INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

An Affect Prediction Approach through Depression Severity Parameter Incorporation in Neural Networks

Rahul Gupta[*], *Saurabh Sahu*[+], *Carol Espy-Wilson*[+], *Shrikanth Narayanan*[o]

[*]Amazon.com, USA
[+]Speech Communication Laboratory, University of Maryland, College Park, MD, USA
[o]Signal Analysis and Interpretation Lab, University of Southern California, Los Angeles, CA, USA

Abstract

Humans use emotional expressions to communicate their internal affective states. These behavioral expressions are often multi-modal (e.g. facial expression, voice and gestures) and researchers have proposed several schemes to predict the latent affective states based on these expressions. The relationship between the latent affective states and their expression is hypothesized to be affected by several factors; depression disorder being one of them. Despite a wide interest in affect prediction, and several studies linking the effect of depression on affective expressions, only a limited number of affect prediction models account for the depression severity. In this work, we present a novel scheme that incorporates depression severity as a parameter in Deep Neural Networks (DNNs). In order to predict affective dimensions for an individual at hand, our scheme alters the DNN activation function based on the subject's depression severity. We perform experiments on affect prediction in two different sessions of the Audio-Visual Depressive language Corpus, which involves patients with varying degree of depression. Our results show improvements in arousal and valence prediction on both the sessions using the proposed DNN modeling. We also present analysis of the impact of such an alteration in DNNs during training and testing.

Index Terms: Depression, Affect prediction, Deep Neural Networks

1. Introduction

Humans use a variety of cues such as facial expressions, gestures and non-verbal vocalizations to express their internal affective states. The prediction of affective dimensions based on such cues is a classical problem in the field of emotion recognition [1,2]. Research also links the impact of several factors onto the affective expression of a person, such as mental health [3], family [4] and social factors [5]. Depression [6, 7] is one such factor that influences emotion expressions. The depression disorder has been shown to impact emotion regulation [8], social information processing [9] and emotional reactivity [10] of a person. Tracking affective states of subjects suffering from depression hence is a crucial problem as it is often associated with events such as extreme emotional expression as well as emotion insensitivity [11, 12]. Despite the existence of several studies correlating depression with affective expression, a limited number of affect prediction models account for depression severity during inference. For instance, in [13], the authors perform a feature transformation conditioned on depression severity before training/testing the affect prediction model. However such a design is ad-hoc as the model optimization is carried out independently of the depression severity incorporation into the features. In this work, we propose a novel Deep Neural Network (DNN) architecture to incorporate the severity of depression within the prediction of affective dimensions (valence, arousal and dominance). The training scheme allows for a feature transformation, but also performs the model optimization using depression severity as a model parameter. In addition to the incorporation of the constant depression severity index value into the DNN model, we further aim to understand the impact of the depression severity on the DNN weights during training as well as the final predictions.

Previoius work: Affective dimension prediction is a widely studied problem in the emotion research community [14, 15]. These works focus on the prediction [16] as well as analysis of emotion along the dimensions of valence, arousal and dominance [17]. Previously proposed prediction models include the use of ensemble methods [18], Kalman filter [19] as well as the use of deep learning models [20]. Studies have also jointly investigated depression and affective expression in multi-cultural settings [21], for the purpose of understanding emotion regulation [6] as well as understanding the perception of emotion expression [22]. The Audio/Visual Emotion Challenge (AVEC) [23, 24] has promoted the study of emotions and depression, with specific focus on affect and depression prediction. Proposed methods for affect prediction include the use ensemble CCA [25] and Recurrent Neural Networks [26], while methods for depression prediction include Fisher vector encoding [27] and use of application dependent meta knowledge [28]. Gupta et al. [13] proposed feature transformation based on depression severity to train an affect prediction system. However, the model training is performed independently after the feature transformation. In this work, we propose a DNN model that uses depression severity as a parameter and the model learning is directly dependent on the depression severity (unlike through feature transformation as in [13]). Furthermore, the proposed model could accept transformed features as inputs. Apart from these contributions, another novelty of our work is the incorporation of a scalar parameter into the DNN architecture to predict a time series. This scheme could be applied more generally to similar problems involving time series prediction conditioned on a set of hyper-parameters.

We perform our experiments on the Audio-Visual Depressive language Corpus (AViD-Corpus) [23] involving subjects with an available depression severity assessment. Correlation analysis in [13] (Table 1) shows that there exist significant correlations between the BDI-II depression index and affect ratings on the AViD-Corpus. Given that depression severity does carry information about the affective state of a person, we test several schemes for incorporating depression severity in this work. Initially, we develop a model to predict affective dimensions from the audio-visual cues from the subject at hand. Then we incorporate the depression severity assessment into affect prediction using feature transformation and propose a method to include depression severity as a model parameter. Our re-

http://dx.doi.org/10.21437/Interspeech.2017-120

sults indicate that incorporation of the depression severity as a model parameter obtains significant improvements (over baseline models without the incorporation of depression severity or with feature transformation only) for valence and arousal prediction on two separate sessions of the AViD-Corpus, each with a different recording protocol. In the next section, we provide further details about the AViD-Corpus followed by a description of the features and methodology.

2. Database

The AViD-Corpus was also used as a part of the Audio/Visual Emotion and Depression Recognition (AVEC) challenges 2013 and 2014. We use the data split used in the AVEC 2014 challenge consisting of two sets of sessions: Freeform and Northwind sessions. Both these sessions involve a human-computer interaction with the Freeform sessions consisting of unscripted response to a question on part of the human; while the Northwind sessions require the participant to recite a predefined excerpt. Each of these sessions contain 150 videos, with a training/development/testing split consisting of 50 videos in each partition. Each video is continuously rated for three affective dimensions of valence, arousal and dominance at a frame rate of 30 Frames Per Second (FPS) by a set of 3-5 annotators. The final ground truth affect ratings are computed as the frame-wise mean over the annotator ratings for a given session. Apart from the continuous time-series affect ratings, the subjects in the sessions also complete the standardized self-assessment based Beck Depression Inventory-II (BDI-II) questionnaire [29]. The questionnaire contains a set of 21 questions and a final BDI-II index score is computed for each subject based on his/her responses. The score ranges between 0-63, with a higher score implying more severe depression. The self-assessment protocol for BDI-II is crucial as this makes our approach scalable, without the requirement for a specialized/professional depression severity assessment. In the next section, we describe the set of features used in our study followed by the modeling schemes.

3. Multi-modal features

The goal of our study is to predict the continuous affective dimensions conditioned on the availability of a set of audio-visual cues and the depression severity assessment in the form of BDI-II index. We provide a description of the set of audio-visual features used in our study below.

3.1. Audio features

We use a set of energy, spectral and voicing related features, as were used in AVEC 2013-2014 challenges. Table 2 in [23] provides a detailed list of these features. These features are extracted at a frame rate of 100 FPS.

3.2. Video features

We use a set of Local Binary Pattern on Three Orthogonal Planes (LBP-TOP) features [30] along with optical flow based motion vector features and features derived from facial landmarks using CSIRO Face analysis SDK [31]. The LBP-TOP features are extracted in the pixel domain per frame along three planes: the spatial (xy) and two temporal planes (xt and yt). Note that the LBP-TOP features used in our experiments are different from the LGBP-TOP features used in the AVEC challenges. The LGBP-TOP features are computed in the Gabor domain leading to a high feature dimensionality. Since our experimentation is based on training neural networks and a high feature dimensionality leads to a model with larger number of parameters, we stick to the LBP in the pixel domain. More details on the video features used in our work can be found in [16].

Note that the audio features are obtained at a higher frame rate than the annotation and video feature frame rate. We downsample the audio features to the annotation/video features frame rate (30 FPS) by averaging their value over every $1/30$ seconds. In the next section, we discuss our experimental methodology.

4. Experiments

We test several experimental methodologies to predict the affective dimensions based on the multi-modal features and the BDI-II index. Our baseline model is a neural network trained to predict the affective dimensions based on the multi-modal features only. We then test two schemes to introduce the depression assessment in the prediction models: (i) using feature transformation based on the BDI-II score and, (ii) altering the neural network architecture to introduce the BDI-II score. We use the correlation coefficient (ρ) between the predicted and true ratings for each affective dimension on the testing set as our evaluation metric (also used in the AVEC challenges 2013-2014 [23, 24]). We discuss our baseline affect prediction model along with the proposed models in detail below.

4.1. Baseline: Affect prediction based on multi-modal features

Our baseline system is a DNN regressor trained on the multi-modal features to predict the three affective dimensions. The model is optimized on the training set to minimize the squared error loss between true and predicted affect ratings. The number of hidden layers and neurons in the hidden layers are tuned on the development set. The hidden layers have a $\tanh$ activation, while the output layer contains linear activation units. As the affective ratings evolve smoothly over time, we further smooth the predictions from the neural network using a low pass temporal filter. The smoothing operation has been shown to improve the prediction in affective dimensions in [16]. We use a moving average filter in our work, with the length of the filter also tuned on the development set. Next, we describe the introduction of depression severity in prediction of affective ratings.

4.2. Introducing depression severity in the feature space

In order to introduce the scalar BDI-II index score in continuous tracking of the affective dimensions, we first experiment with modifying/augmenting the multi-modal features themselves based on the depression severity. We briefly describe our methods for introduction of depression severity score in the feature space below.

4.2.1. T0: Adding depression score as a feature

In this method, we add the depression score as an additional feature to the existing set of frame-wise multi-modal features. Therefore, the depression score is directly added as a source of information in the feature space. Note that the additional feature will be constant per frame for a given session, and varies only across different sessions.

4.2.2. Feature transformation based on the depression score

In this method, we transform the existing multi-modal feature space based on the depression severity index. Previously, Gupta et al. [13] have proposed a few feature transformation techniques for the same application. Although one could use one of the several feature transformation techniques [32], the authors proposed transforming feature means and variances based on the depression severity. In our work, we test three schemes to

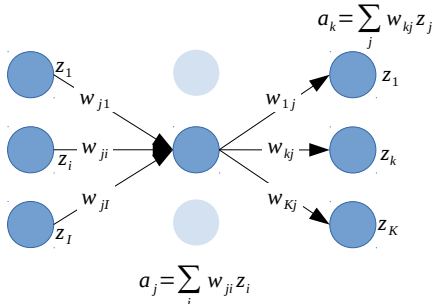

Figure 1: Figure setting the notation for backpropagation derivation as borrowed from [36] (section 5.3)

transform features mean and/or variances, identical to the ones in [13], as discussed below.

(a) **T1; Transforming feature means based on the depression severity score**: In this transformation, we alter the feature means based on the depression severity. Given the BDI-II index d_s and multi-modal feature vector $\boldsymbol{x}_s(t)$ for a session s at the t^{th} analysis frame, we compute the transformed features $\boldsymbol{x}'_s(t)$ as shown in equation 1. This transformation changes the features means for a session s by a factor of $k_m * d_s$. $\mathbf{1}$ is a vector of 1's, of the same dimensionality as the feature vector $\boldsymbol{x}_s(t)$ and k_m is a constant to scale the depression score. The neural network is trained on the transformed features $x'_s(t)$ and k_m is tuned for the best performance on the development set.

$$\boldsymbol{x}'_s(t) = \boldsymbol{x}_s(t) + k_m * d_s \times \mathbf{1} \tag{1}$$

(b) **T2; Transforming feature variances based on the depression severity score**: In this transformation, we alter the feature variances based on the depression score, by performing the operation shown in equation 2. The parameter k_{d1} scales the depression score, while the parameter k_{d2} is tuned between $\{-1, 1\}$ so as to inversely or directly scale the features based on the depression score. For a session s, the variance for its features are scaled by $(k_{d1} * d_s)^{2*k_{d2}}$. The neural network is optimized on the transformed features from the training set and the parameters k_{d1}, k_{d2} are tuned for the best performance on the development set.

$$\boldsymbol{x}'_s(t) = (k_{d1} * d_s)^{k_{d2}} \times \boldsymbol{x}_s(t) \tag{2}$$

(c) **T3; Transforming feature means and variances based on the depression severity score**: This transformation alters both the means and variances of the feature set based on the depression severity as depicted in equation 3. The neural network is optimized on the transformed features from the training set, while the parameters k_m, k_{d1}, k_{d2} are tuned for the best performance on the development set.

$$\boldsymbol{x}'_s(t) = (k_{d1} * d_s)^{k_{d2}} \times \boldsymbol{x}_s(t) + k_m * d_s \times \mathbf{1} \tag{3}$$

After training the DNN based on the extended/transformed feature space, we perform the smoothing operation as discussed in the baseline section 4.1. The length of the MA filter is separately tuned on the development set for each modeling scheme.

4.3. M1: Incorporating depression severity as a DNN parameter

In the schemes presented above, the depression score was used to alter the feature and the model was trained on these features. Once trained, the model parameters remain constant for every session in the testing set. In this section, we propose a scheme

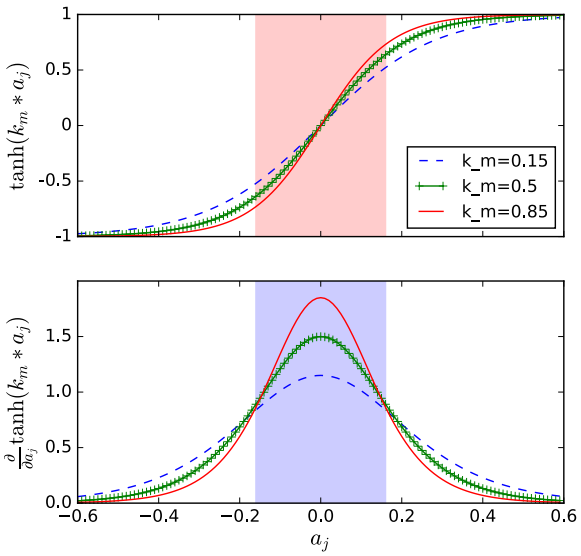

Figure 2: Figure representing the activation function (equation 4) and it's derivative for different values of k_m

for the transformation of the neural network model based on the depression severity score. Specifically, we alter the activations of the hidden layers based on the depression severity d_s for the session s. This scheme is motivated from studies investigating neuro-physiological changes due to depression, leading to alterations in motor control [33, 34]. This has an impact on the affective expressions of a person as depression is hypothesized to effect mechanisms controlling facial and vocal expressions. We aim to capture these neuro-physiological changes in this scheme by incorporating the depression severity score and then optimizing the affect prediction model. We also acknowledge that our method for incorporating a parameter in DNNs is also inspired from schemes such as incorporating I-vector inputs in Convolutional Neural Networks for acoustic modeling [35].

Borrowing the notations for error backpropagation from [36] (Section 5.3), the activation z_j for a neural unit j is given as: $z_j = h(a_j)$, where a_j is input the unit j (depicted in Figure 1). a_j is computed as $a_j = \sum_i w_{ji} z_i$, where z_i are the activation of the units from the previous layer and w_{ji} are the weights for incoming connections to neural unit j. In our previous models, we chose $h(a_j) = \tanh(a_j)$ for the hidden layer. During model transformation, we modify the neural network such that the activation is contingent on the depression score d_s for the session s, as shown in equation 4.

$$y = \tanh(k_m * z); \text{where } \; k_m = 1 + (k_{m1} * d_s)^{k_{m2}} \tag{4}$$

We briefly discuss the implications of this alterations during training and testing in the sections below.

(a) Effect on training: We train the DNN based on the backpropagation algorithm [36], involving iterative forward propagation and a weight update steps. During the forward propagation, the model output based on the current weight values and the features is computed. We plot the $\tanh$ activation function with different values of k_m in Figure 2 and divide the plot into two sections: (i) a dynamic section (shaded red), and (ii) an asymptotic section (unshaded). With the modified DNN activation functions, an increase in k_m leads to a steeper change in the activation function from -1 to 1 in the dynamic section of the ac-

tivation function. Therefore, as k_m increases, a small change in a_j leads to a greater change in z_j, if a_j is in the dynamic section of the $\tanh$ function. Outside the dynamic section, z_j tends to be closer to -1/+1 with a higher k_m. On the other hand, during weight updates, the weights for connections linking a layer i to the layer j are updated using gradient descent (for more details, please refer to [36]). In order to update the weights connecting node i to node j , the derivative of the Mean Squared Error E is computed with respect to the w_{ij} as follows:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial a_j}\frac{\partial a_j}{\partial w_{ji}} = h'(a_j)\sum_k w_{kj}\delta_k \frac{\partial a_j}{\partial w_{ji}}$$
$$\text{where} \quad \delta_k = \frac{\partial E}{\partial a_k}; h'(a_j) = \frac{\partial}{\partial a_j}\tanh(k_m * a_j) \tag{5}$$

The derivative is subtracted from the current weight estimates to get updated weights. Figure 2 plots the derivative with various values of k_m. We again divide the region into two sections corresponding to the (i) dynamic section, and (ii) asymptotic section. We observe that the derivative $h'(a_j)$ is higher for a larger k_m in the dynamic region. Therefore, when a_j lies in the dynamic section of the activation function, a higher k_m encourages a higher weight update contribution for the corresponding training data point. In the asymptotic regions, $h'(a_j)$ is larger for lower k_m, however the differences in the magnitude of $h'(a_j)$ is not as large as in the dynamic section. a_j values in the asymptotic domain of the activation functions have a marginally greater impacts during weight updates, as k_m decreases.

(b) Effect during testing: During testing, a DNN produces predictions on the inputs by performing one step of forward propagation. As discussed above, a higher k_m leads the model to be more sensitive to changes in inputs region of the activation function. In the asymptotic section, outputs are closer to -1/+1.

Note that it is possible to train the transformed DNN model on the original features as well as after applying one of the feature transformations discussed in section 4.2. We tune for the best feature transformation/ no feature transformation on the development set for the Freeform and Northwind sessions. We also smooth the outputs prediction from the transformed DNN model, with the length of the MA filter tuned on the development set. The parameters k_{m1} and k_{m2} are also tuned on the development set. In the next section, we present results using the various modeling schemes.

5. Results

We perform separate training and testing for the Northwind and Freeform datasets due to the inherent differences in the nature of these datasets. We present our results in Table 1. Our evaluation metric is the correlation coefficient ρ between the true and predicted ratings for the three affective dimensions.

From the results, we observe that not all the methods for introduction of depression severity in affect prediction models beat the baseline. However, the model transformation scheme is the best in valence and arousal prediction. The results on prediction of dominance are however not significantly different from the baseline system. Overall, the improvement in prediction of valence and arousal using model transformation is encouraging and we further discuss the results and model settings in the next section.

5.1. Discussion

During parameter tuning, we observed that the parameter k_{m2} is tuned to be -1. We also observed that this configuration is also significantly better than the one with k_{m2} equaling 1. This implies that for the subjects with a lower depression value the output from the hidden nodes tends to be more sensitive to the inputs (in the dynamic region of inputs). Table 1 in [13] suggests that the variance of affective dimension time series is negatively correlated with depression severity. The tuned value of k_{m2} is consistent with this observation wherein as the depression severity increases, the output varies lesser with change in the input features. For the transformation T2 also, the value of k_{d2} is also tuned to be -1. Therefore the variance of input feature decreases as the depression severity increases. Although the DNN model is a non-linear model, lowering the feature variance for high depression helps lowering the variance of predicted affect, thereby improving the performance.

Table 1: Results (ρ) for the baseline and various depression severity incorporation schemes for affect prediction. Bold numbers show the best performance for a dimension and also significantly better than the baseline (Student's t-statistics test, p-value$< 5\%$ with number of samples = number of frames).

	Freeform			Northwind		
	Aro.	Val.	Dom.	Aro.	Val.	Dom.
Baseline	0.48	0.38	0.31	0.38	0.26	0.21
T0	0.50	0.30	0.16	0.41	0.30	0.21
T1	0.42	0.35	0.26	0.32	0.27	0.20
T2	0.46	0.38	0.30	0.43	0.32	0.21
T3	0.45	0.35	0.22	0.36	0.24	0.20
M1	**0.51**	**0.40**	0.30	**0.45**	**0.32**	0.21

We also observe that our model does not improve upon the baseline for dominance. Table 1 in [13] shows that depression has no significant correlation with dominance statistics for the Northwind data, while the correlation is the weakest in Freeform dataset amongst the three affective dimensions. The proposed model transformation method is not significantly different from baseline for dominance prediction, however does perform significantly better for the valence and arousal. This observation is consistent across the two datasets entailing different recording protocols, indicative of the capability of the proposed model across datasets.

6. Conclusion

Despite affect prediction being a widely investigated problem, there is only a limited amount of work on affect prediction conditioned on depression severity. In this work, we proposed a novel DNN scheme that incorporates depression severity as a parameter during affect prediction. We use the depression severity score to alter the activation function of the hidden layer nodes during affect prediction for each subject (patient) at hand. We analyze the effects of high depression values during DNN training and prediction. Our results show that we obtain better performance in arousal and valence prediction on the AViD-Corpus: Freeform and Northwind sessions. The consistent increase in performance across these sessions promises the efficacy of our approach with further potential in similar tasks.

In the future, we aim to test a similar scheme in sequence prediction model. Presently, we have trained a DNN model with framewise inputs. We aim to test a similar schemes applicable to Recurrent Neural Networks [15]. We would also like to test our scheme on time series prediction in presence of multiple factors (e.g. affect time series prediction conditioned on depression severity and medical history). Finally, one could also extend the current scheme to other similar problems involving a time series generation conditioned on a global constant.

7. References

[1] E. Mower, A. Metallinou, C.-C. Lee, A. Kazemzadeh, C. Busso, S. Lee, and S. Narayanan, "Interpreting ambiguous emotional expressions," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. IEEE, 2009, pp. 1–8.

[2] H. Gunes and M. Pantic, "Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners," in *International conference on intelligent virtual agents*. Springer, 2010, pp. 371–377.

[3] J. Ciarrochi, F. P. Deane, and S. Anderson, "Emotional intelligence moderates the relationship between stress and mental health," *Personality and individual differences*, vol. 32, no. 2, pp. 197–209, 2002.

[4] E. D. Hibbs, S. D. Hamburger, M. J. Kruesi, and M. Lenane, "Factors affecting expressed emotion in parents of ill and normal children." *American Journal of Orthopsychiatry*, vol. 63, no. 1, p. 103, 1993.

[5] D. J. Rickwood and V. A. Braithwaite, "Social-psychological factors affecting help-seeking for emotional problems," *Social science & medicine*, vol. 39, no. 4, pp. 563–572, 1994.

[6] J. J. Gross and R. F. Muñoz, "Emotion regulation and mental health," *Clinical psychology: Science and practice*, vol. 2, no. 2, pp. 151–164, 1995.

[7] R. J. Davidson, *Anxiety, depression, and emotion*. Oxford University Press, 2000.

[8] T. Ehring, B. Tuschen-Caffier, J. Schnülle, S. Fischer, and J. J. Gross, "Emotion regulation and vulnerability to depression: spontaneous versus instructed use of emotion suppression and reappraisal." *Emotion*, vol. 10, no. 4, p. 563, 2010.

[9] A. M. Luebbe, D. J. Bell, M. A. Allwood, L. P. Swenson, and M. C. Early, "Social information processing in children: Specific relations to anxiety, depression, and affect," *Journal of Clinical Child & Adolescent Psychology*, vol. 39, no. 3, pp. 386–399, 2010.

[10] L. M. Bylsma, B. H. Morris, and J. Rottenberg, "A meta-analysis of emotional reactivity in major depressive disorder," *Clinical psychology review*, vol. 28, no. 4, pp. 676–691, 2008.

[11] J. Rottenberg and C. Vaughan, "Emotion expression in depression: Emerging evidence for emotion context-insensitivity," in *Emotion regulation*. Springer, 2008, pp. 125–139.

[12] J. Rottenberg, J. J. Gross, and I. H. Gotlib, "Emotion context insensitivity in major depressive disorder." *Journal of abnormal psychology*, vol. 114, no. 4, p. 627, 2005.

[13] R. Gupta and S. Narayanan, "Predicting affective dimensions based on self assessed depression severity," *Interspeech 2016*, pp. 1427–1431, 2016.

[14] M. A. Nicolaou, H. Gunes, and M. Pantic, "Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space," *IEEE Transactions on Affective Computing*, vol. 2, no. 2, pp. 92–105, 2011.

[15] L. F. Barrett, "Discrete emotions or dimensions? the role of valence focus and arousal focus," *Cognition & Emotion*, vol. 12, no. 4, pp. 579–599, 1998.

[16] R. Gupta, N. Malandrakis, B. Xiao, T. Guha, M. Van Segbroeck, M. Black, A. Potamianos, and S. Narayanan, "Multimodal prediction of affective dimensions and depression in human-computer interactions," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*. ACM, 2014, pp. 33–40.

[17] H. Gunes and B. Schuller, "Categorical and dimensional affect analysis in continuous input: Current trends and future directions," *Image and Vision Computing*, vol. 31, no. 2, pp. 120–136, 2013.

[18] M. Kächele, P. Thiam, G. Palm, F. Schwenker, and M. Schels, "Ensemble methods for continuous affect recognition: multimodality, temporality, and challenges," in *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge*. ACM, 2015, pp. 9–16.

[19] K. Somandepalli, R. Gupta, M. Nasir, B. M. Booth, S. Lee, and S. S. Narayanan, "Online affect tracking with multimodal kalman filters," in *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*. ACM, 2016, pp. 59–66.

[20] Y. Kim, H. Lee, and E. M. Provost, "Deep learning for robust feature generation in audiovisual emotion recognition," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013.

[21] M. Brandt and J. D. Boucher, "Concepts of depression in emotion lexicons of eight cultures," *International Journal of Intercultural Relations*, vol. 10, no. 3, 1986.

[22] 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," *Biological psychiatry*, vol. 40, no. 8, pp. 697–705, 1996.

[23] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, and M. Pantic, "Avec 2014: 3d dimensional affect and depression recognition challenge," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*. ACM, 2014, pp. 3–10.

[24] M. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, S. Schnieder, R. Cowie, and M. Pantic, "Avec 2013: the continuous audio/visual emotion and depression recognition challenge," in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*. ACM, 2013, pp. 3–10.

[25] H. Kaya, F. Çilli, and A. A. Salah, "Ensemble cca for continuous emotion prediction," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*. ACM, 2014, pp. 19–26.

[26] L. He, D. Jiang, L. Yang, E. Pei, P. Wu, and H. Sahli, "Multimodal affective dimension prediction using deep bidirectional long short-term memory recurrent neural networks," in *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge*. ACM, 2015.

[27] V. Jain, J. L. Crowley, A. K. Dey, and A. Lux, "Depression estimation using audiovisual features and fisher vector encoding," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*. ACM, 2014.

[28] M. Kächele, M. Schels, and F. Schwenker, "Inferring depression and affect from application dependent meta knowledge," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*. ACM, 2014.

[29] A. T. Beck, R. A. Steer, G. K. Brown *et al.*, "Manual for the beck depression inventory-ii," *San Antonio, TX: Psychological Corporation*, vol. 1, p. 82, 1996.

[30] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 6, 2007.

[31] M. Cox, J. Nuevo-Chiquero, J. Saragih, and S. Lucey, "Csiro face analysis sdk," *Brisbane, Australia*, 2013.

[32] A. Kusiak, "Feature transformation methods in data mining," *IEEE Transactions on Electronics packaging manufacturing*, vol. 24, no. 3, pp. 214–221, 2001.

[33] J. R. Williamson, T. F. Quatieri, B. S. Helfer, G. Ciccarelli, and D. D. Mehta, "Vocal and facial biomarkers of depression based on motor incoordination and timing," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*. ACM, 2014, pp. 65–72.

[34] J. K. Darby, N. Simmons, and P. A. Berger, "Speech and voice parameters of depression: A pilot study," *Journal of Communication Disorders*, vol. 17, no. 2, 1984.

[35] W. Xiong, J. Droppo, X. Huang, F. Seide, M. Seltzer, A. Stolcke, D. Yu, and G. Zweig, "The microsoft 2016 conversational speech recognition system," *arXiv preprint arXiv:1609.03528*, 2016.

[36] C. M. Bishop, "Pattern recognition," *Machine Learning*, vol. 128, 2006.