Population-specific Detection of Couples' Interpersonal Conflict using Multi-task Learning

Aditya Gujral
Texas A&M University
aditya.gujral@tamu.edu

Theodora Chaspari
Texas A&M University
chaspari@tamu.edu

Adela C. Timmons
University of Southern California
adelatim@usc.edu

Yehsong Kim
University of Southern California
yehsongk@usc.edu

Sarah Barrett
University of Southern California
skbarret@usc.edu

Gayla Margolin
University of Southern California
margolin@usc.edu

ABSTRACT

The inherent diversity of human behavior limits the capabilities of general large-scale machine learning systems, that require ample data to provide robust descriptors of the outcomes of interest. Population-specific models comprise a promising line of work for representing human behavior, since they make decisions for clusters of people with common characteristics, reducing the amount of data needed for training. We propose a multi-task learning (MTL) framework for developing population-specific models of interpersonal conflict between couples using ambulatory sensor and mobile data from real-life interactions. The criteria for population clustering include global indices related to couples' relationship quality and attachment style, person-specific factors of partners' positivity, negativity, and stress levels, as well as fluctuating factors of daily emotional arousal obtained from acoustic and physiological indices. Population-specific information is incorporated through a MTL feed-forward neural network (FF-NN), whose first layers capture the common information across all data samples, while its last layers are specific to the unique characteristics of each population. Our results indicate that the proposed MTL FF-NN trained solely on the sensor-based modalities provides unweighted and weighted F1-scores of 0.51 and 0.75, respectively, outperforming a single general FF-NN trained on the entire dataset and separate FF-NNs trained on each population cluster individually, highlighting the importance of taking into account the inherent diversity of different populations for the development of human-centered machine learning models.

CCS CONCEPTS

• **Computing methodologies → supervised learning**;

KEYWORDS

Ambulatory sensing; conflict detection; multi-task learning

ACM Reference Format:
Aditya Gujral, Theodora Chaspari, Adela C. Timmons, Yehsong Kim, Sarah Barrett, and Gayla Margolin. 2018. Population-specific Detection of Couples' Interpersonal Conflict using Multi-task Learning. In *2018 International Conference on Multimodal Interaction (ICMI '18), October 16–20, 2018, Boulder, CO,*

1 INTRODUCTION

Emerging machine learning systems can leverage the large amount of information obtained on a constant basis and generate automatic decisions for outcomes of interest. Such systems have been very successful in traditional computer vision and speech processing applications, where the labels of the corresponding outcomes are well-defined and human annotations are easy to obtain. However, this is not always the case in emerging human-centered applications, since the high diversity of human behavior and interaction imposes additional constraints to the considered problem [27]. The tedious data collection protocols [22], the inherent noise [13], and the limited availability of psychological-science experts [22] are some of the challenges in acquiring large amounts of data, limiting conventional machine learning systems and creating the need for alternative models to more effectively address the data sparsity issues in such applications.

Psychological science has long recognized that human behavior is affected by a variety of person- and context- specific indices [34]. Incorporating such indices in computational models can potentially yield more accurate descriptors of the underlying diversity of human behavior. The above observations have motivated a new line of research that aims to leverage personalized and population-specific information and develop machine learning models specific to a certain individual or population. This paper focuses on population-specific models for detecting behaviors of interest in real-life, where population refers to a group of subjects from the original pool of participants that share common characteristics. Our application focuses on detecting interpersonal conflict between romantic couples in real-life through ambulatory data. Conflict is known for its deleterious effects in every-day and beyond [5, 12]. Especially in romantic relationships, conflict is experienced and expressed differently between distressed and non-distressed couples as well as between people with different relationship attachment characteristics [15]. We propose a multi-task learning (MTL) framework that is able to leverage useful information common across all couples and use population-specific information to further refine the system output. The MTL approach is operationalized as a multi-task feedforward neural network (FF-NN), where the first layers represent common feature embeddings, while the last layers are separately trained based on pre-defined populations from the data of interest. We have further experimented with different clustering criteria to define populations. Our results indicate that the proposed approach outperforms a single FF-NN trained on the entire data sample, as well as separate FF-NNs trained on each population individually, highlighting the importance of incorporating prior information from various global and fluctuating factors for reliably modeling human behavior.

2 PREVIOUS WORK

Early work on personalized and population-specific machine learning has examined the learning of separate models for groups of participants or for each participant separately. Despite its simplicity and intuitive nature, this approach might not be generalizable, as it does not incorporate general knowledge from the entire data sample, while each population might not have enough data related to all outcomes of interest [18, 21]. Adaptive models, such as modified support vector machine, have been proposed as an alternative approach, according to which corpus-wide information provides initial estimates of the model parameters, while the personalized or population-specific information refines these initial estimates [9, 17]. Ensemble methods, such as weighted regression and community similarity networks, have been further used to provide outcome estimates based on the subset of data that are most relevant to the sample of interest [4, 19, 24]. Finally, hierarchical models, such as multitask neural networks, have recently gained attention, by modeling corpus-wide feature representations through the initial network layers and individual-specific representations through the subsequent layers [25, 26]. Despite their promising results, the majority of work has focused on personalized models for detecting and predicting human behavior, while the problem of population-specific representations has not been extensively addressed. Previous studies have proposed the use of a K-nearest neighbor regression [4] and ensemble methods [21], while FF-NNs have not been examined for population-specific approaches. Given the fact that fully personalized models might fail to adequately generalize even within the same person, since each individual might not have enough training samples for each outcome, especially those related to extreme conditions that are highly relevant to health, the current paper focuses on population-specific models for quantifying human behavior.

Previous work has also used audio signals to detect conflict in political debates [20], while other studies have attempted to detect interpersonal conflict between couples in uncontrolled real-life environments through the use of general machine learning models [16, 36].

3 DATA DESCRIPTION

Data were collected through the University of Southern California (USC) Couple Mobile Sensing Project (homedata.github.io) and include 87 couples aged between 18-25 years. Collection was performed from 9am to bedtime and resulted in 1126 (90.5%) non-conflict and 117 conflict samples. The Nexus 5 phone was used to administer hourly ecological momentary assessments (EMA) with self-reports of mood and quality of interactions (MQI) between participants. The same phone also recorded GPS coordinates and 3 minutes of audio data every 12 minutes. Physiological sensors included the Actiwave [1] for the electrocardiogram (ECG) signal and the Q sensor [31] for the electrodermal activity (EDA), temperature, and acceleration recordings.

Acoustic, linguistic, physiological, contextual, and MQI features were extracted over an hourly basis. Automatic voice activity detection and speaker diarization was performed on the audio signals, based on which the mean, median, maximum, minimum, standard deviation, and range of speech loudness and fundamental frequency measures were computed, resulting in 12 *acoustic* features per partner. Manual transcripts were generated from the audio recordings for 53 couples, based on which *linguistic* features were extracted using the Linguistic Inquiry and Word Count (LIWC) [29] software. These include 24 linguistic indices (e.g., personal pronouns, verbs), 32 psychological construct measures (e.g., positivity, negativity, swearing), 7 features of personal concern (e.g., home-, work-, health-related words), and 3 paralinguistic factors (i.e., assent, fillers, disfluencies) per partner. The linguistic features of the couples for which transcripts were not available

were treated as missing values. 54 *physiological* features were computed per partner. Time- and frequency-based ECG features were extracted that include average, standard deviation, minimum, and maximum of the interbeat interval (IBI), average beats per minute, heart rate variability, R-R interval, as well as the very-low (<0.04Hz), low (0.04-0.15Hz), and high (0.15-0.4Hz) frequency component of the IBI. EDA features included the mean skin conductance level, and the number, frequency, and amplitude of skin conductance responses detected with thresholds at 0.01 and 0.02μS. Measures of EDA synchrony were further computed using signal similarity approaches in order to capture inter-partner physiological co-regulation [7]. Body activity count was measured as the l^2-norm of the 3-axis acceleration signals, while body temperature was averaged over the corresponding hour. *Contextual* features include the hourly consumption of caffeine, alcohol, tobacco, and drugs, as well as whether the participants were exercising, interacting with others, or driving, resulting in 7 measures per person. The stress, happiness, sadness, nervousness, and anger are the 5 *self-reported* measures per partner.

Due to the unbalanced nature of the dataset, we report weighted and unweighted evaluation metrics, the latter taking the number of samples per class into account. Conflict labels were based on momentary self-reports, where individuals reported "whether they have expressed annoyance or irritation toward their partner". Missing values were replaced with feature mean.

Prior to the daily data collection, each participant had completed various self-assessment questionnaires, including the Quality of Marriage Index (QMI) [28], the Experiences in Close Relationships-Revised (ECR-R) [33], the Positive and Negative Affect Schedule (PANAS) [37] and the Perceived Stress Scale (PSS) [10]. QMI is a 6-item questionnaire that contains items related to the degree of satisfaction in various areas of the relationship. ECR-R captures relationship anxiety and avoidance through an 18-item questionnaire related to how secure people are about their romantic partners and how comfortable they are depending on others. PANAS is a 10-item scale that measures an individual's positive and negative mood, while PSS captures the degree to which situations in ones' life are appraised as stressful through 10 items. These comprise the prior information that will be used for creating population clusters (Section 4.1).

4 METHODOLOGY
4.1 Population clustering

Population clustering is motivated by findings in psychological sciences, suggesting that the way couples experience and express conflict depends on their relationship quality and attachment styles, as well as to the individual person-dependent characteristics. For example, research findings indicate that various relationship attachment styles (e.g., anxious, avoidant) are related to different types conflict resolution (e.g., compromising, obliging, integrating), each depicted differently on individuals' bio-behavioral signals [30, 32]. Moreover couples that are not satisfied with their relationship are found to engage in a negative conflict style (e.g., feeling irritated towards partner, avoiding discussion) [11, 23]. Clustering was performed using global couple and person-dependent descriptors obtained through the QMI, ECR-R, PANAS, and PSS, as explained in detail in Section 3, as well as fluctuating factors obtained from the ambulatory data collection. The fluctuating factors that were used include signal-derived measures of emotional arousal (i.e., the mean skin conductance level and the mean acoustic fundamental frequency) computed over the entire day. These have been successfully used in relevant studies to capture the daily "microprocesses" in couples' lives [35] and are tightly connected to relationship outcomes and various facets of communication [3, 38]. The
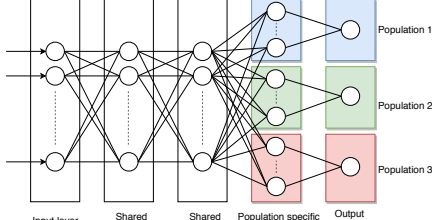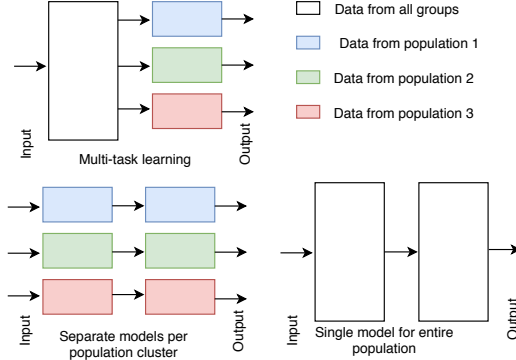
Figure 1: Multi-task learning model.

Figure 2: Multi-task, single and separate models.

Figure 3: Number of (non-)conflict samples per couple for populations 1, 2, and 3 (P1, P2, P3).

Figure 4: The first two dimensions resulting from Principal Component Analysis with the Quality of Marriage Index and Experiences in Close Relationships-Revised scores as an input.

Table 1: Set of values in hyper-parameter tuning.

Hyper-parameter	Values
# neurons in first two hidden layers	60, 80, 120
# neurons in last hidden layer	60, 80, 120
dropout	0, 0.2, 0.3
optimization algorithm	adam, sgd, rmsprop
class weights	{non-conflict:1, conflict:15}, {non-conflict:1, conflict:25}

items of these questionnaires comprise the input to principal component analysis (PCA) to minimize correlations and reduce the dimensionality of the input space. Population clustering was performed with K-means having as input the first three PCA components.

4.2 Population-specific models

Our goal is to classify between conflict and non-conflict from sensory-based and self-reported features using the proposed population- specific machine learning models. While previous work has mostly examined the training of separate models per population [4, 18, 21], we propose the use of MTL for the same task. MTL is inspired by human learning to leverage useful information present in related tasks and apply this for the learning of another proximal task [40, 41]. Instead of learning patterns over similar tasks, MTL can be applied to learn patterns over similar populations. By operationalizing MTL as a feature learning approach, we can assume that the entire group of participants shares generalized feature representations, while specific representations can be subsequently learned for each population. Multi-layer FF-NNs provide an ideal structure for implementing this, since they enable common information among all samples to be captured in the initial layers of the network, while the final layers are specific to each population, capturing its unique characteristics (Fig. 1). We compared the proposed feedforward MTL (denoted as *MTL*) to a single general model trained for entire population (*Single*) and separate unique models trained and tested using data from each population cluster separately (*Separate*), as depicted in Fig. 2.

5 RESULTS

Results are reported using a 5-layer FF-NN with 1 input, 1 output, and 3 hidden layers. The *Single* model consisted of a 5-layer FF-NN, the *Separate* models consisted of separate 5-layer FF-NNs, while the *MTL* had the input and first 2 hidden layers shared among all data samples and the subsequent 2 layers separately trained for each population. During the *MTL* training, weights of the first 3 layers were updated based on all data, while each group of weights in the last two layers was updated using the samples of the corresponding population. Experiments were performed with a nested cross-validation [6], according to which the outer fold was used to assess the performance of
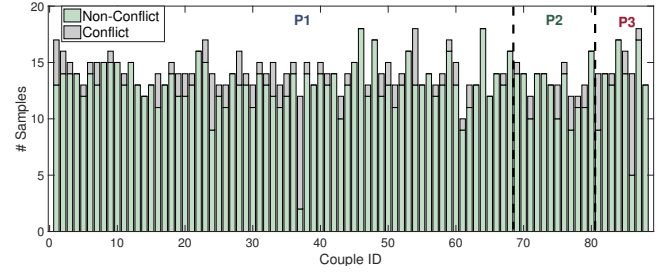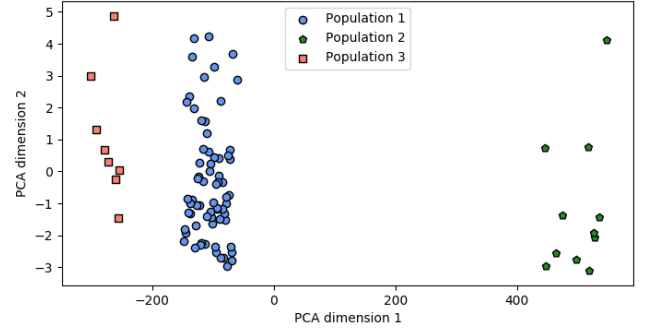
the model, while the inner-fold was used as validation set to tune the model hyper-parameters. Leave-$\frac{1}{5}th$-couples-out cross validation was used to avoid any couple-dependent bias. The tested hyper- parameters include the number of neurons in the hidden layers, the dropout rate applied over all hidden layers, the optimization algorithm for training, and the weight values of the cross-entropy loss function for each class (Table 1), resulting in 162 hyper-parameter combinations. All systems were implemented with the Keras toolbox [2].

Various global and fluctuating factors served as an input for population clustering (Section 4.1). We empirically chose K=3 clusters for the K-means, since this resulted in separable population groups. An indicative 2-dimensional plot of the first and second PCA dimensions resulting from the QMI, anxiety and avoidance scores for the 87 couples is shown in Figure 4. In this figure, we observe highly separable clusters indicating the presence of diversity in the original couples. The distribution of conflict and non-conflict samples per couple and population (Figure 3) indicates that all populations include samples from both classes, i.e., 11.9%, 8.5%, and 7.9% of the samples belong to the conflict class for populations 1, 2, and 3, respectively. Classification results further suggest that the most effective clustering criterion included the QMI and ECR-R scores (Table 2), which were used in the remaining experiments. The same results further highlight the importance of population clustering criteria, and specifically the significance of couple-specific information for detecting conflict.

Table 2: Conflict detection with the proposed multi-task feed-forward neural network using various clustering criteria.

Clustering Features	Weighted F1	Unweighted F1
QMI, ECR-R	0.75	0.51
QMI, ECR-R, PSS, PANAS	0.51	0.36
QMI, ECR-R, skin conductance, pitch	0.68	0.46

QMI: Quality of Marriage Index, ECR-R: Experiences in Close Relationships-Revised
PANAS: Positive and Negative Affect Schedule, PSS: Perceived Stress Scale

Table 3: Conflict classification with multi-task learning (MTL), separate models per population cluster (Separate) and single general model for entire population (Single) trained using the sensor-based data (acoustic, linguistic, physiological).

Model	Class	Precision	Recall	F1
MTL	Non-Conflict	0.93	0.71	0.8
	Conflict	0.15	0.48	0.22
	Weighted Average	**0.86**	**0.69**	**0.75**
	Unweighted Average	**0.54**	**0.60**	**0.51**
Separate	Non-Conflict	0.93	0.61	0.74
	Conflict	0.13	0.57	0.22
	Weighted Average	0.86	0.61	0.69
	Unweighted Average	0.53	.59	0.48
Single	Non-Conflict	0.94	0.54	0.68
	Conflict	0.13	0.67	0.22
	Weighted Average	0.86	0.55	0.64
	Unweighted Average	0.54	.61	0.45

We further compared the three different models using the sensor-based input features only, i.e., acoustic, linguistic, physiological. Our results indicate that the MTL outperforms the baseline models, suggesting that incorporating population-specific information can benefit performance (Table 3). Acoustic and linguistic features outperformed the physiological ones (Table 4), with acoustic features achieving accuracy close to that of the MQI. System performance improved using all three sensory-based modalities (Table 3), indicating the importance of integrating overt and covert information. We finally experimented by including the best two sensor-based modalities (i.e., acoustic, linguistic) per system, and the best two sensor-based modalities with contextual indices and self-reports. Contextual information did not seem to help, while the combination of all modalities performed best (Table 5).

6 DISCUSSION

This paper indicates the usefulness of incorporating indices of relationship quality and measures related to an individual's anxious or avoidant attachment style for modeling couples' conflict. Such prior information is collected from self-reports, that typically do not require a lot of time from individuals to fill in (e.g., QMI is a 6-item questionnaire [28]). These questionnaires are commonly administered by psychological science researchers with the corresponding questions and scoring methods being publicly available. Although further examination is called for, these findings can further stratify the various ways in which couples experience and express conflict, significantly helping the detection of this phenomenon in real-life. One limitation of our approach lies in the fact that the system used manual transcription of the audio data. In our future work, we plan to use transcripts generated as the output of automated speech recognition systems. Previous work indicates the feasibility of automated transcripts for performing linguistic content analysis in similar tasks and for detecting

Table 4: Conflict classification using data from each modality.

Modality of features	Model	Weighted average F1	Unweighted average F1
Linguistic	MTL	0.58	0.41
	Separate	0.61	0.43
	Single	0.59	0.41
Acoustic	MTL	0.61	0.43
	Separate	0.56	0.39
	Single	0.47	0.35
Physiological	MTL	0.52	0.35
	Separate	0.48	0.35
	Single	0.42	0.32
Conextual	MTL	0.42	0.31
	Separate	0.43	0.32
	Single	0.3	0.24
Self-Report	MTL	0.66	0.46
	Separate	0.46	0.43
	Single	0.61	0.43

Table 5: Conflict classification with combination of modalities.

Modalities of features	Model	Weighted average F1	Unweighted average F1
Best 2	MTL	**0.78**	**0.52**
	Separate	0.67	0.47
	Single	0.61	0.41
Best 2 + contexutual	MTL	**0.79**	**0.52**
	Separate	0.7	0.49
	Single	0.71	0.49
All (including self-report)	MTL	0.79	0.56
	Separate	**0.8**	**0.57**
	Single	0.72	0.51

human behaviors in real-life [14, 39]. Finally, another limitation lies in the low precision for detecting the conflict class, that stems from the highly unbalanced class distribution. Given the nature of the application, low precision measures are preferred over low recall, since the risk of missing a conflict sample is higher than the risk of misclassifying a non-conflict sample, i.e., it is safer to warn couples regarding a conflict episode, even if this is not occurring, rather than the opposite. As part of our future work, we will explore oversampling techniques, that might have be able to alleviate this problem [8].

7 CONCLUSION

We proposed population-specific machine learning models for detecting couples' conflict in real life. We experimented with a MTL FF-NN, whose first layers correspond to feature representations of the general sample, while the subsequent layers were trained separately on each population. Population clustering was performed through couple- and partner-specific descriptors and fluctuating factors of emotional arousal. Our results suggest that the proposed MTL FF-NN outperforms the baseline models, achieving weighted and unweighted F1 scores of 0.75 and 0.51, respectively. This highlights the importance of incorporating population diversity when modeling human behavior. As part of our future work, we plan to compare the proposed MTL FF-NN with ensemble methods and employ transfer learning techniques using data from similar domains.

REFERENCES

[1] [n. d.]. Actiwave Cardio. https://www.camntech.com/products/actiwave-cardio/actiwave-cardio-overview. Accessed: 2018-05-01.

[2] [n. d.]. Actiwave Cardio. https://github.com/hadim/keras-toolbox. Accessed: 2018-05-01.

[3] Brian R Baucom, Elisa Sheng, Andrew Christensen, Panayiotis G Georgiou, Shrikanth S Narayanan, and David C Atkins. 2015. Behaviorally-based couple therapies reduce emotional arousal during couple conflict. *Behaviour research and therapy* 72 (2015), 49–55.

[4] Dimitris Bertsimas, Nathan Kallus, Alexander M Weinstein, and Ying Daisy Zhuo. 2017. Personalized diabetes management using electronic medical records. *Diabetes care* 40, 2 (2017), 210–217.

[5] Kira S Birditt, Edna Brown, Terri L Orbuch, and Jessica M McIlvane. 2010. Marital conflict behaviors and implications for divorce over 16 years. *Journal of Marriage and Family* 72, 5 (2010), 1188–1204.

[6] Gavin C Cawley and Nicola LC Talbot. 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research* 11, Jul (2010), 2079–2107.

[7] Theodora Chaspari, Adela C Timmons, Brian R Baucom, Laura Perrone, Katherine JW Baucom, Panayiotis Georgiou, Gayla Margolin, and Shrikanth S Narayanan. 2017. Exploring sparse representation measures of physiological synchrony for romantic couples. In *Affective Computing and Intelligent Interaction (ACII), 2017 Seventh International Conference on.* IEEE, 267–272.

[8] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357.

[9] Lei Clifton, David A Clifton, Marco AF Pimentel, Peter J Watkinson, and Lionel Tarassenko. 2013. Gaussian processes for personalized e-health monitoring with wearable sensors. *IEEE Transactions on Biomedical Engineering* 60, 1 (2013), 193–197.

[10] Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. 1983. A global measure of perceived stress. *Journal of health and social behavior* (1983), 385–396.

[11] Duncan Cramer. 2000. Relationship satisfaction and conflict style in romantic relationships. *The Journal of Psychology* 134, 3 (2000), 337–341.

[12] Morton Deutsch. 2008. Cooperation and Conflict: A personal perspective on the history of the social psychological study of conflict resolution. In *International handbook of organizational teamwork and cooperative working*, Michael A. West, Dean Tjosvold, and Ken G. Smith (Eds.). John Wiley & Sons, Chapter 2, 9–43.

[13] Amir Gandomi and Murtaza Haider. 2015. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management* 35, 2 (2015), 137–144.

[14] Panayiotis G Georgiou, Matthew P Black, Adam C Lammert, Brian R Baucom, and Shrikanth S Narayanan. 2011. That's Aggravating, Very Aggravating: Is It Possible to Classify Behaviors in Couple Interactions Using Automatically Derived Lexical Features?. In *International Conference on Affective Computing and Intelligent Interaction.* Springer, 87–96.

[15] Dara Greenwood. 2016. Attachment and Conflict in Adult Relationships. *The International Encyclopedia of Interpersonal Communication* (2016).

[16] Aditya Gujral, Theodora Chaspari, Adela C Timmons, Sohyun C Han, and Gayla Margolin. 2018. Real-life detection of interpersonal conflict in couples through the use of wearable and mobile technology. In *IEEE Body Sensor Networks (BSN) Conference.* IEEE.

[17] Javier Hernandez, Rob R Morris, and Rosalind W Picard. 2011. Call center stress recognition with person-specific models. In *International Conference on Affective Computing and Intelligent Interaction.* Springer, 125–134.

[18] Markus Kächele, Patrick Thiam, Mohammadreza Amirian, Friedhelm Schwenker, and Günther Palm. 2016. Methods for person-centered continuous pain intensity assessment from bio-physiological channels. *IEEE Journal of Selected Topics in Signal Processing* 10, 5 (2016), 854–864.

[19] Nathan Kallus. 2017. Recursive partitioning for personalization using observational data. In *International Conference on Machine Learning.* 1789–1798.

[20] Samuel Kim, Fabio Valente, and Alessandro Vinciarelli. 2012. Automatic detection of conflicts in spoken conversations: Ratings and analysis of broadcast political debates. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on.* IEEE, 5089–5092.

[21] Saskia Koldijk, Mark A Neerincx, and Wessel Kraaij. 2016. Detecting work stress in offices by combining unobtrusive sensors. *IEEE Transactions on Affective Computing* (2016).

[22] Clemens Scott Kruse, Rishi Goswamy, Yesha Raval, and Sarah Marawi. 2016. Challenges and opportunities of big data in health care: a systematic review. *JMIR medical informatics* 4, 4 (2016).

[23] Lawrence A Kurdek. 1995. Predicting change in marital satisfaction from husbands' and wives' conflict resolution styles. *Journal of Marriage and the Family* (1995), 153–164.

[24] Nicholas D Lane, Ye Xu, Hong Lu, Shaohan Hu, Tanzeem Choudhury, Andrew T Campbell, and Feng Zhao. 2014. Community similarity networks. *Personal and ubiquitous computing* 18, 2 (2014), 355–368.

[25] Daniel Lopez-Martinez and Rosalind Picard. 2017. Multi-task Neural Networks for Personalized Pain Recognition from Physiological Signals. In *IEEE International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW).*

[26] Daniel Lopez Martinez, Ognjen Rudovic, and Rosalind Picard. 2017. Personalized automatic estimation of self-reported pain intensity from facial expressions. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on.* IEEE, 2318–2327.

[27] Shrikanth Narayanan and Panayiotis G Georgiou. 2013. Behavioral signal processing: Deriving human behavioral informatics from speech and language. *Proc. IEEE* 101, 5 (2013), 1203–1233.

[28] Robert Norton. 1983. Measuring marital quality: A critical look at the dependent variable. *Journal of Marriage and the Family* (1983), 141–151.

[29] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The development and psychometric properties of LIWC2015.* Technical Report.

[30] M Carole Pistole. 1989. Attachment in adult romantic relationships: Style of conflict resolution and relationship satisfaction. *Journal of Social and Personal Relationships* 6, 4 (1989), 505–510.

[31] Ming-Zher Poh, Nicholas C Swenson, and Rosalind W Picard. 2010. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE transactions on Biomedical engineering* 57, 5 (2010), 1243–1252.

[32] Lin Shi. 2003. The association between adult attachment styles and conflict resolution in romantic relationships. *American Journal of Family Therapy* 31, 3 (2003), 143–157.

[33] Jeffry A Simpson, W Steven Rholes, and Dede Phillips. 1996. Conflict in close relationships: An attachment perspective. *Journal of personality and social psychology* 71, 5 (1996), 899.

[34] Adela C Timmons, Reout Arbel, and Gayla Margolin. 2017. Daily patterns of stress and conflict in couples: Associations with marital aggression and family-of-origin aggression. *Journal of family psychology* 31, 1 (2017), 93.

[35] Adela C Timmons, Brian R Baucom, Sohyun C Han, Laura Perrone, Theodora Chaspari, Shrikanth S Narayanan, and Gayla Margolin. 2017. New Frontiers in Ambulatory Assessment: Big Data Methods for Capturing CouplesÃćÂĂÂŹ Emotions, Vocalizations, and Physiology in Daily Life. *Social Psychological and Personality Science* 8, 5 (2017), 552–563.

[36] Adela C Timmons, Theodora Chaspari, Sohyun C Han, Laura Perrone, Shrikanth S Narayanan, and Gayla Margolin. 2017. Using multimodal wearable technology to detect conflict among couples. *Computer* 50, 3 (2017), 50–59.

[37] David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* 54, 6 (1988), 1063.

[38] Sarah Weusthoff, Brian R Baucom, and Kurt Hahlweg. 2013. Fundamental frequency during couple conflict: An analysis of physiological, behavioral, and sex-linked information encoded in vocal expression. *Journal of Family Psychology* 27, 2 (2013), 212.

[39] Bo Xiao, Chewei Huang, Zac E Imel, David C Atkins, Panayiotis Georgiou, and Shrikanth S Narayanan. 2016. A technology prototype system for rating therapist empathy from audio recordings in addiction counseling. *PeerJ Computer Science* 2 (2016), e59.

[40] Qian Xu and Qiang Yang. 2011. A survey of transfer and multitask learning in bioinformatics. *Journal of Computing Science and Engineering* 5, 3 (2011), 257–268.

[41] Yu Zhang and Qiang Yang. 2017. A survey on multi-task learning. *arXiv preprint arXiv:1707.08114* (2017).