17]. These methods suffer from subjectivity and bias as they depend upon the honesty and willingness of the patient during interaction, and the clinicians ability to interpret the subjects response [25]. According to WHO, more than half of those affected by depression are misdiagnosed, thereby, increasing the false

positives and false negatives [1]. Hence, it is desirable to build a decision support system to aid the clinicians in making an accurate depression assessment, based on objective behavioral markers such as, the speech properties, facial expressions and body gestures which are less likely to be suppressed by the patient [10].

Research community has studied the role of non-verbal behavior like facial emotions, speech and semantic information for depression diagnosis and have established their correlation with depression [4,5,15]. Some of the suggested methods for depression detection are unimodal [19,29], while others are multimodal, that combine two or more modalities [5,15]. Though the multimodal systems perform better, they entail high time and space complexity. Moreover, acquisition of data from multiple modalities incurs high cost. Hence, it is desirable to build a unimodal depression detection system that is cost effective, simple and efficient, in terms of acquisition as well as building the model.

Many patients suffering from depression are either unable to articulate their feelings or hesitate in discussing them with the clinician. In such situations, facial expressions used for determining the human emotions [24] can be helpful. Mehrabian et al. [14] stated that in our day to day interaction, facial expressions are responsible for 55% of the total information exchange, while language is responsible for only 7% of the daily interaction. Hence, in our work, we propose to strengthen the unimodal depression detection system based on only the facial cues obtained through video recordings.

In literature, various features from the face have been extracted and studied for their correlation with depression [21]. Recently, Pampouchidou et al. [20], proposed a novel method: Landmark Motion History Image (LMHI) to represent the movement of facial landmarks across the video frames. They concluded that Histogram of Oriented Gradient (HOG) features extracted from the LMHI: FaceHOGs are the most relevant features for depression classification, with an F1 score of 0.5 and 0.9, for detecting the depressed and the non-depressed category respectively. The experiment had been performed on the Distress Analysis Interview Corpus - Wizard of Oz (DAIC-WOZ) dataset [23]. The high dimensionality of the FaceHOG feature set may be the reason for the low F1 depressed score. This can be handled effectively using feature selection techniques.

To the best of our knowledge, application of feature selection has not been explored much for depression detection using videos. In our work, we investigate three popular univariate filter feature selection techniques: Fisher Discriminant Ratio (FDR) [8], Mutual Information (MI) [7] and Pearson Correlation (PC) [22], to find the relevant set of FaceHOG features [20]. It is well-known that the learning algorithm plays a key role in the development of the decision model to achieve high performance. Since filter feature selection determines the relevant features independent of the learning algorithm, it becomes important to investigate which combination of univariate feature selection method and learning algorithm provides the maximum performance. Hence, we explored four well-known classifiers: Decision Tree (DT), Linear Discriminant Analysis (LDA), k Nearest Neighbor (KNN) and Support Vector Machine (SVM) [8]. In literature, many of the depression detection models are built as regression problem. Hence, we have also investigated four well-known regression techniques for the

determination of depression severity level: Decision Tree (DTR), Linear (LR), Partial Least Square (PLSR) and Support Vector (SVR) [8]. Through exhaustive experiments, we determine the most suitable feature selection technique, the best performing classification/regression technique and the best combination of the univariate feature selection method and the classification/regression technique. Section 2 summarizes the related work and Sect. 3 presents the experiments and results. Finally, Sect. 4 concludes our work and gives future directions.

2   Related Work

Facial cues have been widely studied in correlation with the mental state of the person [9]. Cohn et al. [4] extracted features using facial action units and Active Appearance Model for prediction of depression. Meng et al. [15] captured facial dynamics from videos in a Motion History Histogram (MHH) image and computed Local Binary Pattern (LBP) and Edge Oriented Histogram (EOH) features for the depression prediction. Cummins et al. [5] extracted Space-Time Interest Points and Pyramid of HOG features from the videos to estimate depression. Jan et al. [12] extracted LBP, EOH and Local Phase Quantization (LPQ) features for each video frame and captured their change across frames using 1-D MHH. Dimension of the resulting 1-D MHH features was reduced using Principal Component Analysis. In [6], facial and head movements were computed from the facial landmarks of a video and Min-Redundancy Max-Relevance method was used to select the relevant features. Nasir et al. [16] computed polynomial parameterization of the visual features and reduced their dimensionality using Mutual Information Maximization, for depression classification. Yang et al. [30] proposed Histogram of Displacement range method to compute features from the facial landmarks. Combination of Deep Convolutional Neural Network (DCNN) and deep neural network model was used for depression detection. Pampouchidou et al. [19] compared the performance of LMHI with Motion History Image and Gabor-inhibited LMHI. LBP, HOG and LPQ features were extracted from these images for depression prediction. Zhu et al. [31] extracted appearance features from static frames of video and motion features from the optical flow images using DCNN for depression detection. Jazaery et al. [2] used 3D CNN and Recurrent Neural Network to learn the spatio-temporal features from videos for depression detection.

Fig. 1. Landmark Motion History Image of a healthy person with Id 310

Fig. 2. Landmark Motion History Image of a depressed person with Id 321

3   Combination of Univariate Feature Selection and the Learning Algorithms

Pampouchidou et al. [20] extracted many visual feature sets viz. FaceLBP, Face-HOG, head motion, blinking rate etc. using the 2D facial landmarks given for each video frame. Using the method [20], we construct the LMHI of size 252×248 pixels, that represents the motion of 2D facial landmarks. Figures 1 and 2 are two example LMHI. To extract the HOG features from the image, it is divided into cells of size 32×32 pixels each and, gradient is computed for each pixel of the cell. A 9 bin histogram is then created to represent the contribution of all the pixels in a cell. To account for illumination variance, histogram normalization is done on blocks of size 2×2 cells. Cell histograms from all the blocks are concatenated to form the FaceHOG feature set of dimension 1296. Pampouchidou et al. concluded that FaceHOG features are the most relevant for classifying depression with an F1 Score of 0.5 and 0.9 for the depressed and non-depressed category respectively. However, dimensionality of the FaceHOG features is high in relation to the number of samples in the DAIC-WOZ dataset. This may cause overfitting of the decision model [3] and can possibly be the reason for a low F1 score for identifying the depressed individuals. To circumvent this problem, it is imperative to reduce the dimension of FaceHOG feature set. Using only a relevant subset of the features, helps to enhance the performance of the decision system. Several filter and wrapper methods are suggested in literature for feature selection [11]. Filter feature selection methods are much simpler than the wrapper approaches and help build a cost-effective system in terms of time and space. To our best knowledge, the univariate filter feature selection techniques have not been explored much for video-based depression detection. In this work, we determine a relevant subset of the FaceHOG features by using the three univariate filter feature selection techniques: Fisher Discriminant Ratio (FDR) [8], Mutual Information (MI) [7] and Pearson Correlation (PC) [22]. After obtaining the subset of relevant features, we apply four well-known classification techniques: DT, LDA, KNN and SVM, and four regression techniques: DTR, LR, PLSR and SVR, to determine the best combination of feature selection and learning algorithm, for effective depression identification (classification) and its severity estimation (regression) respectively.

4   Experiments and Results

All the experiments have been performed on the DAIC-WOZ dataset [23]. The data given had been partitioned into the training, development and test sets. Classification labels (depressed or non-depressed) and PHQ-8 scores (for regression) were given for all the sets except the test set, hence we train our model on the training set (107 samples) and test its efficacy on the development set (35 samples). Each univariate filter feature selection method, computes the relevance of each FaceHOG feature w.r.t. the response variable and ranks them in the descending order of their relevance score. The decision model is learned in the order of the ranked features incrementally and the minimum number of features that give the best performance are finally selected (#).

Table 1. Classification comparison

	Metric	WFS[a]	FDR	MI	PC
DT	F1 dep	0.40	**0.54**	0.44	0.44
	F1 ndep	0.67	**0.79**	0.72	0.72
	#	1296	125	**11**	39
LDA	F1 dep	0.18	**0.70**	0.62	0.64
	F1 ndep	0.62	**0.85**	0.73	0.76
	#	1296	**116**	209	495
KNN	F1 dep	0.23	**0.61**	0.40	0.54
	F1 ndep	0.73	**0.81**	0.76	0.79
	#	1296	**23**	90	685
SVM	F1 dep	0.60	**0.61**	0.50	0.58
	F1 ndep	**0.80**	0.77	0.73	0.78
	#	1296	**4**	43	5

[a] Without Feature Selection

Table 2. Regression comparison

	Metric	WFS[a]	FDR	MI	PC
DTR	MAE	6.58	**4.97**	5.12	5.05
	RMSE	7.99	**6.38**	6.61	6.8
	#	1296	216	**9**	33
LR	MAE	11.5	4.93	5.45	**4.64**
	RMSE	14.6	6.06	6.64	**5.98**
	#	1296	10	**3**	18
PLSR	MAE	5.53	**5.09**	5.19	5.15
	RMSE	6.84	6.41	6.76	**6.31**
	#	1296	24	61	**1**
SVR	MAE	5.45	**5.11**	5.2	5.15
	RMSE	6.73	**6.5**	6.62	6.53
	#	1296	792	364	**154**

[a] Without Feature Selection

Performances of the three feature selection techniques (FDR, MI, PC), in conjunction with the four classification techniques (DT, LDA, KNN, SVM) and the four regression techniques (DTR, LR, PLSR, SVR) are compared in Tables 1 and 2 respectively. The classification performance is shown in terms of F1 depressed score (F1 dep), F1 non-depressed score (F1 ndep) and the regression performance is shown in terms of Mean Absolute Error (MAE), Root Mean Square Error (RMSE). On application of the three univariate methods, the minimum number of features (#) for which maximum performance is achieved (F1 dep for classification and MAE for regression), is recorded. With each learning method, feature selection technique that gives the best performance is highlighted in bold. Following are the observations based on Table 1:

- Application of feature selection improves the performance of the model, except in the case of MI and PC with SVM.
- For each classification technique, FDR outperforms MI and PC in terms of F1 dep, F1 ndep and the number of selected features (except for DT).
- For each feature selection technique, LDA outperforms all other classifiers.
- FDR followed by LDA gives the best F1 Score (depressed and non-depressed). Both FDR and LDA are based on the Fisher criterion. However, FDR is unable to perform feature combination like LDA and LDA is unable to discard the irrelevant features like FDR. They complement each other and the combination gives better performance.

Following observations are based on Table 2:

- Use of feature selection improves the performance of all regression methods.
- FDR outperforms MI and PC in combination with all the regression techniques except with LR.

Table 3. Comparison of the proposed model with the state-of-the-art

Classification F1 depressed (F1 Non-Depressed)	Regression MAE (RMSE)
0.70 (0.85) Proposed	4.60 (5.90) [26]
0.63 (0.89) [16]	**4.64 (5.98) Proposed**
0.62 (0.77) [26]	5.33 (6.45) [28]
0.53 (mean) [28]	5.88 (7.13) [23]
0.50 (0.90) [20]	6.47 (7.86) [16]
0.50 (0.90) [27]	

– PC based feature selection followed by LR gives the best performance in terms MAE and RMSE. PC selects those features which have high degree of linear correlation with the response variable, and LR models the linear relationship between features and the response variable. Hence, the combination of the two provides better performance.

Table 3 compares the best results of the proposed combination technique with the existing methods for depression detection based on visual cues. The proposed combination of FDR and LDA outperforms all the classification models suggested in literature. For, depression severity estimation, the combination of PC and LR outperforms most of the existing regression models.

5 Conclusion

Features captured from video data are relevant for depression detection. They are a strong contender as a unimodal technique, that is capable of supporting clinicians for monitoring patients and correctly assessing the severity of their problem. Due to the high dimensionality of the extracted features from videos, the complexity of the decision models built for depression detection is high. Also, it provides an overfitted learning model. To circumvent this, we have employed univariate filter feature selection methods to reduce the dimensionality of features required to build the depression detection systems. Four well known classifiers and four regression methods have been successfully explored in combination with the feature selection techniques, and the role of feature selection has been emphasized. The relevant features obtained using FDR are transformed by the LDA classifier making the combination of FDR and LDA most appropriate for video-based depression classification. To diagnose the depression severity, PC in combination with the LR has been found to be the most suitable as both PC and LR are based on the linear correlation between the features and the response variable. The proposed combinations for classification and regression for video-based depression detection, outperform most of the existing results.

Future work will focus on reducing the dimension of other visual features obtained from the video data, and identify those features which are relevant for

the task of depression detection. We will also explore advanced feature selection techniques which will not only eliminate irrelevant features with respect to the response variable, but also remove the redundant/correlated features.

References

1. Depression. https://www.who.int/news-room/fact-sheets/detail/depression. Accessed 29 Apr 2019
2. Al Jazaery, M., Guo, G.: Video-based depression level analysis by encoding deep spatiotemporal features. IEEE Trans. Affect. Comput. (2018). https://doi.org/10.1109/TAFFC.2018.2870884
3. Bellman, R.: Curse of Dimensionality. Adaptive Control Processes: A Guided Tour. Princeton University Press, Princeton (1961)
4. Cohn, J.F., et al.: Detecting depression from facial actions and vocal prosody. In: 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, pp. 1–7. IEEE (2009)
5. Cummins, N., Joshi, J., Dhall, A., Sethu, V., Goecke, R., Epps, J.: Diagnosis of depression by behavioural signals: a multimodal approach. In: Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge, pp. 11–20. ACM (2013)
6. Dibeklioğlu, H., Hammal, Z., Yang, Y., Cohn, J.F.: Multimodal detection of depression in clinical interviews. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 307–310. ACM (2015)
7. Ding, C., Peng, H.: Minimum redundancy feature selection from microarray gene expression data. J. Bioinf. Comput. Biol. **3**(02), 185–205 (2005)
8. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. Wiley, Hoboken (2012)
9. Ekman, R.: What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). Oxford University Press, Oxford (1997)
10. Ellgring, H.: Non-Verbal Communication in Depression. Cambridge University Press, Cambridge (2007)
11. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**(Mar), 1157–1182 (2003)
12. Jan, A., Meng, H., Gaus, Y.F.A., Zhang, F., Turabzadeh, S.: Automatic depression scale prediction using facial expression dynamics and regression. In: Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge, pp. 73–80. ACM (2014)
13. Kroenke, K., Spitzer, R.L., Williams, J.B.: The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. **16**(9), 606–613 (2001)
14. Mehrabian, A., Russell, J.A.: An Approach to Environmental Psychology. The MIT Press, Cambridge (1974)
15. Meng, H., Huang, D., Wang, H., Yang, H., Ai-Shuraifi, M., Wang, Y.: 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)
16. Nasir, M., Jati, A., Shivakumar, P.G., Nallan Chakravarthula, S., Georgiou, P.: Multimodal and multiresolution depression detection from speech and facial landmark features. In: Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, pp. 43–50. ACM (2016)

17. Nutt, D.: The Hamilton depression scale- accelerator or break on antidepressant drug discovery? J. Neurol. Neurosurg. Psychiatry **85**, 119–120 (2014). https://doi.org/10.1136/jnnp-2013-306984

18. Organization, W.H., et al.: Depression and other common mental disorders: global health estimates. Technical report, World Health Organization (2017)

19. Pampouchidou, A., et al.: Quantitative comparison of motion history image variants for video-based depression assessment. EURASIP J. Image Video Process. **2017**(1), 64 (2017)

20. Pampouchidou, A., et al.: Depression assessment by fusing high and low level features from audio, video, and text. In: Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, pp. 27–34. ACM (2016)

21. Pampouchidou, A., et al.: Automatic assessment of depression based on visual cues: a systematic review. IEEE Trans. Affect. Comput. (2017). https://doi.org/10.1109/TAFFC.2017.2724035

22. Pearson, K.: Notes on the history of correlation. Biometrika **13**(1), 25–45 (1920)

23. Ringeval, F., et al.: AVEC 2017: real-life depression, and affect recognition workshop and challenge. In: Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge, pp. 3–9. ACM (2017)

24. Sariyanidi, E., Gunes, H., Cavallaro, A.: Automatic analysis of facial affect: a survey of registration, representation, and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **37**(6), 1113–1133 (2014)

25. Schumann, I., Schneider, A., Kantert, C., Löwe, B., Linde, K.: Physicians attitudes, diagnostic process and barriers regarding depression diagnosis in primary care: a systematic review of qualitative studies. Fam. Pract. **29**(3), 255–263 (2011)

26. Sun, B., et al.: A random forest regression method with selected-text feature for depression assessment. In: Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge, pp. 61–68. ACM (2017)

27. Valstar, M., et al.: AVEC 2016: depression, mood, and emotion recognition workshop and challenge. In: Proceedings of the 6th International Workshop on Audio/visual Emotion Challenge, pp. 3–10. ACM (2016)

28. Williamson, J.R., et al.: Detecting depression using vocal, facial and semantic communication cues. In: Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, pp. 11–18. ACM (2016)

29. Williamson, J.R., Quatieri, T.F., Helfer, B.S., Horwitz, R., Yu, B., Mehta, D.D.: Vocal biomarkers of depression based on motor incoordination. In: Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge, pp. 41–48. ACM (2013)

30. Yang, L., Jiang, D., Xia, X., Pei, E., Oveneke, M.C., Sahli, H.: Multimodal measurement of depression using deep learning models. In: Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge, pp. 53–59. ACM (2017)

31. Zhu, Y., Shang, Y., Shao, Z., Guo, G.: Automated depression diagnosis based on deep networks to encode facial appearance and dynamics. IEEE Trans. Affect. Comput. **9**(4), 578–584 (2017)