Skin Conductance Response to Gradual-Increasing Experimental Pain

Elise Syrjälä[1], Mingzhe Jiang[1], Tapio Pahikkala[1], Sanna Salanterä[2] and Pasi Liljeberg[1]

Abstract— Patient self-reporting of pain is not always possible, in those cases automated objective pain assessment could lead to reliable pain assessment. In this context, physiological measurements have been studied and one of the promising signals is skin conductance (SC). In this study, 1Hz SC signal acquisition is performed while gradually increasing heat and electrical pain stimuli are induced. Three labeled study periods are defined based on pain stimuli presence, self-reported pain threshold and pain tolerance. Different classification and regression models are compared, together with selected SC features. The model performances are evaluated using c-index. Results show good predictability, especially for the slow tonic component decomposed from the SC signal.

I. INTRODUCTION

Pain is an unpleasant sensation associated with a subjective emotional experience. Pain self-reporting requires both communicative and cognitive capabilities. When a person is unable to communicate verbally for example due to language barrier, age, Alzheimer's, or even unconsciousness, pain assessment is provided based on observed behaviour and autonomic activity. For effective and safe pain treatment, accurate pain assessment, including regular reassessment, is required.

Objective pain assessment studies have raised discussion that physiological biosignals can offer a quantitative method to assess the emotional distress related to pain [1], [2]. Among the physiological parameters, skin conductance (SC) has showed good performance compared to others in indicating pain in the scenarios with postoperative or experimentally induced pain [3]–[6]. As a result of the distress caused by a painful stimulus, the autonomic nervous system is activated and thus causing increased sweat secretion of the skin, electrodermal activity (EDA), which a produces a measurable SC value. The SC signal is composed of two parts: the slowly changing tonic component skin conductance level (SCL), the spectrum of which is below 0.05 Hz [7], and the fast changing phasic component skin conductance responses (SCRs).

In objective pain assessment studies, the phasic feature, number of skin conductance fluctuations, is mostly studied as a pain intensity index [2], [3], [8]. Comparatively, SCL has shown its potential [2] but was not fully studied for automatic pain assessment using machine learning methods, which is hard to implement when the stimulation duration is short.

In this work, the SC response to continuous and gradual-increasing experimental pain stimulus is analyzed using both tonic and phasic components on top of our previous work, the SpaExp pain database [4]. The database has physiological parameters from healthy volunteers under electrical or heat pain stimulation.

In this study, the estimation of the uncomfortableness and distress evoked by experimental pain stimulation is formulated as a machine learning labeling problem in a continuous estimation setting. The problem representations potential of different SC feature sets is evaluated in the form of predictability performance with different machine learning algorithms.

II. METHODS

A. SpaExp pain database

The SpaExp pain database involved 31 healthy volunteers (16 females). The study was approved by the Ethics Committee of the Hospital District of South West Finland (ETMK:83/1801/2015). Two experimental stimuli, electrical TENS pulses (peak-to-peak amplitude: 2V/level, pulse width: $250\mu s$, pulse frequency: 100 Hz) and heat ($1\,°C$/level) were adopted to trigger non-harmful pain perception. Each stimulus was applied to the left and right side of the body, respectively.

The data labelling in each of the tests is illustrated in Fig. 1. The *test start* was considered as a 30 second period before *stimulus start*. The stimulus started from the minimum intensity (level 1 or $30\,°C$)) at t_1 and increased 1 level or $1\,°C$ every 3 seconds until t_3 the *pain tolerance* was reported by the study subject or the maximum safety intensity was reached (level 50 or $55\,°C$).
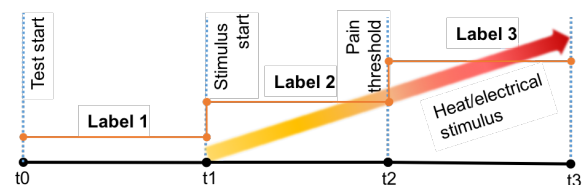

Fig. 1. Data labelling in one test

The EDA was measured with the exosomatic method between the medial phalanx of the middle and ring fingers on the no-stimulation side. The SC signal was sampled at a low frequency of 1 Hz. The outliers in the signal were manually checked and corrected. One example of the pain stimulation and z-score standardized SC signal within tests are presented

in Fig. 2 (a-b). Eight tests (1 subject and 4 other heat tests) were excluded from the analysis due to incomplete SC signal.

B. Feature extraction

The SC signal was first decomposed into SCL and SCRs with cvxEDA algorithm [7], where the SC signal is considered as a superposition of tonic component (t) with spectrum below 0.05 Hz, phasic component (r) and an additive Gaussian noise term (ϵ). Moreover, this decomposition method models the SC signal as a convolution process between SudoMotor Nerve Activity (SMNA) and the Impulse Response Function (IRF) [9], where

$$SC = r + t + \epsilon = SMNA * IRF + \epsilon$$

$$= (Driver_{tonic} + Driver_{phasic}) * IRF + \epsilon$$

The $Driver_{phasic}$ denoted as p, and it were used for extracting features relating to SCRs peaks and rise time in the next step. The chosen model parameters were: $\tau_0 = 2$, $\tau_1 = 0.7$, $\delta = 10$, $\alpha = 0.016$, $\gamma = 0.01$ and solver = quadprog. Most of the parameters follow the setting in [7], whereas α was adjusted to the sparser sampled signal in this database which probably results in sparser spikes. Fig. 2 (c-e) illustrates the components after decomposition.

SCL was considered as a SC feature extracted within a 1 s wide sliding time window. In addition, more features were extracted from t, r and p within another 15 s wide sliding time window. The extracted features are listed in Table I, and are sorted into three groups: SCL, tonic features (Tonic_ftr) and phasic features (Phasic_ftr). Among them, the rise time of a SCRs was estimated from the duration of a peak-to-trough in p.

G1: SCL (= t, window width = 1 s, step =1 s)	
G2: Tonic_ftr (extracted from t, window width = 15 s, step = 1 s)	
tonic_avg:	mean value of t
tonic_auc:	area under the t curve
tonic_std:	standard deviation of t
G3-1: Phasic_ftr (extracted from r, window width = 15 s, step = 1 s)	
phasic_auc:	area under the r curve
phasic_std:	standard deviation of r
G3-2: Phasic_ftr (extracted from p, window width = 15 s, step = 1 s)	
phasic_driver_num_pks:	number of peaks in p (per min)
phasic_driver_pks_amp_max:	maximum p peak amplitude
phasic_rise_time_avg:	mean rise time of SCRs (second)

TABLE I

THREE FEATURE GROUPS AND LIST OF THE EXTRACTED FEATURES

C. Machine learning methods

The machine learning algorithms used in modeling included traditional machine learning algorithms k-nearest neighbour (KNN) and support vector machine (SVM) with different kernels, and a more recent method XGBoost [10]. KNN is a useful method for both linear and non-linear patterns and for SVM, linear and RBF kernels were chosen. The XGBoost is a boosted ensemble learning method utilizing decision tree learners, and has gained popularity in machine learning competitions and biomedical modeling [11].
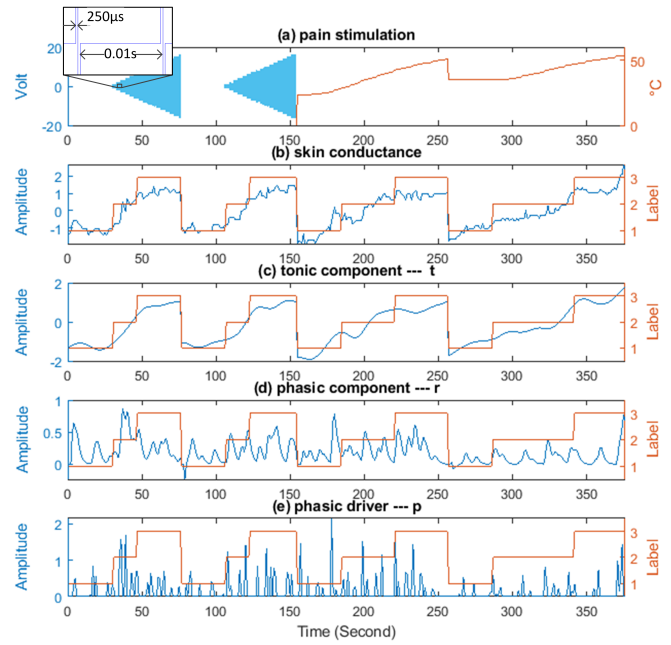

Fig. 2. An example of pain stimulation (a), original SC signal and its decomposed signals ((b-e) left y-axis) with the labels ((b-e) right y-axis).

Analysis included both categorical and regression modeling, as in this study context the categories are ordinal: KNN classification (KNNc) and regression (KNNr), SVM linear kernel classification (LinSVC) and regression (LinSVR), SVM radial basis function kernel regression (RBF SVR), and Extreme Gradient Boost tree ensemble (XGBoost). Different combinations of extracted features were separately analyzed, instead of feature selection: SCL alone, other tonic features extracted from SCL, and the phasic elements extracted from the phasic element SCRs.

D. Leave-subject-out cross-validation

When building a predictive machine learning model, the training and testing must be done with separate independent data. Additionally, hyper-parameters must be tuned outside of the training procedure, with separate train and validation sets. If this is not done, the model over-performs in evaluation and consequently under-performs with new unseen data.

In the study analysis, the meta-learning model (training, validation and test), was performed with nested leave-subject-out (LSO) cross-validation (visualized in Fig. 3). To obtain an unbiased estimate of the prediction performance, all the hyper-parameter optimization and feature selection were carried out using an inner cross-validation (train and validation sets) and the overall prediction performance was evaluated with an outer cross-validation (test sets) [12], [13]. Consecutive subject data in this setup is considered to be dependent, hence leave-subject-out cross-validation [14], [15] was used.

For each model, hyper-parameter optimization (k in KNNc and KNNr, C in LinSVC, C and ϵ in LinSVR, C, ϵ and γ in RBF SVR, and *eta*, *min_child_weight*, *max_depth*, *subsample*, *colsample* and *scale_pos_weight* in XGBoost) was performed

using grid search, with inner and outer leave-one-subject-out (LOSO) meta-learning models. The inner-loops in RBF SVR and XGBoost were split into five group folds of independent subject data, instead of LOSO, due to the computational complexity caused by the higher amount of hyper-parameters.

Finally, each model was evaluated with LSO cross-validation, using the optimal hyper-parameters from the meta-learning model.
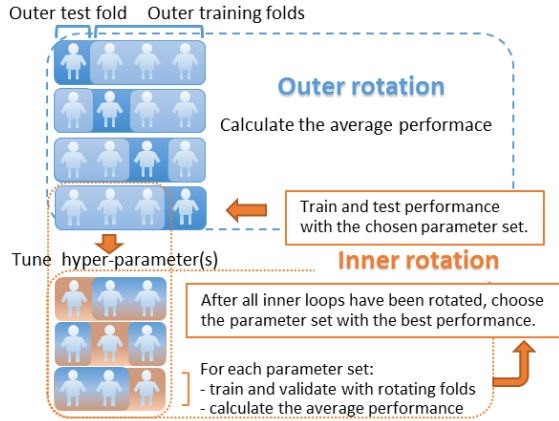

Fig. 3. The meta-learning model: tuning hyper-parameters with nested leave-subject-out cross-validation algorithm.

E. Model evaluation with c-index

Self-reporting based labels used in our study were unevenly distributed. In such situation, calculating accuracy based on actual accuracy (classification) or residuals (regression) could result too good result by predicting over-presented label values. C-index (also known as Harrell's c and c-statistics) is more suitable for performance evaluation with uneven categories, for both regression and ordinal classification. This is as it measures the consistence concordance between real and predicted responses. [16]

C-index values range from 0.0 to 1.0 with an example interpretation of predictions: 0.9–1.0 as excellent, 0.8–0.9 good, 0.7–0.8 fair, 0.6–0.7 poor, 0.5–0.6 random, and values below 0.5 discordant.

In the c-index algorithm real-prediction pairs are compared with each other and the proportional amount of the concordant pairs and half of the proportion of tied pairs are added together (Equation (1)-(2)).

$$\text{c-index}(\boldsymbol{y}, \hat{\boldsymbol{y}}) = \frac{1}{n_{samples}} \sum_{y_i > y_j} H(\hat{y}_i - \hat{y}_j), \qquad (1)$$

where H is a Heaviside step function for which

$$H(d) = \begin{cases} 1, & \text{if } d > 0. \\ 0.5, & \text{if } d = 0. \\ 0, & \text{if } d < 0. \end{cases} \qquad (2)$$

III. RESULTS

A. Ordinal classification and regression

The SC signal features (Table I) were grouped into seven different sets (Table II). Each of the three feature groups

was tested with each model individually and combined with the other one or two groups. Machine models were performed with six different methods, in which the ordinal label was treated as either ordinal class or numeric label. Analysis result performance metrics, the average c-index of all subjects, are showed in Table II.

The tonic features, including plain SCL, outperformed the phasic features clearly. When modeling only with Phasic_ftr, the prediction performed from random to poor. Predictions of models containing Tonic_ftr or SCL, on average showed at least fair concordant predictability. Near excellent average performance was reached in Linear SVR, and in RBF kernel SVR, when SCL or it's derivatives Tonic_ftr were used as features.

B. Performance in subject level

The resulting performance of subject level predictions by SCL+Tonic_ftr, RBF SVR varies from poor to excellent (Fig. 4). The majority of the predictions are clearly concordant with the self-report based test labels. The pain experience of two subjects (29th and 30th) are poorly predicted with the training data collected from other subjects. Fig. 5 gives an example of the tonic features and the predictions from the 9th subject who has excellent prediction performance in both classification and regression, and an example from 30th subject whose results are poor. The tonic_avg curve in Fig. 5(a) is representative in this study with a monotonically increase trend within one test, especially for those have excellent or good results. Comparatively, the example in Fig. 5(b) does not follow the common pattern, which may explain its unpredictability of pain.
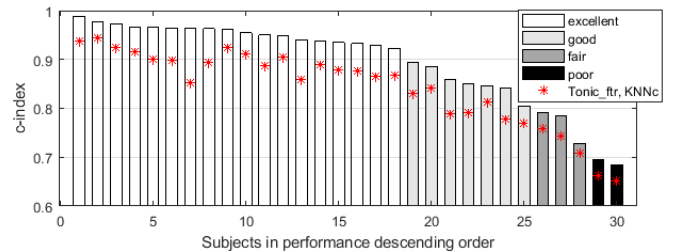

Fig. 4. Subject level performance with SCL+Tonic_ftr, RBF SVR, the corresponding c-index with Tonic_ftr, KNNc are marked for comparison

IV. DISCUSSION

In the gradual-increasing pain stimulation study setup, the tonic component of SC, and features extracted from it, but not the phasic features, contribute to good predictions. This could be due to the low frequency sampling, which produces a SC signal that represent more information on the low frequency element and, subsequently also on corresponding features. On the other hand, it could also be a result of the continuous response of SC to the continuous stimulation. Predictions fit very well to the pain stimuli periods, despite that the length differs from subject to another, and within subjects. Additionally, the overall performance showed the models could generalize over two difference pain stimuli.

	KNNc	KNNr	LinSVC	LinSVR	RBF SVR	XGBoost	Best method
SCL	0.820(0.10)	0.819(0.10)	0.757(0.09)	0.862(0.12)	0.862(0.12)	0.818(0.10)	LinSVR, RBF SVR
Tonic_ftr	**0.843(0.08)**	**0.841(0.08)**	0.766(0.09)	**0.883(0.10)**	0.892(0.09)	0.837(0.08)	RBF SVR
Phasic_ftr	0.599(0.07)	0.594(0.07)	0.546(0.09)	0.623(0.07)	0.632(0.08)	0.587(0.09)	RBF SVR
SCL+Tonic_ftr	0.835(0.09)	0.834(0.09)	0.783(0.08)	0.882(0.10)	**0.895(0.09)**	**0.842(0.09)**	RBF SVR
SCL+Phasic_ftr	0.805(0.10)	0.806(0.10)	0.765(0.09)	0.862(0.12)	0.865(0.12)	0.816(0.10)	RBF SVR
Tonic+Phasic_ftr	0.813(0.09)	0.811(0.09)	0.786(0.07)	0.882(0.10)	0.887(0.10)	0.830(0.09)	RBF SVR
All	0.820(0.09)	0.820(0.09)	**0.788(0.07)**	0.880(0.10)	0.889(0.10)	0.834(0.09)	RBF SVR
Best features	Tonic_ftr	Tonic_ftr	All	Tonic_ftr	SCL+Tonic_ftr	SCL+Tonic_ftr	

TABLE II

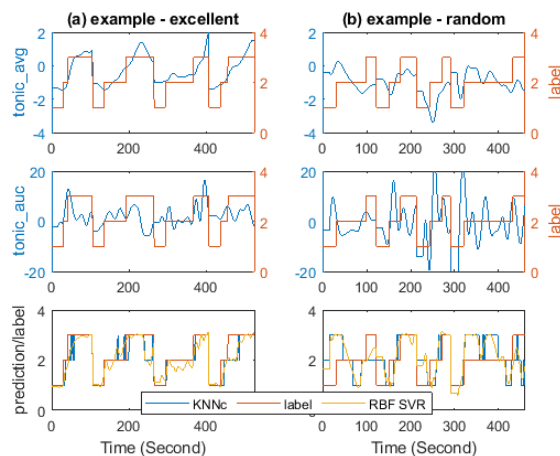LEAVE-SUBJECT-OUT C-INDEX AVERAGE PERFORMANCE, MEAN(STD)

Fig. 5. An example of excellent predictions (a) and an example of poor predictions (b) with Tonic_ftr, KNNc and RBF SVR

One of the limitations of unimodal use of SC signal in pain estimation in general is, that the electrodermal activity follows the emotional experience with an identical trend [17] identifying the active emotional response, but not distinguishing between negative and positive emotional arousal. In this study context, the positive arousal is very unlikely. However, the negative arousal may be a mixture of pain and emotional stress.

Pain estimation on individual subject level could be improved when using personalized or stratified modeling. This could be incorporated by adding clinical and contextual information [18]. With a larger group of study subjects, the variables related to different pain experience in a controlled experimental pain stimuli test should be further studied. These could be more detailed demographics and pain history questionnaires.

V. CONCLUSION

The tonic element of skin conductance represents well the experience under gradual-increasing pain stimuli. The majority of the subjects respond to the experience very similarly, hence the generalized model performs very well with the signal acquired from those individuals. However, the results show that there are subjects whose pain experience differs from the majority, in terms of physiological response in the form of the electrodermal activity.

REFERENCES

[1] M. L. Loggia *et al.*, "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.

[2] R. Treister *et al.*, "Differentiating between heat pain intensities: The combined effect of multiple autonomic parameters," *Pain*, vol. 153, no. 9, pp. 1807–1814, 2012.

[3] H. Storm, "Changes in skin conductance as a tool to monitor nociceptive stimulation and pain," *Current Opinion in Anaesthesiology*, vol. 21, no. 6, pp. 796–804, 2008.

[4] M. Jiang *et al.*, "Acute pain intensity monitoring with classification of multiple physiological parameters," *Journal of Clinical Monitoring and Computing*, vol. 0, no. 1, pp. 1–12, 2018.

[5] D. Lopez-Martinez *et al.*, "Continuous pain intensity estimation from autonomic signals with recurrent neural networks," *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5624–5627, 2018.

[6] B. T. Susam *et al.*, "Automated pain assessment using electrodermal activity data and machine learning," *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 372–375, 2018.

[7] A. Greco *et al.*, "CvxEDA: A convex optimization approach to electrodermal activity processing," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 797–804, 2016.

[8] T. Ledowski *et al.*, "Monitoring of sympathetic tone to assess postoperative pain: Skin conductance vs surgical stress index," *Anaesthesia*, vol. 64, no. 7, pp. 727–731, 2009.

[9] A. Greco *et al.*, "Modeling for the analysis of the EDA," in *Advances in Electrodermal Activity Processing with Applications for Mental Health*, pp. 19–33, Springer International Publishing, 2016.

[10] T. Chen *et al.*, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, ACM Press, 2016.

[11] A. S. Benjamin *et al.*, "Modern machine learning as a benchmark for fitting neural responses," *Frontiers in Computational Neuroscience*, vol. 12, p. 56, 2018.

[12] S. Varma *et al.*, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformatics*, vol. 7, pp. 1–8, 2006.

[13] G. Cawley *et al.*, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *The Journal of Machine Learning Research*, vol. 11, pp. 2079–2107, 2010.

[14] G. Xu *et al.*, "Asymptotic optimality and efficient computation of the leave-subject-out cross-validation," *Annals of Statistics*, vol. 40, no. 6, pp. 3003–3030, 2012.

[15] M. Esterman *et al.*, "Avoiding non-independence in fMRI data analysis: Leave one subject out," *NeuroImage*, vol. 50, no. 2, pp. 572–576, 2010.

[16] F. E. Harrell *et al.*, "Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors.," *Statistics in Medicine*, vol. 15, no. 4, pp. 361–387, 1996.

[17] L. Shu *et al.*, "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, p. 2074, 2018.

[18] G. Zamzmi *et al.*, "A review of automated pain assessment in infants: features, classification tasks, and databases," *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 77–96, 2018.