A Multimodal Approach for Predicting Changes in PTSD Symptom Severity

Adria Mallol-Ragolta, Svati Dhamija, Terrance E. Boult
Vision And Security Technology (VAST) Lab
University of Colorado Colorado Springs
Colorado Springs, Colorado, USA
{amallolr,sdhamija,tboult}@uccs.edu

ABSTRACT
The rising prevalence of mental illnesses is increasing the demand for new digital tools to support mental wellbeing. Numerous collaborations spanning the fields of psychology, machine learning and health are building such tools. Machine learning models that estimate effects of mental health interventions currently rely on either self-reports or measurements of user physiology. In this paper, we present a multimodal approach that combines self-reports from questionnaires and skin conductance physiology in a web-based trauma-recovery regime. We evaluate our models on the EASE multimodal dataset and create PTSD symptom severity change estimators at both total and cluster-level. We demonstrate that modeling the PTSD symptom severity change at the total-level with self-reports can be statistically significantly improved by the combination of physiology and self-reports or just skin conductance measurements. Our experiments show that PTSD symptom cluster severity changes using our novel multimodal approach are significantly better modeled than using self-reports and skin conductance alone when extracting skin conductance features from triggers modules for avoidance, negative alterations in cognition & mood and alterations in arousal & reactivity symptoms, while it performs statistically similar for intrusion symptom.

CCS CONCEPTS
• Applied computing → Health care information systems; Consumer health; Psychology; • Human-centered computing → Empirical studies in HCI;

KEYWORDS
Trauma recovery; PCL-5 questionnaire; CSE-T questionnaire; machine learning; signal processing; skin conductance.

1 INTRODUCTION
A post-traumatic stress disorder (PTSD) is an ongoing life-affecting disorder suffered by people who have experienced a traumatic event, which may be caused by natural disasters or human actions [61]. An epidemiology of mental disorders revealed that 68 out of 1000 adult men and women in the United States may suffer from PTSD at least once in their lives [45]. In order to cope with PTSD, patients undergo recovery treatments that target the re-creation of psychological faculties damaged or deformed by the traumatic experience [34]. Such treatments are usually based on cognitive and behavioral interventions [18, 69]. Recent studies have shown evidence highlighting the use of the internet to deliver PTSD treatment [15] through web-based interventions [11, 52, 54, 74]. A widespread protocol amongst professionals is the posttraumatic assessment using self-reports [30]. However, answering trauma-related questions during the treatment can be challenging for patients [13].

Trauma psychologists have conducted multiple control tests to identify diverse sets of potential contributors to posttraumatic recovery. In these different multivariate analyses, Benight et al. [8] found that perceived trauma-focused coping self-efficacy (CSE-T), explained in Section 2, emerged as a focal mediator of posttraumatic recovery. This finding of CSE-T being critical to trauma recovery makes it generalizable for trauma treatments irrespective of the kind of psychological trauma, whether from natural disasters, terrorist attacks, military combat, sexual or criminal assaults.

An analogous research in biological psychology revealed that skin conductance (SC) response post-trauma was significantly associated with PTSD diagnosis 6-months later [38]. SC responses are often measured to make statements about psycho-physiological processes such as cognitive load, emotional arousal, or threat prediction [16]. Monitoring and modeling the SC responses can thereby aid in non-intrusive measurements of PTSD symptom severity and reduce self-reports. The automatic prediction of PTSD symptom severity will allow web-based trauma-recovery treatments to automatically monitor subjects’ symptom severity as the treatment progresses and adapt interventions to their needs.

In order to estimate PTSD symptom severity changes, in this work we explore the feasibility of replacing CSE-T questionnaires with SC responses and also the fusion of both modalities as shown in Figure 1. Our aim is not to replace clinicians, but to estimate PTSD symptom-severity changes, while patients are interacting with a trauma-recovery-focused website for the potential adaptation of its content. The contributions of this paper are as follows:
The American Psychiatric Association (APA) in the fifth edition of Diagnostic and Statistical Manual of Mental Disorders (DSM-5) defines trauma as the psychological distress experienced by individuals after being exposed to stressful events such as war, disaster, or terrorism, among others [1]. These traumatic experiences might lead to PTSD: a psychiatric disorder that causes subjects to have negative thoughts and feelings related to their traumatic experience after its occurrence, in such a way that affects their daily life.

PTSD Checklist (PCL-5): Trauma severity can be assessed using multiple tests [30]. One of the most widely used tests to evaluate PTSD symptom severity is the PTSD Checklist (PCL) [76], which asks subjects about how much they were bothered by different situations in the past month. On its revised version for DSM-5 [14, 77], PCL-5 consists of a 20-item questionnaire that measures the 4 symptoms of PTSD, which are intrusion, avoidance, negative alterations in cognition & mood, and alterations in arousal & reactivity. Each item in the PCL-5 questionnaire has a 5-point likert scale (from 0 = "Not at all true" to 4 = "Extremely true"). PTSD symptom severity from the PCL-5 questionnaire is scored in 2 different ways. The first method computes a total PTSD symptom severity score, which is a summation of the numerical answers provided to the 20-item questionnaire. The second method involves the computation of PTSD symptom cluster severity scores by summing the numerical answers provided to specific questionnaire items: intrusion symptom is captured by items 1-5, avoidance symptom, by items 6-7, negative alterations in cognition & mood symptom, by items 8-14, and alterations in arousal & reactivity, by items 15-20.

Trauma-Focused Coping Self-Efficacy (CSE-T): Self-efficacy is a key mechanism to approach goals, tasks and face challenges [5]. When undergoing trauma-recovery treatments, self-perception about one’s capabilities to manage intense recovery demands becomes important [8]. Coping self-efficacy (CSE) is defined as a person’s subjective appraisal of his or her ability to cope with the environmental demands of threatening situations [4]. Over the last two decades, CSE has been shown to be predictive of PTSD recovery in different contexts [7, 9, 10, 21, 39, 41, 46, 49, 55, 68]. In order to contextualize self-efficacy measurements [5] related to trauma recovery, a trauma-focused CSE measure (CSE-T) has recently been developed [12]. CSE-T, which asks subjects how capable they are to cope with different situations, is a 9-item psychometric measure used to evaluate subjects perceived capability to cope with challenges and demands from traumatic events. Each item in the CSE-T questionnaire presents choices to subjects from 1 = “very incapable” to 7 = “very capable”. The overall CSE-T score is an average of all items in the questionnaire, resulting in an absolute value ranging from a minimum value of 1 to a maximum of 7.

2 BACKGROUND
The American Psychiatric Association (APA) in the fifth edition of its Diagnostic and Statistical Manual of Mental Disorders (DSM-5) defines trauma as the psychological distress experienced by individuals after being exposed to stressful events such as war, disaster, or terrorism, among others [1]. These traumatic experiences might lead to PTSD: a psychiatric disorder that causes subjects to have negative thoughts and feelings related to their traumatic experience after its occurrence, in such a way that affects their daily life.

PTSD Checklist (PCL-5): Trauma severity can be assessed using multiple tests [30]. One of the most widely used tests to evaluate PTSD symptom severity is the PTSD Checklist (PCL) [76], which asks subjects about how much they were bothered by different situations in the past month. On its revised version for DSM-5 [14, 77], PCL-5 consists of a 20-item questionnaire that measures the 4 symptoms of PTSD, which are intrusion, avoidance, negative alterations in cognition & mood, and alterations in arousal & reactivity. Each item in the PCL-5 questionnaire has a 5-point likert scale (from 0 = “Not at all true” to 4 = “Extremely true”). PTSD symptom severity from the PCL-5 questionnaire is scored in 2 different ways. The first method computes a total PTSD symptom severity score, which is a summation of the numerical answers provided to the 20-item questionnaire. The second method involves the computation of PTSD symptom cluster severity scores by summing the numerical answers provided to specific questionnaire items: intrusion symptom is captured by items 1-5, avoidance symptom, by items 6-7, negative alterations in cognition & mood symptom, by items 8-14, and alterations in arousal & reactivity, by items 15-20.

Trauma-Focused Coping Self-Efficacy (CSE-T): Self-efficacy is a key mechanism to approach goals, tasks and face challenges [5]. When undergoing trauma-recovery treatments, self-perception about one’s capabilities to manage intense recovery demands becomes important [8]. Coping self-efficacy (CSE) is defined as a person’s subjective appraisal of his or her ability to cope with the environmental demands of threatening situations [4]. Over the last two decades, CSE has been shown to be predictive of PTSD recovery in different contexts [7, 9, 10, 21, 39, 41, 46, 49, 55, 68]. In order to contextualize self-efficacy measurements [5] related to trauma recovery, a trauma-focused CSE measure (CSE-T) has recently been developed [12]. CSE-T, which asks subjects how capable they are to cope with different situations, is a 9-item psychometric measure used to evaluate subjects perceived capability to cope with challenges and demands from traumatic events. Each item in the CSE-T questionnaire presents choices to subjects from 1 = “very incapable” to 7 = “very capable”. The overall CSE-T score is an average of all items in the questionnaire, resulting in an absolute value ranging from a minimum value of 1 to a maximum of 7.

3 RELATED WORK
Various researchers have studied how new technologies can be used to estimate and evaluate PTSD [17]. For instance, Hinrichs et al. [37] analyzed skin conductance signals captured with electrodes via a mobile device to show the existence of a positive correlation between skin conductance and PTSD symptoms. Webb et al. [78] used virtual reality videos to trigger subjects while evaluating physiology to identify subjects with and without PTSD. Nonetheless, to the best of our knowledge, this is the first work that employs a multimodal machine learning-based approach using CSE-T questionnaires and skin conductance to predict changes in PTSD symptom severity.
Figure 2: Illustration of the protocol followed in the first 2 sessions of the web-based trauma-recovery treatment, which EASE database was gathered from. This visualization illustrates when self-reports were answered, and when physiological signals were recorded during the treatment.

The key to design an efficient machine learning-based approach for PTSD prediction is the different information and modalities captured from the subjects themselves. Kessler et al. [44] used subjects information related to socio-demographics, trauma experiences, and prior mental disorders obtained through face-to-face interviews. Galatzer-Levy et al. [31] and Karstoft et al. [42] employed information gathered during subjects Emergency Department admissions and posterior phone interviews regarding traumatic experience details, admission data, and trauma-severity scores, among others. On the other hand, Rozgic et al. [63] used neurophysiological features as electroencephalogram (EEG), electrocardiogram (ECG), galvanic skin response (GSR) and head motion signals, while Zhuang et al. [82] explored speech and EEG signals.

Physiology in general and skin conductance in particular are analyzed in stress detection problems due to the correlation between stress and galvanic skin response [47, 70]. For instance, Healey and Picard [33] used physiological signals to predict stress levels of subjects while driving. Sano and Picard [64] investigated the automatic binary prediction of stress from subjects using subjective measurements and physiological, behavioral and mobile phone usage data. Hernandez et al. [35] aimed to predict call center workers stress levels while helping customers using galvanic skin response, call logs, and stress self-reports after each call.

One of the main challenges to deal with when combining physiology and machine learning in automatic human emotion and behavior understanding is the feature extraction from physiological signals. As a result, multiple sets of features from physiological signals are employed in the literature: some of the works extracted features based on signal statistics [3, 32, 36, 43, 51, 56, 59, 79], while others worked with nonlinear features [22].

Physiological signals can be captured using either medical equipment or wearable sensors [48, 60]. The main advantage wearable sensors offer is the possibility to perform in-home measurements, which allow, among other applications, the post-surgical recovery of patients from their own homes [58]. Nevertheless, either artifacts or noise can be induced to physiological signals due to undesired displacements of the sensors while measuring, among other causes. Hence, noise identification and cancellation deserves attention: Sweeney and Ward [71] described algorithms to remove artifacts, while Taylor et al. [72] and Xia et al. [80] explored the use of machine-learning techniques to identify noisy segments in skin conductance signals.

4 EASE DATASET

The Engagement Arousal Self-Efficacy (EASE) multimodal dataset [27, 53] contains facial and audio data, physiological signals and self-reports. The data was collected from subjects while undergoing a web-based trauma-recovery treatment. The web-based intervention consisted of 3 sessions, with 2 out of 6 possible modules in each. However, we focus our analysis on the first 2 sessions of the treatment, in which triggers (TR) and relaxation (RX) modules were assigned. TR modules are aimed at educating participants about PTSD symptoms and their prevention, while RX modules include video demonstrations about breathing and muscle relaxation exercises. Module order in the first session of the treatment was randomly assigned, but reversed in the second session, as shown in Figure 2. In other words, subjects who were assigned triggers followed by relaxation modules in the first session of the treatment were assigned relaxation followed by triggers modules in the second session.

EASE dataset was gathered from 110 participants over 3 sessions – 88 females, 15 males and 7 unspecified subjects – with ages between 18 and 79. This work uses data from skin conductance, CSE-T self-reports and PCL-5 questionnaires. Due to incomplete self-reports and corrupted sensor data, the usable data is reduced to 66 participants – 57 females and 9 males. From the total 110 participants, 10 subjects dropped out of treatment, i.e. before Session 2, 9 did not completely answer CSE-T and/or PCL-5 questionnaires, and 25 were discarded due to sensor data corruption.

The EASE protocol required participants to sit in front of a computer and interact with the web-based trauma-recovery intervention. At the beginning of each session, while data collection was running, participants watched a neutral introductory video, and then, participants were asked to remain quiet for a minute. After this quiet time, they worked on modules corresponding to triggers or relaxation in the specific order they were assigned. Self-reported questionnaires were acquired at different moments during the treatment. PCL-5 questionnaires were answered by participants...
at the very beginning and at the very end of the treatment. CSE-T questionnaires were answered before and after modules assigned during each session of the treatment (3 reports per session), as shown in Figure 2.

Raw questionnaires, data, and code are accessible through the URL http://vast.ucsd.edu/multimodal-predicting-ptsd for specific details on patients’ assessment and results reproducibility.

5 METHODOLOGY: PTSD SYMPTOM SEVERITY PREDICTOR SYSTEM

The system implemented aims to quantify the increment or regression of PTSD symptom severity using not only monomodal data with skin conductance signals or data from CSE-T questionnaires, but also multimodal data through the combination of both data types. Although the use of other physiological signals, such as heart rate or respiration rate, could also be explored, skin conductance is the only physiological measure employed in our work because of its relationship to psycho-physiological processes, such as stress, pain and depression [16]. Since the goal is to quantify the change in PTSD symptom severity as a result of the treatment, we define our target as the difference between PTSD symptom severity scores computed from PCL-5 questionnaires, which will be referred as APCL scores. The remaining of this section addresses the different stages of the PTSD symptom severity predictor system implemented, which is illustrated in Figure 1.

5.1 Signal Preprocessing

Skin conductance signals from the EASE dataset were recorded using electrodes at a sampling rate of $f_s = 256$ Hz with a 16-bit quantizer. Since quantization noise or artifacts can be present in the physiological signals collected, our first step is to filter them using a 10th order low-pass Butterworth filter [57]. The Butterworth filter has been widely used in the literature to preprocess physiological signals [73, 81]. It has constant gain in the passband, a quick roll-off around the cut-off frequency, and no ripple. Particularly, as Hernando-Gallego et al. [36] suggested that skin conductance signals can be downsampled to $f_s = 8$ Hz or even $f_s = 4$ Hz without losing signal quality, we set the cut-off frequency of the 10th order low-pass Butterworth filter to $f_c = 8$ Hz.

In EASE, Sessions 1 and 2 of the web-based treatment were conducted on different days. Hence, external stimuli to the treatment itself, like changes in ambient temperature, can impact skin conductance measurements. We remove inter-session variability by normalizing each noise-free skin conductance signal with the information recorded during the quiet time from the corresponding session [20], as mathematically formulated in the equation below:

$$SC_{norm} = \frac{SC_{noise-free} - \min(\text{SC}_{\text{quiet time}})}{\max(\text{SC}_{\text{quiet time}}) - \min(\text{SC}_{\text{quiet time}})}.$$ (1)

SC response signal can be decomposed into its tonic and phasic components [16]. While the tonic component is a slowly varying signal that changes over the course of minutes, the phasic component, which is impacted when subjects were involved in mental activities [2], is a fast varying signal that changes over the course of seconds. Hence, the last stage of the preprocessing involves the decomposition of normalized noise-free skin conductance signals into their tonic and phasic components, which is computationally performed through the MATLAB open-source software Ledalab [6].

5.2 Feature Extraction

After preprocessing SC signals, we then proceed to extract features from them. Since either combined or specific treatment-module models have been explored in previous works on the EASE dataset [26, 28, 53], we aim to compare PTSD symptom severity models trained with skin conductance features extracted from TR+RX, TR and RX treatment modules. Although no statistical evidence was found to support the use of specific treatment-module models (TR/RX) over a combined model of TR+RX [53], we believe that the analysis of TR+RX treatment-module models is necessary, since works using facial data found differences [26, 28]. Features are extracted from the corresponding portions of skin conductance signals and concatenated in a chronological order. In this work, we compare 2 skin conductance feature sets, which are inspired by the features extracted from the galvanic skin response defined in the Toolbox for Emotional feAture extraction from Physiological signals (TEAP) [67], but with some particularities.

$SC_{org}$: This set extracts the following features from the normalized noise-free skin conductance signals: mean and standard deviation of the signal, in addition to the number of peaks per second, the mean of the peaks magnitude, which is computed as the difference between a maximum and its preceding minimum, and the mean of the rise times. Based on the threshold defined by TEAP, only those peaks whose magnitude is greater than 0.01 $\mu S$ are considered valid peaks for our computations. The abbreviation org in the set name stands for original, since features are extracted from the original normalized noise-free skin conductance signals.

$SC_{dec}$: Oppositely, this set extracts mean and standard deviation from the tonic component of the normalized noise-free skin conductance signals, and the number of peaks per second, the mean of the peaks magnitude and the mean of the rise times from their phasic component. Only those peaks from the phasic component whose magnitude is greater than 0.1 $\mu S$ are considered valid peaks for our computations [25]. The abbreviation dec in the set name stands for decomposed, since features are extracted from the tonic and phasic components, which skin conductance signals are decomposed into.

Regardless of the skin conductance feature set employed, features extracted are 2-normalized [23], so they are zero-mean and unit-variance before being used as input to the machine-learning stage of our pipeline.

In our analysis, CSE-T questionnaires are also used as system input. Since they are reported 6 times in the first 2 sessions of the treatment, we compute the 6 overall CSE-T scores and use them as CSE-T features to be used as input for the machine-learning stage of our pipeline. Due to the agreement among psychologists regarding the use of CSE-T as a valid predictor of PTSD [7, 9, 10, 21, 39, 41, 46, 49, 55, 68], PTSD symptom severity models trained with CSE-T questionnaires are set as a baseline.

5.3 Machine Learning

The aim of this work is to quantify the PTSD symptom severity change experienced by subjects who interacted with a web-based
We divide the original SC $\Delta$MSE between actual and predicted scores with SC models are lower than the MSEs mean measured with results reported in Table 1, we can state that both SC feature sets, performance of CSE-T models is used as the baseline. According to the statistical significance threshold is reduced to 0.01. Correlation [24]. After including the Bonferroni correction term, the statistical significance threshold is initialized with a $t$-value of 0.05 by 3 to apply the Bonferroni correction. After including the Bonferroni correction term, the statistical significance threshold is initialized with a $t$-value of 0.05. We perform the baseline significantly when the features are extracted from TR+RX treatment modules. Therefore, for all monomodal and multimodal experiments in the following sections, we build PTSD symptom severity models with SC features extracted from SC signal portions and SC vs CSE-T+SC) performed independently for TR+RX, TR and RX treatment-module models. SC models trained with SC$_{dec}$ significantly outperform the baseline for TR+RX and RX treatment-module models, which skin conductance features are extracted from, and SC$_{dec}$ models significantly outperform SC$_{org}$ models when skin conductance features are extracted from RX treatment modules.

<table>
<thead>
<tr>
<th>Input Data</th>
<th>CSE-T</th>
<th>TR+RX</th>
<th>SC</th>
<th>TR</th>
<th>RX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature Set</td>
<td>$SC_{org}$</td>
<td>$SC_{dec}$</td>
<td>$SC_{org}$</td>
<td>$SC_{dec}$</td>
<td>$SC_{org}$</td>
</tr>
<tr>
<td>Mean</td>
<td>294.9</td>
<td>138.3</td>
<td>135.1</td>
<td>178.8</td>
<td>179.9</td>
</tr>
<tr>
<td>Median</td>
<td>137.3</td>
<td>34.8</td>
<td>46.4</td>
<td>47.7</td>
<td>50.1</td>
</tr>
<tr>
<td>Std</td>
<td>466.1</td>
<td>254.3</td>
<td>231.7</td>
<td>320.1</td>
<td>294.4</td>
</tr>
<tr>
<td>BL $p$-value</td>
<td>$&lt;0.01$</td>
<td>$&lt;0.01$</td>
<td>$0.06$</td>
<td>$0.05$</td>
<td>$0.21$</td>
</tr>
<tr>
<td>FEAT $p$-value</td>
<td>0.77</td>
<td>0.96</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: MSEs comparison of SC models trained with $SC_{org}$ and $SC_{dec}$ feature sets when predicting total $\Delta$PCL scores. BL $p$-value compares SC models with the baseline, while FEAT $p$-value compares SC models trained with $SC_{org}$ and $SC_{dec}$ feature sets extracted from the same treatment modules. SC models trained with $SC_{dec}$ significantly outperform the baseline for TR+RX and RX treatment-module models, which skin conductance features are extracted from, and SC$_{dec}$ models significantly outperform SC$_{org}$ models when skin conductance features are extracted from RX treatment modules.

### 6.1 Skin Conductance Feature Sets Comparison

In this section, we compare the predictive capability of PTSD total symptom severity models trained with $SC_{org}$ and $SC_{dec}$ features extracted from SC signals corresponding to TR+RX, TR and RX treatment modules. The purpose of this experiment is to assess the suitability of extracting features from the original noise-free SC signal or from their tonic & phasic components. Monomodal models trained with either CSE-T or SC are validated using LOSO cross-validation method. The MSEs computed fold-wise are synthesized in Table 1. The predictive capability of monomodal models are compared using 2-tailed paired $t$-tests also reported in Table 1. The statistical significance threshold is initialized with a $p$-value of 0.05. In order to minimize the family-wise error due to the 3 comparisons (CSE-T vs $SC_{org}$, CSE-T vs $SC_{dec}$ and $SC_{org}$ vs $SC_{dec}$) performed independently for TR+RX, TR and RX treatment-module models, we divide the original $p$-value of 0.05 by 3 to apply the Bonferroni correction [24]. After including the Bonferroni correction term, the statistical significance threshold is reduced to 0.01.

As mentioned earlier, for all analyses in this paper, the performance of CSE-T models is used as the baseline. According to the results reported in Table 1, we can state that both SC feature sets, $SC_{org}$ and $SC_{dec}$, outperform the baseline. The MSEs mean measured with SC models are lower than the MSEs mean measured with CSE-T models, for all treatment-module models. The $p$-values computed as a result of comparing the baseline performance with SC models are presented as BL $p$-value in the table. The results demonstrate that SC models trained with $SC_{dec}$ features significantly outperform the baseline for TR+RX and RX treatment-module models. However, SC models trained with $SC_{org}$ features only outperform the baseline significantly when the features are extracted from TR+RX treatment modules. The $p$-values computed when comparing the SC models trained with $SC_{org}$ and $SC_{dec}$ features, FEAT $p$-value, indicate that models trained with $SC_{dec}$ features significantly outperform those trained with $SC_{org}$ for RX treatment modules. Thus, this result supports the hypothesis that the feature extraction from the tonic & phasic components of normalized noise-free skin conductance signals is advantageous. Also, observe that the MSEs for $SC_{dec}$ are lower or at par with $SC_{org}$ for the different treatment modules. Therefore, for all monomodal and multimodal experiments in the following sections, we build PTSD symptom severity models with SC features extracted from $SC_{dec}$ set.

### 6.2 Monomodal and Multimodal PTSD Symptom Severity Models Analysis

This section evaluates performance differences of PTSD symptom severity models when trained with monomodal and multimodal data at both total and cluster-level. The statistical significance threshold is set to 0.01, as a result of minimizing the family-wise error due to the 3 comparisons (CSE-T vs SC, CSE-T vs CSE-T+SC and SC vs CSE-T+SC) performed independently for TR+RX, TR and RX treatment-module models after applying the Bonferroni correction [24] to the initial statistical significance threshold of 0.05.

#### 6.2.1 Total-level $\Delta$PCL Scores Prediction

The predictive capability of monomodal and multimodal PTSD symptom severity models when predicting total-level $\Delta$PCL scores is assessed in this passage. To perform this analysis, CSE-T model performance is set as the baseline, and $SC_{dec}$ features are extracted from SC signal portions.
T+SC models significantly outperform the baseline for TR+RX and with 2-tailed paired

Table 2: MSEs comparison of PTSD total symptom severity models trained with monomodal data, CSE-T questionnaires (left) and skin conductance (middle), and multimodal data (right). SCdec skin conductance features are extracted from TR treatment modules.

<table>
<thead>
<tr>
<th>Modules</th>
<th>TR-RX</th>
<th>TR</th>
<th>RX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input Data</td>
<td>CSE-T</td>
<td>SC</td>
<td>CSE-T-SC</td>
</tr>
<tr>
<td>Mean</td>
<td>294.9</td>
<td>135.1</td>
<td>133.0</td>
</tr>
<tr>
<td>Median</td>
<td>137.9</td>
<td>46.4</td>
<td>41.1</td>
</tr>
<tr>
<td>Std</td>
<td>466.1</td>
<td>231.7</td>
<td>221.7</td>
</tr>
<tr>
<td>BL p-value</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>MOD p-value</td>
<td>0.79</td>
<td>0.07</td>
<td>0.36</td>
</tr>
</tbody>
</table>

6.2.2 Cluster-level ΔPCL Scores Prediction. This subsection focuses on the evaluation of monomodal and multimodal PTSD symptom cluster severity models validated using LOSO cross-validation method. As in prior sections, to perform this analysis, the performance of CSE-T models is set as the baseline. For SC and CSE-T+SC models, SCdec feature set is extracted from SC signal portions corresponding to TR-RX, TR and RX treatment modules. Statistics extracted from the MSEs computed fold-wise regarding the prediction of intrusion, avoidance, negative alterations in cognition & mood, and alterations in arousal & reactivity symptoms are synthesized in Table 3. Performance differences between monomodal and multimodal models are evaluated using 2-tailed paired t-tests, which are also reported in Table 3. BL p-value reports performance differences between the baseline and SC or CSE-T+SC models. MOD p-value reports performance differences between SC and CSE-T+SC modalities.

The first observation to highlight from the results in Table 3 is that, for all symptoms and most treatment modules, SC and CSE-T+SC models outperform the baseline. Both SC and CSE-T+SC cause significant reduction in MSE from the baseline CSE-T. The BL p-values obtained as a result of comparing the performance of SC PTSD symptom cluster severity models and the baseline CSE-T indicate that although SC models do not statistically outperform the baseline for intrusion symptom, they do so for avoidance, negative alterations in cognition & mood, and alterations in arousal & reactivity symptoms for TR-RX treatment modules.

Similarly, according to the BL p-values computed when comparing the performance between multimodal model CSE-T+SC and the baseline CSE-T, we realize that, although multimodal models are statistically similar to the baseline for intrusion symptom, they significantly outperform the baseline for avoidance, negative alterations in cognition & mood, and alterations in arousal & reactivity symptoms, irrespective of the treatment modules.

The MOD p-values obtained by comparing performance differences between SC and CSE-T+SC PTSD symptom cluster severity models demonstrate that SC models perform statistically similar to multimodal models for intrusion symptom, regardless of the treatment modules. However, the results show compelling evidence that multimodal models significantly outperform SC models for avoidance, negative alterations in cognition & mood, and alterations in arousal & reactivity symptoms, when SCdec features are extracted from SC signals corresponding to TR treatment modules.

Figure 4 compares the distributions of actual and predicted total-level ΔPCL scores, when SCdec features are extracted from
Figure 4: Actual and predicted cluster-level \( \Delta PCL \) scores distributions obtained from the evaluation of PTSD symptom cluster severity models trained with monomodal and multimodal data. BL \( p \)-value compares SC with CSE-T+SC models. The results show that SC and CSE-T+SC models reduce the baseline MSE in most of the cases irrespective of the treatment modules, which \( SC_{dec} \) features are extracted from. For intrusion symptom, neither SC nor CSE-T models significantly outperform the baseline. On the other hand, for avoidance, negative alterations in cognition & mood, and alterations in arousal & reactivity symptoms, CSE-T+SC models significantly outperform not only the baseline but also SC models when \( SC_{dec} \) features are extracted from TR treatment modules.

Table 3: MSEs comparison of PTSD symptom cluster severity models trained with monomodal and multimodal data. BL \( p \)-value compares the baseline with SC or CSE-T+SC models, while MOD \( p \)-value compares SC with CSE-T+SC models. The results convey that SC and CSE-T+SC models reduce the baseline MSE in most of the cases irrespective of the treatment modules, which \( SC_{dec} \) features are extracted from. For intrusion symptom, neither SC nor CSE-T models significantly outperform the baseline. On the other hand, for avoidance, negative alterations in cognition & mood, and alterations in arousal & reactivity symptoms, CSE-T+SC models significantly outperform not only the baseline but also SC models when \( SC_{dec} \) features are extracted from TR treatment modules.

SC signals. The histograms correspond to TR treatment-module models with the aim to understand the predictive capability of the monomodal and multimodal PTSD symptom cluster severity models trained. We observe that the distributions of scores predicted with multimodal models are close to normal for all 4 PTSD symptoms and the actual cluster-level \( \Delta PCL \) scores.

### 7 DISCUSSION AND CONCLUSION

The results obtained in Section 6.1 state that the performance of SC models trained with \( SC_{dec} \) features extracted from RX treatment modules significantly outperform SC models trained with \( SC_{arg} \) features. This result supports the effectiveness of extracting skin...
conductance features from tonic and phasic components of normalized noise-free skin conductance signals over extracting features from the normalized noise-free skin conductance signals directly. It also justifies the use of $SC_{dec}$ features to assess the performance of monomodal and multimodal PTSD symptom severity models at both total and cluster-level.

The evaluation of PTSD symptom severity models at a total-level, in Section 6.2.1, indicates that SC models significantly outperform the baseline CSE-T for TR+RX and RX treatment modules. The results obtained also show that our novel multimodal model, CSE-T+SC, significantly outperforms the baseline for TR+RX, TR and RX treatment modules. Therefore, both SC and the fusion of CSE-T & SC provide significant improvements for estimating total-level $PCL$ over just CSE-T questionnaires. However, SC and CSE-T+SC PTSD symptom severity models perform statistically similarly when predicting total $PCL$ scores.

Moreover, the evaluation of monomodal and multimodal PTSD symptom severity models when predicting cluster-level $PCL$ scores, in Section 6.2.2, reveal 2 common trends amongst symptoms: the first one involves modeling changes in intrusion symptom, while the other one in modeling changes in severity for avoidance, negative alterations in cognition & mood, and alterations in arousal & reactivity symptoms. For estimating changes in intrusion symptom severity, we notice that, although there is MSE reduction from CSE-T using either CSE-T+SC or SC alone, neither SC nor our multimodal model of CSE-T+SC statistically outperform the baseline, i.e. CSE-T models. Furthermore, the results allow us to assert that SC and multimodal models perform statistically similarly for intrusion symptom. Hence, the replacement of CSE-T for estimating changes in intrusion is a debatable choice between CSE-T+SC and SC alone and thereby contingent on the treatment module being modeled.

The estimation of changes in PTSD symptom severity for avoidance, negative alterations in cognition & mood, and alterations in arousal & reactivity symptoms shows that both SC and CSE-T+SC models significantly outperform the baseline when $SC_{dec}$ features are extracted from TR+RX treatment modules.

The statistical analysis of $p$-values when comparing the performances between SC and our novel multimodal model CSE-T+SC indicates that PTSD symptom cluster severity models trained with multimodal data significantly outperform the models trained with SC data only, when $SC_{dec}$ features are extracted from portions of skin conductance signals corresponding to TR treatment modules. Thus, we can highlight that the multimodal approach significantly outperforms the use of either SC or CSE-T individually when predicting PTSD symptom severity changes for avoidance, negative alterations in cognition & mood, and alterations in arousal & reactivity symptoms. Our novel methodology of extracting SC features by employing signal processing techniques and decomposing the original signal into tonic & phasic components, and then fusing the extracted SC features with CSE-T self-reports surpasses all other approaches to model changes in PTSD symptom cluster severity.

8 FUTURE WORK

This work evaluated the use of self-reported questionnaires and skin conductance to train monomodal and multimodal PTSD symptom severity models to estimate changes at both total and cluster-level.

This research was targeted at post-traumatic stress disorder (PTSD) recovery and a specific type of self-efficacy, i.e. coping self-efficacy for trauma (CSE-T). Other psychometric measures, such as subjective units of distress scale (SUDS) [50], responses to script-driven imagery (RSI) [40], etc., can also be combined with CSE-T measures to get a better estimate of distress for psychological trauma and subjects’ emotional state. Moreover, other physiological signals, such as heart rate or respiration rate, could also be included in the future to train PTSD symptom severity models with the aim to provide a complete study on the interaction of physiology and PTSD symptom severity. The pattern of physiological arousal defines physiological toughness and, in interaction with psychological coping, corresponds to positive performance in even complex tasks with emotional stability and with immune system enhancement [29, 83]. Also, facial expressions provide valuable information regarding the emotional state of a subject [19, 62, 65]. Thus, sophisticated methods that use facial features and combine self-reports to train both monomodal and multimodal PTSD symptom severity models could also be a possible extension of this work. Furthermore, different combinations between these suggested data sources and those already used should be analyzed to assess the impact that these different input data sources have on both monomodal and multimodal PTSD symptom severity models.

ACKNOWLEDGEMENT

This work is partially supported by NSF Research Grant SCH-INT 1418520 “Learning and Sensory-based Modeling for Adaptive Web-Empowerment Trauma Treatment”. A special thanks to Dr.Charles C. Benight, Dr. Kotaro Shoji, Ms. Carolyn Yeager and the UCCS Trauma Health and Hazards Center for collecting the EASE dataset.

REFERENCES


Preprint of paper to appear ACM International Conference on Multimodal Interaction (ICMI), 2018