# Learning based Visual Engagement and Self-Efficacy

Svati Dhamija
*University of Colorado Colorado Springs*
*1420 Austin Bluffs Parkway, Colorado Springs, CO 80918*
*sdhamija@vast.uccs.edu*

***Abstract*—Self-help web interventions for mental health effectively follow a *one-size-fits-all* approach lacking the personalization of regular psychotherapy sessions and the effectiveness associated with the treatment. A scalable *adaptive person-centered approach* is therefore essential to non-invasively monitor symptom severity, enhance coping capabilities and increase engagement levels for maximal impact. In this work, we propose a novel approach to empower mental trauma patients, improve outcomes and *EASE* their suffering while reducing health-care costs. We develop machine learning algorithms that use both voluntary and involuntary feedback encapsulating the interactions of brain and body in a non-intrusive setting, by calibrating physiological arousal and engagement from face videos leading towards a hidden state of self-efficacy.**

## 1. Introduction

Mental trauma following disasters, military service, accidents, domestic violence and other traumatic events is often associated with adversarial symptoms like avoidance of treatment, mood disorders, and cognitive impairments. Lack of treatment for serious mental health illnesses annually cost $193.2 billion in lost earnings [1]. Providing proactive, scalable and cost-effective web-based treatments to traumatized people is, therefore, a problem with significant societal impact [2]. Though self-help websites for trauma-recovery exist, they are usually generic, based on one-size-fits-all models, instead of being person-specific.

Interventions are required that can be used repeatedly without losing their therapeutic power, that can reach people even if local health care systems do not provide them with needed care or recommend costly procedures or are simply unapproachable. Such interventions require adaptive models that are tailored to individual needs and can be shared globally without taking resources away from the populations where the interventions were developed. Research suggests that personalization and automated adaption in self-help websites can positively aid people with mental health issues and advance mental health computing [3].

Like the numerous tasks we work on daily, our outcomes are a factor of how persuasive we are in the endeavor and this is especially true for coping with trauma. The worst a person is at self-regulation, has direct impact on the positive or negative impact of the treatment. According to Social Cognitive Theory (SCT), perceived coping self-efficacy emerges as a focal mediator of post traumatic recovery [4]. Amongst the currently available e-health interventions, evidence to support the clinical effectiveness of most interventions exists, however, patient engagement with these interventions is still a major concern [5], [6], [7]. Such interventions measure user engagement from infrequent questionnaire's. Self-reported user engagement has been found, in many psychology studies, to be highly correlated with outcomes [8], [9], [10]. Estimating engagement is important as various psychological studies indicate that engagement is a key component to measure the effectiveness of treatment and can be predictive of behavioral outcomes in many applications.

Owing to the uncertain nature of trauma recovery which includes frequent mood-swings, it is essential to look for self-efficacy and engagement values over shorter time periods than the self-reports. Self-reports are limited by the frequency at which the user can be asked for a response without significantly annoying them. Self-reports are also impossible to collect on a per-second basis. Listed below are the research questions that drive our work:

1) Can advanced computer vision based learning techniques help in creating adaptive interventions?
2) How do we use infrequent self-reports?
3) Can self-efficacy be predicted from videos or physiological data or both?
4) What is the relationship between self-efficacy and engagement?
5) Is engagement or change in self-efficacy dependent on the task at hand?
6) How do we integrate a priori information about the user to predict user-response?
7) What features should we use to predict engagement from video?
8) Are all psychometric measures of a user independent from each other?

## 2. Background and Related Work

In recent years various researchers have explored Behavioral Intervention Technologies [11], [12] to augment traditional methods of delivering psychological interventions, face-to-face in one-to-one psychotherapy sessions, in order to expand delivery models and/or increase the outcomes of

978-1-5386-0563-9/17/$31.00 ©2017 IEEE

therapy. Influenced by these works, we postulate the need for the development of computer vision and machine learning-based methods for automated engagement prediction, mood prediction and self-efficacy assessment in the domain of web-based trauma recovery.

Social cognitive theory prescribes mastery experiences as the principal means of personality change [13]. In this social learning analysis, expectations of personal efficacy are based on four major sources of information: performance accomplishments, vicarious experience, verbal persuasion, and physiological states [13]. Operative competence requires orchestration and continuous improvisation of multiple subskills to manage ever-changing circumstances [14]. In our research where we present trauma subjects with 2 training modules of relaxation and triggers and 4 selection modules post-training (self-talk, social-support, unhelpful-coping and professional-help), the goal is to predict the post module self-efficacy value based on the heuristics of the pre-training self-efficacy measurements or Coping Self-Efficacy Trauma (CSE-T) value(s), keeping in mind the content of the modules. Current gold-standard to measure self-efficacy is from self-reports. Self-reports are limited by the frequency at which the user can be asked for a response without significantly annoying them. Self-reports are also impossible to collect on a per-second basis. However, various psychology studies have shown that self-efficacy and engagement are correlated [15].

Engagement is a critical component of student learning, web-based interventions, commercial applications for marketing, etc. and face-based analysis is the most successful non-invasive approach for engagement estimation [5], [16], [7], [17], [18]. The majority of research to predict automated engagement has been limited to the field of education where learning algorithms are built to determine student engagement from behavioral cues like facial expressions, gaze, head-pose, etc. [19], [17], [18], [20]. These works primarily rely on extracting facial features and developing machine-learning-based approaches to identify engagement activity of students in classroom performing various tasks, e.g., reading/writing, etc. The subjects are often assumed to be co-operative with control over their emotions and monitored by an external actor e.g. the teacher. One of the notable differences in these works and data collected from trauma subjects is that subject co-operativeness varies significantly, depending on the severity of mental illness and the task (self-regulation exercises) that they are assigned, leading to multiple challenges in applying methods from student learning directly [21]. In our work [22], [23], we show that computer vision and deep-learning-based techniques can be used to predict user engagement from webcam feeds with content. Once we have tools for reliable engagement measurement during an intervention, the website and task can adapt to enhance or maintain engagement and recovery. Such experimentation requires datasets that contains adequate amount of engagement, arousal and self-efficacy data. Unfortunately, datasets where humans are experimental subjects is not readily available to researchers and is ethically restricted. The availability of such datasets, till date, has been confined to the affective space of valence and arousal where ground truth is available in form of self-reports or post-processing limited annotations [24], [25], [26] etc.

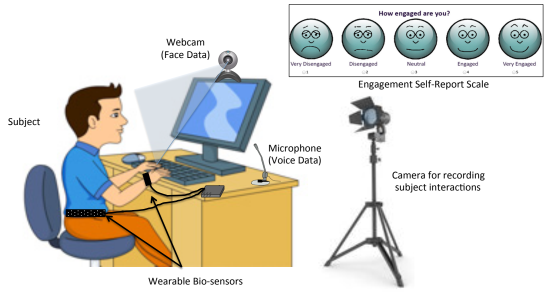

**Figure 1: Experimental setup for EASE data collection:** Subjects interacted with the website while performing self-regulation exercises. Face data was captured using an external webcam; voice data was captured using a microphone. Additional data such as skin-conductance, respiration, and ECG signals were also recorded using wearable sensors. All the interactions were recorded using an external camera. Finally, while the subjects were viewing the trauma-recovery website, the system asks them about their engagement level, with self-report on a scale of 1-5 (top right corner) where 1 is “Very Disengaged” and 5 “Very Engaged”

According to social psychology, human experience is an intertwined outcome of behavior (interactions), cognition (thoughts) and affect (feelings) [27], where behavior and affect are generally detected through a series of facial expressions, gestures, body movements, speech, and other physiological signals, such as heart rate, respiration, sweat, etc. The purpose of this research is to explore a machine learning based approach towards analyzing and predicting cognitive human behavior through behavioral modeling from face videos. *Since human behavior is complex, we target a closed space of EASE (Engagement, Arousal and Self Efficacy)*. The strategy is to employ an integrated set of psychological principles that have cognitive effects and cause behavioral changes.

# 3. Current Results

## 3.1. Data collection procedure:

The web-intervention used to collect the data was based on the findings of Social Cognitive Theory [4] and consisted of subjects undergoing six tasks (modules) namely: social-support, self-talk, relaxation, unhelpful coping, professional help and triggers. The broader study was divided into three sessions/visits in the form of a Randomized Control Trial (RCT). Each participant was assigned 2 out of the six modules in each visit. The first two visits were restricted to “Relaxation” and “Triggers” modules only, and in the third visit, the participants were free to choose from the remaining four modules. Each visit lasted for approx. 30 minutes - 1.5 hours. In the first visit, subjects were randomly allocated Relaxation or Triggers as the first module and a reverse order during the second visit and second module. At the beginning of each visit, the subjects listened to a neutral introductory video. During these sessions, a Logitech webcam with a resolution of 640x480 at 30 fps was placed on top of the monitor. It records video of the participants face along with audio. Physiological data was also recorded for the entire session. The participants could freely interact with the trauma recovery website, and their interactions were recorded in the form of Picture in Picture video using a Camtasia recorder (with screen and webcam recording simultaneously). During the module, participants provided self-reports about their engagement level [22]. For all experiments in the following sections we use the same dataset.

## 3.2. Contextual Engagement Prediction from video:

A wide range of research has used face data to estimate a person‘s engagement, in applications from advertising to student learning. An interesting and important question not addressed in prior work is if face-based models of engagement are generalizable and context-free, or do engagement models depend on context and task. Our research shows that context-sensitive face-based engagement models are more accurate, at least in the space of web-based tools for trauma recovery. In our work [22], we analyze user engagement in a trauma-recovery regime during two separate modules/tasks: relaxation and triggers. The dataset comprises of 8M+ frames from multiple videos collected from 110 subjects, with engagement data coming from 800+ subject self-reports. We build an engagement prediction model as sequence learning from facial Action Units (AUs) using Long Short Term Memory (LSTMs). Our experiments demonstrate that engagement prediction is contextual and depends significantly on the allocated task. Models trained to predict engagement on one task are only weak predictors for another and are statistically significantly less accurate than context-specific models.

## 3.3. Mood-Aware Engagement Prediction:

Since our experiments demonstrate that engagement prediction models are contextual, we take this a step further and ask the question: *“Does using current mood as context improve engagement prediction for a given task?”*. In order to answer this question, we use Profile Of Mood States (POMS) data that was collected before and after the session from each subject. Our POMS questionnaire has first 24 questions from POMS-SF [28], which are clustered into five negative sentiments (tension: 5 questions, depression: 6 questions, anger: 5 questions, fatigue: 2 questions, confusion: 2 questions) and one positive sentiment, vigor: 4 questions. The final POMStmd(total mood disturbance) level is computed as difference of sum of negative n(x) and positive p(x) sentiments:

$$\text{POMStmd} = \frac{1}{21.1} \sum_{x \in \text{neg. senti.}} n(x) - \sum_{x \in \text{pos. senti.}} p(x)$$

here we scaled the POMStmd scores by the observed value, so that the range is between [0,1].

The POMStmd score is then used to condition each AU input to obtain mood-aware engagement prediction results. We precondition the basic engagement multi-class LSTM with POMStmd values obtained using self-reports by adding the normalized to the AU representation. Since the engagement scores are ordinal, not categorical, for testing of mood-aware modeling we use the more traditional Leave-One-Subject-Out (LOSO) methodology, reporting root-mean-squared-error. Even though the LSTM model was optimized for categorical correctness, we notice a significant improvement in performance by augmenting AUs with POMS data. Mood-aware engagement model for triggers showed a significant reduction in error (p=.0007).

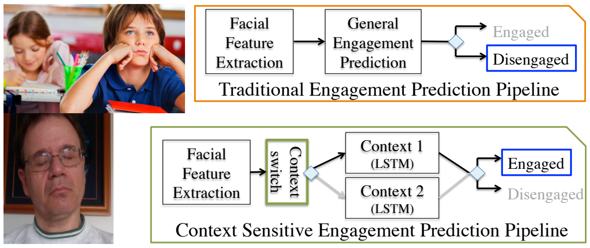
Facial Feature Extraction
General Engagement Prediction
Engaged
Disengaged
Traditional Engagement Prediction Pipeline
Facial Feature Extraction
Context switch
Context 1 (LSTM)
Context 2 (LSTM)
Engaged
Disengaged
Context Sensitive Engagement Prediction Pipeline

**Figure 2:** Consider the images on the left. Which subjects are engaged and which are disengaged? Would you change your answer if you knew one had a task of doing a relaxation exercise? What if it was reading web content, watching a video or taking a test? **We contend that face-based engagement assessment is context sensitive.** Traditional engagement prediction pipelines based on facial feature extraction and machine learning techniques learn a generic engagement model, and would consider the face in lower left disengaged. In trauma recovery, individuals are often advised to do particular exercises, e.g. self-regulation exercises where the task involves the subject to “close your eyes, relax and breathe”. The image on the lower left is a highly engaged subject. Hence, there is a need to revisit existing facial-expression-based engagement prediction techniques and augment them with the context of the task at hand. As shown in bottom right, this work develops context-sensitive engagement prediction methods based on facial expressions and temporal deep learning.

## 3.4. Automated Mood-Aware Engagement Prediction:

Developing intelligent machines that recognize facial expressions, detect spontaneous emotions and infer affective states of an individual are all challenging problems. While significant amount of work in recent years has focussed on advancing machine learning techniques for affect recognition and affect classification, the prediction of mood from facial analysis and the usage of mood data have received less attention. Questionnaires for psychometric measurement of mood-states are common, but using them during interventions that target psychological well-being of people are arduous and may burden an already troubled population. In our work [23], we create two separate automated LSTM models: a total mood disturbance predictor and a mood sub-scale predictor, and then use them to aid predictions of subject engagement levels. Our mood-aware engagement predictor uses total mood disturbance score, and our analysis compares both mood sub-scale predictors and an overall mood disturbance predictor for engagement prediction. Our experiments show that mood-aware engagement predictor using our novel visual analysis approach performs significantly better or on par with using self-reports.

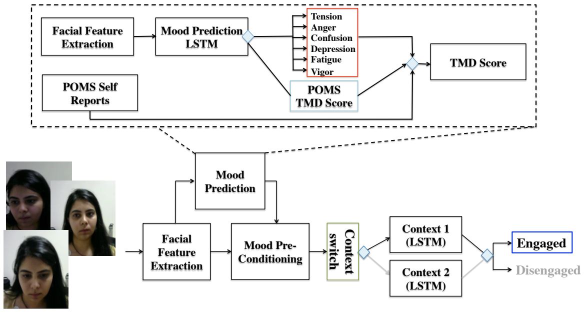

**Figure 3: Automated Mood Prediction for Mood-Aware Context Sensitive Engagement Prediction:** Trauma patients are often reluctant to express themselves openly, suffer from mood changes that last an extended period, which in turn affects their cognitive abilities. We propose Mood-Aware Contextual Engagement prediction for trauma subjects. The top part of the figure shows mood prediction pipeline aimed at predicting the mood of trauma patients from facial videos. The mood estimates are then used to pre-condition learning for context sensitive engagement prediction models. Temporal deep learning methods are used to learn long-term dependencies to estimate mood and its interplay with contextual engagement.

# 4. Future Work

## 4.1. Predicting change in Self-Efficacy:

Bandura (1997) reviewed substantial empirical evidence for increasing CSE (Coping Self Efficacy) perceptions by promoting mastery experiences, opportunities for vicarious success modeling, positive verbal persuasion, and reductions in physiological arousal [29]. CSE is quantified from individual responses to standard questionnaires that target at accessing user(s) interpretation/belief of their ability to retaliate to a specific hypothetical situation. Coping Self-Efficacy Trauma (CSE-T) has emerged as a focal mediator of post-traumatic recovery [30]. Neither, the absolute value of self-efficacy nor the changes in self-efficacy, are directly observable from visual inferences. Psychometric measures using CSE-T questionnaire [2] are used to assess coping self-efficacy in EASE. We propose a learning technique that can predict the self efficacy values post-module, using the pre-module self-efficacy value along with derived predictions of engagement from visual/multimodal data and arousal from physiological responses.

## 4.2. Multimodal Self-Efficacy Prediction:

Number of multimodal approaches have been explored and demonstrated to improve accuracy of affect detection methods in various HCI applications. Several psycho-physiological signals such as EEG, skin conductance, respiratory rate and others have found to correlate with affective states of subjects. In this work, we propose to analyze multimodal approaches to determine relationship between facial video data and sensory signals on trauma patient's coping self-efficacy level while performing a particular task. We wish to develop a multimodal LSTM which can explicitly model the long-term dependencies both within the modality of facial video data and across other sensory modalities in a single multimodal LSTM.

# 5. Challenges and Broader impact

The work proposed in this paper is the first step towards building EASE models for trauma-recovery; we expect others will be able to further improve on the models herein. It is impractical to expect uniform sampling across engagement levels from PTSD subjects, so an issue that will need to be addressed in future efforts is to build machine-learning models that are aware of the data imbalances. Moreover, the CSE-T and engagement values from standardized questionnaires are sparse i.e. one value per ten-thousand video frames and approx. 250K physiology data samples at a minimum. For simplicity, we considered fixed segment-lengths for engagement and mood-aware engagement prediction. However, advanced learning techniques like sparse-label propagation [31], multi-label transfer [32], [33] need to explored to approach the problem efficiently. The most natural way of building better classifiers is training with an even larger dataset and performing parameter optimization of LSTMs. Sophisticated methods like feature (AUs) pooling over space and time, jointly using additional tracking data such as head-pose, gaze, expressions and other emotional states would also likely improve accuracy. Additionally, there is a need to consider different timescales to create a real-time self-efficacy or engagement predictor, shorter durations may not provide enough context for user's affective state, while longer video segments tend to be harder to evaluate, possibly because of the increase in data mixes from self-efficacy changes and levels of engagement. Another possibility is to use more contextual information like, who the user is, what their symptom severity levels are, demographic information etc. the more we know, the better we can predict the affective state of the user.

Though our research is targeted at Post Traumatic Stress Disorder (PTSD) recovery and a specific type of self-efficacy i.e. Coping Self Efficacy for Trauma (CSE-T), objective measurement of self-efficacy is applicable to a broader list of tasks in various fields, like academic performance of students in education, weight loss interventions or diet control in health, recovery treatments from cancer, behavioral parent training, fitness control, developing socially assistive robots, employment training (like public speaking, preparation of job-interviews), designing virtual reality games etc.

# 6. Acknowledgement

This work supported in part by NSF Research Grant SCH-INT 1418520 “Learning and Sensory-based Modeling for Adaptive Web-Empowerment Trauma Treatment”. I would like to thank my advisor Dr.Terrance E.Boult for his continued supervision and support. A special thanks to Dr.Charles C.Benight, Dr.Kotaro Shoji, Ms.Carolyn Yeager and the UCCS Trauma Health and Hazards Center for the trauma-recovery website and for collecting the EASE dataset.

# References

[1] T. R. Insel, “Assessing the economic costs of serious mental illness,” 2008.

[2] C. C. Benight, K. Shoji, L. E. James, E. E. Waldrep, D. L. Delahanty, and R. Cieslak, "Trauma coping self-efficacy: A context-specific self-efficacy measure for traumatic stress." *Psychological trauma: theory, research, practice, and policy*, vol. 7, no. 6, p. 591, 2015.

[3] R. A. Calvo, K. Dinakar, R. Picard, and P. Maes, "Computing in mental health," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2016, pp. 3438–3445.

[4] C. Benight and A. Bandura, "Social cognitive theory of posttraumatic recovery: the role of perceived self-efficacy," *Behaviour Research and Therapy,Elsevier*, 2004.

[5] S. Steinmetz, C. Benight, S. Bishop, and L. James, "My disaster recovery: a pilot randomized controlled trial of an internet intervention," *Anxiety Stress Coping*, vol. 25, no. 5, pp. 593–600, 2012.

[6] S. U. Marks and R. Gersten, "Engagement and disengagement between special and general educators: An application of miles and huberman's cross-case analysis," *Learning Disability Quarterly*, vol. 21, no. 1, pp. 34–56, 1998.

[7] D. Macea, K. Gajos, Y. D. Calil, and F. Fregni, "The efficacy of web-based cognitive behavioral interventions for chronic pain: a systematic review and meta-analysis," *J. of Pain*, vol. 11, no. 10, pp. 917–929, 2010.

[8] M. Couper, G. Alexander, N. Zhang, R. Little, N. Maddy, M. Nowak, J. McClure, J. Calvi, S. Rolnick, and M. Stopponi, "Engagement and retention: measuring breadth and depth of participant use of an online intervention," *Journal of Medical Internet Research*, vol. 12, no. 4, 2010.

[9] G. Eysenbach, "The law of attrition," *Journal of Medical Internet Research*, vol. 7, no. 1, 2005.

[10] L. Donkin and N. Glozier, "Motivators and motivations to persist with online psychological interventions: A qualitative study of treatment completers," *Journal of Medical Internet Research*, vol. 14, no. 3, 2012.

[11] E. Bunge, B. Dickter, M. Jones, G. Alie, A. Spear, and R. Perales, "Behavioral intervention technologies and psychotherapy with youth: A review of the literature," *Current Psychiatry Reviews*, vol. 12, no. 1, pp. 14–28, 2016.

[12] S. M. Schueller, R. F. Muñoz, and D. C. Mohr, "Realizing the potential of behavioral intervention technologies," *Current Directions in Psychological Science*, vol. 22, no. 6, pp. 478–483, 2013.

[13] A. Bandura and S. Wessels, "Self-efficacy," 1994.

[14] A. Bandura, "Self-efficacy mechanism in human agency." *American psychologist*, vol. 37, no. 2, p. 122, 1982.

[15] C. Benight, K. Shoji, C. Yeager, A. Mullings, S. Dhamija, and T. Boult, "The importance of self-appraisals of coping capability in predicting engagement in a web intervention for trauma," 2016. [Online]. Available: http://bluesunsupport.com/wp-bluesun/wp-content/uploads/2016/04/ISRII-Poster-2016-Final.pdf

[16] S. U. Marks and R. Gersten, "Engagement and disengagement between special and general educators: An application of miles and huberman's cross-case analysis," *Learning Disability Quarterly*, vol. 21, no. 1, pp. 32–56, 1998.

[17] H. Monkaresi, P. Bosch, R. Calvo, and S. D'Mello, "Automated detection of engagement using video-based estimation of facial expressions and heart rate," *IEEE Trans. on Affective Computing*, 2017.

[18] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, "Faces of engagement: Automatic recognition of student engagement from facial expressions," *IEEE Trans. on Affective Computing*, vol. 5, no. 3, pp. 86–98, 2014.

[19] H. O'Brien and E. Toms, "The development and evaluation of a survey to measure user engagement," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 1, pp. 50–69, 2010.

[20] M. N. Giannakos, L. Jaccheri, and J. Krogstie, "How video usage styles affect student engagement? implications for video-based learning environments," in *State-of-the-Art and Future Directions of Smart Learning*. Springer, 2016, pp. 157–163.

[21] S. Scherer, G. Lucas, J. Gratch, A. Rizzo, and L. P. Morency, "Self-reported symptoms of depression and ptsd are associated with reduced vowel space in screening interview," *IEEE Trans. on Affective Computing*, 2016.

[22] S. Dhamija and T. Boult, "Exploring contextual engagement for trauma recovery," *CVPR Workshop on Deep Affective Learning and Context Modelling*, 2017.

[23] S. Dhamija and T. Boult, "Automated mood-aware engagement prediction," *Affective Computing and Intelligent Interaction*, 2017.

[24] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[25] Y. Baveye, E. Dellandrea, C. Chamaret, and L. Chen, "Liris-accede: A video database for affective content analysis," *IEEE Transactions on Affective Computing*, vol. 6, no. 1, pp. 43–55, 2015.

[26] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2012.

[27] R. Jhangiani, H. Tarry, and C. Stangor, "Principles of social psychology-1st international edition," 2015.

[28] S. Shacham, "A shortened version of the profile of mood states," *Journal of Personality Assesment*, vol. 47, no. 3, pp. 305–306, 1983.

[29] A. Bandura, *Self-efficacy: The exercise of control*. Macmillan, 1997.

[30] A. Luszczynska, C. C. Benight, and R. Cieslak, "Self-efficacy and health-related outcomes of collective trauma: A systematic review," *European Psychologist*, vol. 14, no. 1, pp. 51–62, 2009.

[31] G. Lin, K. Liao, B. Sun, Y. Chen, and F. Zhao, "Dynamic graph fusion label propagation for semi-supervised multi-modality classification," *Pattern Recognition*, vol. 68, pp. 14–23, 2017.

[32] Q. Wu, H. Wu, X. Zhou, M. Tan, Y. Xu, Y. Yan, and T. Hao, "Online transfer learning with multiple homogeneous or heterogeneous sources," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 7, pp. 1494–1507, 2017.

[33] H. Chang, J. Han, C. Zhong, A. Snijders, and J.-H. Mao, "Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.