Automated Pain Assessment in Children Using Electrodermal Activity and Video Data Fusion via Machine Learning

Busra T. Susam, Nathan T. Riek, Murat Akcakaya, Xiaojing Xu, Virginia R. de Sa, Hooman Nezamfar, Damaris Diaz, Kenneth D. Craig, Matthew S. Goodwin, and Jeannie S. Huang

***Abstract*— *Objective:* Pain assessment in children continues to challenge clinicians and researchers, as subjective experiences of pain require inference through observable behaviors, both involuntary and deliberate. The presented approach supplements the subjective self-report-based method by fusing electrodermal activity (EDA) recordings with video facial expressions to develop an objective pain assessment metric. Such an approach is specifically important for assessing pain in children who are not capable of providing accurate self-pain reports, requiring nonverbal pain assessment. We demonstrate the performance of our approach using data recorded from children in post-operative recovery following laparoscopic appendectomy. We examined separately and combined the usefulness of EDA and video facial expression data as predictors of children's self-reports of pain following surgery through recovery. Findings indicate that EDA and facial expression data independently provide above chance sensitivities and specificities, but their fusion for classifying clinically significant pain vs. clinically nonsignificant pain achieved substantial improvement, yielding 90.91% accuracy, with 100% sensitivity and 81.82% specificity. The multimodal measures capitalize upon different features of the complex pain response. Thus, this paper presents both evidence for the utility of a weighted maximum likelihood algorithm as a novel feature selection method for EDA and video facial expression data and an accurate and objective automated classification algorithm capable of discriminating clinically significant pain from clinically nonsignificant pain in children.**

***Index Terms*—Electrodermal activity (EDA), galvanic skin response (GSR), computer vision, facial expression, pain assessment.**

I. Introduction

CONSIDERABLE investments are being made to advance pain measurement for clinical and research purposes.[1] Current definitions of pain primarily focus on subjective experience, necessitating inference through observable manifestations, either biomarkers or behavioral expression [1]. Pain is complex, comprised of sensory-discriminative, cognitive-evaluative and emotional-motivational components [2], with physiological and behavioral correlates associated with voluntary (e.g., self-report) and involuntary responses (e.g., reflexive withdrawal, autonomic activity, and facial expression) that have functional adaptive importance [3], [4].

Pain evaluation can be achieved through self-report, nonverbal behavior, or physiological responses [5], with knowledge about sources of pain in tissue damage [6], [7] also important [6], [8]. Self-report can be sufficiently nuanced to tap multiple dimensions of pain in adults and in children [9]. Children as young as 5 years of age can discriminate sensory and emotional features of the pain experience [10]. However, self-report requires substantial cognitive, linguistic, and social competencies not always available in children <8 years of age and other populations (e.g., infants, people with developmental delays, brain damage or dementia) [11]. Efforts to generate more objective pain assessment techniques using brain imaging [12] or biomarkers [13] require patients to be transported to another site for measurement or bedside sampling and laboratory analysis – both of which do not provide immediate results for timely clinical intervention. Instead, a time-sensitive technique to objectively assess pain at the bedside is needed. Such an assessment might be possible using nonverbal and physiological responses that provide understanding of emotional, sensory, and cognitive modulation of pain experiences.

[1]NIH Pain Consortium call for proposals: [Online]. Available: https://painconsortium.nih.gov/Funding_Research/Funding_Opportunities).

Manuscript received September 2, 2020; revised December 22, 2020, June 15, 2021, and June 22, 2021; accepted June 28, 2021. Date of publication July 9, 2021; date of current version December 23, 2021. This work was supported by the National Institutes of Health under Grant R01 NR013500. *(Corresponding author: Busra T. Susam.)*

Busra T. Susam is with the Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA 15260 USA (e-mail: bts38@pitt.edu).

Nathan T. Riek and Murat Akcakaya are with the Department of Electrical and Computer Engineering, University of Pittsburgh, USA.

Xiaojing Xu is with the Department of Electrical and Computer Engineering, University of California, USA.

Virginia R. de Sa is with the Department of Cognitive Science, Halicioglu Data Science Institute, University of California San Diego, USA.

Hooman Nezamfar is with the Department of Electrical and Computer Engineering, Northeastern University, USA.

Damaris Diaz is with the Department of Pediatrics, University of California San Diego, USA.

Kenneth D. Craig is with the Department of Psychology, University of British Columbia, Canada.

Matthew S. Goodwin is with the Department of Health Sciences, Northeastern University, USA.

Jeannie S. Huang is with the Division of Gastroenterology, Rady Children's Hospital and Department of Pediatrics, University of California, USA

This article has supplementary downloadable material available at https://doi.org/10.1109/TBME.2021.3096137, provided by the authors.

Digital Object Identifier 10.1109/TBME.2021.3096137

Nonverbal pain responses include facial expressions, body movements, and vocalizations that closely associate with emotional distress [14]–[16]. Physiological responses such as electrodermal activity (EDA), electrocardiography (ECG), and electromyography (EMG) have been shown to provide substantial information about pain given interactions between neural structures and autonomic control in pain sensation [17], [18]. In the current work, we demonstrate a machine-learning method for objectively differentiating clinically relevant pain from clinically nonsignificant pain using weighted Bayesian fusion of EDA and facial video recordings.

Machine learning is a data-driven approach able to generate algorithms for automatically detecting pain phenotypes using physiological and behavioral data [19]. Developments in computer vision and wearable peripheral physiological sensing devices provide technology that can be implemented in research in clinical settings. The data they produce also enable computational science strategies for automating pain assessments.

Prior studies using behavioral expression (facial and bodily activity) have explored machine learning algorithms to recognize pain in newborn infants [20], [21] and children [22], differentiating affective and painful vocal and facial expressions [23]. High pain detection accuracy (85%) using similar measures has also been reported in studies involving adults [24]–[30]. Studies using peripheral physiological measures have identified patterns in ECG, EMG, and EDA that associate with pain [31], [32], including differentiating pain from emotional states of surprise and boredom [33].

An increasing number of studies suggests that EDA accessed through ambulatory sensors can provide an objective means of measuring emotional distress associated with pain [31], [34]–[37]. In previous studies, different physiological modalities have been analyzed and salient features for pain assessment have been extracted from such modalities [38], [39]. Findings demonstrate that EDA signals mostly outperform other physiological signals including EMG and ECG in terms of accurate pain assessment. However, tonic fluctuations and phasic response delays in EDA can present challenges in making accurate inferences [22]. To overcome these difficulties, we developed a framework using timescale decomposition (TSD) for statistical feature extraction from EDA signals and a linear support vector machine (SVM) based classifier to distinguish clinically significant pain from clinically nonsignificant pain conditions. This approach decomposes EDA signals into shifting windows of time to enable identification of short (phasic) and long-term (tonic) changes over the signal, with subsequent feature discovery leading to the creation of a pain detection algorithm [40].

The current study uses extended sets of both EDA and facial expression video data and includes a novel fusion and pain inference algorithm to test the hypothesis that EDA and video combined will better predict self-reported pain in children recovering from surgery.

II. Related Work

Existing studies in the literature on pain assessment have combined different measurement modalities with various data fusion approaches [41], [42]; specifically, early fusion [43]–[46] and late fusion [38], [39], [47]. Early fusion is the training of a single classifier using a high dimensional data set, which involves concatenation of features extracted from each modality [39]. Late fusion combines the output of these modalities, including output of classifiers trained in each modality in order to compute a final decision architecture using majority vote or averaged classification scores of base classifiers [48]. While the early fusion approach is simple, it suffers from the limitation of high dimension, and it cannot deal with missing data [49]. It also combines non-comparable dimensions from different modalities which is problematic as changing the units or scale function for one modality will change its relative weighting. (This problem cannot be fixed by normalizing (e.g., by z-scoring) individual dimensions as previously comparable dimensions (e.g., neighboring pixels) lose their useful structure [50]) .Late fusion does not suffer from these problems, but it can be complex and computationally intensive. Therefore, in the present study we introduce a novel fusion approach that utilizes a weighted Bayesian fusion over a clinically significant/nonsignificant pain conditional probability distribution of linear support vector machine (Linear SVM) classifier scores obtained for EDA and video data separately.

Most automated pain recognition studies involve healthy adults and experimental pain stimulation (See Table I). Heat stimulation is the most commonly used due to its precise control of stimulus intensity and timing [51]. Lopez *et al.* [46], Werner *et al.* [45], and Kachele *et al.* [52] applied different early fusion architectures on the Bio Vid heat data set which includes both video and physiological signals (i.e., EMG, EDA, ECG) to classify pain. Findings across these studies demonstrate that pain induced by heat differ from a nonsignificant pain condition with an accuracy rate of 82.7% in [46], 80.6% in [45], and 78.9% in [52]. Kessler *et al.* [53] evaluated the benefit of fusing photoplethysmography (rPPG) and respiratory data (RSP) with video signals in the SenseEmotion Database. The results of their fusion algorithm yielded an accuracy rate of 71.85%. Using SenseEmotion Database with a late fusion approach. Thiam *et al.* [39] achieved an accuracy of 82% differentiating the highest pain elicitation temperature from the lowest pain elicitation temperature.

While existing multi-modality automated pain assessment studies utilize well-controlled stimuli in highly controlled laboratory environments, few examine clinical populations, even fewer focus on children, and no other groups target assessments made at the clinical bedside that we are aware of. There are compelling reasons to study pain in children and adolescents at the bedside, strongest among them being: (1) there are substantial correlations between pain syndromes in children and pain syndromes in adulthood; and (2) intervention in children/adolescents offers a chance for prevention of future pathology [54]. Hence, in the current study we focus on children recovering from surgical appendectomy to assess automated pain recognition. To the best of our knowledge, there is currently no other published work on pediatric pain evaluation using multimodal data fusion besides our own. In our previous work [55], we obtained promising results through late fusion detecting pediatric pain in children. In [55], we first applied Linear SVM on extracted EDA and video data to obtain Linear SVM scores.

TABLE I
STUDIES USING FUSION OF MULTI MODALITIES IN ACUTE PAIN STIMULI

Author	Studied Population	Modalities	Fusion architecture
Kachele et al. [38]	Heathy subjects (N=86; age range 18-35)	Fusion of Video, EMG, EDA, ECG	Best: 0 vs 4 pain threshold -Random Forest -LOPO* -Late fusion ACC**: 83.1%
Thiam et al. [39]	Healthy subjects (N=40; age range ≥ 18)	Fusion of Video, audio, EMG, ECG, RSP, EDA	Best: 0 vs 3 pain threshold -Random forest -LOPO -Late Fusion +Mean fixed mapping ACC: 82%
Werner et al. [45]	Heathy subjects (N=87; age range: 18-35)	Fusion of Video, EDA, ECG	Best BSLN vs 4 pain -Random forest -LOPO -Early fusion ACC:80.6 %
Lopez et al. [46]	Healthy subjects (N=87; age range: 18-35)	Fusion of ECG and EDA	Best: 0 vs 4 pain threshold -Multitask neural network - 10-fold cross validation -Early fusion ACC: 82.75%
Kachele et al. [52]	Healthy subjects (N=87; age range: 18-35)	Fusion of Video, EMG, EDA, ECG	Best: 0 vs 4 pain threshold -Early Fusion -Random Forest -LOPO + SFFS ACC:78.90%
Kessler et al. [53]	Healthy subjects (N=40; age range ≥ 18)	Fusion of Video, RSP, ECG, rPPG	Two class classification -Random Forest -LOPO -Score Fusion -Pseudo Inverse Hierarchical ACC: 71.85%
Xu et al. [55]	Pediatric research participant (N=42; age range ≥10)	Fusion of EDA and Video	Pain vs No-pain -Late Fusion -Linear SVM scores -LOPO -LDA ACC=75%

*Leave-one-participant-out.
**Accuracy

Then, the Linear SVM scores of EDA and video data were fused using Linear Discriminant analysis (LDA), followed by leave-one-participant-out (LOPO) cross validation. The results of late fusion yielded an accuracy rate of 75%.

In the present study, different than our previous preliminary approach [55], we develop a novel fusion method consisting of weighted Bayesian fusion of EDA and video class conditional distributions for maximum likelihood classification to identify unique contributions of each modality to clinically significant pain assessment. More specifically, in addition to extending our previous approach through Bayesian fusion, the method presented here enables us to optimize the contribution of EDA and video components in the fusion process for each participant. Thus, we hypothesize that EDA and video data will independently contribute to pain detection, and that combining these modalities will significantly improve discrimination between significant pain and nonsignificant pain in children in the clinical arena. We compare our novel weighted Bayesian fusion algorithm against our previous approach.

III. METHODS

A. Participants and Experimental Setup

Define abbreviations Fifty-eight participants (41 male (71%): 17 female (29%); 95% Caucasian; 79% Hispanic; and median age 13 years (5-17 years range)) recovering from laparoscopic appendectomy participated in a study examining automated assessment of children's post-operative pain using wearable biosensors and videotaped facial expressions. Ninety-seven interpretable EDA and video data collection sessions were acquired. EDA recordings were collected from the non-dominant hand of study participants. EDA signals were automatically screened for data quality using previously developed software from our group [56] and confirmatory expert visual inspection. From this initial data processing, 83 sessions from 45 participants (31 male (69%); 14 female (31%); 98% Caucasian; 80% Hispanic; and median age 12 years (5-17 years range)) remained. Demographic data did not differ between participants whose data were included in this study as compared to participants whose data were excluded ($p > 0.05$).

Participants were assessed for pain across three study visits. Visit 1 (V1) occurred within 24 hours following surgery (inpatient setting), Visit 2 (V2) occurred one day later (inpatient setting), and Visit 3 (V3) occurred up to 42 days later (outpatient setting). At each study visit, physiological (EDA) and behavioral reactions (video) were recorded during manual abdominal pressure applied adjacent to the surgical incision site for a 10-second interval, whereupon youth scored their experienced pain during the pressure stimulus from 0 (no-pain) to 10 (worst pain ever). Acceptability of this protocol has been previously demonstrated [57]. A pain score threshold of 4 is a commonly used pain score criterion in clinical settings to categorize between clinically significant pain and clinically non-significant pain [58]. In the present paper, we report on our analysis of 83 EDA and video datasets: V1 Data ($N = 25$) and V2 Data ($N = 11$) from participants whose pain score of ≥ 4 were considered a clinically significant pain sample, and V3 Data ($N = 36$) and V2 Data ($N = 11$) from participants with pain score of <4 were considered as a clinically non-significant pain sample. Using pain ratings used by clinicians [59], [60], we denoted EDA and video with pain ratings of 0–3 as clinically nonsignificant pain, and EDA and video with pain ratings of 4–10 as clinically significant pain.

B. Feature Extraction

Our aim in this investigation was to evaluate the accuracy of automatically differentiating clinically significant pain from

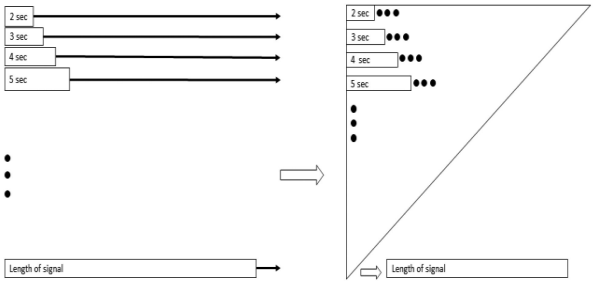

Fig. 1. Time scale decomposition framework.

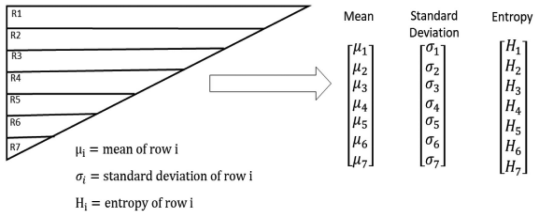

Fig. 2. Feature extraction of TSD output matrix.

clinically nonsignificant pain (two-class classification) through multimodal fusion of EDA and video data. A pain score cut-off of 4 on a 0-to-10 Numerical Rating Scale was used to segment EDA and video data into clinically significant pain and clinically nonsignificant pain classes, as was done in our previous work classifying pain with unimodal EDA [40] and Video [61] data. A pain score criterion threshold of 4 is an agreed upon standard traditionally used to differentiate mild pain from moderate pain in clinical settings [58]. The segmented data were then processed as described below to extract EDA and video features.

C. EDA Feature Extraction

EDA was collected via the Affectiva Q sensor, which was worn on the non-dominant wrist and modified for finger-tip data collection. Gelled adhesive electrodes were used for signal collection. The sensor wirelessly records EDA, skin surface temperature, and 3D motion at 16Hz [62], [63]. All EDA signals were evaluated with automated quality assessment software [56]. Resulting high quality EDA sig-nals corresponding with the time of noxious stimuli (com-prising 10 seconds before, 10 seconds during, and 10 sec-onds after the pressure stimulus) were selected and then low pass filtered using a 0.35 Hz FIR. After filtering, data were down sampled to 1 Hz to reduce analysis time. Next, data were normalized to overcome baseline differences in EDA levels using z-score transformation. Finally, EDA signals were trimmed to a fixed length of 30 seconds sur-rounding the time of noxious stimuli.

In order to extract EDA features associated with pain dis-tress, we applied time scale decomposition (TSD) [64], exploiting temporal characteristic of EDA at different time scales as we did in [40]. The TSD procedure is a simple extension of the sliding window approach, which decom-poses EDA signals into consec-utive, overlapping windows, and calculates a given metric. TSD iterates this procedure, calculating a given metric at all window lengths at all starting points, and systematically organizes results into a single matrix. The TSD method is summarized in Fig. 1.

Assuming a window length of 2 seconds, in the resulting TSD triangular matrix, consecutive rows differ by a window length of 1 second, with progressively increasing window lengths. Similar to our previous work [40], we computed the standard deviation of EDA in each window in order to capture fluctuations in tonic response and delays in phasic responses. We then computed the mean, standard deviation, and entropy of each row of the TSD output matrix for use in machine learning classification and entered them into a single feature matrix (See Fig. 2).

D. Automated Facial Action Unit Detection From Video

For each 10-second video sample recorded during the press, filmed at 60 frames per second, we extracted facial action unit (AU) codes per frame to obtain a sequence of AUs. Facial actions are anatomically-based differentiable movements of the facial musculature that can be objectively coded manually [65], as well with computer vision and pattern recognition [24]. Here, this was done automatically using iMotions software (www.imotions.com). Twenty of the 44 discrete AUs in the Facial Action Coding System[65] (AU 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 18, 20, 23, 24, 25, 26, 28, 43) and three head pose indicators (yaw, pitch, and roll) were extracted from each frame. These AUs and head pose indicators have previously been associated with pain [15], [66]. The values of these codes are the estimated log probabilities of AUs, ranging from -4 to 4.

We then extracted features from the sequence of AUs, applying 11 statistics (mean, max, min, standard deviation, 95th, 85th, 75th, 50th, 25th percentiles, half-rectified mean, and max-min) to each AU over all frames to obtain 11x23 features for each sample. The range of each feature was rescaled to 01 to normalize features over the training data.

In [61], the problem of domain adaptation [67] was discussed with automatically detected AU features. Since our data were collected across two different environmental domains, inpatient (V1 & V2) and outpatient (V3), and since iMotions is sensitive to environmental factors such as lighting and facial pose (which differs between inpatient and outpatient settings), inherent differences exist between samples in V1 & V2 and V3. Therefore, when we train a pain recognition machine learning model using clinically significant pain data from V1 & V2 and clinically nonsignificant pain data from V3, we end up with a model classifying inpatient versus outpatient environments more than clinically significant pain versus clinically nonsignificant pain. Such a model would fail on test data with both classes from V1 & V2.

In [61], we had a human expert AU coder manually annotate videos along with iMotions, and discovered that the human was not as sensitive to environmental changes as iMotions (as shown by comparing classifiers trained using pathway (2) compared to (1) in Fig. 3) and then used the data to learn a mapping from iMotions features to human features (following (5-4) in Fig. 3).

E. Further Processing of Video and EDA Features

Video features attained via iMotions software (AUs) and EDA features obtained through TSD were further z-score normalized separately to overcome participant-specific baseline differences. This normalization is performed using the entire data set of EDA

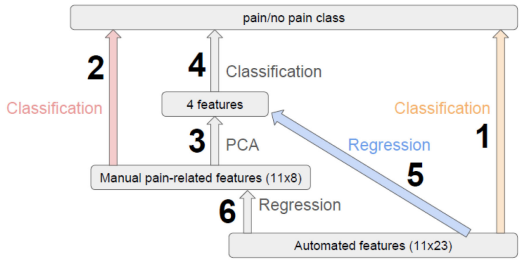

Fig. 3. Illustration of Machine Learning Models. 1 and 2 are classifications using automated/or human manual pain features, in which 2 does better than 1. 3 and 4 can be undertaken to reduce feature dimensions while maintaining performance. 6-2 and 5-4 are our transfer learning models, training a regression network to map automated features to a subspace of human manually coded features.

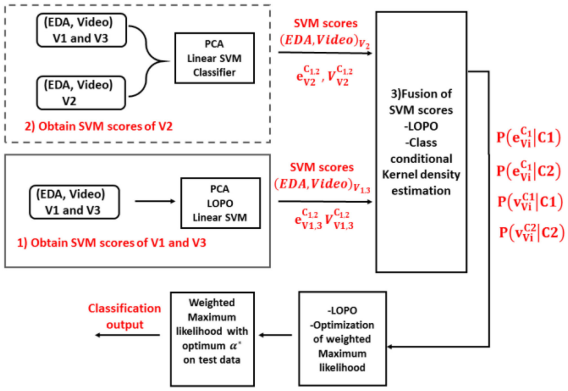

Fig. 4. Classification procedure framework.

and video separately. Then, principal component analysis (PCA) was applied to the EDA and video data to reduce feature dimensionality separately. In order to prevent over-fitting, we separated data into training and testing components through leave- LOPO (details of training and testing components are provided below), wherein PCA was applied to the training set and the subsequent testing feature vector was projected onto principal components to generate reduced feature vectors. LOPO is applied specifically on V2 data, and V1 and V3 were always used for training. In this step, PCA was set to capture between 90% and 99% of the variance in the V1 and V3 data, removing all other data. For further analyses, the PCA threshold providing maximum performance on the training data (V1 and V3 data) was selected for application to V2 data. The PCA process is presented in Fig. 4 as the initial step before dimensionality reduction, data fusion, and classification steps. Please see below for details on all steps that lead to classification of pain.

IV. Data Fusion for Classification

An overview of our data fusion classification scheme is presented in Fig. 4. The scheme includes three main parts: (i) obtaining linear SVM scores; (ii) kernel density estimation (KDE) using SVM scores; and (iii) weighted Bayesian fusion of class conditional distributions for maximum likelihood classification. Using this scheme, we first considered EDA and video features separately to extract SVM scores assuming clinically significant pain and clinically nonsignificant pain conditions are two distinct classes. We then performed KDE with radial basis function kernel and optimized kernel bandwidth (based on Silverman's rule of thumb [68]) to learn the class conditional distribution of EDA and video SVM scores separately. Then, assuming conditional independence between observed EDA and video score distributions conditioned on the given class, we performed Bayesian fusion, which resulted in maximum likelihood classification. Our approach deviates from conventional Bayesian fusion in that we take the convex combination of the class conditional log-likelihoods to obtain optimal contribution from each modality (EDA and video) for pain classification. We refer to this approach as "weighted maximum likelihood classification."

Since V1 and V2 data were collected in the inpatient arena, and V3 data in the outpatient arena, we used V1 and V3 data for training and V2 data for testing to avoid confounds of dominant but irrelevant (to pain) environmental differences that may affect classification outcomes. For training, V1 data were selected to be in the clinically significant pain class and V3 data were selected for the clinically nonsignificant pain class. Finally, V2 data were separated into clinically significant pain and clinically nonsignificant pain classes for testing based on the same pain score threshold of 4.

In the following, for class c (c = 1, 2: significant pain, nonsignificant pain) and visit i (i = 1, 2, or 3), we denote $e_{v_i}^C$ *and* $v_{v_i}^C$ as EDA and video scores used for training. In order to distinguish our training and test data, we denote $e_{v_i}^t$ *and* $v_{v_i}^t$ as EDA and video scores respectively, with t indicating those used for testing. The classification scheme is provided next and summarized in Fig. 4. (For details of the weighted Bayesian fusion algorithm, see Supplementary figure S1.)

In the first step of our proposed data fusion and classification scheme, linear SVM was employed for dimensionality reduction. More specifically, V1 and V3 training data $e_{v_i}^C$ and $v_{v_i}^C$, for i = 13 and C = 12 were used through LOPO to learn SVM scores for V1 and V3 training data (as shown by 1 in Fig. 4). Then, all V1 and V3 data were used to train a final linear SVM that was applied to V2 testing data to learn the testing SVM scores (as shown by 2 in Fig. 4). In the second step, we used all SVM scores from the training sets (V1 and V3) and a subset of SVM scores from V2 (this subset is selected through nested LOPO to avoid overfitting) to learn class conditional distributions of EDA and video scores through kernel density estimation. That is, we learn:

$$P\left(e_{v_i}^C|C\right), \text{ and } P\left(v_{v_i}^C|C\right) \text{ for } i = 1,2,3 \text{ and } C = 1,2. \quad (1)$$

Using these conditional distributions, we then find the scores under the learned densities for the nested left out subject denoting these scores as $e_{v_2}^t$ and $v_{v_2}^t$ for EDA and video SVM scores. For a given alpha, class C corresponding to each EDA and video score pair is then given the following proposed weighted-score maximum likelihood classification scheme:

$$\hat{C} = \underset{C}{\text{argmax}}(\alpha \times \ln\left(P\left(e_{v_2}^t|C\right)\right) + (1-\alpha) \times \ln\left(P\left(v_{v_2}^t|C\right)\right) \quad (2)$$

For each left out participant t, one optimum α value which we denote as α^* was learned using sensitivity maximization

TABLE II
THE WEIGHTED MAXIMUM LIKELIHOOD ALGORITHM RESULTS

Modalities	Accuracy	Sensitivity	Specificity
EDA-Only	68.18%	54.55%	81.82%
Video-Only	77.27%	90.91%	63.64%
EDA-Video Fusion	90.91%	100%	81.82%

(pain detection accuracy) obtained by solving Equation 2 and calculating sensitivity through all participants in V2 excluding participant t. The optimum α^*. was then used in Equation 2 together with participant t data to obtain classification results for participant t. During optimization, we searched for α values in the range 0 to 1 with increments of 0.1.

V. RESULTS

As described above, the proposed framework summarized in Fig. 4 was used to classify V2 data into clinically significant pain or clinically nonsignificant pain conditions. V1 and V3 data were used for training. The data in V1, V2, and V3 were categorized into clinically significant pain or clinically nonsignificant pain conditions based on patient-provided pain scores. Through nested cross-validation, the contribution/weights α of each modality were optimized. Results of the weighted maximum likelihood classification algorithm are summarized in Table II. In this Table, EDA-video Fusion corresponds to Equation 2, which considers optimum contributions from both modalities based on the training data.

Overall, after fusion, we observe that clinically significant pain vs. clinically nonsignificant pain classification achieves 90.91% accuracy, with 100% sensitivity (accuracy of significant pain detection), and 81.82% specificity (accuracy of nonsignificant pain detection). In the same table, we also report clinically significant pain vs. clinically nonsignificant pain classification results when only a single modality is used. EDA-Only means to set $\alpha = 1$ and Video-Only corresponds to set $\alpha = 0$ in Equation 2. We observe that for EDA-Only, overall accuracy is 68.18%, sensitivity is 54.55%, and specificity is 81.82%. For Video-Only, we observe overall accuracy is 77.27%, sensitivity is 90.91%, and specificity is 63.64%.

To provide further insight into the weighted maximum likelihood algorithm, we illustrate the optimum α values identified during classification for each participant in Fig. 5, which graphically displays each modality's contribution to sensitivity for pain classification.

In this figure, the y-axis represents the optimum α (note that $1 - \alpha$ corresponds to that of the EDA features while α corresponds to the optimum contribution of the video features) and the x-axis represents each participant. From this figure, we observe that EDA and video features contribute equally to improving sensitivity of clinically significant pain vs clinically nonsignificant pain classification in half of the cases. For the other half, video features contribute more to clinically significant pain vs clinically nonsignificant pain classification.

These observations are consistent with our fusion results presented in Table II. In contrast, the specificity of the fusion model

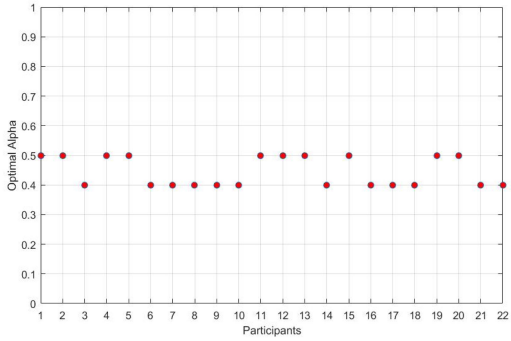

Fig. 5. Video vs. EDA modality contribution to the fusion model's clinically significant/nonsignificant pain classification.

TABLE III
THE PERFORMANCE MEASUREMENTS OF THE LATE FUSION ALGORITHM PROPOSED BY THIAM WITH RESPECT TO 150 AND 80 TREES

Num=150	Accuracy	Sensitivity	Specificity
EDA-Only	45.45%	0.09%	81.82%
Video-Only	59.09%	63.64%	54.55%
EDA-Video Fusion	68.18%	63.64%	72.73%

Num=80	Accuracy	Sensitivity	Specificity
EDA-Only	50%	18%	81.82%
Video-Only	63.64%	72.73%	54.555
EDA-Video Fusion	63.64%	63.64%	63.64%

Sensitivity: Correct identification of clinically significant pain
Specificity: Correct identification of clinically nonsignificant pain

remains equal to EDA-Only specificity (i.e., entirety of detection contribution from EDA). We also compared our proposed algorithm with late fusion and early fusion algorithms proposed by Thiam [39] and Werner [45] and our previous work [55]. We used these algorithms for comparison because they achieve a high pain detection rate with reasonable complexity and run time. Our proposed weighted Bayesian algorithm shows clear advantage over the late and early fusion algorithms proposed by Thiam *et al.* [39] Werner *et al.*. [45], and Xu *et al.* [55] at automatically discriminating pediatric pain in children.

The late fusion algorithm proposed by Thiam [39] was implemented and applied on the above-mentioned EDA and video features in Section II-B. First, a Random Forest (RF) classifier was used to obtain RF scores of EDA and video features from V1, V2, and V3 in the fusion classification. Then, Linear Discriminant Analysis (LDA) was used to fuse the RF scores of EDA and video, followed by LOPO cross validation. Second, we performed EDA-Only and Video-Only classification using RF scores of EDA and video data separately as input to an LDA classifier. Of note, we sought the maximum performance of the proposed algorithm. Therefore, we overfit the model using the rest of the RF scores of EDA and video from V2 in the training set. In addition, we specified the range of the number of trees from 10 to 150 with an increment of 10 and provided the best two performance measurements in the range of the number of trees. Results of the LDA classifier are demonstrated in Table III with respect to the corresponding number of trees.

TABLE IV
THE PERFORMANCE MEASUREMENTS OF THE EARLY FUSION ALGORITHM PROPOSED BY WERNER

Modality	Accuracy	Sensitivity	Specificity
EDA-Only	48.18%	41.82%	54.55%
Video-Only	87.27%	83.64%	90.91%
EDA-Video Fusion	76.36%	78.18%	74.55%

Sensitivity: Correct identification of clinically significant pain
Specificity: Correct identification of clinically nonsignificant pain

TABLE V
THE PERFORMANCE MEASUREMENTS OF OUR PREVIOUS LATE FUSION METHOD

Modality	Accuracy	Sensitivity	Specificity
EDA-Only	55.5%	70%	41%
Video-Only	80%	70%	90%
EDA-Video Fusion	80%	80%	80%

Sensitivity: Correct identification of clinically significant pain
Specificity: Correct identification of clinically nonsignificant pain

As a result, EDA is found to be more informative in detecting clinically nonsignificant pain while video is more sensitive with clinically significant pain detection. These findings provide incipient validity of our results obtained through a weighted Bayesian fusion algorithm. However, details of this algorithm (e.g., number of trees used in the random forest) necessary to replicate it were not provided.

In the implementation of the early fusion technique proposed by Werner [45], we first concatenated EDA and video features. Then the RF classifier was applied on these features followed by LOPO cross validation. To select the optimal values for the RF, we applied 5-fold cross validation on our training set including the EDA and video data from V1, V2, and V3. After a parameter search, we set the number of trees to 100, the maximum depth to 11, and minimum sample counts for node splitting to 5. To maintain comparability, we conducted EDA-Only and Video-Only classification. The RF classifier with selected parameters was directly applied to EDA features followed by LOPO cross validation. We applied the same strategy to our video features. Our results suggest that video is superior at detecting clinically nonsignificant pain, whereas fusion of EDA and video data did not improve performance measurements in clinically significant pain classification (See Table IV).

We also applied our previous late fusion technique on this data set [55]. We first obtained Linear SVM scores of EDA and video data as explained in [55]. Then, the LDA classifier was used in fusion of Linear SVM scores of EDA and video data using LOPO cross validation. To obtain the maximum performance of the classifier, we trained the model using the rest of the EDA and video data from V2. In addition, we conducted EDA-Only and Video-Only classification. Finally, LOPO cross validation was applied on each modality. Results of the LDA classifier are illustrated in Table V. Overall, our proposed weighted Bayesian algorithm demonstrated superior performance over our previous work.

TABLE VI
THE PERFORMANCE MEASUREMENTS OF WEIGHTED BAYESIAN ALGORITHM USING THE MORPHOLOGICAL EDA FEATURES OBTAINED BY BIO-SP TOOL

Modality	Accuracy	Sensitivity	Specificity
EDA-Only	45.45%	45.45%	45.45%
Video-Only	77.27%	**90.91%**	63.64%
EDA-Video Fusion	36.36%	45.45%	27.27%

Sensitivity: Correct identification of clinically significant pain
Specificity: Correct identification of clinically nonsignificant pain

Additionally, we extracted morphological features of EDA signals using the Biosignal-Specific Processing Tool (Bio-SP Tool) [69] and compared them with the TSD-based features. Specifically, these features are used separately in our fusion algorithm and the classification results are compared. Bio-SP toolbox allowed us to detect SCRs and extract the following features: (1) SCR duration mean; (2) mean SCR amplitude; (3) mean SCR rise-time (where rise-time of an SCR is defined as the time between the initial rise and the peak of an SCR); (4) mean of signal; and (5) number of detected SCRs. After extracting the features, a Linear SVM classifier was applied to obtain SVM scores of V1, V2, and V3 EDA data as described above. Note that we did not use PCA in this step, since higher dimensionality is no longer a problem using these five features. Finally, the weighted Bayesian fusion algorithm was applied on Linear SVM scores of EDA related features, obtained through the Bio-SP Tool, and Linear SVM scores of video features as described in Section II-C. The performance measurements of weighted Bayesian algorithm using morphological features of EDA are illustrated in Table VI.

The weighted Bayesian algorithm using the morphological features indicated poor performance compared to the performance measurements of the classification using TSD features, see Table VI and Table II.

VI. DISCUSSION

This study demonstrates that EDA and video data independently make substantial contributions to pain detection but combining these modalities significantly improves discrimination of clinically significant pain from clinically nonsignificant pain in post-operative children. The primary contributions of this paper are twofold. First, we illustrate the utility of a weighted maximum likelihood algorithm as a novel feature selection method for EDA and video signals for classifying clinically significant pain vs clinically nonsignificant pain. Second, we present evidence for an accurate classification algorithm capable of automatically discriminating clinically significant pain and clinically nonsignificant pain using EDA and video feature fusion. These classification findings discriminating clinically moderate-to-severe pain and nonsignificant pain conditions are promising. More specifically, the novelty of our proposed weighted Bayesian fusion approach is to identify the unique and shared contribution of EDA and video to automatically detect pain levels in children. Compared to the early fusion algorithm [45] and late fusion algorithm [39], our proposed weighted Bayesian fusion algorithm indicated superior performance such

that EDA is more sensitive to nonsignificant pain conditions while video is more sensitive to significant pain conditions. Additionally, the utility of TSD in the weighted Bayesian fusion algorithm demonstrated a clear advantage over traditional physiological morphological features. We believe this indicates that EDA might carry information about pain beyond previously established morphological features. Another possibility is that automated SCR detection and identification of morphological SCR related features includes errors. Overall, these comparisons have further strengthened our confidence that our proposed framework demonstrates an accurate and objective automated classification algorithm capable of distinguishing clinically significant pain and clinically nonsignificant pain by fusing EDA and video signals in children.

Both EDA and video input provided useful signals for categorizing pain. As presented in Table II, using the method described in this paper, it was shown that EDA was superior to video in terms of identifying clinically nonsignificant pain, whereas video was more informative for detecting clinically significant pain. The weighted maximum likelihood fusion algorithm used optimized information from both modalities to differentiate clinically significant pain vs. clinically nonsignificant pain, capitalizing on the unique strengths of each modality. In other words, using both data modalities, we demonstrate higher accuracies both for clinically significant pain and clinically nonsignificant pain identification. For example, EDA sensitivity alone only modestly exceeded chance, suggesting high false negative error rates, whereas facial activity (video data) provided a high level of sensitivity. In combination, they proved fully sensitive; no event defined as painful escaped detection.

The reactions captured by EDA and video appear to reflect different features of the complex pain response. EDA is perhaps best conceptualized as a measure of indiscriminate emotional reactivity to the extent that it indexes generalized peripheral sympathetic nervous system activation. Facial expressions can be conceptualized as reflecting emotional arousal, the typical conception, but also cognitive states, including meaningful interpretation, expectation, and attention [70]. Here, we show that differences between EDA and video data can be used in a complementary manner to distinguish painful and non-painful levels. While there are no benchmarks in the clinical arena to determine the magnitude of sensitivity and specificity of global judgments of pain in others, these are commonly described as error prone and vulnerable to observer bias [6]. Our automated maximum likelihood approach using fused EDA and facial activity data substantially improves on this.

We used self-reported pain, the gold standard for clinical pain assessment, as the criterion index for the presence or absence of clinically significant pain. As youth participating in the presented work were otherwise healthy and without significant prior pain history, medication-seeking behaviors were not felt to contribute to potential intentional or self-interest bias in reported pain. Nevertheless, variation in stoicism across the population could have contributed to report bias. Furthermore, the automated approach described here has the potential to avoid observer biases that bedevil judgments of pain. Observers often use salient but irrelevant patient characteristics (e.g., sex and gender, age, socioeconomic status, ethnicity or race) in formulating judgments rather than explicit and objective manifestations of pain in the behavioral and physiological domains [71].

The research reported here also has substantial ecological validity. Children experiencing post-operative pain were recorded shortly following surgery in the setting where pain assessment is crucial to delivery of care. The algorithms were trained using dynamic data in the clinical context, under conditions in which the algorithm is intended to operate. It is noteworthy that the video assessment of facial activity focuses upon nonverbal behavioral expression manifestly available to clinicians and other observers. While this information can contribute to accurate pain judgments, particularly when the context confirms the expression is associated with pain, as would be the case in the post-operative setting, availability of physiological data, in this case EDA, substantially improves accuracy in detecting pain.

Biosensing technologies have long been used in clinical medicine (e.g., glucometers [72]) and include recent expansion in video technologies [73] and other wearable sensors [74], [75], wherein video improves provider-patient communication through telemedicine transactions [76], documents clinical events and conditions (e.g., seizures [77]), and demonstrates value in education, performance assessment, quality improvement, and clinical care [73]. EDA, assessed via wearable sensors on easily accessible body parts including wrists or ankles, has demonstrated promise in monitoring clinical status and events (e.g., sleep [74], seizures [75]). Therefore, it would appear worthwhile to consider both measurement strategies for automatic assessment of pain level in clinical settings, with post-operative pain a good candidate. Use of computer vision and ambulatory peripheral physiological signals minimizes cumbersome demands on patients yet permits a continuous record of objective signals that can be readily subjected to computational analysis and automated generation of pain information. There is also potential for continual updating of such algorithms using further data to enhance generalized and person-dependent performance.

VII. Limitations

The generalizability of this work to other clinical and patient populations may be limited by the population represented in this study: otherwise, healthy children in acute post-surgical pain. However, while the exact findings may be specific to participants in this context, underlying pain physiology and resulting behaviors have been well described across both acute and chronic pain populations [78], [79] and speak to the potential utility and application of the methods described herein to other clinical pain scenarios

We do recognize that prior work has evaluated EDA to be nonspecific in its ability to differentiate between clinically significant pain and clinically nonsignificant pain levels. However, utilizing rigorous data quality control checks and a novel method of evaluation (Time Scale Decomposition), we were able to extract new features that allow increased specificity, which, in conjunction with video data, were able to improve its accuracy for specificity in clinically significant pain vs clinically nonsignificant pain level detection.

Discrepancies between computer vision-enabled facial coding and manual coding of facial expressions have been identified as a consequence of the differing technologies and methodologies involved in their study [24], [80]. Consequently, automated computer vision coding does not fully correlate with manual, time-intensive coding. We have previously shown [61] that computer vision pixel methods may be uniquely sensitive to environmental external factors, including rotation, scaling, and illumination [61]. Given our data sampling was across two different environments, our current work is also subject to this limitation. However, we addressed this issue through utilizing transfer learning [61], which enabled us to take advantage of the best aspects of both (i.e., automation of computer vision and validity of manual facial action coding). In parallel, we are studying the properties of human-labeled action units [81] as well as developing better automated methods for facial action unit coding for pain detection [82].

With respect to EDA, there are currently no published studies on the psychological significance of bilateral asymmetry in recordings as a biological marker of the pain experience that we are aware of. However, future work could explore whether bilateral asymmetry in EDA recordings are a useful biomarker of the pain experience.

VIII. Conclusion

In summary, we present results demonstrating the feasibility, validity, and utility of combining EDA and video data for measuring and discriminating pain-related behaviors. Our approach may increase automation and accuracy of pain assessment in the clinical arena in an at-risk population (youth following surgery). The portability, accessibility, and increasing use of wireless biosensors and video capture devices in the general population enable scalability in the healthcare setting. Furthermore, the portability and utility of these sensors/devices outside of the healthcare setting and in the home further enhance the potential usage and applicability of this approach where healthcare resources are scarce or distantly located, and/or social isolation is required (as evidenced in the recent Covid pandemic). As our healthcare settings become increasingly more instrumented with such monitoring technology, there is a growing opportunity and need for theoretically driven and empirically demonstrated data-rich methods such as those presented here to improve our ability to detect and thus better manage patients in pain.

Acknowledgment

The authors would like to thanks Dr. Goodwin is on the Scientific Advisory Board of Affectiva who manufactures the Q sensor and receives stock options for his service. All other authors have no conflicts of interest to report.

References

[1] A. C. D. C. Williams and K. D. Craig, "Updating the definition of pain," *Pain*, vol. 157, no. 11, pp. 2420–2423, 2016.
[2] R. Melzack and K. L. Casey, "Sensory, motivational, and central control determinants of pain: A new conceptual model," *Skin Senses*, vol. 1, pp. 423–443, 1968.
[3] K. D. Craig *et al.*, "Perceiving pain in others: Automatic and controlled mechanisms," *J. Pain*, vol. 11, no. 2, pp. 101–108, 2010.
[4] T. Hadjistavropoulos and K. Craig, "A theoretical framework for understanding self-report and observational measures of pain: A communications model," *Behav. Res. Ther.*, vol. 40, no. 5, pp. 551–570, 2002.
[5] C. L. Von Baeyer and L. J. Spagrud, "Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years," *Pain*, vol. 127, no. 1/2, pp. 140–150, 2007.
[6] L. De Ruddere *et al.*, "Health care professionals' reactions to patient pain: Impact of knowledge about medical evidence and psychosocial influences," *J. Pain*, vol. 15, no. 3, pp. 262–270, 2014.
[7] P. L. Jackson, A. N. Meltzoff, and J. Decety, "How do we perceive the pain of others? A window into the neural processes involved in empathy," *Neuroimage*, vol. 24, no. 3, pp. 771–779, 2005.
[8] M. Martel *et al.*, "Contextual determinants of pain judgments," *Pain*, vol. 139, no. 3, pp. 562–568, 2008.
[9] L. L. Cohen *et al.*, "Topical review: State of the field of child self-report of acute pain," *J. Pediatr. Psychol.*, vol. 45, pp. 239–246, 2019.
[10] J. Y.-C. Chan and C. L. von Baeyer, "Cognitive developmental influences on the ability of preschool-aged children to self-report their pain intensity," *Pain*, vol. 157, no. 5, pp. 997–1001, 2016.
[11] T. Oberlander and F. J. Symons, "The problem of pain in developmental disability," in *Pain Developmental Disabilities*. Baltimore, MD, USA: Paul H. Brookes, 2006.
[12] T. D. Wager *et al.*, "An fMRI-based neurologic signature of physical pain," *New England J. Med.*, vol. 368, no. 15, pp. 1388–1397, 2013.
[13] L. Arendt-Nielsen *et al.*, "Association between experimental pain biomarkers and serologic markers in patients with different degrees of painful knee osteoarthritis," *Arthritis Rheumatol.*, vol. 66, no. 12, pp. 3317–3326, 2014.
[14] A. C. D. C. Williams, "Facial expression of pain: An evolutionary account," *Behav. Brain Sci.*, vol. 25, no. 4, pp. 439–455, 2002.
[15] D. C. Turk and R. Melzack, *Handbook of Pain Assessment*. New York, NY, USA: Guilford Press, 2011.
[16] M. Kunz *et al.*, "Are both the sensory and the affective dimensions of pain encoded in the face?," *Pain*, vol. 153, no. 2, pp. 350–358, 2012.
[17] G. Stewart and A. Panickar, "Role of the sympathetic nervous system in pain," *Anaesth. Intensive Care Med.*, vol. 14, no. 12, pp. 524–527, 2013.
[18] E. E. Benarroch, "Pain-autonomic interactions: A selective review," *Clin. Autonomic Res.*, vol. 11, no. 6, pp. 343–349, 2001.
[19] J. Lötsch and A. Ultsch, "Machine learning in pain research," *Pain*, vol. 159, no. 4, 2018, Art. no. 623.
[20] B. Gholami, W. M. Haddad, and A. R. Tannenbaum, "Relevance vector machine learning for neonate pain intensity assessment using digital imaging," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 6, pp. 1457–1466, Jun. 2010.
[21] L. Nanni, A. Lumini, and S. Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis," *Artif. Intell. Med.*, vol. 49, no. 2, pp. 117–125, 2010.
[22] K. Sikka *et al.*, "Automated assessment of children's postoperative pain using computer vision," *Pediatrics*, vol. 136, no. 1, pp. 2015–0029, 2015.
[23] H. Meng and N. Bianchi-Berthouze, "Affective state level recognition in naturalistic facial and vocal expressions," *IEEE Trans. Cybern.*, vol. 44, no. 3, pp. 315–328, Mar. 2014.
[24] M. S. Bartlett *et al.*, "Automatic decoding of facial movements reveals deceptive pain expressions," *Curr. Biol.*, vol. 24, no. 7, pp. 738–743, 2014.
[25] P. Lucey *et al.*, "Automatically detecting pain in video through facial action units," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 3, pp. 664–674, Jun. 2011.
[26] D. Huang *et al.*, "Pain-attentive network: A deep spatio-temporal attention model for pain estimation," *Multimedia Tools Appl.*, vol. 79, pp. 1–26, 2020.
[27] P. Thiam, H. A. Kestler, and F. Schwenker, "Two-Stream attention network for pain recognition from video sequences," *Sensors*, vol. 20, no. 3, 2020, Art. no. 839.
[28] R. A. Virrey *et al.*, "Visual data of facial expressions for automatic pain detection," *J. Vis. Commun. Image Representation*, vol. 61, pp. 209–217, 2019.
[29] M. Bellantonio *et al.*, "Spatio-temporal pain recognition in cnn-based super-resolved facial images," in *Video Analytics. Face and Facial Expression Recognition and Audience Measurement*, Berlin, Germany: Springer, 2016, pp. 151–162.
[30] D. Liu, P. Fengjiao, and R. Picard, "DeepFaceLIFT: Interpretable personalized models for automatic estimation of self-reported pain," in *Proc. Workshop Artif. Intell. Affect. Comput.*, 2017, pp. 1–16.

[31] E. K. Choo *et al.*, "Skin conductance fluctuations correlate poorly with postoperative self-report pain measures in school-aged children," *Anesthesiol.: J. Amer. Soc. Anesthesiologists*, vol. 113, no. 1, pp. 175–182, 2010.

[32] S. Gruss *et al.*, "Pain intensity recognition rates via biopotential feature patterns with support vector machines," *PloS One*, vol. 10, no. 10, 2015, Art. no. e0140330.

[33] E.-H. Jang *et al.*, "Analysis of physiological signals for recognition of boredom, pain, and surprise emotions," *J. Physiol. Anthropol.*, vol. 34, no. 1, 2015, Art. no. 25.

[34] M. M. Bradley, T. Silakowski, and P. J. Lang, "Fear of pain and defensive activation," *PAIN*, vol. 137, no. 1, pp. 156–163, 2008.

[35] M. Eriksson *et al.*, "Skin conductance compared to a combined behavioural and physiological pain measure in newborn infants," *Acta Paediatrica*, vol. 97, no. 1, pp. 27–30, 2008.

[36] M. L. Loggia, M. Juneau, and M. C. Bushnell, "Autonomic responses to heat pain: Heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity," *PAIN*, vol. 152, no. 3, pp. 592–598, 2011.

[37] P. Schestatsky *et al.*, "Skin autonomic reactivity to thermoalgesic stimuli," *Clin. Autonomic Res.*, vol. 17, no. 6, pp. 349–355, 2007.

[38] M. Kächele *et al.*, "Multimodal data fusion for person-independent, continuous estimation of pain intensity," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, 2015, pp. 275–285.

[39] P. Thiam and F. Schwenker, "Multi-modal data fusion for pain intensity assessment and classification," in *Proc. 7th Int. Conf. Image Process. Theory Tools Appl.*, 2017, pp. 1–6.

[40] B. T. Susam *et al.*, "Automated pain assessment using electrodermal activity data and machine learning," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2018, pp. 372–375.

[41] D. Lopez-Martinez and R. Picard, "Continuous pain intensity estimation from autonomic signals with recurrent neural networks," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2018, pp. 5624–5627.

[42] P. Werner *et al.*, "Automatic recognition methods supporting pain assessment: A survey," *IEEE Trans. Affect. Comput.*, to be published, doi: 10.1109/TAFFC.2019.2946774.

[43] J. Chen, Z. Chi, and H. Fu, "A new framework with multiple tasks for detecting and locating pain events in video," *Comput. Vis. Image Understanding*, vol. 155, pp. 113–123, 2017.

[44] M. A. Haque *et al.*, "Deep multimodal pain recognition: A database and comparison of spatio-temporal visual modalities," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2018, pp. 250–257.

[45] P. Werner *et al.*, "Automatic pain recognition from video and biomedical signals," in *Proc. 22nd Int. Conf. Pattern Recognit.*, 2014, pp. 4582–4587.

[46] D. Lopez-Martinez, O. Rudovic, and R. Picard, "Physiological and behavioral profiling for nociceptive pain estimation using personalized multitask learning," in *Proc. Neural Inf. Process. Syst. Workshop Mach. Learn. Health*, Long Beach, CA, USA, 2017.

[47] M. Amirian, M. Kächele, and F. Schwenker, "Using radial basis function neural networks for continuous and discrete pain estimation from bio-physiological signals," in *Proc. IAPR Workshop Artif. Neural Netw. Pattern Recognit.*, 2016, pp. 269–284.

[48] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[49] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed., Cambridge, MA, USA: Academic., 2008, p. 900.

[50] V. R. De Sa *et al.*, "Multi-view kernel construction," *Mach. Learn.*, vol. 79, no. 1/2, pp. 47–71, 2010.

[51] A. E. Olesen *et al.*, "Human experimental pain models for assessing the therapeutic efficacy of analgesic drugs," *Pharmacological Rev.*, vol. 64, no. 3, pp. 722–779, 2012.

[52] M. Kächele *et al.*, "Bio-visual fusion for person-independent recognition of pain intensity," in *Proc. Int. Workshop Mult. Classifier Syst.*, 2015, pp. 220–230.

[53] V. Kessler *et al.*, "Multimodal fusion including camera photoplethysmography for pain recognition," in *Proc. Int. Conf. Companion Technol.*, 2017, pp. 1–4.

[54] M. S. Rathleff and T. Graven-Nielsen, "Transition from acute to chronic pain in children: Novel pieces of the puzzle," *Pain*, vol. 158, no. 5, pp. 767–768, 2017.

[55] X. Xu *et al.*, "Towards automated pain detection in children using facial and electrodermal activity," in *Artificial Intelligence in Health*, Berlin, Germany: Springer, 2018, pp. 181–189.

[56] I. R. Kleckner *et al.*, "Simple, transparent, and flexible automated quality assessment procedures for ambulatory electrodermal activity data," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 7, pp. 1460–1467, Jul. 2018.

[57] K. Hawley *et al.*, "Youth and parent appraisals of participation in a study of spontaneous and induced pediatric clinical pain," *Ethics Behav.*, vol. 29, pp. 259–273, 2019.

[58] D. L. Hoffman *et al.*, "How do changes in pain severity levels correspond to changes in health status and function in patients with painful diabetic peripheral neuropathy?," *PAIN*, vol. 149, no. 2, pp. 194–201, 2010.

[59] A. M. Boonstra *et al.*, "Cut-off points for mild, moderate, and severe pain on the visual analogue scale for pain in patients with chronic musculoskeletal pain," *Pain*, vol. 155, no. 12, pp. 2545–2550, 2014.

[60] A. M. Boonstra *et al.*, "Cut-off points for mild, moderate, and severe pain on the numeric rating scale for pain in patients with chronic musculoskeletal pain: Variability and influence of sex and catastrophizing," *Front. Psychol.*, vol. 7, 2016, Art. no. 1466.

[61] X. Xu *et al.*, "Automated pain detection in facial videos of children using human-assisted transfer learning," in *Artificial Intelligence in Health*, Cham, Switzerland: Springer, 2018, pp. 162–180.

[62] R. R. Fletcher *et al.*, "iCalm: Wearable sensor and network architecture for wirelessly communicating and logging autonomic activity," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 215–223, Mar. 2010.

[63] M. - Z. Poh, N. C. Swenson, and R. W. Picard, "A wearable sensor for unobtrusive, long-term assessment of electrodermal activity," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 5, pp. 1243–1252, May 2010.

[64] M. Sips *et al.*, "A visual analytics approach to multiscale exploration of environmental time series," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2899–2907, Dec. 2012.

[65] P. Ekman and W. Friesen, *Facial Action Coding System*. Palo Alto, CA, USA: Consulting Psychol Press, 1978.

[66] K. D. Craig, K. M. Prkachin, and R. V. Grunau, "The facial expression of pain," *Handbook Pain Assessment*, vol. 2, pp. 257–276, 1992.

[67] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[68] H. Läuter and B. W. Silverman, "Density estimation for statistics and data analysis," *Biometrical Journal*, vol. 30, no. 7, London, U.K.: Chapman & Hall, 1986, p. 175, £12.—.

[69] M. Nabian *et al.*, "A biosignal-specific processing tool for machine learning and pattern recognition," in *Proc. IEEE Healthcare Innovations Point Care Technol.*, 2017, pp. 76–80.

[70] P. Ekman and E. L. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. New York, NY, USA: Oxford Univ. Press, 1997.

[71] K. M. Prkachin, P. E. Solomon, and J. Ross, "Underestimation of pain by health-care providers: Towards a model of the process of inferring pain in others," *Can. J. Nurs. Res.*, vol. 39, no. 2, pp. 88–106, 2007.

[72] D. Rodbard, "Continuous glucose monitoring: A review of recent studies demonstrating improved glycemic outcomes," *Diabetes Technol. Therapeutics*, vol. 19, no. S3, pp. S–25–S–37, 2017.

[73] K. Taylor *et al.*, "Prevalence and indications for video recording in the health care setting in north american and british paediatric hospitals," *Paediatrics Child Health*, vol. 16, no. 7, pp. e57–e60, 2011.

[74] A. Herlan *et al.*, "Electrodermal activity patterns in sleep stages and their utility for sleep versus wake classification," *J. Sleep Res.*, vol. 28, 2019, Art. no. e12694.

[75] S. Ramgopal *et al.*, "Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy," *Epilepsy Behav.*, vol. 37, pp. 291–307, 2014.

[76] B. L. Burke and R. Hall, "Telemedicine: Pediatric applications," *Pediatrics*, vol. 136, no. 1, pp. 2015–1517, 2015.

[77] D. F. Ghougassian *et al.*, "Evaluating the utility of inpatient video-EEG monitoring," *Epilepsia*, vol. 45, no. 8, pp. 928–932, 2004.

[78] M. G. Pagé *et al.*, "Acute pain trajectories and the persistence of post-surgical pain: A longitudinal study after total hip arthroplasty," *J. Anesth.*, vol. 30, no. 4, pp. 568–577, 2016.

[79] A. Z. Weinrib *et al.*, "The psychology of chronic post-surgical pain: New frontiers in risk factor identification, prevention and management," *Brit. J. Pain*, vol. 11, no. 4, pp. 169–177, 2017.

[80] M. Kunz *et al.*, "Problems of video-based pain detection in patients with dementia: A road map to an interdisciplinary solution," *BMC Geriatrics*, vol. 17, no. 1, 2017, Art. no. 33.

[81] X. Xu and V. R. de Sa, "Exploring multidimensional measurements for pain evaluation using facial action units," in *Proc. 15th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2020, pp. 559–565.

[82] X. Xu, J. S. Huang, and V. R. De Sa, "Pain evaluation in video using extended multitask learning from multidimensional measurements," in *Proc. Mach. Learn. Health NeurIPS Workshop*, 2019, pp. 141–154.