



Mingzhe Jiang

Automatic Pain Assessment by Learning from Multiple Biopotentials

TURKU CENTRE *for* COMPUTER SCIENCE

TUCS Dissertations
No 248, November 2019

Automatic Pain Assessment by Learning from Multiple Biopotentials

Mingzhe Jiang

*To be presented, with the permission of the Faculty of Science and
Engineering of the University of Turku in MED D1024 Säätiö-Sali on
November 28, 2019, at 12 noon*

University of Turku
Department of Future Technologies

2019

Supervisors

Professor Pasi Liljeberg

Department of Future Technologies, University of Turku, Finland

Adjunct Professor Amir M. Rahmani

Department of Future Technologies, University of Turku, Finland

Assistant Professor

University of California, Irvine, USA

Reviewers

Associate Professor Raquel Bailón

Department of Electronic Engineering and Communications

University of Zaragoza, Spain

Associate Professor Steven Su

School of Biomedical Engineering

University of Technology Sydney, Australia

Opponent

Adjunct Professor Heli Koskimäki

Biomimetics and Intelligent Systems Group

University of Oulu, Finland

Senior Data Scientist

OURA Health Ltd, Finland

Painosalama Oy, Turku, Finland

ISBN 978-952-12-3889-5

ISSN 1239-1883

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service

Abstract

Accurate pain assessment plays an important role in proper pain management, especially among hospitalized people experience acute pain. Pain is subjective in nature which is not only a sensory feeling but could also combine affective factors. Therefore self-report pain scales are the main assessment tools as long as patients are able to self-report. However, it remains a challenge to assess the pain from the patients who cannot self-report. In clinical practice, physiological parameters like heart rate and pain behaviors including facial expressions are observed as empirical references to infer pain objectively. The main aim of this study is to automate such process by leveraging machine learning methods and biosignal processing.

To achieve this goal, biopotentials reflecting autonomic nervous system activities including electrocardiogram and galvanic skin response, and facial expressions measured with facial electromyograms were recorded from healthy volunteers undergoing experimental pain stimulus. IoT-enabled biopotential acquisition systems were developed to build the database aiming at providing compact and wearable solutions. Using the database, a biosignal processing flow was developed for continuous pain estimation. Signal features were extracted with customized time window lengths and updated every second. The extracted features were visualized and fed into multiple classifiers trained to estimate the presence of pain and pain intensity separately. Among the tested classifiers, the best pain presence estimating sensitivity achieved was 90% (specificity 84%) and the best pain intensity estimation accuracy achieved was 62.5%.

The results show the validity of the proposed processing flow, especially in pain presence estimation at window-level. This study adds one more piece of evidence on the feasibility of developing an automatic pain assessment tool from biopotentials, thus providing the confidence to move forward to real pain cases. In addition to the method development, the similarities and differences between automatic pain assessment studies were compared and summarized. It was found that in addition to the diversity of signals, the estimation goals also differed as a result of different study designs which made cross dataset comparison challenging. We also tried to discuss which parts in the classical processing flow would limit or boost the prediction performance and whether optimization can bring a breakthrough from the system's perspective.

Tiivistelmä

Kivun täsmällinen arvointi on tärkeää kivunhallinnassa, erityisesti sairaanhoitoa vaativille kipupotilaille. Kipu on subjektiivista, sillä se ei ole pelkästään aistituntemus, vaan siihen saattaa liittyä myös tunnekokemuksia. Tällöin itsearvointiin perustuvat kipuasteikot ovat tärkein työkalu, niin kauan kun potilas pystyy kokemuksensa arvioimaan. Arvointi on kuitenkin haasteellista potilailla, jotka eivät itse pysty kertomaan kivustaan. Klinisessä hoitotyössä kipua pyritään objektiivisesti arvioimaan esimerkiksi havainnoimalla fysiologisia muuttujia kuten sykettä ja käyttäytymistä esimerkiksi potilaan kasvonilmeiden perusteella. Tutkimuksen päätavoitteena on automatisoida arvointiprosessi hyödyntämällä koneoppimismenetelmiä yhdessä biosignaalien prosessointin kanssa.

Tavoitteineen saavuttamiseksi mitattiin autonomista keskushermoston toimintaa kuvastavia biopotentiaaleja: sydänsähkökäyrää, galvaanista ihoreaktiota ja kasvolihasliikkeitä mittaavaa lihassähkökäyrää. Mittaukset tehtiin terveillä vapaaehtoisilla, joille aiheutettiin kokeellista kipuärsykettä. Järjestelmän kehittämiseen tarvittavaa tietokantaa varten rakennettiin biopotentiaaleja keräävä Internet of Things -pohjainen tallennusjärjestelmä. Koostetun tietokannan avulla kehitettiin biosignaaleille prosessointimenetelmä jatkuvaan kivun arvointiin. Signaaleista eroteltiin piirteitä sekuntitasoon mukaanluulla aikaikkunoilla. Piirteet visualisoitiin ja tarkasteltiin eri luokittelijoilla kivun ja kiputason tunnistamiseksi. Parhailla luokittelumenetelmillä saavutettiin kivuntunnistukseen 90% herkkyykskyky (sensitivity) ja 84% erottelukyky (specificity) ja kivun voimakkuuden arvointiin 62,5% tarkkuus (accuracy).

Tulokset vahvistavat kyseisen käsitellytavan käyttökelpoisuuden erityisesti tunnistettaessa kipua yksittäisessä arvointi-ikkunassa. Tutkimus vahvistaa biopotentiaalien avulla kehitettävän automatisoidun kivun arvioinnin toteutettavuuden kokeellisella kivulla, rohkaisten etenemään todellisen kivun tutkimiseen samoilla menetelmillä. Menetelmää kehitettäessä suoritettiin lisäksi vertailua ja yhteenvetoa automaattiseen kivuntunnistukseen kehitetyjen eri tutkimusten välisistä samankaltaisuksista ja eroista. Tarkastelussa löytyi signaalien eroavaisuuksien lisäksi tutkimusmuotojen aiheuttamaa eroa arvointitavoitteisiin, mikä hankaloitti tutkimusten vertailua. Lisäksi pohdittiin mitkä perinteisten prosessointitapojen osiot rajoittavat tai edistävät en-nustekykyä ja miten, sekä tuoko optimointi läpimurtoa järjestelmän näkökulmasta.

Acknowledgements

When I stepped on the land of Finland for the first time, I had no idea how amazing the coming years would be. This peaceful nature land seems to have the power of amplifying joy and healing sorrows. A degree is only one of the milestones throughout life. The growth gained while reaching there is the real treasure to me. All my academic growth and achievements cannot be made without supports from my supervisors and the inter-disciplinary research team that I'm in.

I would like to express my sincere thanks to my principal supervisor Prof. Pasi Liljeberg for the inclusiveness, flexibility, academic knowledge and resources always provided with wise guidance. I would like to thank my supervisor Adjunct Prof. Amir M. Rahmani, who offered sharp publication and polishing ideas with generosity in time and wisdom sharing. I want to thank Prof. Sanna Salanterä in the team from the Nursing Science department. Her strong supports throughout my doctoral study helped me filling my knowledge gap in the pain assessment field and in clinical research. I want to thank the doctor experts, Prof. Riku Aantaa and Adjunct Prof. Nora Hagelberg, who were in the team and brought inspiring and intellectual discussions. Also, I would like to express my gratitude for receiving help and encouragement, especially before and at the beginning of my study, from Research Prof. Geng Yang, Adjunct Prof. Tomi Westerlund, and Prof. Hannu Tenhunen.

I'm grateful to have Associate Prof. Raquel Bailón and Associate Prof. Steven Su as dissertation reviewers, for both valuable comments and acceptance with honours. My appreciation is also extended to Adjunct Prof. Heli Koskimäki for consent to act as the opponent in the public examination.

This study was supported by the Academy of Finland, China Scholarship Council, and HPY Research Foundation. The academic trips during my doctoral study were partially supported by Turku University Foundation, MATTI travel grant from the doctoral program, and IEEE Sensors Applications Symposium. I want to thank these organizations for supporting young researchers in building their early careers.

Both of the presented thesis work and related publications cannot be them without the creative and precious contributions from my talented and

hardworking colleague co-authors. I want to express my deep gratitude to Tuan Nguyen Gia, Victor Kathan Sarker, and Arman Anzanpour, who always have impressive solutions to small everyday challenges in research. I would especially thank Elise Syrjälä, who brought me lots of encouragement with positive attitude, self-motivation, and beneficial discussions on research during my toughest doctoral time in the middle. My thanks are also for my co-workers with different experiences and backgrounds in the Smart Pain Assessment team, Hanna-Maria Matinolli, Riitta Mieronkoski, Mikko Koivumäki, and Virpi Terävä. There were many moments we cover each other's back, and I do appreciate the mutual support and knowledge sharing at each moment. I would like to thank all the colleagues in the same office environment and across the corridor in different periods. The interesting stories, enthusiasm in hobbies, and positive attitudes colored my doctoral life.

I would like to thank all my great friends, those who were in Turku, and those who are still in Turku, especially Ping Wang, who lightened the dark days of winter in Finland like bright sunshine. It is also my luck to have Xueying Ma and Wei Yang, who inspired me with their passion to life. I'm thankful to have many lovely friends from different parts of China sharing laughs and worries, playing sports, and cooking together. Last but not least, endless thanks to my parents, and my husband Hao Niu, for backing me up unconditionally and always.

List of original publications

Publications related to this thesis

1. **Jiang, M.**, Mieronkoski, R., Syrjälä, E., Anzanpour, A., Terävä, V., Rahmani, A.M., Salanterä, S., Aantaa, R., Hagelberg, N., Liljeberg, P., 2019. Acute pain intensity monitoring with the classification of multiple physiological parameters, *Journal of Clinical Monitoring and Computing*, 33(3), pp.493-507. [study 1]
2. **Jiang, M.**, Mieronkoski, R., Rahmani, A.M., Hagelberg, N., Salanterä, S., Liljeberg, P., 2017, November. Ultra-short-term analysis of heart rate variability for real-time acute pain monitoring with wearable electronics, In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1025-1032). IEEE. [study 2]
3. **Jiang, M.**, Gia, T.N., Anzanpour, A., Rahmani, A.M., Westerlund, T., Salanterä, S., Liljeberg, P. and Tenhunen, H., 2016, April. IoT-based remote facial expression monitoring system with sEMG signal. In *2016 IEEE Sensors Applications Symposium (SAS)* (pp. 1-6). IEEE.
4. **Jiang, M.**, Rahmani, A.M., Westerlund, T., Liljeberg, P. and Tenhunen, H., 2015, October. Facial Expression Recognition with sEMG Method. In *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM)* (pp. 981-988). IEEE.
5. Syrjälä, E., **Jiang, M.**, Pahikkala, T., Salanterä, S. and Liljeberg P., 2019, July. Skin conductance response to gradual-increasing experimental pain. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. (Accepted) [study 3]
6. Yang G., **Jiang, M.**, Ouyang, W., Ji, G., Rahmani, A.M., Liljeberg, P., Tenhunen, H., 2018. IoT-based remote pain monitoring system: from device to cloud platform, *IEEE Journal of Biomedical and Health Informatics*, 22(6), pp.1711-1719.
7. Sarker, V.K., **Jiang, M.**, Gia, T.N., Anzanpour, A., Rahmani, A.M. and Liljeberg, P., 2017, March. Portable multipurpose bio-signal acquisition and wireless streaming device for wearables. In *2017 Sensors Applications Symposium (SAS)* (pp. 1-6). IEEE.

Other publications

8. Gia, T.N., **Jiang, M.**, Sarker, V.K., Rahmani, A.M., Westerlund, T., Liljeberg, P. and Tenhunen, H., 2017, June. Low-cost fog-assisted health-care IoT system with energy-efficient sensor nodes. In *2017*

- 13th International Wireless Communications and Mobile Computing Conference (IWCMC)* (pp. 1765-1770). IEEE.
9. Gia, T.N., **Jiang, M.**, Rahmani, A.M., Westerlund, T., Liljeberg, P. and Tenhunen, H., 2015, October. Fog computing in healthcare internet of things: A case study on ecg feature extraction. In *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM)* (pp. 356-363). IEEE.
 10. Negash, B., Gia, T.N., Anzanpour, A., Azimi, I., **Jiang, M.**, Westerlund, T., Rahmani, A.M., Liljeberg, P. and Tenhunen, H., 2018. Leveraging Fog Computing for Healthcare IoT. In *Fog Computing in the Internet of Things* (pp. 145-169). Springer International Publishing.
 11. Rahmani, A.M., Gia, T.N., Negash, B., Anzanpour, A., Azimi, I., **Jiang, M.** and Liljeberg, P., 2018. Exploiting smart e-health gateways at the edge of healthcare internet-of-things: A fog computing approach. *Future Generation Computer Systems*, 78, pp.641-658.
 12. Yang, G., Deng, J., Pang, G., Zhang, H., Li, J., Deng, B., Pang, Z., Xu, J., **Jiang, M.**, Liljeberg, P. and Xie, H., 2018. An IoT-Enabled Stroke Rehabilitation System Based On Smart Wearable Armband And Machine Learning. In *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 6, pp. 1-10.

List of Abbreviations

μS micro-Siemens

ABP arterial blood pressure

AC alternating current

ACC accuracy

AdaBoost adaptive boosting

ADC analog-to-digital converter

Ag silver

AgCl silver chloride

ANFIS adaptive-network-based fuzzy inference system

ANI analgesia/nociception index

ANS autonomic nervous system

AR autoregressive model

AU action unit

AUC area under the GSR/ROC curve

Bag bagging, bootstrap aggregation

BIAS bias drive output in biopotential acquisition

BPAT Behavior Pain Assessment Tool

BPS Behavioral Pain Scale

BVP blood volume pulse

CH channel

CM common-mode voltage

CMRR common-mode rejection ratio

CNAP continuous noninvasive arterial pressure

cor facial muscle-corrugator supercilii

CPOT Critical-Care Pain Observation Tool

DAMV difference absolute mean value

dB decibel

DC direct current

E4 E4 wristband from empatica

ECG electrocardiogram

EDA Electrodermal activity

EDR ECG-derived respiration

EEG electroencephalogram

EMG electromyogram

FACS Facial Action Coding System
FFT fast Fourier transform
FLACC Face, Legs, Activity, Cry, and Consolability Behavioral Assessment Tool
FMD median frequency
FMN mean frequency
fMRI functional magnetic resonance imaging
FPK peak frequency

GSR galvanic skin response

H124SG KendallTMECG electrodes, round 24mm
HF HRV-the high-frequency power component of NN series (0.15-0.40 Hz)
HRV heart rate variability

IASP International Association for the Study of Pain
ICU intensive care unit
INxN negative input in biopotential acquisition
INxP positive input in biopotential acquisition
IoT Internet-of-Things

knn k-nearest neighbor

LA ECG measurement-left arm
LD log detector
lev facial muscle-levator labii superioris
LF HRV-the low-frequency power component of NN series (0.04-0.15 Hz)
LL ECG measurement-left leg

MAE mean absolute error
MAV mean absolute value
MAVS mean absolute value slope
MDS multidimensional scaling
MEG magnetoencephalography
MPV maximum absolute value
MSE mean square error

NAN nociception/antinociception
NaN not a number
NN normal-to-normal
NPAT Nonverbal Pain Assessment Tool
NRS numeric rating scale
NSCF number of skin conductance fluctuations
NVPS Non-verbal Pain Scale

orb facial muscle-orbicularis oculi

PBAT Pain Behavioral Assessment Tool

PET positron emission tomography

PPG photoplethysmogram

PPGA PPG amplitude

PSPI Prkachin and Solomon Pain Intensity Metric

R² R squared

RA ECG measurement-right arm

REF common reference in biopotential acquisition

ris facial muscle-risorius

RMS root mean square

RMSE root mean square error

RMSSD HRV-root mean square of the successive differences

RNN recurrent convolutional neural network

ROC receiver operating characteristic

RUSBoost random undersampling boosting

S1 the first step classification

S2 the second step classification

SaO₂ arterial oxygen saturation

SBP systolic blood pressure

SCL skin conductance level

SCR skin conductance response

SD standard deviation

SDNN HRV-standard deviation of all NN intervals

sEMG surface electromyogram

SMNA SudoMotor Nerve Activity

SNR signal-to-noise ratio

SPA Smart Pain Assessment

SpaExp Smart Pain Assessment - Experimental Pain Database

SPI surgical pleth index

SpO₂ Peripheral oxygen saturation

SPS samples per second

SSC slope sign change

SSI surgical stress index

SVM support vector machine

TNR true negative rate

TPR true positive rate

VAS visual analog scale

VRS verbal categorical rating scale

WL wave length

ZC zero crossing

zyg facial muscle-zygomaticus major

List of Figures

2.1	Remote pain monitoring in hospital for inpatients and ICU patients ©2018 IEEE	23
3.1	Mechanisms of difference interference caused by the electric field in ECG measurement [1, 2]	28
3.2	Driven right leg circuit to decrease common-mode voltage	28
3.3	Part of the ADS1299 configurations including lead-off detection, common reference, and bias drive	30
3.4	Comparison of power line interference in ADS1299's different working modes	31
3.5	The block diagram of the designed sEMG acquisition device	32
3.6	Software workflow	34
3.7	The firmware structure of the AVR processor and its communication with the software	35
3.8	File operations in the software	37
3.9	A piece of ECG signal collected in Section 3.1.1	38
3.10	The envelopes of the sEMG signals acquired by the system developed in this study (UTU-BASD) and by ME6000 Biomonitor	38
3.11	The respiratory modulation of the waveform amplitude	42
3.12	Two GSR exosomatic direct voltage recording circuits	43
4.1	The timeline and pain stimulus intensity in one test	47
4.2	The waveform of the electrical stimulation	48
4.3	The potential scald injury caused by heat and the temperature curve of the heat pain stimulation in the test	49
4.4	The distribution of the reported VAS scores at t_2 in all tests and the tests without an "intolerable" report	51
4.5	Biosignals acquisition software platform	52
4.6	GSR electrode sites on the volar surface	54
4.7	The signals in one electrical test	56
5.1	Adaptive noise cancellation to denoise electrical pulses	61
5.2	ECG and sEMG denosing, signal processing flow	62

5.3	The frequency response of the filters and their time response on ECG signal	63
5.4	An illustration of ECG waves [3]	64
5.5	Some manual corrections for the wrong or missed R peak detections	65
5.6	Signal processing flow in 1) R peak detection and 2) respiratory cycles extraction from the detected R peaks	67
5.7	Visualization of the key steps in EDR extraction (a-c). Capacitive pressure sensor waveform (d). The comparison of the estimated breathing rate and heart rate between the reported values by Bioharness and the extracted values in this study (e-f)	68
5.8	The distribution of the mean difference between the reported values and the extracted values in all the study subjects	69
5.9	Decompose GSR into tonic SCL and phasic SCR	74
6.1	The GSR feature list and the linear correlation (Pearson's r) between every two features	81
6.2	The tonic component of all the tests in each pain category normalized to t_1 and their average curve (a-c); a comparison of the average curves (d).	82
6.3	Area under the tonic component (t_auc)	82
6.4	Standard deviation of the tonic component (t_std) in each pain category and a comparison of their averages	83
6.5	Number of phasic driver peaks per minute (p_num_pks)	83
6.6	The ECG feature list and the linear correlation (Pearson's r) between every two features	85
6.7	Heart rate (ecg_hr) rescaled with λ_1	86
6.8	SDNN (ecg_nsdnn) rescaled with λ_2	86
6.9	RMSSD (ecg_nrmsd) rescaled with λ_2	87
6.10	LF (ecg_nlf) rescaled with λ_3	87
6.11	HF (ecg_nhf) rescaled with λ_3	88
6.12	LF/HF (ecg_lf/hf) normalized to t_1	88
6.13	Sample entropy (ecg_sampen) normalized to t_1	89
6.14	Stimulus intensity distribution in each pain category	89
6.15	ECG derived respiration rate ($respr_edr$) rescaled with λ_4	91
6.16	Bioharness respiration rate ($respr_bha$) rescaled with λ_4	91
6.17	Mean absolute Spearman's ρ of the subjects between adaptively filtered sEMG signals, $std \in (0.13, 0.22)$	92
6.18	Mean absolute Spearman's ρ of the subjects between fifteen sEMG features of muscle corrugator and muscle zygomaticus major, $std \in (0, 0.25)$	93
6.19	Root mean square of corrugator sEMG (emg_cor_rms)	94

6.20	The first coefficient in the 4th autoregressive model of corrugator sEMG (<i>emg_cor_ar_1</i>)	94
6.21	The first coefficient in the 4th autoregressive model of zygomaticus major sEMG (<i>emg_zyg_ar_1</i>)	95
6.22	Multidimensional scaling plot of the Euclidean distance between the representative features	95
7.1	Study design interpretation	99
7.2	Signal processing & feature extraction flow summary and the composition of the feature matrix	100
7.3	Confusion matrix in S1 binary classification	102
7.4	S1 - relative importance of the features in tree-based classifiers and linear discriminant analysis classifiers, with average test sensitivity and specificity	106
7.5	S1 - Model performance without hyperparameter optimization	107
7.6	S1 - The performance in each leave-subject-out cross-validation test fold (subjects in ensemble_RUSBoost sensitivity descending order)	109
7.7	S1 - Classification performance with default model settings and optimized settings	109
7.8	Two S1 RUSBoost estimation examples	112
7.9	S2 - relative importance of the features in tree-based classifiers with average train accuracy and test accuracy	114
7.10	S2 - Model performance without hyperparameter optimization	115
7.11	S2 - Training accuracy and test accuracy change after hyperparameter optimization	116
7.12	S2 - Pain intensity labels and predictions in optimized_svm_linear 3-class classification	119

List of Tables

2.1	The biosignals reflecting ANS activities and some NAN indexes [4, 5]	12
2.2	The description on the facial expressions/emotion in the behavioural pain assessment tools	16
2.3	The facial action units involved in pain facial expressions in adults	17
2.4	The description on the movements/guarding/muscle tension in the behavioural pain assessment tools	19
2.5	The description on the vocalization/compliance with ventilation/consolability/phsiological signs in the behavioural pain assessment tools	20
2.6	Example devices and systems providing raw data for research	22
3.1	The lead wire configuration and key registers configuration in the four modes	31
3.2	Firmware's responses to the software commands	36
3.3	Frequency and amplitude characteristics of the EEG, ECG and EMG	37
4.1	Information lookup table, item list	50
4.2	Timestamp lookup table, item list	51
4.3	Electrodes placement and their connections to the device	53
4.4	Signal summary	55
5.1	Common biosignal contaminants [6, 7]	58
5.2	R peak detection results and manual correction	65
5.3	HRV features and their definitions	70
5.4	sEMG features and their descriptions	72
5.5	GSR features and their descriptions	73
5.6	The recommended minimum time window length in feature extraction	75
6.1	A summary of features of interest	79

6.2	The RMSE and Spearman's ρ between the respiration rate calculated from difference sources, mean (std). *One extremely large value is excluded in the averaging.	90
7.1	Cost matrix for binary classification	103
7.2	Specifications of the classifiers	104
7.3	S1 - Four selected models	108
7.4	S1 - The best sensitivity and the corresponding model for each test subject fold and the sensitivity standard deviation among all the 26 models (subjects in best sensitivity descending order)	111
7.5	S2 - The best test accuracy and the corresponding model for each test subject fold and the test accuracy standard deviation among the implemented 26 models (subjects in best test accuracy descending order)	117
7.6	A list of representative studies on automatic pain assessment using machine learning	120
7.7	Algorithm and performance comparison (studies in Table 7.6)	130

Contents

1	Introduction	1
1.1	Background	1
1.2	Research focus	3
1.3	Aim and objectives	4
1.4	Contributions	4
1.5	Thesis content summary	5
2	Labels and Signals in Pain Assessment	7
2.1	The complex and subjective nature of pain	7
2.2	Pain scales in clinical practice and algorithms	9
2.3	Signals in pain intensity recognition	10
2.3.1	Autonomic nervous system (ANS) based signals	11
2.3.2	Respiration and oxygen saturation	14
2.3.3	Behavioural signals	14
2.3.4	Neuroimaging signals	21
2.4	Wearable devices and IoT-enabled systems	21
2.5	Summary	23
3	Biosignals Acquisition	25
3.1	Facial surface electromyography	25
3.1.1	The core of the acquisition system	26
3.1.2	sEMG/Biopotential acquisition system design	32
3.1.3	Discussion	36
3.2	Electrocardiography and respiration	40
3.2.1	Wearable device - Bioharness 3	40
3.2.2	Patient monitor	41
3.2.3	Respiration derived from other measurements	41
3.3	Galvanic skin response	43
3.4	Summary	44
4	Study Design and the SpaExp Database	45
4.1	Experiment protocol	45
4.2	Study design	46

4.2.1	Experimental pain stimulation	47
4.2.2	Biosignals and the acquisition system	52
4.3	Summary of the collected signals	55
5	Biosignal Processing and Feature Extraction	57
5.1	Signal denoising	57
5.1.1	Motion artifact	57
5.1.2	Instrumentation contaminants	58
5.1.3	Interference	60
5.1.4	Implementation of denoising	61
5.2	Feature extraction	63
5.2.1	R peak detection in ECG	64
5.2.2	ECG derived respiration	66
5.2.3	Heart rate variability	69
5.2.4	sEMG	71
5.2.5	GSR	73
5.2.6	Window length for feature extraction	73
5.2.7	Inter-subject and intra-subject variability, and signal or feature normalization	76
5.3	Summary and future work	77
6	Features and Visualization	79
6.1	GSR features	80
6.1.1	Tonic component	81
6.1.2	Phasic component	83
6.2	ECG features	84
6.2.1	Heart rate	85
6.2.2	Heart rate variability	86
6.2.3	Respiration rate	90
6.3	sEMG features	91
6.4	Distance in representative features	96
6.5	Summary and discussion	96
6.5.1	GSR features	96
6.5.2	ECG features	96
6.5.3	sEMG features	97
7	Pain Estimation from Physiological Features	99
7.1	Methods	100
7.1.1	Leave-subject-out cross-validation	101
7.1.2	Performance metrics	101
7.1.3	Hyperparameter or Model tuning	102
7.1.4	Misclassification cost matrix in cost-sensitive learning	102
7.1.5	Feature importance	103

7.1.6	Classifiers	103
7.2	Results	105
7.2.1	Step 1 - Estimating the presence of the pain or inadequate pain control	105
7.2.2	Step 2 - Estimating pain level when pain is present or pain control is inadequate	113
7.3	Discussion of automatic pain assessment algorithms	118
7.3.1	Signal processing time window and the pain assessment frequency	123
7.3.2	Labels and proxy/objective pain assessment	124
7.3.3	Data exclusion	126
7.3.4	Machine learning algorithms, cross-validation and performance evaluation	127
7.3.5	Future work	128
8	Conclusions	131
8.1	Main findings	131
8.2	Significance of the study	133
8.3	Future research	134
References		135

Chapter 1

Introduction

1.1 Background

Proper pain management plays an important role in assessing the quality of care among hospitalized people experience acute pain. Inadequately managed pain negatively affects patients' physical and psychological health as well as hospital's performance, while overtreatment may lead to serious and even life-threatening consequences. Efforts to avoiding both undertreatment and overtreatment of acute pain are being made from the perspective of analgesic medications and techniques, and both analgesic efficacy and safety are being improved [8]. Meanwhile, pain assessment by physicians and nurses is a critical part of reaching optimal pain management. Pain assessment by physicians and nurses is also a critical part of achieving good pain management. Pain assessment mainly refers to an assessment of pain intensity which helps to decide the type of intervention that will be used including the type of analgesic to be administered and the dosage [9]. In addition to intensity, the location and quality (e.g., aching and burning) of pain are also the assessment aspects.

Pain is defined as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage" by the International Association for the Study of Pain (IASP)¹. The definition of pain reveals that pain has a sensory dimension and an affective or emotional dimension. Although pain, especially acute pain, originates from the sensory neurons in response to damaging or potentially damaging stimuli, the pain sensation is actually processed by the brain [10, 11]. Due to the subjective nature of pain, self-reporting is considered to be the gold standard when assessing pain. Pain intensity is assessed with a pain scale in one of several forms. For example, in acute postoperative pain, a score 4 in the 11-point scale from 0 to 10 or between 30 and 40 in a 100 mm scale is clinically important borderline of receiving adequate pain control [12, 13].

¹<https://www.iasp-pain.org/Education/Content.aspx?ItemNumber=1698#Pain>

Pain assessment is usually documented several times a day for patients with acute pain and the document frequency is among the indicators of pain management quality [14]. Depending on the defined nursing protocol, the assessment of postoperative acute pain could be regularly every 4 h plus 1 h after the intervention [15], or every 15 min immediately after surgery [16]. However, such manual inquiries can never produce continuous monitoring and would have "missing values" when a patient cannot self-report (e.g., in sleep).

Moreover, several patient groups cannot communicate by any means (e.g., critically ill patients and people with limited cognitive ability) and thus cannot self-report. In these cases, the alternative tools in clinical practice are pain behavior observation tools [17–22] where the levels of pain behaviors are observed by a trained caregiver. The pain behaviors cover pain facial expression, body movement, vocalization or compliance with ventilation. Additionally, in one tool, some physiological signs such as systolic blood pressure, respiration rate, heart rate, and oxygen saturation rate are considered nonverbal signs of pain.

The previous observations and work on pain behaviors and physiological signs of pain built the foundation and intuition for the later studies that tried to quantitate the reactions to pain. Then more studies arise with the proliferation of miniaturized sensing systems and machine learning techniques. Now, several of the pain behaviors and physiological signals are within the scope of automatic pain assessment studies aiming at pain intensity estimation in a pain event or in a time-continuous manner. The automatic assessment "machine" is expected to work as accurately and reliably as self-report, and at least as well as expert human observers. This study is part of the efforts to reach this ultimate goal. The study was initialized and conducted within the Smart Pain Assessment (SPA) multidisciplinary research group.

The biopotentials included in this study are several physiological signals including

- electrocardiogram (ECG) - a graph of the electrical activity of the heart,
- galvanic skin response (GSR) - the skin resistance change due to the activity of sweat glands in the skin,
- facial electromyogram (EMG) - a record of the electrical activity produced by facial skeletal muscles which are also a pain behavior indicator corresponding to facial expression.

Additionally, the respiration rate is also part of the analysis. The reactions of these signals in response to experimental pain were recorded from healthy volunteers and were analyzed. Despite using a simulated scenario with experimental pain, a broader scope of knowledge on pain (e.g., postoperative pain and surgical stress) and pain reactions were used to ascertain information and discussed. We believe this study will add value back to this field.

1.2 Research focus

This study focuses on modeling self-report pain from chosen physiological signals to provide a continuous pain estimation, by recognizing the pattern of the signals. The rationale behind the study is twofold. One is the autonomic nervous system (ANS) activities [23], which act unconsciously and regulate the body functions to fight-or-flight or rest-and-digest. Such regulations are reflected in the physiological parameters such as heart rate, respiration rate, and sweat. The other one is the pain behaviors introduced above.

One difficulty in achieving an automatic pain assessment tool is the individual difference in pain perception which is influenced by the inherent past and the present states such as genetic, physical, social and psychological factors. The fact that any single potential objective pain indicator does not respond only to pain also makes the development difficult. For example, heart rate may vary due to age, depth of anesthesia, medications, and emotions, and this is the same as regards pain behaviors. Usually, an assessment method is narrowed down to a specific type of pain and scenario and must be validated before moving to a new one (e.g., surgical stress index). Moreover, among the behavior observation tools, there are indications that a tool developed for one patient group may not apply or be generalized to another group (e.g., from children to adults) [24].

From the technical point of view, the main challenge is how to adapt the existing biosignal processing and machine learning techniques to the case of pain measurement. In a classic pattern recognition flow, the techniques such as signal acquisition, noise cancellation, feature extraction, and machine learning models have been and continue to be developed for either a general purpose or specific applications. For example, EMG feature extraction and selection have been studied for myoelectric control applications, and heart rate variability analysis is used for mental stress assessment or as a risk factor for sudden cardiac death. Meanwhile, it would be interesting to see how far these techniques can aid the progress of this application. Finally, this study was inspired by the upsurge in wearable technology, the growing number of published databases for automatic pain assessment using machine learning methods¹, as well as the finding that multiple parameters or a multimodal is better than one [30, 31].

¹1) EmoPain [25] (2015), chronic pain; 2) UNBC-McMaster [26] (2011), shoulder pain; 3) Infant COPE [27] (2007), neonatal pain; 4) BioVid [28] (2013), experimental heat pain; 5) X-ITE [29] (2019), experimental heat and electrical pain.

1.3 Aim and objectives

The study aims to model self-report pain from ECG, GSR, and facial surface EMG signals in order to provide a continuous estimation of pain with the Smart Pain Assessment - Experimental Pain Database (SpaExp). With adjustable stimulus intensity and duration, experimental pain stimuli evoke pain in a controllable way to unify the subjective pain report (i.e., pain threshold ¹, pain tolerance ², and pain intensity) and are commonly used in pain studies. As a study learning from experimental pain, this work is expected to provide a system design and answer questions regarding validity, reliability, and the limitations of the approach. The results will give some indications regarding real acute pain cases such as postoperative pain and the pain in intensive care unit (ICU). To reach the study goal, the system design is divided into several sub-designs linked to one another but also relatively independent. The step-by-step objectives are

- To identify the target biosignals and assistive pain assessment tools that have been used or potentially can provide noninvasive and continuous monitoring of pain;
- To build and evaluate a biosignal acquisition system for data collection and build the SpaExp database by collaborating with medical doctor and nurse researchers;
- To develop a processing flow for continuous pain estimation based on the attribute of each signal or feature and the SpaExp study design whilst also considering the compatibility to different databases or study design with the same signal;
- To examine the potential capability of each extracted feature on estimating pain by visualizing and observing the average response to pain;
- To assess pain (intensity) by multimodal machine learning;
- To evaluate the importance of the features and the model performance, and compare the performances across multiple classifier models.

1.4 Contributions

Developing automatic pain assessment methods is an interdisciplinary topic. The existing studies either involve experts in clinical medicine and computer science or in psychology and computer science due to the affective dimension of pain. The increasing number of pain databases with different signals or study designs encouraged the desire of many computer science researchers to improve the estimation performance within a database from the perspective

¹<https://www.iasp-pain.org/terminology?navItemNumber=576#Painthreshold>

²<https://www.iasp-pain.org/terminology?navItemNumber=576#Paintolerancelevel>

of algorithm optimization (i.e., trying designs with different face descriptors, features extractions, and machine learning or deep learning models). Within the same database (i.e., study design), the performance could be improved through algorithm optimization. However, the steps forward have been subtle rather than revolutionary to make a valid and reliable tool. Moreover, few studies analyzed data across different databases. The similarity and difference among the study designs have hardly been discussed, especially between the pain in hospital and the stimulated pain in an experimental environment. This study widely reviewed the studies from clinical assessment tools to the efforts made to develop automatic pain assessment tools. Untangling the similarities among the study designs and finding the gaps is the first contribution of this study.

Secondly, the endeavors of this study and my other works using the same database [32–34] were aimed at answering the questions of: 1) which parts in the classical processing flow for window-level pain classification limit the prediction performance, and 2) whether it is valuable to put significant efforts on optimizing signal processing, feature extraction as well as machine learning so as to make a better prediction within one database. Multiple attempts and efforts were made on analysis time window, biosignal features, feature normalization, machine learning models, performance evaluation, and study design interpretation. The answers although they may be limited to the database, do however contribute some new thinking to the field.

1.5 Thesis content summary

As pain estimation is defined as a pattern recognition problem, Chapter 2 first reviews the corresponding outputs/labels and input signals. The potential labels are reviewed from the existing pain scales in clinical practice and the study designs in automatic pain assessment studies. The pain scales cover the rating scales used in self-reporting, objective observation, and nociception/antinociception (NAN) balance. The reviewed signals include ANS system based signals, other biosignals including respiration, oxygen saturation, and neuroimaging signals, as well as pain behaviors summarized from seven behavioral pain assessment tools. At the end of Chapter 2, some example devices or systems which can continuously record one or several of the mentioned biosignals are briefly introduced.

Chapter 3 provides detailed technical specifications of the devices for the SpaExp data collection and the next phase of the study. The devices include the developed biopotential acquisition system for multi-channel facial surface electromyogram (sEMG) recording [35], Bioharness 3 for ECG and respiration recording, and the eHealth v2.0 platform for GSR recording. In addition, a more advanced patient monitor and wearable GSR solution - E4 wristband from empatica (E4) - are presented.

Chapter 4 presents the study design and the SpaExp database specifications. The details of the study design are experimental protocol, pain stimulation, and biosignal measurements. The collected data and biosignals/parameters are summarized and listed.

Chapter 5 reviews, discusses, and presents the implementation of signal processing and feature extraction with the collected ECG, GSR, and sEMG signals. A total of 103 features are derived as a result of signal denoising, processing, feature extraction from sliding time windows with customized lengths, and feature rescaling.

Chapter 6 visualized sixteen of the total 103 features. They are chosen due to their weak correlation with the others. Each feature is a time series and is plotted as a line graph. In total 120 pain tests are divided into three groups according to the self-reported pain intensities. In each group, the feature time series of the tests are aligned to the time when the gradual-increasing pain stimulus started to be perceived as pain. The average line in each group is observed and compared so as to answer the three questions: 1) How does the feature react to the pain stimulus since stimulation start? 2) How does the feature react to the sensation of pain after the pain stimulus is perceived as pain? 3) Does the feature have the potential to differentiate pain intensities across subjects? At the end of the chapter, the similarity among all the representative features is checked with multidimensional scaling.

Chapter 7 introduces the final machine learning models. By analyzing and interpreting the study design, the pain intensity recognition is designed as a two-step estimation. The first step is binary classification estimating whether the pain was present or whether the pain control was inadequate as a clinical significance. The second step was an estimation of three classes of pain intensity - when the pain is present or pain control is inadequate. More than twenty models are built in both classification steps, and the top best models are selected to observe any performance change if the hyperparameters of one model are optimized. The results in each step are separately presented and discussed. Another important part of this chapter is a review of automatic pain assessment studies. The review lists the similarity and difference between the related studies and databases from four perspectives, and the difficulties in the generalization of the methods and results are discussed. Finally, Chapter 8 concludes the study.

Chapter 2

Labels and Signals in Pain Assessment

The aim of this study is to build a model mapping from non-self-reported signs of pain as a substitute for pain assessment in self-reporting or checklists. This chapter gives a comprehensive review of the methods used for pain assessment, first from the angle of assessment targets - labels of pain, and then from the perspective of observed signs of pain - potential input signals. The labels and signals indicate the learning aim and the materials for learning, respectively, when using machine learning methods.

2.1 The complex and subjective nature of pain

The perception of pain results from the interplay of the three motivational components: sensory, affective and cognitive, based on Melzack and Casey's model of pain [36]. In addition to the physical, social and psychological factors that could influence the individual pain experience, other sensory perceptions and visions could also impact the perceptual experience of pain because they are integrated into the brain and nervous system.

Due to the complex and subjective nature of pain, self-report is considered to be the gold standard for assessing pain and is widely applied to assess pain in clinical trials and pain management. Valid and reliable assessment of pain is essential in pain management decision-making. The assessment of pain location and pain intensity is sufficient for acute pain that lasts less than three months and in cases where the pain is a symptom of trauma or disease [37]. While in the cases of chronic pain and some specific diseases or pain conditions, some corresponding tests and questionnaires are designed for each case assessing the impact of pain on the patients' physical and emotional functioning as well as their life.

However, self-report is not applicable to the patients who cannot communicate verbally, in writing or by other means such as finger span or blinking eyes to answer yes or no questions. The patient groups are i) older adults with advanced dementia, ii) infants and preverbal toddlers, iii) critically ill/unconscious patients, iv) persons with intellectual disabilities and v) patients at the end of the life [24]. In clinical practice, patient behaviors are observed as an indication of pain and various behavioral pain assessment tools have been designed for each population of patients. Unfortunately, no behavioral pain assessment tool applies to all because the patient population and given context differs. The reliability and validity of a tool should be ensured in each case.

Considering the complex nature of pain as described above, the scope of the study regarding pain is narrowed down to the intensity of acute pain derived from the healthy adult volunteers, which is expected to be transferred to postoperative patients in the future and ultimately to the critically ill/unconscious adults. Although this study only includes data from experimental pain with healthy volunteers, studies with a broad scope of pain and patient groups (e.g., intensity recognition of chronic pain) are also involved in the review due to the insight they give on algorithm design and the scarcity of related works on the topic. Properly managed acute pain could prevent pain from becoming chronic [38]. It is therefore essential to deal with acute pain such as postoperative pain effectively. One important reason to only focus on acute pain is that the cause of acute pain is mainly tissue damage while for chronic pain it could also be nerve damage or a consequence of depression/anxiety. For this reason, the pain management strategies for acute and chronic pain are different. Another important reason is that the pattern of the body's response to acute pain and chronic pain could be different. For example, acute pain may elevate heart rate, blood pressure, and respiration rate, while chronic pain is not usually accompanied by behavioral changes except exacerbation [39].

The motivations for developing automatic pain monitoring systems are twofold. First, pain assessment and the adjustments on pain medication in a hospital requires routine checks so that adequate pain management is highly reliant on manpower. Second, a gap may exist between patients and healthcare providers when inferring pain in others and interpreting pain intensity scores, which could result in undertreatment or overtreatment of pain [40, 41]. In other words, automatic pain assessment could be a solution optimizing the existing pain assessment approaches and pain management in terms of efficiency, reliability, and validity.

2.2 Pain scales in clinical practice and algorithms

Pain intensity, as one of the dimensions in pain assessment, is commonly used in clinical practice to characterize pain and evaluate the effects of pain treatment. In the research on pain intensity pattern recognition, the existing pain assessment tools and concepts act as the ground truth in the system development and validation.

The two well-known pain intensity scales are the visual analog scale (VAS) [42] and the numeric rating scale (NRS) [43]. The main difference between VAS and NRS is that VAS [44] is in continuous numbers within a range between no pain and the worst imaginable pain, usually, one point on a 100 mm ruler, while NRS is a scale of integers between 0 and 10. The third scale is more coarse-grained, and is called the verbal categorical rating scale (VRS), in a four-point VRS: "mild" is equal to the numbers from 1 to 3 in NRS; "moderate" is equal to the numbers in NRS from 4 to 6; and "severe" is equal to the numbers larger than 6 in NRS. Among these three scales, NRS and VAS are considered to be the most practical to use. In contrast, pain can be underestimated when using VRS [37]. In addition to the descriptive words "mild," "moderate," and "severe," other words such as "unbearable", "intense pain" and "maximum pain" are also found in the literature to describe pain intensity [45, 46].

For the patients who are unable to self-report, the pain scores of some behavioral tools are derived from adding up the scores of several indicators. Therefore, the range of the scale could be different from the standard pain intensity rating. In this case, the scores in different scales are incomparable. For example, in the Critical-Care Pain Observation Tool (CPOT) [45], facial expression, body movements, muscle tension and compliance with the ventilator/vocalization are scored separately from 0 to 2, so the range of the overall score is between 0 and 8. Another similar example is in the UNBC-McMaster shoulder pain expression archive database [26]. The score in Prkachin and Solomon Pain Intensity Metric (PSPI) [47] is the sum of the intensities of four action units (AUs) each intensity ranging between 0 and 4. In the same database, ratings in two other scales, VAS (0-10) and observed pain intensity (0-5), are documented. UNBC-McMaster database is one of the open access databases for researchers to develop automatic pain assessment systems.

In addition to the description of pain intensities introduced above, the two terminologies relating to both the pain experience and the pain-inducing stimulus are *pain threshold* and *pain tolerance*. They are commonly found in pain sensitivity studies or/and the studies involving experimental pain stimulation such as electrical stimulation and thermal stimulation [28, 33, 48, 49]. According to the definitions by IASP, *pain threshold* is "the minimum intensity of a stimulus that is perceived as painful", and *pain tolerance* is "the

maximum intensity of a pain-producing stimulus that a subject is willing to accept in a given situation." The notations following the definitions also emphasized that a *pain threshold* and *pain tolerance* are the subjective experiences of the individual rather than the stimulus level itself. The definitions of *pain threshold* and *pain tolerance* are separate from the definition of pain intensity scales. However, the *pain threshold* is considered as the moment when the VAS exceeds 0 [50] and just reached 3 or 4 [33] respectively for the first time during a continuous increase of stimulation intensity.

Another perspective to pain intensity assessment could be NAN balance [51]. However, the index measuring it is mainly applied to cases under general anesthesia, e.g., in an operation, during emergency and postoperative care.

2.3 Signals in pain intensity recognition

Pain is a subjective feeling generated by the brain to protect ourselves. However, pain observations tools, pain intensity recognition algorithms coded from measurable signals, and pain reported by parents or family members are proxy measures and are objective in essence. Although the current technologies already enable us to observe brain activities and the change of many physiological signals, assessing pain objectively is still challenging from several aspects. On one hand, the body responds to pain through numerous and interconnected physiological processes involving multiple body systems including cardiovascular, respiratory, immune, endocrine, gastrointestinal, urinary, musculoskeletal, nervous and brain [39]; nevertheless, there is no one signal(s) reflecting all the responses; On the other hand, each measurable signal does not only react to pain. To ensure the reliability of a developed pain assessment tool, an ideal one should be sensitive and specific to pain. It needs to be observer-independent, not reliant on the patient's ability to communicate and not influenced by disease characteristics [5]. Therefore, the fusion of multiple signals is one trend in developing automatic pain assessment methods.

There are several rationale dimensions found in the literature that automatic pain assessment tools or nociception/antinociception balance tools have. One of them is based on the assumption that pain induces alterations in the sympathetic nervous system [52]. The sympathetic nervous system is one of the two main divisions of the ANS which regulates the body's unconscious actions. The sympathetic nervous system stimulates the body to "fight-or-flight", while the other main division of ANS, the parasympathetic nervous system stimulates the body to "feed and breed" and "rest and digest". Another approach to recognizing pain is to recognize part of pain behaviors such as facial expressions of pain [26]. A third dimension is visualizing the brain in pain through neuroimaging, where both physical

and cognitive influences on pain can be directly observed [53–55]. A fourth dimension could be biomarkers at cellular level measuring genetic or protein responses, or metabolic products [5]. The developed pain assessment tools usually cover one or several of the first three dimensions. These tools and their signals are reviewed in the following part of this subsection. Besides the signals within the dimensions mentioned above, two physiological parameters, respiration and blood oxygen saturation are also introduced below because they are both found being discussed in pain assessment studies and breathing techniques are used to alleviate pain.

2.3.1 Autonomic nervous system (ANS) based signals

The literature showed that many existing NAN balance tools are based on analyzing the reactions of ANS during general anesthesia. Even though there are already several NAN balance measurement commercialized solutions, no gold standard exists [4]. The ANS-based signals used in the NAN balance tools are also commonly found in a bedside monitor such as blood pressure, heartbeat interval and photoplethysmogram (PPG). PPG is also referred to as blood volume pulse (BVP). However, it is different in the case of pain assessment of patients who are unable to self-report as the ANS-based signals are poor in specificity when differentiating pain from other sources of distress. These indexes could be influenced by numerous other physiological and psychological conditions as well such as age, co-morbidities, depth of anaesthesia, surgical stimulation, medications, and emotions. Moreover, the absence of a change in vital signs does not indicate the absence of pain. For these reasons, the ANS-based signals are not recommended in assessing the pain of patients who are unable to self-report in clinical practice [24]. In spite of this, ANS-based signals, especially the fusion of them, are active in the discussion of automatic pain assessment studies [28, 30, 33, 56].

Inter-beat interval

The signals reflecting ANS activities are listed in Table 2.1 together with some NAN balance indexes. Among them, the pulse interval is commonly used in the NAN balance models either as original or transformed parameters. With the help of heart rate variability (HRV) analysis, the sympathetic activity is traceable from the sympathetic cardiac regulation and the same applies to parasympathetic activity. The inter-beat intervals are extracted from electrocardiography mostly and from PPG in some cases. The normal inter-beat intervals are called normal-to-normal (NN) series.

Further analysis, which is called HRV analysis, can be made within a time period in the time domain, frequency domain and in other forms. When HRV is used to trace sympathetic or parasympathetic activity, the analysis is usu-

Table 2.1: The biosignals reflecting ANS activities and some NAN indexes [4, 5]

Signal	Index name	Parameter(s)	Measure	Model
Pupil dilation	pupillometry	pupillary diameter and the light-induced pupillary dilation reflex	sympathetic activity	-
PPG	Surgical pleth index (SPI)	pulse wave amplitude (PPGA) and heart beat interval (HBI) number & amplitude of SCR fluctuations (NSCF & ASCF)	sympathetic activity	$SPI=100-(0.7 \times PPGAn_{norm} + 0.3 \times HBI_{norm})$
GSR	skin conductance		sympathetic activity	painful stimuli induce an immediate increase in peaks per second
ECG	Analgesia/Nociception index (ANI)	0.15Hz-0.4Hz bandpassed RR series _{norm}	parasympathetic activity	$ANI=100 \times (5.1 \times AUC_{min} + 1.2) / 12.8$, where AUC is the area between the envelope of local maxima and local minima
ECG and continuous blood pressure	CARDiovascular DEpth of ANalgosia(CARDEN)	RR under curve and systolic blood pressure under curve	sympathetic activity	a minor elevation in blood pressure followed by minor tachycardia

ally within less than 5 minutes or several tens of seconds depending on the study design. Among the HRV features, NN50, the number of NN interval differences that larger than 50 ms and HRV-the high-frequency power component of NN series (0.15-0.40 Hz), the high-frequency power component of NN series (HF: 0.15 to 0.40 Hz), are considered as indices of parasympathetic activity. In addition, the ratio of LF/HF (LF: 0.04 to 0.15 Hz) is an index of sympathetic activity [57]. As Table 2.1 shows, HF is the primary parameter to observe in an analgesia/nociception index (ANI) [58].

PPG amplitude

In addition to the inter-beat intervals, the PPG amplitude (PPGA) is also part of the surgical pleth index (SPI) or surgical stress index (SSI) by another name [59, 60]. The up or down of PPGA can be regulated by anesthesia, sympathetic activation, arterial blood pressure (ABP) increase and some other factors. The sympathetic activation could lower PPG, while anesthetics could increase PPGA, and the rise of ABP may lead to either PPGA up or down due to different causes [61].

Blood pressure

In the CARDEAN index [62], the change of continuous systolic blood pressure (SBP, in mmHg) is taken into account along with the change of NN series. Meanwhile, it has been observed that pain can raise blood pressure and therefore an increase in systolic blood pressure is considered one of the signs indicating intraoperative nociception [51, 63].

Electrodermal activity/Galvanic skin response

Electrodermal activity (EDA) monitors resistance variations in the skin due to the autonomic activation of sweat glands in the skin. As EDA reflects activity only within the sympathetic axis of the autonomic nervous system, it is widely used in indexing emotional processing and sympathetic activity [64]. EDA is also referred to as skin conductance, or GSR and GSR will be the term used throughout the thesis. GSR is recorded as the conductance in micro-Siemens (μ S) between two measuring points on hand palms, fingers, or foot soles. The GSR signal has two components to show change: one is in the tonic level, changing within tens of seconds to minutes; the other one is the phasic response on top of the tonic changes where obvious peaks can be observed within one or several seconds. The former component is the skin conductance level (SCL). The latter component is called skin conductance response (SCR), which is more informative and is found sensitive to emotionally arousing stimulus events [65]. Similarly, SCR is also emphasized in pain assessment studies, for example, as peaks per second [66].

2.3.2 Respiration and oxygen saturation

Respiration

The signals introduced above are regulated unconsciously by the body, while respiration is different because it can be adjusted manually. Breathing techniques such as slow breathing and paced slow deep breathing are commonly applied to alleviate pain as a routine procedure in the hospital, although the physiological mechanisms behind it are not fully known yet. On the other hand, respiratory parameters are also taken as potential indicators of pain and were observed in both experimental and clinical studies. According to Jafari H. *et al*'s review of pain and respiration [67], the respiratory pattern does change along with a painful procedure or pain relief. For example, inspiratory flow is found increased in experimental studies due to sudden cutaneous pain. The process of respiration consists of inspiration (or inhalation) where air rushes into the lungs and expiration (or exhalation) where the air is forced out. Inspiratory flow is the ratio of inspiratory volume over inspiratory time. The increase of inspiratory flow in response to sudden cutaneous pain is concluded to be similar to the startle reflex. This finding is summarized to be consistent in the cases of sustained pain in experimental studies and painful procedures in clinical studies, where increased ventilation or hyperventilation is observed as a result of deeper breathing, faster breathing (respiration rate increase) or a combination of both. However, these observations are concluded from the mean value out of a group of subjects, and may not be consistent across individuals.

Oxygen saturation

In medicine, oxygen saturation is the fraction of oxygen-saturated hemoglobin relative to total hemoglobin in the blood. Peripheral oxygen saturation (SpO_2) can be derived from PPG, and it is an approximation to arterial oxygen saturation (SaO_2). Normal SpO_2 is between 95% and 100%. The observations are inconsistent among the related studies with different patient populations and pain scenarios. SpO_2 was observed to decrease in number or/and increase in variability during a painful procedure in some studies [68–70], while it was not found to be useful in some other studies [71, 72].

2.3.3 Behavioural signals

...However, the appropriateness of a tool must be assessed patient by patient, and no one tool should be an institutional mandate for all patients. For example, a behavior pain tool developed for persons with dementia may not be appropriate for patients in the ICU who are unable to communicate, and tools for children are not generalizable to adults. [24]

There are many behavioral pain assessment tools in use for different patient populations. As critically ill/unconscious adults in ICU are the patient population this study will eventually serve, the behavioral pain assessment tools that can be applied to critically ill/unconscious adults according to [24], together with a tool validated in [22] are reviewed below. The tools are: 1) BPS: Behavioral Pain Scale [17], 2) CPOT: Critical-Care Pain Observation Tool [73], 3) FLACC: Face, Legs, Activity, Cry, and Consolability Behavioral Assessment Tool [18], 4) PBAT: Pain Behavioral Assessment Tool [19], 5) NPAT: Nonverbal Pain Assessment Tool [20], 6) NVPS: Nonverbal Pain Scale [21], and 7) BPAT: 8-item Behavior Pain Assessment Tool [22].

Each tool observes behavioral signs from at least three perspectives. These are facial expressions, movements, and vocalization/compliance with ventilation. There are two other categories not rigidly fitting into these three parts. Following the principle of proximity, the emotion category in NPAT is merged into facial expressions in the review below and the physiologic category in NVPS is discussed with vocalization/compliance with ventilation. In most of the tools, the output is a final score from a sum of sub-scores from each category. In each category/indicator/item, the sub-score is defined by either the presence or degree of the behavior. A larger score of each tool may indicate higher pain intensity, however, the score number and pain intensity number are not highly correlated. For example, it is reported that BPAT showed a moderate ability to discriminate severe levels of pain intensity ($NRS \geq 8$), and there is a moderate correlation between pain distress and behavioral scores during common procedures performed in ICU patients which also supports the interrelation between the affective and behavioral dimensions of pain [22].

Facial expressions

From the descriptions of facial expressions summarized in Table 2.2, it can be seen that the minimum score is consistently described as *relaxed* or *calm*. However, the descriptions for the middle score(s) and the maximum score demonstrate some differences across the tools, for example, in terms of the target facial expressions. In tools such as BPS, CPOT, and NPAT, the presence of some facial expression determines the score number, while in FLACC and NVPS, while in FLACC and NVPS, the concern is more about the frequency of shown facial expressions. The most frequent description shown in Table 2.2 is *grimacing*, which is followed by *frown* and *eyelid closed*. In BPAT, the involved AUs are listed for each behavior. AUs are defined in the Facial Action Coding System (FACS), which describes all visually distinguishable facial activity from 44 AUs [74], and each AU is dominated by one or two facial muscles. FACS is an observational coding scheme serving both manually and computer-based automated coding. In BPAT, grimace is

Table 2.2: The description on the facial expressions/emotion in the behavioural pain assessment tools

	Min score	Middle score(s)	Max score
BPS	1- Relaxed	2- Partially tightened (e.g., brow lowering) 3- Fully tightened (e.g., eyelid closing)	4- Grimacing
CPOT	No muscular tension observed	Presence of frowning, brow lowering, orbit tightening, and levator contraction	All of the above facial movements plus eyelid tightly closed
	0- Relaxed, neutral	1- Tense	2- Grimacing
FLACC	0- No particular expression or smile	1- Occasional grimace, frown, withdrawn or disinterested	2- Frequent to constant frown, clenched jaw, quivering chin
NVPS	0- No particular expression or smile	1- Occasional grimace, tearing frown or wrinkled forehead	2- Frequent grimace, tearing, frown or wrinkled forehead
NPAT (Emotion)	0- Relaxed, calm expression 0- Smiling, calm, relaxed, exhibiting no emotion	1- Drawn around mouth and eyes; tense 1- Anxious, irritable, withdrawn, closes eyes, does not engage with physical environment	2- Facial frowning, wincing, grimacing 2- Tearful or uncooperative
PBAT	Grimace, frown, wince, eyes closed, eyes wide open with eyebrows raised, looking away in opposite direction of the pain, grin/smile, mouth wide open to expose teeth and tongue, clenched teeth exposing slightly open mouth, unable to assess, other		
BPAT		Neutral expression, grimace, wince, eyes closed	

defined as a combination of AU4 *brow lowering*, AU6 *cheek raising*, AU7 *lid tightening*, AU20 *mouth stretching* and AU43 *eye closing*.

Table 2.3: The facial action units involved in pain facial expressions in adults

Facial action unit	Muscular basis	Description/Scale	Grimace	Wince	PSPI	CPOT
AU4 brow lower	Corrugator supercilii	✓		✓	✓	
AU6 lids tighten	Orbicularis oculi	✓	✓	✓	✓	
AU7 cheek raise	Orbicularis oculi	✓	✓	✓	✓	
AU9 nose wrinkle	Levator labii superioris			✓	✓	
AU10 upper lip raiser	Levator labii superioris			✓	✓	
AU12 lip corner pull	Zygomaticus major					
AU20 horizontal mouth stretch	Risorius	✓				
AU43 eyes closed	Relaxation of Levator palpebrae superioris, Orbicularis oculi	✓		✓	✓	

The FACS and its muscular basis are the theoretical foundations of automatic emotion or pain recognition from facial expressions, as FACS provides the possibility of quantitative measures on facial expressions. The AUs involved in pain facial expressions among adults are summarized in the review [75] and are listed in Table 2.3 together with some descriptions of pain facial expressions and the PSPI metric. In the PSPI metric, pain intensity is mapped from the intensity (0=absent, 5=maximum) of three AUs and the binary intensity of AU43, which is calculated based on Equation 2.1.

$$\text{Pain intensity} = \text{Intensity(AU4)} + (\text{Max intensity AU6 or AU7}) + (\text{Max intensity AU9 or AU10}) + \text{Intensity(AU43)} \quad (2.1)$$

As Table 2.3 shows and according to Prkachin and Solomon [47], the four actions, *brow lowering* (AU4), *orbital tightening* (AU6 and AU7), *levator contraction* (AU9 and AU10) and *eyes closed* (AU43) are the core actions that carry most of the information about pain. However, the use of AU43 needs to be further considered because *eyes closed* could be the second most frequently observed facial expression among sedated patients at rest after *neutral*. Therefore, AU43 may not be effective in indicating the presence of pain in non-communicative critically ill patients. This may explain why the description of *eyelid tightly closed* is used instead in CPOT.

The PSPI metric is implemented mainly in the UNBC-McMaster video sequences database for researching automatic pain facial expression recognition. In the database, the video frames were coded manually with AUs and were labeled with both self-report VAS scores and observer ratings. Meanwhile, models can be built to detect facial movements automatically so that the pattern of each AU and its intensity could be recognized in order to

calculate PSPI. Another approach to recording facial expressions is facial sEMG, where surface electrodes are placed on the muscle area of interest to capture the electric potential generated by muscle cells during muscle contraction. In the detection of pain facial expressions, the facial muscles listed in Table 2.3 are the muscle areas of interest.

Movements

Compared to facial expressions, the observed perceptions of how the body reacts to pain are more versatile in behavioral pain assessment tools. According to the descriptions on movements in Table 2.4, the body's reactions can be summarized into three types, the first type is voluntary movements (e.g., legs movements in FLACC and body movements or activity in CPOT, NVPS and NPAT); the second type is reactions to passive movements (e.g., muscle tension in CPOT and guarding in NPAT); the third type is the posture or the static state of the body described, for example, *rigid, clenched fists, and fetal position*. To date, studies on automatic pain assessment by recognizing pain movements are rarely found in the literature. However, hand and body gesture recognition using a camera has been under development during the past decades for human-computer interaction (e.g., [76] and [77]). An endeavor was especially made in [76] to recognize the pattern of emotions with a fusion of facial expression, body gesture, and acoustic analysis, which could be used as a reference to pain assessment studies in the near future.

Vocalization/Compliance with ventilation

In ICU, the patients could be intubated or mechanically ventilated and the equipment does not allow them to make verbal sounds. In this case, compliance with the ventilator or not is observed instead of vocalization, which is described as the level of patient-ventilator asynchrony or dyssynchrony as Table 2.5 shows. Descriptions such as *tolerating movement* and *coughing but tolerating* are also found in the minimum and middle scores of BPS and CPOT. In terms of vocalization, the middle scores are described as *sighing, moaning and whimpering*, while the maximum scores are described including *sobbing, crying out and screaming*. In addition, *verbal complains of pain* is also considered.

Among the discussed assessment tools, NVPS is the only one considering physiological signs, and it covers most of the signals mentioned in Section 2.3.1 and Section 2.3.2. The relative changes on vital signs and respiratory parameters are observed, either within a long time-window of 4 hours or comparison with the defined baseline.

Table 2.4: The description on the movements/guarding/muscle tension in the behavioural pain assessment tools

	Min score	Middle score(s)	Max score
BPS (Upper limbs)	1- No movement	2- Partially bent 3- Fully bent with finger flexion	4- Permanently retracted
CPOT (Body)	Does not move at all (<i>does not necessarily mean absence of pain</i>)	Slow, cautious movements, touching or rubbing the pain site, seeking attention through movements	Pulling tube, attempting to sit up, moving limbs/thrashing, not following commands, striking at staff, trying to climb out of bed
(Muscle tension)	0- Absence of movements No resistance to passive movements 0- Relaxed	1- Protection Resistance to passive movements 1- Tense, rigid	2- Restlessness Strong resistance to passive movements, inability to complete them 2- Very tense or rigid
FLACC (Legs) (Activity)	0- Normal position or relaxed 0- Lying quietly, normal position, moves easily	1- Uneasy, restless, or tense 1- Squirming, shifting back and forth, or tense	2- Kicking, or legs drawn up Arched, rigid or jerking
NVPS (Activity)	0- Lying quietly, normal position	1- Seeking attention through movement of slow cautious movements	2- Restless activity and/or withdrawal reflexes
NPAT (Body) (Positioning /Guarding)	0- None, sleeping comfortably 0- Relaxed body	1- Restless or slow, decreased movement 1- Guarding, tense	2- Immobile, afraid to move or increased motion 2- Fetal position, jumpy when touched, withdraws when touched
PBAT	No movement, rigid, arching, clenched fists, shaking, withdrawing, splinting, flailing, picking/touching site, restlessness, Rubbing/massaging, repetitive movements, defensive grabbing, pushing, guarding, unable to assess, other		
BPAT		Rigid, clenched fists	

Table 2.5: The description on the vocalization/compliance with ventilation/consolability/physiological signs in the behavioural pain assessment tools

	Min score	Middle score(s)	Max score
BPS (compliance with ventilation)			
1- Tolerating movement	2- Coughing but tolerating ventilation for the most of the time	4- Unable to control ventilation	
CPOT (compliance with the ventilator)			
Alarms not activated, easy ventilation	Alarms stop spontaneously	Asynchrony: blocking ventilation, alarms frequently activated	
0- Tolerating ventilator or movement	1- Coughing but tolerating	2- Fighting ventilator	
Talking in normal tone or no sound	Sighing, moaning	Crying out, sobbing	
0- Talking in normal tone or no sound	1- Sighing, moaning	2- Crying out, sobbing	
FLACC (Cry)			
0- No cry	1- Moans, whimpers, or occasional complaint	2- Crying steadily, screams or sobs, frequency complaints	
NVPS (Vital signs)			
0- Stable vital signs, no change in past 4 hours	1- Change over past 4 hours in any of the following: SBP>20, HR>20, RR>10 1- RR>10 above baseline or SpO ₂ decrease 5%	2- Change over past 4 hours in any of the following: SBP>30, HR>25, RR>20 2- RR>20 above baseline or SpO ₂ decrease 20%	
(A. Respiratory, compliant with ventilator)			
Complaint with ventilator	Mild asynchrony with ventilator	Severe asynchrony with ventilator	
(B. Physiological signs)			
NPAT	0- Warm, dry skin	1- Dilated pupils, perspiring, flushing 2- Diaphoretic, pallor	
PBAT	0- Intubated, no verbalization	1- Whimpering, moaning, sighing 2- Screaming, crying out	
BPAT		Moaning, screaming, whimpering, crying, using protest words, verbal complaints of pain, none, unable to assess, other	
		Moaning, verbal complains of pain	

2.3.4 Neuroimaging signals

It is natural to think about assessing pain directly from brain as it dominates all the dimensions of pain processing. However, due to the nature of each signal, there are limitations on either ability constraints or device size and expense that impede it from being studied and validated in a large scale. In each study, the pattern of pain is learned from a small group of subjects with chronic pain or healthy volunteers with experimental pain stimulus (e.g., [78] and [79]). As is concluded in [55], among the functional imaging techniques, the hemodynamic methods positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), are utilized to ascertain specific points of cerebral activation due to a better spatial resolution. While the electrophysiological methods, electroencephalogram (EEG) and magnetoencephalography (MEG), are often used to observe the brain's temporal response to pain due to their good temporal resolution.

2.4 Wearable devices and IoT-enabled systems

Although the potential indicators of pain have been enumerated in Section 2.3, not all of them are applicable in continuous pain monitoring. For example, most of the neuroimaging methods are expensive with cumbersome machines which cannot be used every day and by patients in beds with various tubes from other machines. Exceptionally, the EEG and near-infrared spectroscopy methods have wearable solutions where a sensing device is made into a headset or a cap.

Regarding the behavioral signals, they could be videoed and recorded as they can all be checked visually. However, as they are measured indirectly from the video, computer vision techniques are needed to recognize each sign of pain. Moreover, videos, especially those including the facial area, violate the privacy of the patients in hospital. The existing data sources of facial expression videos are mainly from healthy people or people with chronic pain rather than patients in hospital. By contrast, the behavioral signs of pain other than facial expressions are rarely found in automatic pain assessment studies. In addition to the video recording, behavioral signs of pain could also be determined from other sensing methods with small and wearable sensing devices. For example, the muscle tension and facial muscle contraction can be recorded with sEMG; body movements and body position could be calculated from a set of accelerometers and gyroscopes when placed in proper positions. The development of a compact facial sEMG device is presented in this study with details in Chapter 3.

Compared to the rest signs of pain, most of the physiological signals are much easier to record for later analysis. Many of them, including ECG, PPG, blood pressure, respiration and oxygen saturation, are vital signs which are

routinely monitored by a bedside monitor in a recovery room or an ICU. Meanwhile, such wearable devices already exist in the market to monitor ECG, PPG, respiration, oxygen saturation or GSR. One key point the researchers in this study concern is that whether the data at original sampling rate is accessible. From the data accessibility point of view, some devices and systems are listed as examples in Table 2.6. E4 wristband and Bioharness 3 chest strap are wearable devices measuring multiple physiological parameters. Both of them support real-time signal waveform display and on-device signal recording. The other two systems in the table are bedside or patient monitors used in hospitals. IntelliVue MMS X2 represents the bedside monitors with comprehensive functions. In terms of noninvasive blood pressure, it could only be measured discretely every 15 minutes whereas CNAP® can provide continuous monitoring.

Table 2.6: Example devices and systems providing raw data for research

Device	Sensors	Platform
E4 wristband (E3 version [80])	PPG, GSR, 3-axis accelerometer, infrared thermopile, internal real-time clock.	Recording: wristband-computer-cloud Streaming: wristband-mobile device-cloud
Bioharness 3 device & chest strap (version 1 [81, 82])	ECG Lead I, respiratory cycles, 3-axis accelerometer.	device-computer
CNAP®monitor	double finger sensor, for monitoring continuous noninvasive arterial pressure (CNAP)	
IntelliVue MMS X2 patient monitor	up to 12-Lead ECG, Noninvasive blood pressure, Invasive arterial blood pressure, PPG, etc.	

Similar to the central patient monitoring systems in hospitals, wearable devices can be integrated into a remote monitoring system within an Internet-of-Things (IoT) architecture. As shown in the example described in [83], the architecture could be in three layers. The bottom layer contains smart devices with the wireless communication function. The middle layer contains gateways and each gateway acts as a bridge between a group of smart devices and the top cloud layer. Finally, the central and remote monitoring is realized through cloud services.

Two demo systems concretizing the proposed architecture were designed and implemented in [32] and [84]. Figure 2.1 presents the remote pain monitoring idea, where the biosignals are transmitted through an IoT network and can be accessible to caregivers at a distance after being interpreted. On the basis of the three-layer IoT architecture especially, a mobile web application that can run inside the web browser of most operating systems was devel-

oped in [84] for caregiver end-users. In the developed web application, the biosignal data stream can be synchronized from the cloud server, processed and presented as signal waveforms.

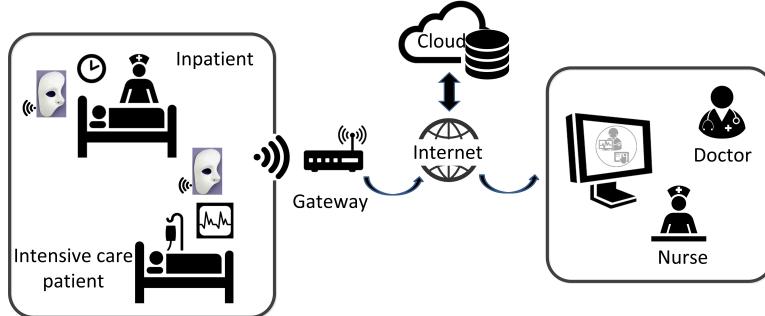


Figure 2.1: Remote pain monitoring in hospital for inpatients and ICU patients © 2018 IEEE

The methods, results, and discussions to be presented in the following chapters are about how accurate and reliable the automatic pain assessment could be from the point of view of the signal quality, processing algorithms, and the inherent data pattern. These aspects are considered to define the boundaries of the application's feasibility. Furthermore, merging the application into IoT systems to realize real-time monitoring and assist decision-making in pain management is also worthy of exploration. On the one hand, IoT systems can provide flexible processing capabilities from the resource-constrained bottom sensor layer to the less resource-constrained top cloud computing layer that can support deep learning streaming [85]. On the other hand, challenges pertaining to technology as well as regarding human factors are to be solved before the concepts could turn into reality [86].

2.5 Summary

This chapter presents the rationale for research on automatic pain assessment tool development using machine learning methods, mainly supervised learning methods. The review covers pain indexes as scales, checklists, and automatic measures in a broad scope. The following presents the signal selection process from the angle of wearable devices and IoT-enabled systems. In the end, the vision of remote pain monitoring within an IoT architecture was introduced, where using wearable devices is one of the original motivations when this study was initialized. The remote pain monitoring could be an application in the future.

Chapter 3

Biosignals Acquisition

In this study, a set of the key signs of pain are observed from healthy volunteers. Among the signs introduced in Section 2.3, facial sEMG, ECG, respiratory cycle and GSR are included in the data collection and analysis. In this chapter, the chosen or developed biosignal acquisition systems are presented and discussed.

3.1 Facial surface electromyography

Facial sEMG is an alternative approach to face images or videos in reading facial expressions. Compared to face images or videos, it conceals facial identity so that privacy is protected. However, on the other hand, the development of its technologies and applications is not as fast as the video approach at this stage due to its comparatively more intrusiveness and the resulting small database. Another challenge in generalizing facial sEMG data is the difference in the different electrode placements covering different muscle areas.

Most of the sEMG studies follow Fridlund and Cacioppo's guidelines [19] on electrode placements where electrode pair positions for ten muscle areas on one side of the face are recommended. The guidelines also recommended miniature electrodes for facial sEMG sites with 0.25 cm diameter silver (Ag)/silver chloride (AgCl) detection surfaces and 1 cm inter-electrode spacing, because facial muscles are small muscles compared to the big trunk muscles. However, the most commonly used disposable pre-gelled surface electrode H124SG has 1 cm diameter detection surface and 2.4 cm total diameter with solid gel and adhesive collar. Smaller electrodes exist, such as 0.1 cm EEG cup electrodes, however, they are expensive to dispose off and need extra conductive paste and tape to fix in place. In the measurement, monopolar electrode configuration is adopted instead of bipolar where in each channel on electrode is placed on the target muscle. It is to cover more muscular areas with fewer electrodes.

The multi-channel sEMG measurement, Table 2.3, shows that at least five facial muscles contribute to pain facial expressions, these are: *corrugator supercilli*, *orbicularis oculi*, *levator labii superioris*, *zygomatic major* and *risorius*. However, existing commercial wearable or wireless sEMG devices each can only measure up to two channels (CHs) (e.g., Shimmer3 EMG unit, 2-CH, @512 samples per second (SPS) per channel; Delsys Trigno™Wireless EMG sensor, 1-CH, @2000 SPS per channel; BIONOMADIX-EMG, 2-CH, @1000 SPS per channel). Therefore, in this study, we also sought to find a compact solution to multi-channel sEMG measurement. On the basis of the device, a wearable facial sEMG sensor can enhance the device's easiness-to-use.

In summary, the aims of the facial sEMG acquisition system design were:

- To have one electrode per muscle area to reduce measurement intrusiveness;
- To develop a compact multi-channel sEMG measurement wireless device with wearable solutions;
- To provide a continuous recording of at least 5-CH sEMG for at least 2 hours.

The designed sEMG acquisition system and its test results are presented in the following parts.

3.1.1 The core of the acquisition system

The developed facial sEMG system was built on a Texas Instruments low-noise analog-to-digital converter (ADC) for biopotential measurements, ADS 1299[87]. ADS 1299 is 24-bit ADC with a flexible configuration on analog amplifier gain (up to 12), sampling rate (250 - 16k SPS), analog inputs and outputs. However, this flexibility complicates its configuration of analog inputs and outputs on the other hand. In addition, some terms (e.g., reference or reference electrode) may refer to different settings in different contexts, and some different terms refer to the same setting. Therefore, the analog configuration of ADS1299 and its comparison with other sensors or systems are introduced below.

Monopolar electrode configuration, single-ended measurement, or referential montage

In a biopotential measurement, especially multi-channel measurement, the electrical potential on each site can be measured by the reference of the same measurement site or an individual paired measurement site. The three terms in this part all describe the first setting of the two mentioned. *Monopolar* is used to describe a electrode configuration, where a single electrode is placed on the skin above the muscle [88]. *Single-ended* is used to describe the analog

inputs to a differential amplifier where the voltage between the input signal and the ground signal is measured. *Referential montage* is mainly used in the context of EEG measurement. This mode reduces the number of connectors used because only one physical connection is required to one input. However, the common point should be carefully chosen to avoid measurement errors. Meanwhile, the signals captured in this mode could be more susceptible to external noise as the external influence coupled to the two separate wires could be non-identical and thus cannot be canceled at a differential amplification [88, 89].

Bipolar electrode configuration, differential measurement, or sequential montage

The terms in this part respectively correspond to the three in the former part. In sEMG measurement, the terms mean the voltage is measured from the electrode pair on the skin above the muscle. *Differential measurement* is more commonly employed compared to *single-ended measurement* due to a better spatial resolution and a better signal-to-noise ratio (SNR) in the measurement by suppressing the common mode noise. However, even the high common-mode rejection ratio (CMRR) of a differential amplifier entirely suppress power line noise due to the impedance mismatches in a pair of electrodes and lead wires. Such mismatches convert common-mode noise voltage into differential interference in the measurement, and thus more solutions occur [1].

Ground, reference, bias drive and right leg drive

These terms are about the methods to reduce the captured interference in recording small biopotential signals; therefore some interference mechanisms caused by the electric field are introduced first.

As the model in Figure 3.1 shows, when the skin-electrode impedance A and B is denoted as Z_{ea} and Z_{eb} and the currents induced in the wires are i_a and i_b , the differential interference voltage caused by the interference currents in to the measurement cables is $V_{ab_cable} = i_a Z_{ea} - i_b Z_{eb}$, which could be $120\mu\text{V}$ (when $i_a \approx i_b = 6\text{nA}$ with 9 m cables, $Z_{ea} - Z_{eb} = 20\text{k}\Omega$). The interference V_{ab_cable} can be minimized by shielding the leads and grounding each shield [2].

Another differential interference caused by the electric field is because of the common-mode voltage on the body generated by the current flows from the power line through the body and ground impedance. This common-mode voltage V_{cm} equals to $i_i Z_G$ in Figure 3.1 and could be 10mV ($0.2\mu\text{V} \times 50\text{k}\Omega$) and even larger [2], and $V_{ab_body} = V_{cm}(\frac{Z_{in}}{Z_{in}+Z_{ea}} - \frac{Z_{in}}{Z_{in}+Z_{eb}}) = V_{cm}(\frac{Z_{eb}-Z_{ea}}{Z_{in}})$ because Z_e is much less than input impedance Z_{in} . A typical

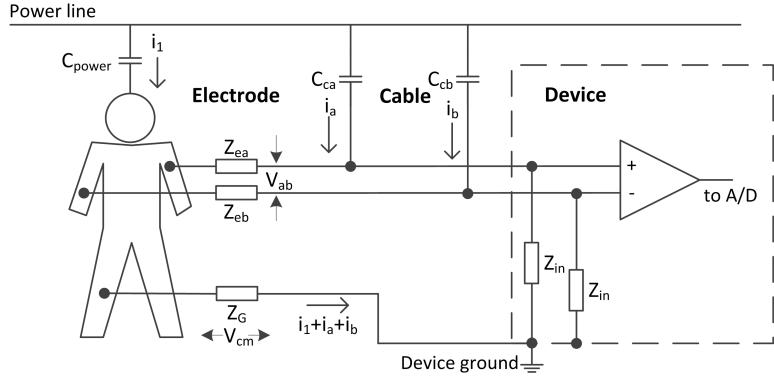


Figure 3.1: Mechanisms of difference interference caused by the electric field in ECG measurement [1, 2]

value of V_{ab_body} is $40 \mu\text{V}$. Therefore, to depress power line interference, lowering input imbalance and raising input impedance are two critical factors because the common-mode voltage is always present.

It is also possible to diminish V_{cm} by adding a driven right leg circuit instead of grounding the patient in ECG measurement, as shown in Figure 3.2. The negative feedback provided by the driven right leg circuit drives the common-mode voltage to a low value, in which way the interference is reduced [1, 2, 90]. Meanwhile, it forces the patient's common voltage close to a dc voltage level (tied to the ground in Figure 3.2) so as to maximize the input dynamic range.

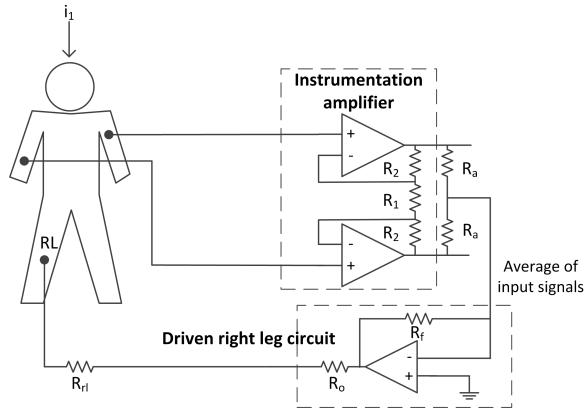


Figure 3.2: Driven right leg circuit to decrease common-mode voltage

A *reference* lead usually appears in the analog sensing part of a biopotential acquisition system. In the monopolar mode, the *reference* lead is connected to the negative inputs of all the channels so that the biopoten-

tial of each channel is measured as the difference between the positive input electrode and the reference electrode. *Reference* could also have a different meaning, especially in a bipolar system, where the potential at each electrode in an electrode pair is first detected with respect to the reference electrode [91, 92]. In this way, the same alternating current (AC) and direct current (DC) noises are subtracted before the differential amplification in the next stage so as to improve CMRR. Such *reference* is also referred to as *ground* in the systems. In addition to these, *reference* could mean the *right leg drive* in Figure 3.2, which is called *bias drive* in a system based on ADS1299 (e.g., Shimmer EMG [93]).

Active electrode and preamplifier

The *active electrode* is used to distinguish from the passive electrode. A passive electrode is the metal discs connected to the amplification circuit through a long and unshielded lead wire. The impedance of a differential amplifier is designed to be high in order to reach a high ratio of amplifier input impedance to electrode impedance in capturing the small biopotential. The capturing sensitivity, on the other hand, makes the unshielded lead wires susceptible to power line noise and any movement artifact (skin-electrode movement and cable movements). One solution to this problem is to place the differential amplifier to the electrode as close as possible [94], which is called the *active electrode* especially when the biopotential is amplified locally on the electrode site. As the output impedance of the amplifier could be as low as 10Ω , the noises have a much smaller influence on the subsequent lead wire. Most of the active electrodes are reusable (e.g., TrignoTMsensors and g.GAMMAsys). In the cases where disposable electrodes are used, the term *preamplifier* is used instead of the *active electrode* (e.g., in Motion Lab Systems and ME6000 biomonitor).

The analog configuration of ADS1299

ADS1299 was designed for biopotential acquisition, especially for EEG applications. Its working mode can be switched between monopolar and bipolar electrode configuration by changing its configuration in firmware. The electrode connection should also be changed accordingly when switching the analog input mode. Driven right leg circuit is integrated on the chip and the *bias drive* can be taken into use or kept separate from the amplification circuit in firmware configuration. Part of the ADS1299 configurations are presented in Figure 3.3, including lead-off detection, common reference, and bias drive.

The analog configurations of ADS1299 in different modes are presented here as examples of ECG lead I measurement (the potential between left

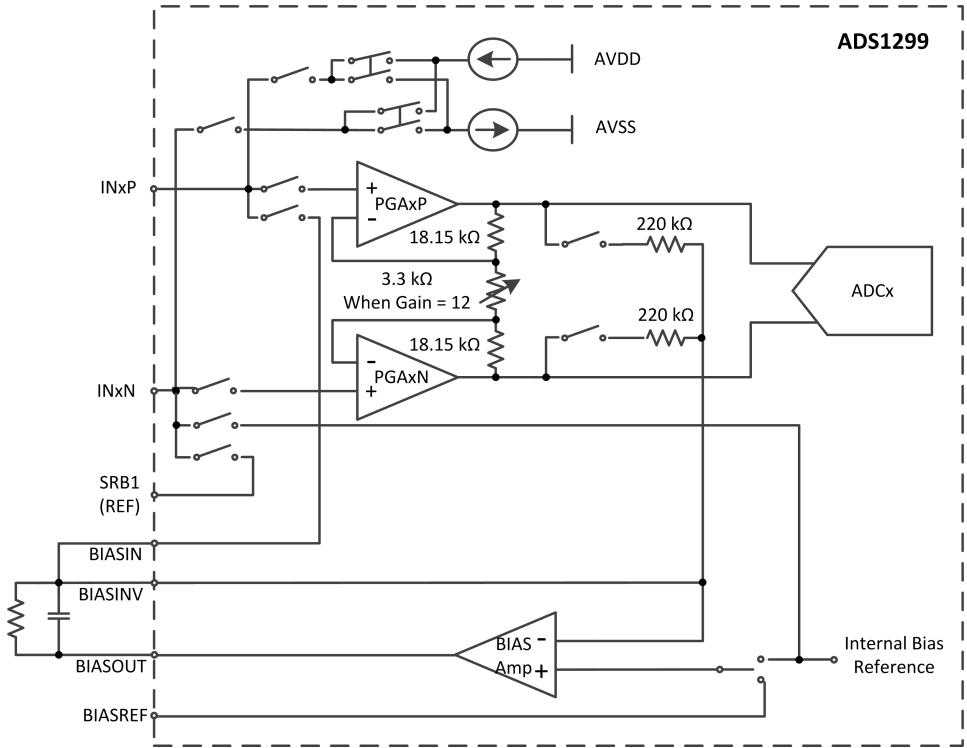


Figure 3.3: Part of the ADS1299 configurations including lead-off detection, common reference, and bias drive

arm, LA and right arm, RA). To compare the power line interference levels in different working modes, the ECG signals (sampled at 500 SPS from CH1) after detrending are analyzed in the frequency domain. The data acquisition system used in the test is the improved version of our prototype presented in [35]. The design of the system in use will be introduced in detail in the next part. Unshielded 1 meter long lead wires and H124SG electrodes are used in the test. The lead wire configuration and the key register configuration in the four working modes are listed in Table 3.1. The input REF is the common reference buffered from SRB1 pin of ADS1299. The SRB1 pin can be configured to connect the negative input of all channels through register MISC1. The driven bias circuit on-chip can be powered on or off by changing the register CONFIG3. The inputs to the drive bias circuit can be configured through the registers BIAS_SENSP and BIAS_SENSN and BIAS_SENSP/N, 0x01 means both the positive and negative inputs of channel one are connected to the drive bias circuit.

As the test took place in a normal office room having many wall-powered computers, a considerable amount of power line noise can be observed in the time domain waveform in any mode. The frequency analysis of the

Table 3.1: The lead wire configuration and key registers configuration in the four modes

Mode	LA	RA	RL	Key registers
single-ended, bias drive powered-off	CH1+	REF	open	CONFIG3, 0xF0 MISC1,0x20
single-ended, bias drive connected	CH1+	REF	BIAS	CONFIG3, 0xFF MISC1,0x20 BIAS_SENSP/N,0x01
differential, bias drive powered-off	CH1+	CH1-	open	CONFIG3, 0xF0
differential, bias drive connected	CH1+	CH1-	BIAS	CONFIG3, 0xFF BIAS_SENSP/N,0x01

captured 10-second ECG signals in Figure 3.4 shows that the differential mode has superior performance than single-ended mode without *bias drive* (93dB@50Hz versus 101dB@50Hz). But when *bias drive* is connected, the influence of power line decreases in both modes, and the improvement is more remarkable in single-ended mode (from 101dB@50Hz to 76dB@50Hz). The overall frequency performance improvement in the whole bandwidth in the single-ended mode with bias may indicate the necessity of using an analog buffer at the analog input end.

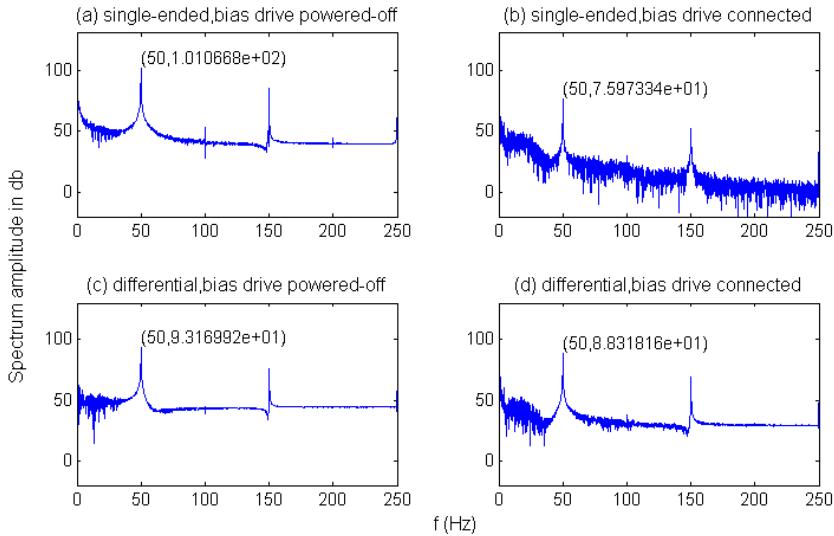


Figure 3.4: Comparison of power line interference in ADS1299's different working modes

3.1.2 sEMG/Bipotential acquisition system design

The designed sEMG acquisition system includes two main parts, 1) the device, i.e., the hardware which captures, pre-conditions, digitizes, and wirelessly transmits the digitized signals; 2) the software, which receives the digitized signals, applies digital signal processing, displays the signals in waveform for checking in real-time, and saves the recorded data in data files. In addition to the signal flow in the system, the software can also change the ADS1299 setting (e.g., the sampling rate and the input mode) by sending defined commands.

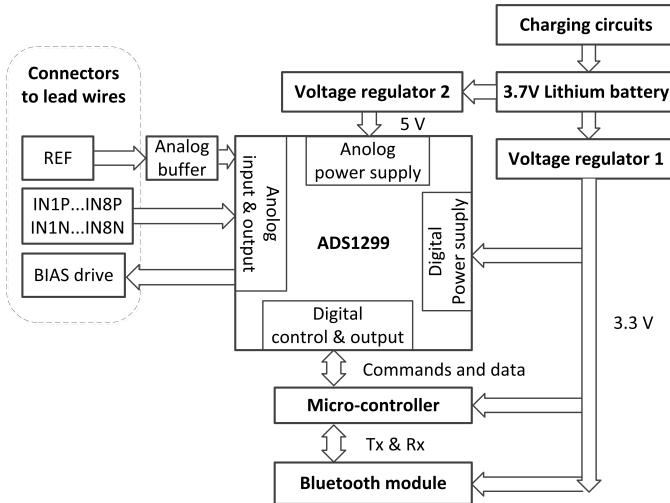


Figure 3.5: The block diagram of the designed sEMG acquisition device

Hardware

The device introduced here is an updated version of our work in [35]. In this version, the power supply part is optimized by adding charging circuits and improving the power supply efficiency. The functional block diagram of the designed device is presented in Figure 3.5. The analog part and digital part of ADS1299 are powered separately by 5 V and 3.3 V and both of them are generated from a 3.7 V lithium battery. The ADS1299 is controlled by an AVR-based 8-bit micro-controller, which meanwhile controls an HC-05 Bluetooth module. Regarding the analog part, there are in total 18 lead wire connectors including 8-channel positive and negative input pairs (INxP, INxN), a buffered common reference (REF), and a bias drive output (BIAS). The REF is analog buffered to reduce the leakage current in the single-ended mode where the currents from all the channels are added at REF [90].

The device is 56 mm × 40 mm × 10 mm in size. When powered at 4.2 V, the working current of the device was on average 58 mA when sending data

continuously (@500 SPS) and was 30 mA when Bluetooth was connected and idle. In addition to the collected biopotential data at the rate of 96 kbit/s, two-byte data flags were inserted as either an indicator of the start of one 8-channel sample (sample_flag) or the start of the first sample in each second (start_flag). The data flags are used to verify the correctness of data transmission in the software. Including the data flags, the total data rate in the transmission is 10.4016 kbit/s. The sampling rate of the system can be configured to be 250 SPS and 500 SPS and this functions well, however, the data transmission can be unstable when the sampling rate is raised to 1000 SPS.

Software

The software in the system is developed with LabVIEW in dataflow and graphical programming. As continuous biopotential data (data_recording) receiving, presenting and recording are the core functions of the software, the classic producer-consumer design pattern is employed in realizing the core functions. In the producer loop, the data at the serial port is taken in continuously and added to a queue. In parallel, the elements in the queue are read out in the consumer loop to be further interpreted, filtered and written into a signal file.

Next, in order to achieve a flexible control of the start and stop recording, a four-state machine structure is added on top of the producer-consumer loops, which is the data acquisition state machine in the software workflow (Figure 3.6). As shown in the software workflow, there are four states (S0-S3) in the state machine. The state transitions are triggered by three click buttons on the front panel, Ctl_1-read continuously, Ctl_2-quit the program, and Ctl_3-stop reading continuously. The start state (S0) is the default state waiting for the action of reading continuously or quit the program. In this state, the estimated value of the battery output voltage (data_batteryVoltage) can be acquired by sending a request command (cmd_batteryVoltage). By clicking Ctl_1, a start command (cmd_recordStart) is sent to the device and the device starts sending, meanwhile, the current computer time is captured and saved as the start time of the recording session. Similarly, when the stop reading continuously action is taken from the control panel, the state transits from S1 to S2. In S2, a stop reading command (cmd_recordEnd) is sent to the device, the current time is marked as the end time of the recording session and it is saved at the end of the signal file together with the session start time.

Outside the data acquisition state machine, as shown in Figure 3.6, is a sequence structure where the state machine is one frame in it. Before entering the state machine, the device can be configured in terms of the sampling rate and the input mode according to the preset parameters in the

software. After sending the corresponding commands to the device, the data acquisition state machine can be entered only after a successful handshake (receiving data_handshake).

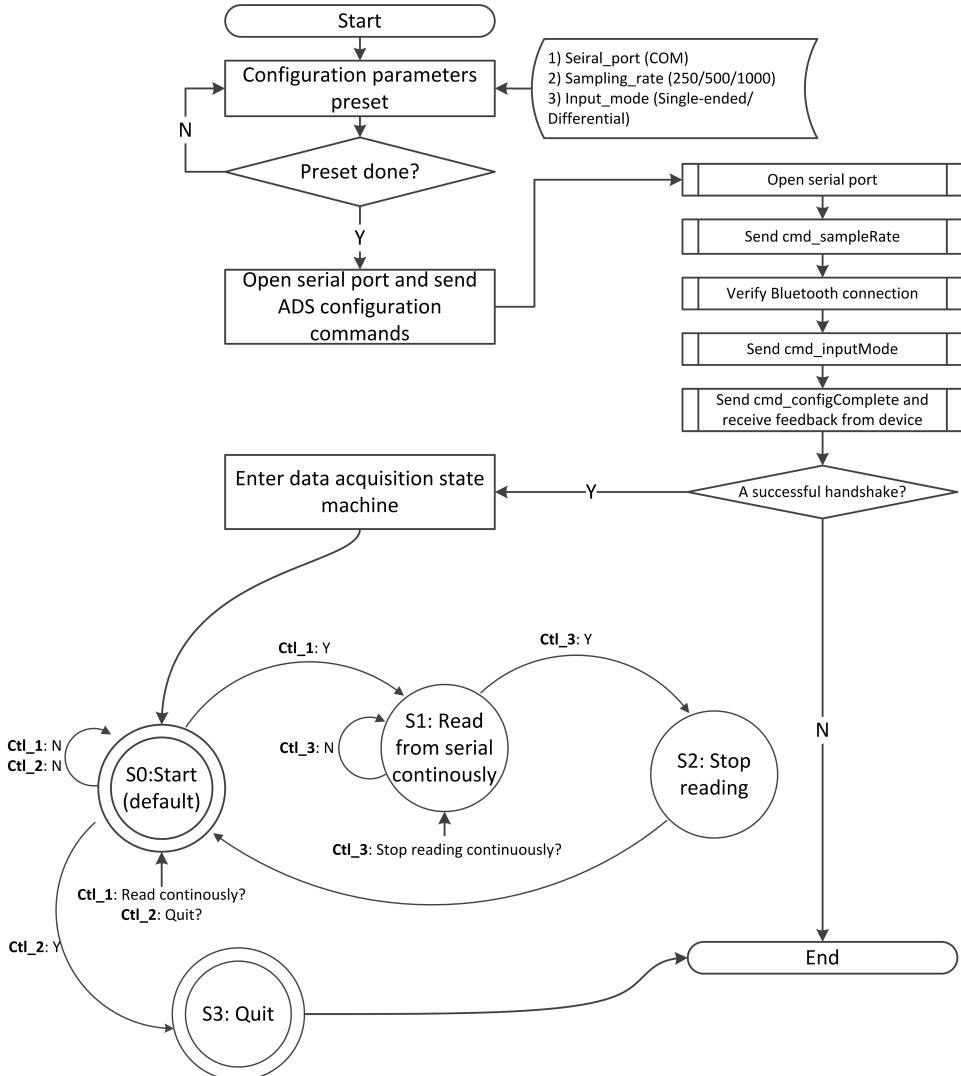


Figure 3.6: Software workflow

The communication between hardware and software

As described above, the software in the system sends commands to the device so that the device can respond by either changing configurations or sending the requested data. The firmware structure is shown in Figure 3.7. After initializing the sub-modules on the device and enabling global

interruptions, it enters the main loop of the firmware, which contains a three-state machine, Bluetooth connection check state (BT_CON_CHK), ADS1299 configuration state (ADS_CONFIG) and biopotential data sending state (ADS_DATA). The full list of input commands and output data are listed in Figure 3.7 as well as in Table 3.2. The state transitions between BT_CON_CHK and two other states are based on the values of two Boolean flags (flag_recordEnable and flag_initialization). Table 3.2 shows how the flags are assigned in response to each command and what actions are taken inside the corresponding state.

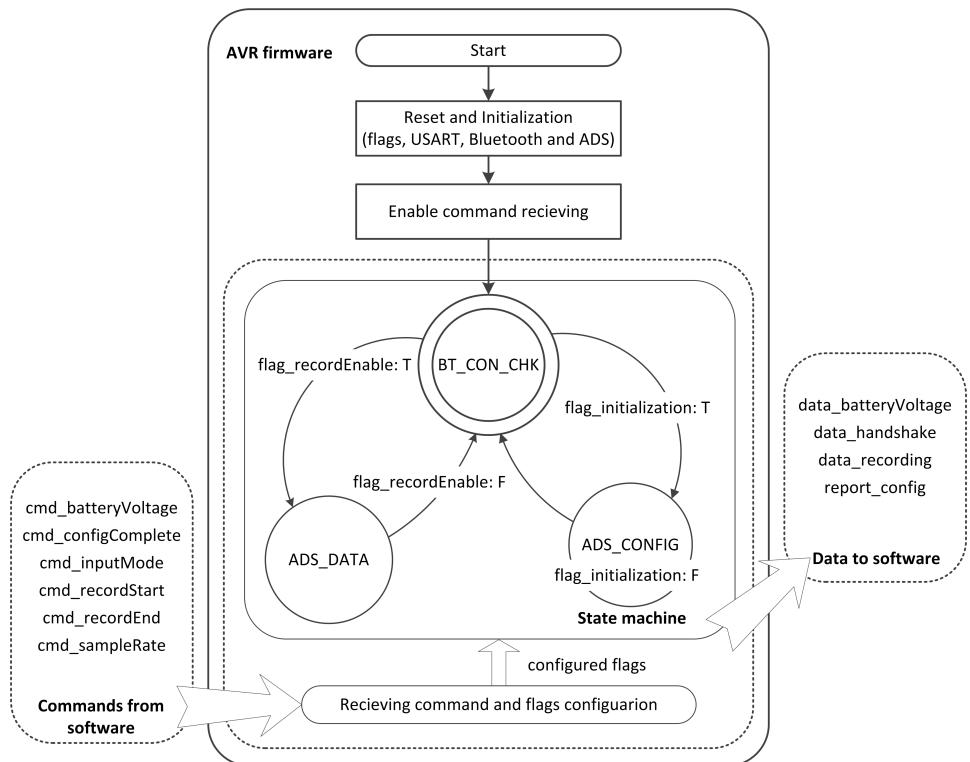


Figure 3.7: The firmware structure of the AVR processor and its communication with the software

In addition to the signal file that is generated in one recording sessions, two other data files are generated as well within the state machine, which are a time point file and a time stamp file. The processes of data saving are presented in Figure 3.8. The time point file is a record of time labels given to the signal in the second resolution. As mentioned above, a 2-byte sample_flag is added in the streaming data to denote the start of an 8-channel sample. In the S1 state in the software, this flag is recognized to cut the sample from the continuous data. Meanwhile, the number of the

samples is counted from 0 when second_flag is received; this is to check the reliability of the received data in case of any data loss. The time at which the second_flag is received and recognized, together with the final count number before being reset, is also saved in the initialized time point file. The time stamp file is independent of the signals. It is to mark the external events during a recording session. Time marks and their numbers in sequence are added when clicking the add time mark button on the front panel of the software.

Table 3.2: Firmware's responses to the software commands

cmd_	flag_initialization	flag_recordEnable	State in Figure 3.7	Action
Initial state	False	False	BT_CON_CHK	Differential, SampleRate500
batteryVoltage	-	-	BT_CON_CHK	data_batteryVoltage
configComplete	True	-	ADS_CONFIG	report_config, data_handshake, flag_initialization←F
inputMode	-	-	BT_CON_CHK	Differential/ Single-ended
recordStart	-	True	ADS_DATA	data_recording
recordEnd	-	False	BT_CON_CHK	exit ADS_DATA
sampleRate	-	-	BT_CON_CHK	SampleRate250/500

3.1.3 Discussion

Signal quality

The quality of the acquired signal is among the main concerns in a biopotential acquisition system. First, half of the sampling rate should cover the main frequency range of the sampled signal due to the Nyquist theorem. Although the biopotentials share the same mechanism, their frequency and amplitude characteristics have some differences, as shown in Table 3.3. Generally, EMG needs a higher sampling frequency than the other two and the optimal setting is 1000 Hz. However, a sampling rate of 500 Hz in this system is high enough to cover the dominant energy range of EMG between 50 Hz and 150 Hz, and is sufficient for collecting EEG and ECG as well. The communication baud rate is set to be 921.6 kbps (ATmega $f_{osc}=14.7456$ MHz). It is in theory high enough to handle full channel data at 1000 SPS. However, the upper limit of the sampling rate in the system is restrained by the wireless data transmission provided by the Bluetooth module HC-05. In the system, as each signal channel can be individually configured, the unused channels can be shut down and excluded in the data transmission to reach a higher sampling rate of the data received in the software.

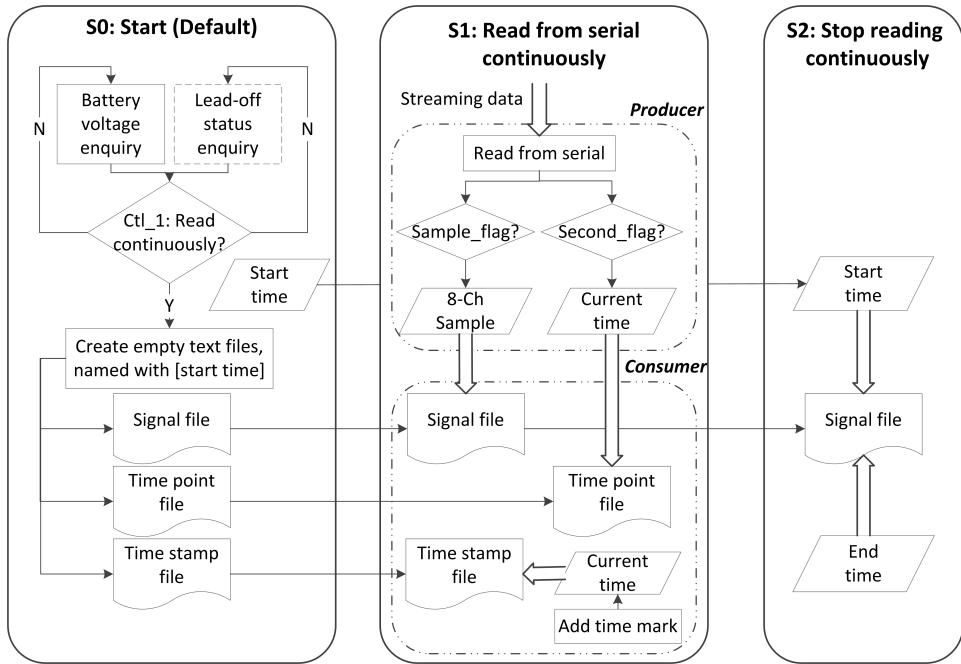


Figure 3.8: File operations in the software

Table 3.3: Frequency and amplitude characteristics of the EEG, ECG and EMG

Biopotential	Amplitude (μV) [95]	Frequency (Hz) [95]	Dominant energy range / Suggest sampling rate (Hz)
EEG	1 - 10	0.5 - 40	- / 128-1024 [91]
ECG	10^3 - 5×10^3	0.05 - 100	0-50 / 100 [96]
EMG	10^3 - 10^4	20 - 2000	50-150 / 1000 [94]

The captured interference in the acquired signal should be low so as to reflect enough biopotential details for analysis, which means a high SNR. Although the subsequent digital signal processing afterward can help improve SNR, the complexity of digital signal processing increases when the raw signals are too poor in quality and it is not a panacea for all the cases with signal distortion. For example, part of the raw and digital filtered ECG signals described in Section 3.1.1 are plotted in Figure 3.9. These signals were not collected under optimal conditions, and skin preparation was not preformed nor was the recording isolated from line interference. The signal with the lowest noise (Figure 3.9 (b)) shows the clearest ECG waves after filtering when same digital filters (2nd Butterworth 0.5-90Hz bandpass and 2nd Butterworth 50Hz notch) are applied to the four signals. However, the same processing is not sufficient for the raw signal in Figure 3.9 (a), where the whole frequency range of the signal is contaminated by noise as well.

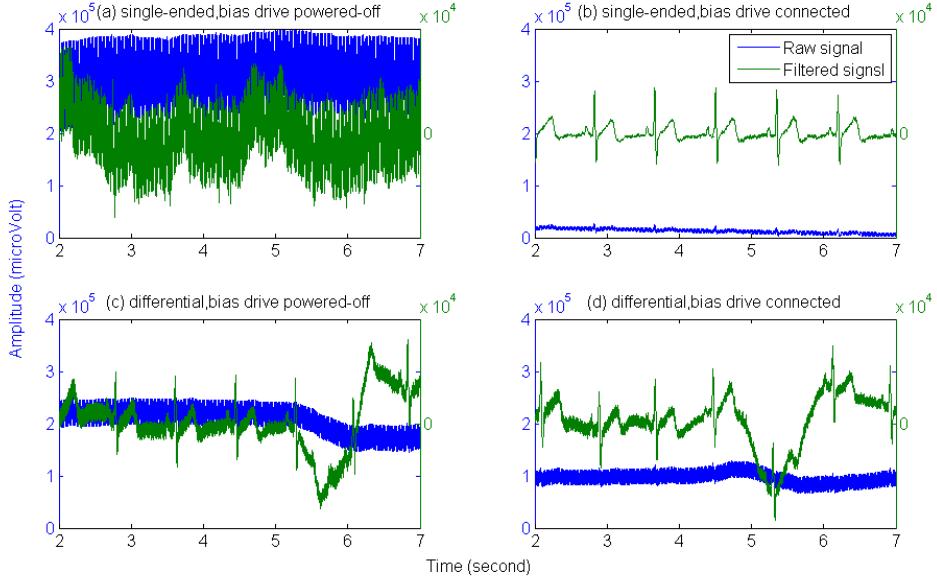


Figure 3.9: A piece of ECG signal collected in Section 3.1.1

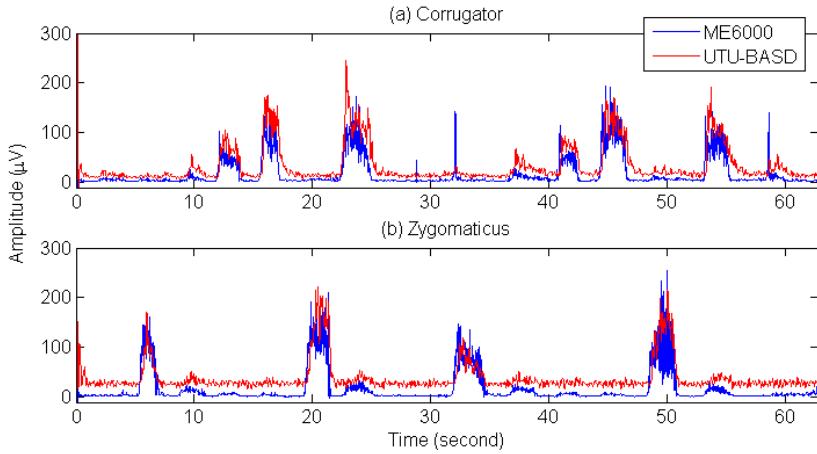


Figure 3.10: The envelopes of the sEMG signals acquired by the system developed in this study (UTU-BASD) and by ME6000 Biomonitor

This system (UTU-BASD) was compared to the ME6000 Biomonitor system in terms of signal quality. The sEMG signals of the two systems were collected from the symmetrical sites of the face on corrugator and zygomaticus muscle areas. ME6000 Biomonitor has a pre-amplifier near the electrodes and a bandpass filter on the device. The data were sampled at 1000 SPS. UTU-BASD was set as differential mode, and the bias drive was placed on

the bony area behind the left ear. The sampling rate of UTU-BASD was 500 SPS and the digital signal processing included a 20 Hz highpass filter and several notch filters. The envelope of each sEMG signal from the two systems is presented in Figure 3.10. In addition to some difference in the envelope shape between the two systems, the zygomaticus sEMG signal in UTU-BASD system shows a higher level of baseline noise. The baseline is the recorded electrical noise when the muscle is not contracting. It could include powerline interference and reflect the stability of the skin-electrode interface. The high baseline noise would hinder the reliable separation of sEMG data from noise. The acceptable baseline noise is concluded to be smaller than $15 \mu\text{V}$ in Delsys' documentation [97].

To obtain high-quality sEMG signals and rely less on the use of digital filters, one improvement that can be done in the system is to use active electrodes or add preamplifiers in the analog part of the device, as introduced in Section 3.1.1. In addition, proper preparation work can help ensure the sEMG quality. According to Delsys' recommendations, the location of the electrode pair is suggested to be aligned along the muscle fibers directions and to be far from tendon origins and innervation zones. This aims to maximize the signal amplitude in improving SNR. The other approaches are to minimize the noise amplitude. Proper skin preparation is suggested where the skin surface is wiped with an alcohol swab to remove oil and debris. The firm attachment of electrodes can help avoid the noise caused by the movement between the electrode and the skin. Finally, keeping the environment clear from power line devices is suggested if the signal contains excessive power line interference.

Further hardware improvements

ADS1299 is equipped with a lead-off detection function, which is to verify a proper electrode-skin connection. In this function, a fixed current is injected between a channel input and the power supply. The voltage at the input is monitored to check the lead connection. The lead-off detection function can provide a debugging reference when checking the signal quality especially in a long-term recording where the patient electrode impedance decay over time. However, the validity of the function itself needs to be confirmed by choosing the proper setting among multiple configurable options. Additionally, its influence on the target signals needs to be evaluated.

Possibilities of being wearable

The signals collected by UTU-BASD were measured with off-the-shelf electrodes and lead wires. The data collection device is compact for portable usage, nevertheless the whole system is not an integrated wearable solution.

Further development on the wearable solution may take the using scenario into consideration in the design (e.g., the motion extent and frequency the wearable device should tolerate). The material to be attached to the skin may also need to be tested in terms of effectiveness and safety. Some referential wearable solutions are clear facial mask with embedded H124SG electrodes [84] and the adhesive electrode sensor strip used in bispectral index measurement [98]. The wearable solutions will also help improve signal quality when the signals are amplified close to the measurement sites and movement artifact is unlikely to happen.

3.2 Electrocardiography and respiration

Pulse rate and respiration rate are both among the vital signs in medical monitoring. Their equipment and techniques for their continuous monitoring are relatively mature. In addition to the advanced patient monitoring systems in the hospital, there are wearable solutions as well. Regarding respiration, it can also be derived from the waveform of other measurements due to the interactions between the heart and lungs in the body system.

3.2.1 Wearable device - Bioharness 3

The wearable device Bioharness 3 consists of a washable chest strap with a monitoring device (BioModule) attached to it. Bioharness 3 integrates several sensors including 1) two electrode sensors housed within the chest strap to measure ECG, 2) a capacitive pressure sensor detecting circumference expansion and contraction of the torso to measure respiratory cycles, 3) a 3-axis accelerometer to measure activity and 4) an inclinometer to measure posture [81].

Bioharness 3 has software that can check the measurements in real-time. Meanwhile, BioModule continuously logs the collected data in the chosen format on its memory. By choosing the General and ECG logging format, breathing waveform at 18 Hz and ECG waveform at 250 Hz together with other extracted parameters are saved on the device memory for up to 140 hours due to its memory capacity. Another logging format, Summary and Waveform, can record the raw data of accelerations at 100 Hz, breathing waveform at 25 Hz and ECG waveform at 250 Hz. In this logging format, the memory can hold up to 55 hours of recordings.

As a wearable device in a system that could be used during exercise and sports, the validity and reliability of the first version Bioharness device were tested in both a laboratory environment [81, 82] and a field-based environment [99]. The device showed good data precision and reproducibility at a low velocity in sports (4-6 km/h) in both scenarios. Comparatively, the precision of heart rate and breathing frequency decreased at a higher velocity

in the field based environment. In this pain assessment study, Bioharness 3 should be adequate when collecting ECG and respiratory cycles data because the users are basically in a relatively static state.

3.2.2 Patient monitor

A patient monitor is usually used at a patient's bedside to provide continuous monitoring of multiple vital signs and alarms during the monitoring. It is a precise and reliable data source to collect raw physiological signals in the hospital for analysis. Usually, the bedside monitor has a screen showing the waveforms, parameters, and alarms in real-time. In addition to the one fixed in one position with a big screen, there are portable solutions as well in the monitoring system supporting a patient being transferred from one place to another without interruption in the monitoring. For example, the portable Philips IntelliVue MMS X2 can be connected to Philips IntelliVue MX800 which has an integrated computer.

Regarding respiration, there are multiple approaches to monitor it. The direct approaches measure air flow into and out of the lungs with devices such as spirometers and nasal thermocouples. Meanwhile, there are approaches measuring respiration indirectly and the capacitive pressure sensor that Bioharness uses is among them. Taking IntelliVue Patient Monitor as an example, the respiration rate can be calculated from the carbon dioxide waveform, which is the airway respiration rate. Meanwhile, respiratory cycles can be monitored by measuring the thoracic impedance between two ECG electrodes, which are RA and LL in standard ECG electrode placement.

Open source solution acquiring raw data for research from a patient monitor are scarce, but openICE is one. The openICE project (open-source Integrated Clinical Environment)¹ is actually aimed to provide a platform networking medical devices by enabling the interoperability in the platform [100]. With a Java-based computer running an openICE Device Adapter from a serial to the Ethernet, a series of medical devices can be visualized and recorded in a computer running an openICE Supervisor. If reading from a single device, direct serial communication could be directly built between a device and a computer where the Device Adapter and Supervisor are on the same host computer.²

3.2.3 Respiration derived from other measurements

The indirect respiration monitoring approaches are mostly based on the phenomenon that short-term changes in thoracic impedance reflect the filling and emptying of the lungs. Impedance pneumography is such an approach.

¹<https://www.openice.info/index.html>

²https://www.openice.info/docs/6_example-setups.html

Texas Instruments integrates respiration impedance measurement in some ECG measurement ADCs (ADS129xR), where respiratory cycles can be measured between RA and LA by injecting a high-frequency ac current [101]. However, RA and LA cannot be used to measure ECG when they are connected to the internal modulation clock output for respiration measurement [102]. In contrast, respiration can be alternatively derived from ECG, ABP, and PPG waveforms without an external current excitation by applying signal processing techniques.

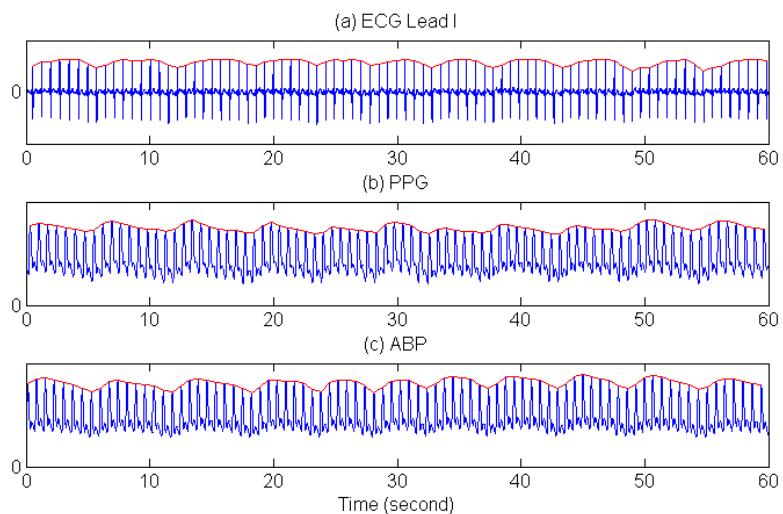


Figure 3.11: The respiratory modulation of the waveform amplitude

The respiration is reflected in ECG as a modulation of QRS amplitude [103]. It is explained as the modulation in the direction of the mean cardiac electrical axis in the dipole model of cardiac electrical activity. Since the heart rate is usually much higher than twice the respiration rate, the respiration rate is measurable from the time series derived from the area under the QRS complexes from Nyquist sampling theorem point of view. In the ECG waveform, the fluctuation of the baseline could also reflect respiration due to the motion of the chest, however, it usually only seen in deep or exaggerated breathing [104]. Similar peak amplitude modulation is also found in ABP and PPG waveforms. The respiratory changes in ventricular stroke volumes and arterial pressure is induced by a series of cardiopulmonary interactions [105]. Figure 3.11 shows the synchronously measured ECG, PPG and ABP without any movement artifact. The respiratory modulation of the waveform amplitude can be observed from the waveform envelope, and the respiration rate can be then estimated.

3.3 Galvanic skin response

The measurement of GSR commonly uses exosomatic direct current or voltage recording. Figure 3.12(a) shows a GSR measurement circuit structure with an operational amplifier for exosomatic direct voltage recording (eHealth v2.0 platform). The given constant voltage is 0 V, and the bias voltage to the positive input of the amplifier is 0.5 V which is divided from V_{ref} . A capacitor is connected in parallel with the feedback resistor and acts as a low-pass filter (cutoff frequency $f_c = 15.9$ Hz). The skin conductance ($1/R_{skin}$) can be calculated according to Equation 3.1. One problem with this circuit is the different amplification requirements SCRs and SCL, where SCRs fluctuate in a small amplitude difference while stacking on SCL. SCL is in a wide range due to inter- and intra individual differences [106]. Another problem with the eHealth v2.0 platform when working with Arduino Uno is the skin conductance resolution, which is $0.098 \mu\text{S}$ (where the ADC resolution, $\Delta V_{out} = \frac{5V}{1023}$) and the SCRs fluctuations smaller than that could not be captured. To cope with the first issue, Poh M. *et al* [107] proposed a solution by adding an integrator circuit (shown in Figure 3.12(b)) so that the bias voltage to the positive input of the amplifier (V_b) can change along with amplifier output (V_o). In this solution, the skin conductance is calculated from Equation 3.2.

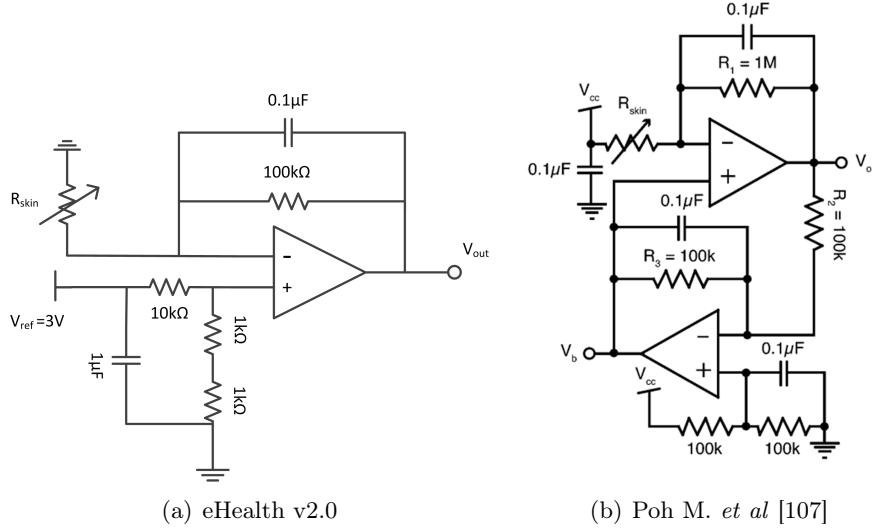


Figure 3.12: Two GSR exosomatic direct voltage recording circuits

$$\frac{0.5V}{R_{skin}} = \frac{V_{out} - 0.5V}{100k\Omega} \quad (3.1)$$

$$\frac{V_{cc} - V_b}{R_{skin}} = \frac{V_o - V_b}{1M\Omega} \quad (3.2)$$

In terms of GSR measurement sites, the middle phalanges of the index and middle fingers for bipolar recording are commonly used. As palmar sites showed distinguished electrodermal activity [108], other similar sites are acceptable if electrodes can be fixed easily. In addition to the palmar sites, the distal forearm sites were also proved to be a viable alternative and have been implemented in the E4 wristband [80].

3.4 Summary

This chapter presents the mechanisms of electronics devices and systems, for biopotentials recording mainly. Further improvement possibilities are discussed

Chapter 4

Study Design and the SpaExp Database

Healthy volunteers are commonly involved in the studies to develop an automatic pain intensity indicator or an equivalent tool. Experimental pain models provide the possibility to explore the pain system under controlled settings. The experimental stimuli could be mechanical, thermal, electrical or chemical [109]. The stimuli can be applied to skin, muscles, and viscera for assessing different pain pathways and pain mechanisms. In this study, two experimental pain modules, thermal and electrical, were employed. The signals recorded in the database included some ANS-based signals (ECG, respiration, GSR) and one behavioral signal, facial expressions using facial sEMG. The analysis was done in a supervised learning fashion, where the samples were firstly labeled before the pattern of the labels were learned by machine learning algorithms. Thirty-one volunteers were recruited by the co-workers in the Department of Nursing Science at the University of Turku. The study was initially designed by Prof. Sanna Salanterä, Prof. Riku Aantaa, Prof. Pasi Liljeberg, and Adjunct Professor Amir M. Rahmani. The previous study results were reported in [33], [34] and [110].

4.1 Experiment protocol

The study subjects recruited to the study were working-age people in a generally healthy condition without any chronic, acute somatic or mental illness. In addition, no regular medication was taken during or in the two weeks prior to the data collection. All the study subjects' cardiovascular parameters were within normal range. They had normal sensations and healthy skin on the face and upper extremities. Pregnant women and individuals with a body mass index larger than 30 kg/m² were excluded from the study. This study was approved by the Ethics Committee of the Hospital District of Southwest

Finland (ETMK:83/1801/2015).

The data collection took place in a quiet room where powerline interference was avoided as much as possible. One technician and one research nurse were also in the room to set the signal acquisition system and instruct the study subject in the Finnish language, respectively. The research nurse first briefly introduced the processes of the study to the study subject, then the study subject read and signed an informed consent. In the next preparation phase, the study subject put on the Bioharness 3 before sitting on an armchair with a footstool. The Bioharness 3 was turned on for recording before being put on. After placing the pre-gelled surface electrodes on the left side of the study subject's face, the sEMG recording started.

It was planned that each study subject took four tests, in each of which the pain stimulation increased gradually in intensity. Two experimental pain models were employed in the study. Each type of stimulus was applied once to each symmetrical side of the body. The relative order of the four pain stimulations was defined beforehand in a side-switching manner (i.e., 1-right heat, 2-left heat, 3-right electrical, 4-left electrical). However, the order of the four tests was randomized by choosing the pain stimulation to start with. Before the pain stimulation was put on each time, enough rest time was ensured for each study subject to return to a relatively steady heart rate level. This was to check whether the heart rate at rest was in the normal range at the very beginning and to ensure enough recovery from the previous excitation. The signal recordings ended after the end of the last test.

4.2 Study design

Each of the four tests was designed as presented in Figure 4.1. The pain stimulation at intensity 0 is attached to the skin surface before the pain stimulus start time t_0 . From t_0 , the intensity of the pain stimulus increases almost linearly until the study subject reports the pain as "intolerable" which is denoted as t_2 on the timeline. The pain stimulation is either stopped or detached from the skin surface after t_2 , and the intensity at this point (i_2) is marked down. Immediately after t_2 , the study subject was asked for a self-report VAS score. To protect the study subject from tissue damages such as skin burn, the upper limit of the stimulus intensity was set. In cases where the maximum intensity was reached without an "intolerable" report, the pain stimulation was ceased and the maximum intensity is recorded as i_2 . Between t_0 and t_2 , the study subject was also instructed to indicate the time t_1 when the sensation turned into "uncomfortable." The self-report time is defined by the study subject when pressing a button that triggered a beeper.

The description "uncomfortable" is interpreted as intensity 3 or 4 in NRS/VAS from a clinical nursing point of view which represents a clinical

importance in pain severity indicating patients' perception of adequate pain control [12, 13]. The self-reported t_1 was instructed as "start to perceive it as pain", and therefore t_1 and t_2 could respectively also be interpreted as personalized *pain threshold* and *pain tolerance*.

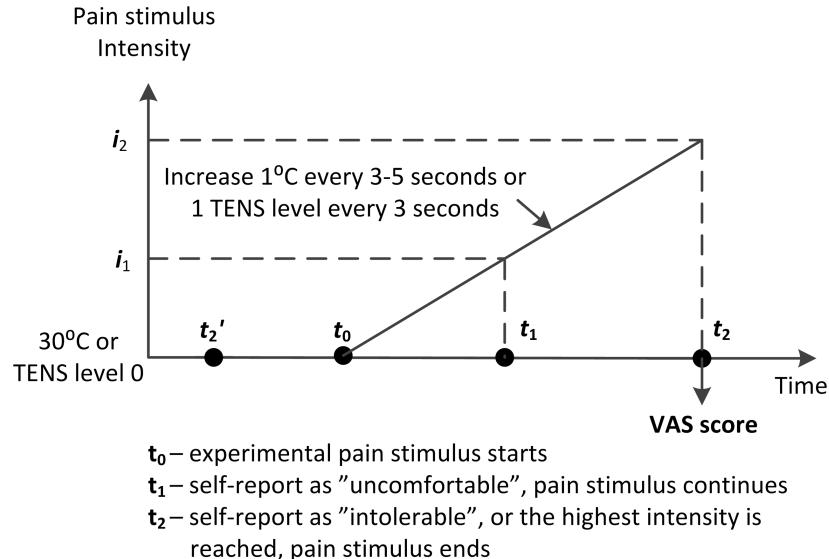


Figure 4.1: The timeline and pain stimulus intensity in one test

4.2.1 Experimental pain stimulation

Acute pain is generated when a stimulus is presented to the body. The perception of pain, called nociception, depends on specifically dedicated receptors and pathways. The nerve cell endings that initiate the sensation of pain are called nociceptors, and they have two types of axons, A δ fibers and C fibers. The two categories of pain perception are described as a sharp *first pain* and a more delayed, diffuse and longer-lasting *second pain* [111]. The tingling sensation or a sharp *first pain* is caused by the activation of A δ fibers, while the longer-lasting *second pain* is sensed via the slowly conducting C fiber axons. Two experimental pain models in the skin were employed in this study to cover more than one pain sensation. It is easier to access and activate the nociceptors in the skin compared to muscle, bone, and viscera. The details of the two pain stimulation settings are described below.

Electrical stimulation of the skin

The advantage of electrical stimulation is that it is easy to control the stimulation device output with a variety of patterns in terms of waveform, fre-

quency, intensity, and duration. The population and activation of the nerve fibers depend on the stimulus intensity where C fibers have a higher activation threshold than A δ fibers [109]. A muscle stimulation device, Sanitas SEM43-Digital EMS/TENS was used as the electrical pain stimulation device in this study. The pair of electrode pads were resized and placed vertically on the fingertip of the ring finger. The electrodes were fixed on the positions with medical tape to prevent them from falling off.

The Sanitas device outputs biphasic rectangular pulses with configurable pulse width, pulse frequency, and pulse intensity. The maximum output current of the device is 200 mA peak-to-peak at 500Ω , and the intensity is adjustable in a scale between 0 and 50. Before the start of the data collection, different settings on the pulse width and frequency were evaluated within the research group regarding the induced pain. The final chosen pulse width and frequency were 250 μ s and 100 Hz because the pulses in this combination could evoke noticeable pain sensations. In the test, the pulse intensity was set manually by pressing the "intensity up" button on the device in every 3 seconds. The waveform of the electrical pulses is visualized in Figure 4.2.

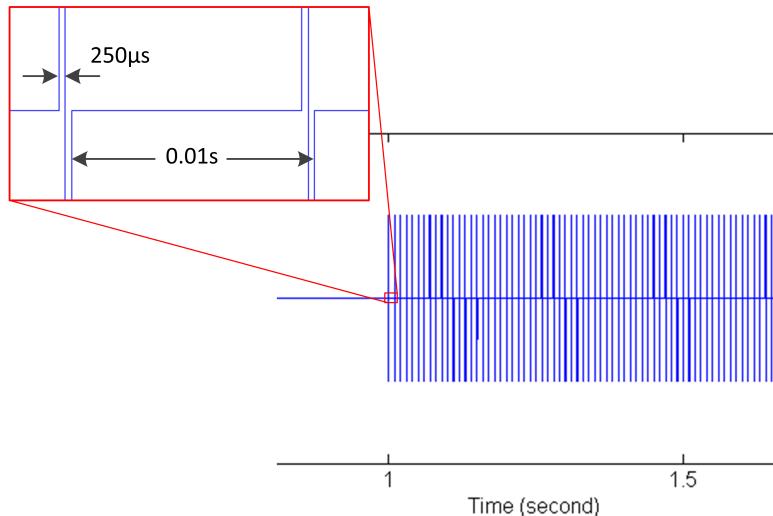


Figure 4.2: The waveform of the electrical stimulation

Thermal stimulation of the skin

Thermal stimulation includes several approaches in practice, which are cold, contact heat and laser. The stimulation used in this study was contact heat where a heating thermode evoked the heat pain. Rapid heating activates first A δ fibers within less than 0.5 s, which is followed by a C-fiber-mediated *second pain*, while slow heating less than $1^{\circ}\text{C}/\text{s}$ gives a preferential activation of C fibers [49]. The slowing heating that activates C fibers was the second

experimental pain stimulation model used in the study. Long exposure to hot tap water may cause scald injury. It is a synthetic effect of temperature and exposure time, and the relationships of both of these to a scald injury are presented in Figure 4.3(a) and (b) [112]. For example, it is concluded that 55°C would lead to second-degree burn in 17 s and third-degree burn in 30 s.

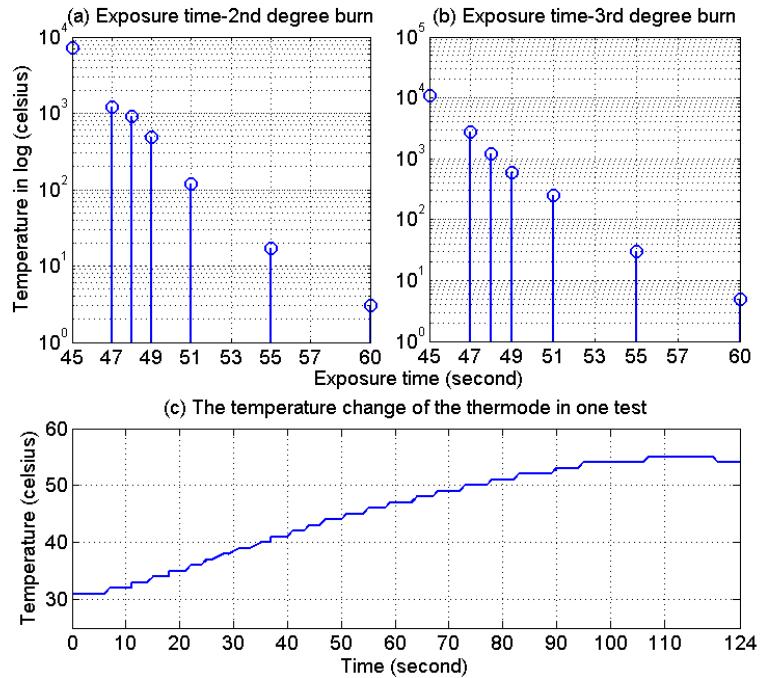


Figure 4.3: The potential scald injury caused by heat and the temperature curve of the heat pain stimulation in the test

The thermode system was developed in the laboratory. The manufactured thermode has a round contact surface of metal with a diameter of 3 cm. The heating speed of the thermode was regulated through setting the switching frequency of the solid state relay between the heating element and its power supply. A digital temperature sensor was used in the thermode for temperature monitoring in real time. The position of the temperature sensor and the contact surface were vertically symmetric to the heating element in order to estimate the temperature of the contact surface. The temperature curve of the thermode in one test is plotted in Figure 4.3(c). It shows the temperature increasing approximately linearly before 50°C. The heating processing slows down after that and the temperature starts to drop from 55°C. In the heat pain tests, the thermode was placed on the study subject's inner

forearm. A cool pad was placed on the heated skin spot right after the end of the test.

Data summary

In total, 122 tests were taken in the data collection. Sixty of them were with electrical pain stimulation, and sixty-two were with heat pain stimulation. Among the thirty-one study subjects that took part in the data collection, thirty of them (fifteen male and fifteen female) took all the four tests. One study subject did not take the two electrical pain stimulation tests. In the other two heat pain stimulation tests, most of the GSR recording of this study subject were invalid either due to the high skin impedance where the skin conductance was too low to read. Therefore, these two heat tests were also be excluded from the analysis that relates to GSR, and the total number of tests was 120 in this case.

The data collected in this part were saved into two lookup tables. The first one, is the information lookup table. Table 4.1 lists the name and data range of each column in the information lookup table file. The second one is the timestamp lookup table. Table 4.2 lists the name and format of the timestamps. The *rest start* and *rest end* took place before test1 started. In both tables, inapplicable values are each denoted as not a number (NaN).

Table 4.1: Information lookup table, item list

No.	Item	Data range/unit
1	subject number	1, 2,..., 31
2	subject gender	male, female
3	test number	1, 2, 3, 4
4	experimental stimulus	elec, heat
5	side of the body	left, right
6	self report VAS score at t_2	[4,10]
7	whether pain tolerance was reached at t_2	yes, no
8	thermode temperature at t_1	°C
9	thermode temperature at t_2	°C
10	TENS device output intensity at t_1	0, 1,..., 50
11	TENS device output intensity at t_2	0, 1,..., 50

The distributions of the VAS scores reported at t_2 in Figure 4.1 in the two types of pain stimulation are presented in Figure 4.4. This figure shows that, within the intensity range, the high-frequency electrical stimulation evoked higher pain intensity than the slowly increasing heat on average (avg.VAS: electrical - 7.3, heat - 6.6). Moreover, pain tolerance was less readily reached within the upper-intensity limit in the tests with heat pain stimulation. The difference between the perceived pain caused by the two experimental pain stimuli is reflected not only in the distribution variety of the overall reported VAS, but also in the variety of the self-reports within the same study subject.

Table 4.2: Timestamp lookup table, item list

No.	Item	Data format
1	record start	
2	rest start	
3	rest end	
4-6	test1 t_0, t_1, t_2	
7-9	test2 t_0, t_1, t_2	HH:MM:SS.FFF
10-12	test3 t_0, t_1, t_2	
13-15	test4 t_0, t_1, t_2	
16	record end	

Among the 11 study subjects without an "intolerable" report, the reported VAS scores in the heat tests were 1.6 lower than the electrical ones on average. Comparatively, this difference was only 0.2 among the subjects that reported "intolerable" in the heat tests. All these differences among the study subjects validate the subjectivity in pain perception in terms of pain tolerance.

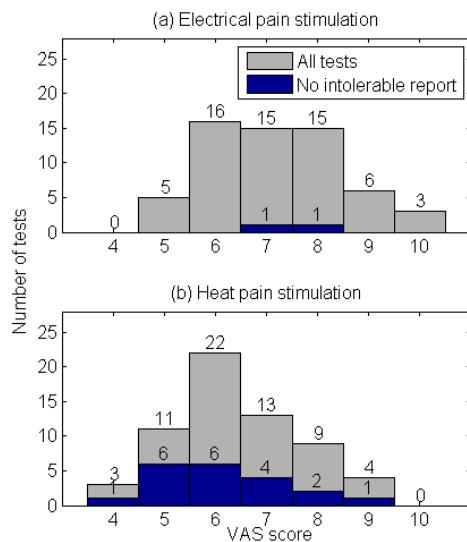


Figure 4.4: The distribution of the reported VAS scores at t_2 in all tests and the tests without an "intolerable" report

Regarding the stimulus intensity at t_1 and t_2 , the electrical tests had a greater variety among the subjects than the heat tests. The average electrical intensity at t_1 is 5/50 (among all tests, SD=3) and is 16/50 (among the "intolerable" reported tests, SD=9) at t_2 . In the heat tests, the average heating temperature at t_1 is 49°C (among all tests, SD=3°C) and is 53°C (among the "intolerable" reported tests, SD=2°C) at t_2 . The temperature reported at *pain threshold* and *pain tolerance* are slightly higher than some

other studies on average with nearly the same variation. With similar or different stimulation trial design (i.e., temperature rise speed and plateau duration if there is), the average reported *pain threshold* temperature was 45.7°C in [56], 44.3°C in [113], and 42.4°C among subjects with dark hair in [48]. While the average reported *pain tolerance* temperature was 47.7°C among subjects with dark hair in [48] and 45.9°C in [114] although the average self-report intensity was close to this study (avg.VAS 6.4 [114] v.s. avg. VAS in this study 6.6). In the electrical tests, the electrical current is estimated as 0.1 mA per intensity when converting the electrical pulses to an average current per second: $250 \mu\text{s}^2 \times 2(\text{biphasic}) \times 100 \text{ Hz} \times (200/50 \text{ mA})$. Then both the intensity at *pain threshold* and *pain tolerance* are lower than the 5 Hz biphasic sinusoid alternating current among subjects with dark hair (threshold - $2.2 \pm 1.1 \text{ mA}$ [48] v.s. $1.0 \pm 0.6 \text{ mA}$ in this study, tolerance - $3.3 \pm 1.9 \text{ mA}$ [48] v.s. $3.2 \pm 1.8 \text{ mA}$ in this study).

4.2.2 Biosignals and the acquisition system

The biosignal waveforms collected in this database included: 1) 6-channel facial sEMG, 2) ECG Lead I, 3) respiration and 4) GSR. As introduced in Chapter 3, a breathing waveform can be derived from an ECG waveform, which can be an alternative when extracting breathing rate. The collected waveforms and parameters were visualized on the laptop in real-time. The signal acquisition software platform working in one data collection is presented in Figure 4.5.

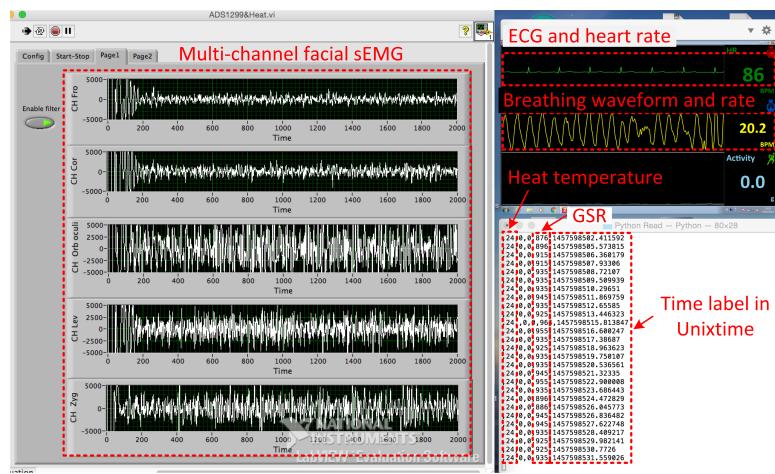


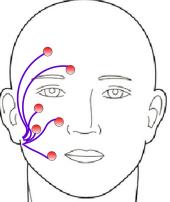
Figure 4.5: Biosignals acquisition software platform

Facial sEMG

The facial sEMG signals were collected with the daughter board in the ADS1299 performance demonstration kit [115]. The data recording software was an earlier version of the LabView software described in Chapter 3. It has the function of continuous data receiving from the hardware via a serial port by clicking the buttons Ctl_1-read continuously and Ctl_3-stop reading continuously on the front panel. The start time and end time of each recording session were saved for adding time labels to the samples. The daughter board was configured in single-ended mode by an Arduino Mega controller, and the bias drive or any bias circuit on the daughter board was not in use.

The facial muscle areas were chosen based on Table 2.3. The single electrode for each area was placed on the right side of the face following the facial EMG electrode placement guidelines [116]. The connections between electrodes and the analog inputs of the daughter board are explained in Table 4.3¹. Muscle frontalis is not involved in pain facial expressions in existing literature. Its signal was taken as a noise reference to all the other sEMG signals due to the electrical pulse interference which was caused by the pain stimulation in the electrical tests. Before attaching the H124SG electrodes to the face, the electrode sites were wiped with an alcohol pad as skin preparation.

Table 4.3: Electrodes placement and their connections to the device

Electrode site	Analog input	Muscle/Head area
	CH1+	Frontalis
	CH2+	Corrugator supercilii
	CH3+	Orbicularis oculi
	CH4+	Levator labii superioris
	CH5+	Zygomatic major
	CH6+	Risorius
	REF	Bony area behind ear

Facial sEMG signals were sampled at 1000 Hz. The ADC resolution of ADS1299 is 24-bit. Its analog input range was between -4.5 V/gain and 4.5V/gain as the internal reference 4.5 V was used and the gain was set to be 24 in the data collection. Therefore, the input voltage range was between -187.5 mV and 187.5 mV. The laptop powered the data acquisition hardware through a USB cable which was also the data transmission channel. The laptop ran on battery power during the data collection to avoid bringing additional powerline interference to sEMG signals.

¹The analog inputs are denoted as AIN1P, ANI1N...REF_ELEC in ADS1299EEG-FE

ECG and respiration

ECG waveform and breathing waveform were collected with Bioharness 3 chest strap. The device's logging format was configured to be *General + ECG*, where the log files reported extracted parameters such as heart rate and breathing rate at 1 Hz, breathing waveform at 18 Hz, and ECG waveform at 250 Hz. The ADC on the device has a resolution of 12-bit. The device has a maximum log duration of 35 hours with new batteries, and the maximum memory capacity of the device can store 140 hours of data with the chosen logging format¹ which are beyond the protocol's requirements. The two sensor pads were moistened before use to aid conductivity, and the strap should be adjusted to be a snug fit².

GSR

The GSR acquisition system was developed from the eHealth v2.0 platform. Its analog circuits for GSR acquisition is plotted in Figure 3.12(a). The eHealth shield is compatible with the Arduino Uno controller which digitizes the analog signal at 10-bit resolution.

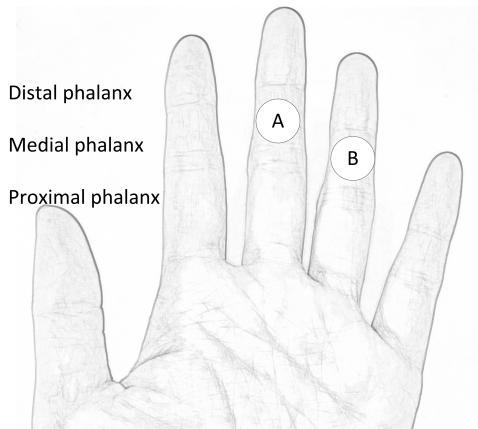


Figure 4.6: GSR electrode sites on the volar surface

Palms and volar surfaces of the fingers are active sites for electrodermal recording. Medial phalanges of the index and middle fingers were recommended for GSR recordings because they have larger areas for electrodes and are less prone to scarring and movements compared to the other finger areas [106]. In this study, the medial phalanges of the middle and ring fingers were chosen instead (as shown in Figure 4.6) as the index finger was occupied by a pulse oximeter for checking heart rate in real time. The same H124SG

¹Bioharness 3, Log Data Descriptions

²Bioharness 3, User Manual

electrodes were used for GSR measurement. The electrodes were placed on the opposite side of the body to the pain stimulation. For example, as a test on the right always followed a test on the left side, the GSR electrodes were moved from the right side to the left side between the two tests.

4.3 Summary of the collected signals

This chapter documents how the database was built and the related technical specifications. The signals collected in the database included waveform signals from which interpretable features could be extracted from (e.g., sEMG, ECG and breathing waveform) and the extract parameters could be provided by Bioharness 3 (e.g., heart rate and breathing rate). Table 4.4 lists all the waveform time series and part of the parameter time series in the database. The six channels of sEMG are named *fro-frontalis*, *cor-corrugator supercilii*, *orb-orbicularis oculi*, *lev-levator labii superioris*, *zyg-zygomatic major* and *ris-risorius*. The digitization resolution of the sEMG signal was $0.54 \mu\text{V}$ approximately. The display resolution of the GSR signal was 10 nS (nano Siemens), which is adequate to display the digitization resolution of the signal ($\approx 98 \text{ nS}$). The signal time series before and during one electrical test were plotted with their data labels in Figure 4.7. The sEMG and ECG signals in the figure have been processed.

Table 4.4: Signal summary

No.	Signal	Time resolution	Data type	Data unit
0	Heat temperature.	$\approx 1 \text{ s}$	integer	$^{\circ}\text{C}$
1-6	sEMG	1 ms	$[-2^{23}+1, 2^{23}-1]$	$4.5/2^{23}*10^6 \mu\text{V}$
7	ECG	4 ms	[0 4095]	$(x-2048)*6.25 \text{ mV}$
8	breathing	1/18 s	[0 4095]	-
9	GSR	$\approx 1 \text{ s}$	integer	10 nS
10	heart rate	1 s	integer	beats per minute (bpm)
11	breathing rate	1 s	integer	breaths per minute
12	acceleration	1 s	[0 16]	VMU ¹ of max.acc (g)
13	activity	1 s	[0 16]	VMU of avg.acc (g)
14	posture	1 s	[-180 180]	degree (vertical = 0°)

¹VMU = $\sqrt{accx^2 + accy^2 + accz^2}$, vector magnitude unit

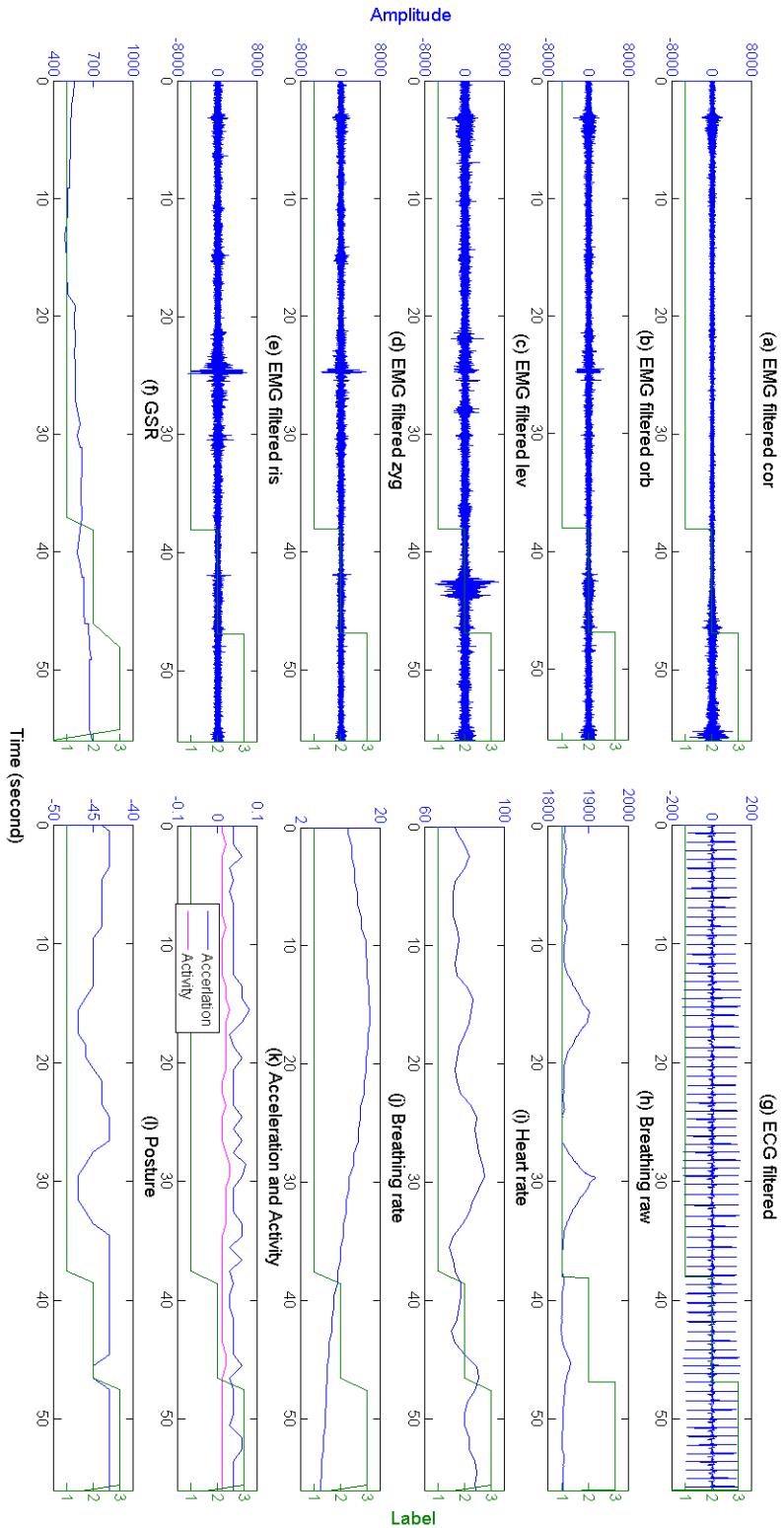


Figure 4.7: The signals in one electrical test

Chapter 5

Biosignal Processing and Feature Extraction

Before feeding the collected biosignals into a classifier for pattern recognition, the signals need first to be processed for three reasons. First, the contaminations in the raw signal could be prominent or even cover the useful information in the signal and therefore should be removed. Second, the parameters in a multi-parameter model could be sampled at different time intervals and unifying the time interval via feature extraction can simplify the modeling process. Third, the constructed model is more interpretable when the inputs are meaningful features instead of waveform segments. The principles and methods used in the signal denoising and feature extraction are explained below.

5.1 Signal denoising

The signal denoising strategies are defined based on the characteristics of the noise sources. Therefore the noise sources are first introduced in this section before explaining their implementation. The common biosignal contaminants are categorized in Table 5.1.

5.1.1 Motion artifact

Motion artifact is caused by voluntary or involuntary patient/subject movement during the data acquisition. The motion here may refer to one or a combination of cable motions, electrode-skin motions and the mechanical stretching of the skin. The cable motion artifact appears when the cable between the electrode and the acquisition device swings or is touched. For the passive electrodes, such artifact can swing the captured signal out of the

Table 5.1: Common biosignal contaminants [6, 7]

Type	Contaminants
Measurement (Motion artifact)	Cable motion artifact
	Electrode to electrolyte/skin motion artifact
	Skin stretch reflex
Instrumentation	Baseline wander
	Amplifier saturation
	ADC over-ranging
	Quantization noise
Interference	Poor electrode contact (including electrode lift)
	Powerline interference
	RF interference
	Unwanted physiological biosignals

ADC's input range, while it has little effect on the active electrodes¹. Because of this the passive electrode cable is usually fixed or shielded to avoid cable motion artifact. In contrast to the cable motion artifact, the other two types of motion artifacts cannot be avoided in dynamic contractions and vigorous activities. The difference between the electrode-skin motion artifact and the skin-stretch reflex motion artifact is that the magnitude of the first one is highly dependent on the nature of the electrode material but the second is not [7]. In addition to the electrode material, the use of gel also influences the level of the motion artifact. For example, compared to the wet Ag/AgCl which has low skin-electrode impedance ($100\text{k}\Omega \parallel 10\text{nF}$), dry and non-contact electrodes have a much higher skin-electrode impedance ($1\text{M}-200\text{M}\Omega \parallel 10\text{nF}$) and are more vulnerable to motion artifact where the skin-electrode impedance is no longer negligible to the amplifier input impedance [117].

A highpass 20 Hz Butterworth filter with a slope of 12dB/oct is recommended for general use in an sEMG signal to de-noise both motion artifact and baseline noise [118]. The same bandwidth was suggested in facial sEMG processing to remove additional artifacts such as eye movements, eye blinks, the activity of neighboring muscles, respiration and swallowing [119]. The motion artifact in an ECG signal is the results of patient movement or respiration and it ranges from 0.5 Hz to several Hz [120]. Therefore, a lowpass filter with a cutoff frequency close to zero can be used to remove such motion artifact from the ECG signal.

5.1.2 Instrumentation contaminants

The instrumentation baseline noise is intrinsic originating in the electronics system and at the skin-electrode interface. The inherent noise in the electronics cannot be eliminated but can be reduced through circuit design.

¹Comparison: active versus passive electrodes. g.tec Medical Engineering

The noise with a buffered common reference input in ADS1299EEG-FE was tested to be $2 \mu\text{V}$ peak-to-peak with a $14 \mu\text{V}$ offset [115]. Among the commonly used electrode materials, gold, platinum, AgCl and sintered Ag/AgCl, gold electrodes are polarizable and provide a good signal quality at frequencies above 0.1 Hz, while Ag/AgCl electrodes are widely used due to its low polarization and low baseline drift (0.13 mV at 25°C) [121]. The inherent noise ranges from 0 Hz to several thousand Hz. This noise is mostly reported rather than being filtered. In addition, the low frequency part including the DC offset or baseline drift especially is already removed when removing the motion artifact.

However, the prerequisite to a good performance for the Ag/AgCl electrodes is a good electrode contact. An extreme case is termed an electrode lift in myoelectric control where an electrode loses contact with the skin completely and the prosthesis is driven by powerline interference instead [7]. The contact can be evaluated via the electrode-skin impedance measured between one electrode pair that is attached to the skin. Although this check is usually done before data acquisition, the impedance can change during the measurement [6].

The remaining three instrumentation contaminants listed in Table 5.1 are more like signal distortions where the signal can hardly be saved via processing. ADC over-ranging or clipping happens when the gain of the amplifier is large or/and a big common-mode voltage deviates the signal far from the center of the input range so that the analog output of the amplifier exceeds the input range of the ADC in the next stage. As regards ADS1299, its ADC input differential dynamic range depends on the analog supply and the reference voltage. In cases where its internal voltage reference ($V_{\text{ADC_REF}}$) is used, the ADC input range is between -4.5 V and 4.5 V. The ADC saturation causes the output to be clipped to the maximum or minimum value. The quantization noise is also related to the ADC. It should be considered when the resolution voltage is too large compared to the small change of the input analog signal so that the small change cannot be captured in the digitization. However, this quantization noise is probably not an issue in a specialized biopotential acquisition system.

Comparatively, the attention should be paid to the amplifier saturation, which causes the nonlinear output of the amplifier and ADC saturation in the next stage. In biopotential acquisition, the peak to peak voltage range of a biopotential signal is small at a microvolt or millivolt level. However, the common-mode voltage (CM) presented at the differential inputs may cause an amplifier saturation problem. According to ADS1299's datasheet, the output of the amplifier cannot swing to the analog power supply closer than 200 mV. Moreover, the amplifier output should stay within the input range of the ADC to avoid ADC saturation. Therefore, the lower limit and upper limit of the CM range (CM_{lower} and CM_{upper}) are described in Equation

5.1 and 5.2, where AVDD and AVSS are an analog supply and an analog ground, and V_{MAX_DIFF} is the maximum differential signal at the input of the amplifier. In theory, when $AVDD = 5$ V, $AVSS = 0$ V, gain = 24, assuming $V_{MAX_DIFF} = 20$ mV, then the CM should stay within (8.3 mV, 177.5 mV). If V_{MAX_DIFF} increases to 200 mV, then the gain should be lowered to 12 at the maximum.

$$CM_{lower} = \frac{\max\{-V_{ADC_REF}, AVSS + 0.2V\}}{gain} + \frac{V_{MAX_DIFF}}{2} \quad (5.1)$$

$$CM_{upper} = \frac{\min\{V_{ADC_REF}, AVDD - 0.2V\}}{gain} - \frac{V_{MAX_DIFF}}{2} \quad (5.2)$$

5.1.3 Interference

The most common interference in an indoor environment is the powerline interference. This consists of the harmonics of sine waves at 50 or 60 Hz, the powerline frequency. In addition to the instrumentation approaches to diminish its influence on the biopotential, several signal processing approaches are feasible in removing the powerline interference in the signal post-processing. Some commonly known approaches are narrow-band notch filtering, adaptive filtering, frequency domain Hampel filtering and the subtraction approach [122]. One starting point of the approaches such as notch filtering and frequency Hampel is to de-noise in the frequency domain as powerline interference can be easily be identified. Another starting point of the approaches such as subtraction and adaptive filtering is to estimate the noise and remove the estimation from the original signal. The latter approaches have been shown to be more effective and introduce less signal distortion though inferior in simplicity compared to the well-known notch filtering approach [122, 123].

Another interference in this study is produced by the electrical stimulation pulses. The electrical pulses have high amplitude and can be captured by the sEMG electrodes. A similar situation is also found in the study where functional electrical stimulation is used to induce muscle contraction and corresponding joint movement [124]. Simple notch filters or a comb filter is not effective in denoising the pulse contamination in sEMG signals because of the frequency overlap of noise and signal (comparing Figure 5.1(d) and (f)). Therefore, adaptive noise cancellation is naturally an approach to be considered. Figure 5.1 shows the results of adaptive noise cancellation with the help of a noise reference. The noise reference in (b) is correlated, but not the same as the noise in the contaminated sEMG signal in (a). It was collected on the skin surface in parallel with the sEMG signal on the site where no biopotential was elicited to introduce additional noise. It can be

seen that either the repeat frequency or the amplitude of the pulses changed during the test. The noise in (a) was estimated at 500 ms step as (c) and was subtracted from it to obtain the estimated sEMG signal in (e). Figure (d), (e) and (f) show that not only was most of the pulse noise eliminated but also most of the 50 Hz powerline noise from the contaminated sEMG signal.

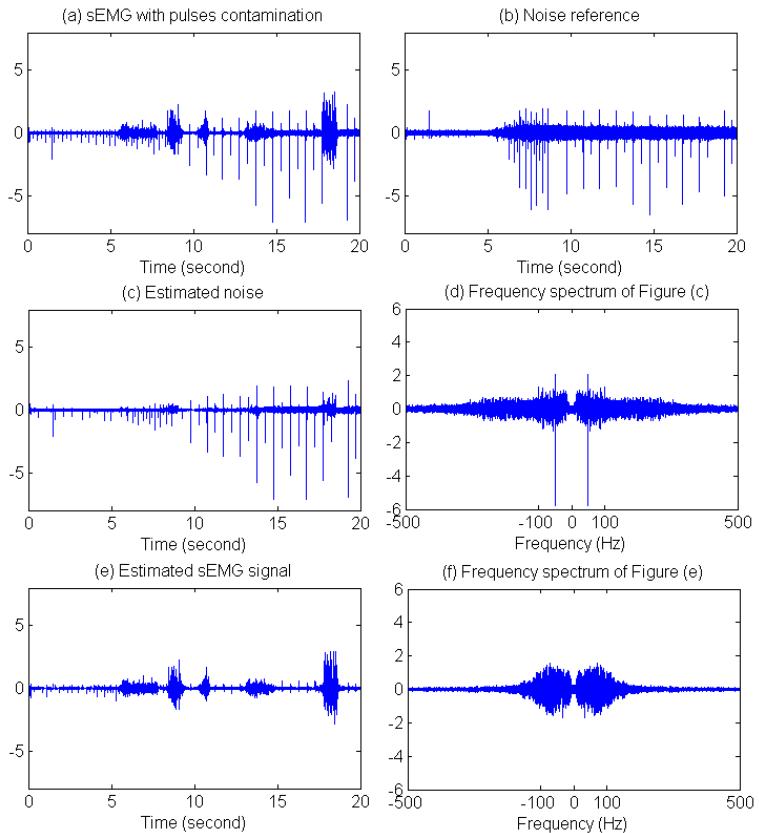


Figure 5.1: Adaptive noise cancellation to denoise electrical pulses

5.1.4 Implementation of denoising

The signal processing flows of ECG and sEMG denoising are presented in Figure 5.2. In the ECG signal, only a 10-point moving average filter was used to remove the motion artifact in the signal baseline as it is the primary artifact. An n -point moving average filter is described as Equation 5.3 and its frequency response is presented in Equation 5.4. A moving average filter is a smoothing filter or a lowpass filter. The more points involved in the averaging, the narrower the pass bandwidth is (as shown in Figure 5.3(a)).

The moving average acts well in the time domain for smoothing but its frequency response is unsatisfactory due to poor stopband attenuation and thus it is a poor low-pass filter [125] (comparing (a) to (c)). However, when applied to ECG signal, the motion artifact derived with a moving average filter is almost the same to that with a 2nd Butterworth low-pass filter when their frequency responses are tuned to be close with 1 Hz cut-off frequency ((a)-(d) in Figure 5.3). Comparatively, a highpass filter removing the motion artifact directly with the same cut-off frequency could introduce P and T wave distortion ((e) and (f)).

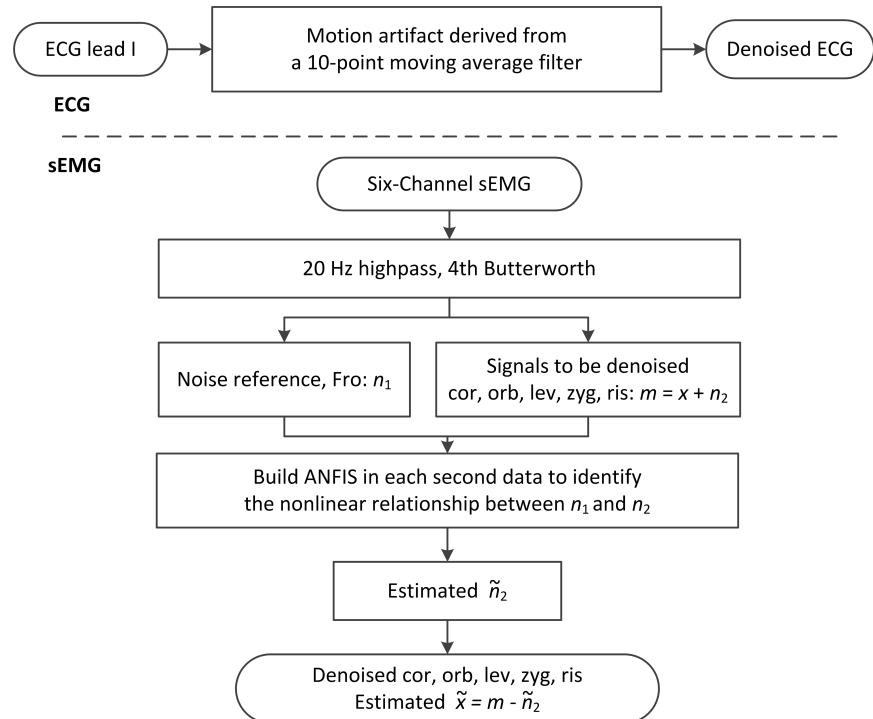


Figure 5.2: ECG and sEMG denosing, signal processing flow

$$y[i] = \frac{1}{M} \sum_{j=0}^{M-1} x[i-j] \quad (5.3)$$

$$H[z] = \frac{1}{M} \sum_{j=0}^{M-1} z^{-j} \quad (5.4)$$

The captured sEMG had motion artifact, powerline interference, and pulse interference. As shown in Figure 5.2, all the six channels first passed through a 20 Hz highpass filter to remove the motion artifact. Then the

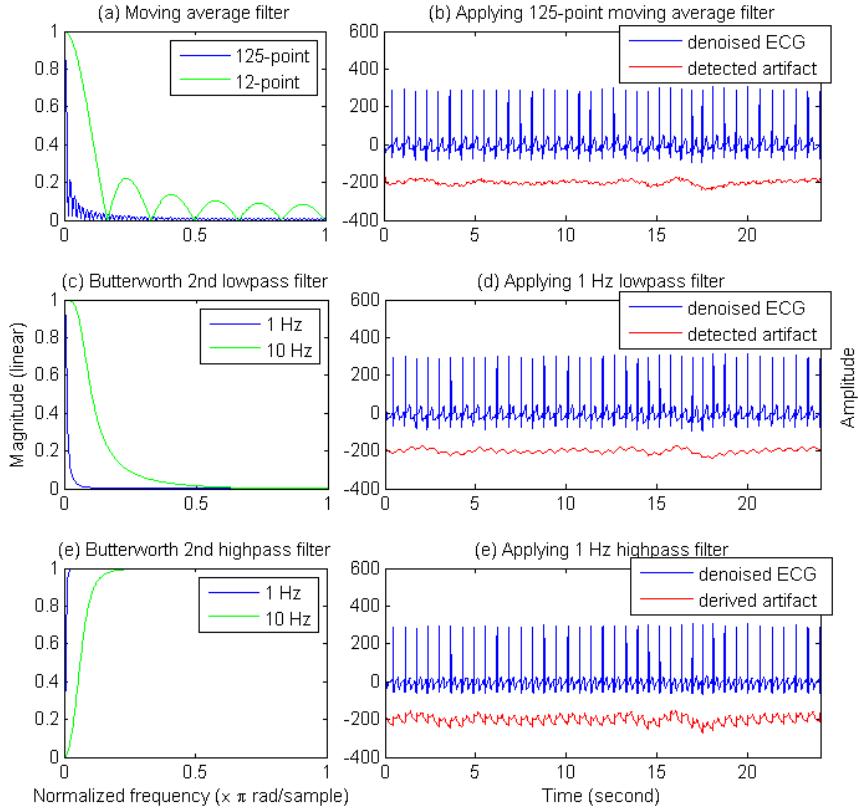


Figure 5.3: The frequency response of the filters and their time response on ECG signal

adaptive-network-based fuzzy inference system (ANFIS) [126] based adaptive noise cancellation was applied to estimate the remaining interference. Each of the five channels sEMG (cor, orb, lev, zyg, and ris) m is considered as the sum of a clean signal x and a noise n_2 . The ANFIS model can identify the nonlinear relationship between two time series. As the noise reference n_1 and the noise n_2 are related, the nonlinear relationship between them can be derived with ANFIS so as to obtain the noise estimation \tilde{n}_2 . Finally, the estimated clean signal \tilde{x} is derived by subtracting \tilde{n}_2 from m .

5.2 Feature extraction

The features extracted from the raw data are the input variables entered into the classifier in pattern recognition. The extract features may present a better data pattern in the feature space compared to the raw data. The features are domain specific and most of the time extracting features involves human expertise. In some cases, such as using an artificial neural network

and deep learning, the extraction or construction of features are integrated into the modeling process. Feature extraction/construction is one of the key steps in the data analysis process which largely impacts the success of any subsequent statistics, and information is not expected to be lost in this stage [127] (" It is always better to err on the side of being too inclusive rather than risking to discard useful information. "). On the other hand, the complexity of the pattern increases with a high dimension and noisy features, which leads to feature selection in the next stage. In this work, the feature extraction centers on the experiential ones introduced in Chapter 2, for example, heartbeat interval and the sEMG features that can quantify muscle activity.

5.2.1 R peak detection in ECG

The ECG signal is repetitive with a distinct shape pattern. Figure 5.4 presents the shape of standard ECG in one cardiac cycle and illustrates the definitions of each wave in the waveform. The wave pattern indicates the depolarization and repolarization of the heart muscle. P wave, QRS complex, and T wave represent atrial depolarization, ventricular depolarization, and ventricular repolarization, respectively. Among these waves, QRS complex detection, or R peak detection is commonly implemented to obtain the RR interval to acquire the heart rate, the HRV and sometimes ECG-derived respiration (EDR). Thus the correct extraction of R peaks is fundamental in deriving the parameters that follow.

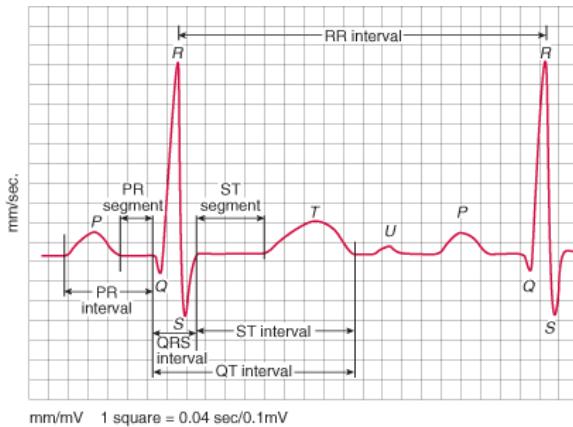


Figure 5.4: An illustration of ECG waves [3]

A QRS or R peak detection algorithm is usually a threshold-based peak detection method which is followed by an average RR interval checking rule to decide whether to keep it as a detected R peak [128, 129]. Single or double amplitude thresholds are defined by the detected noise peak level and signal

peak levels. The choice of the parameters in amplitude threshold definition is crucial to the R peak detection performance. Most of the time they are found by practical experience and their values are provided by the study where the method is proposed. In a case where the ECG shape or heart rates are irregular (e.g., arrhythmia), the parameters should be changed accordingly [128].

Table 5.2: R peak detection results and manual correction

Number of ECG file(s)	Detected peaks	Wrong/Missed detection	Correction: removed	Correction: added	Error % _o percentage
25	38685	0	0	0	0
1	1436	2	1	1	1.4
1	1627	9	4	5	5.5
1	2674	2	0	2	0.7
1	1197	1	1	0	0.8
1	2130	2	1	1	0.9

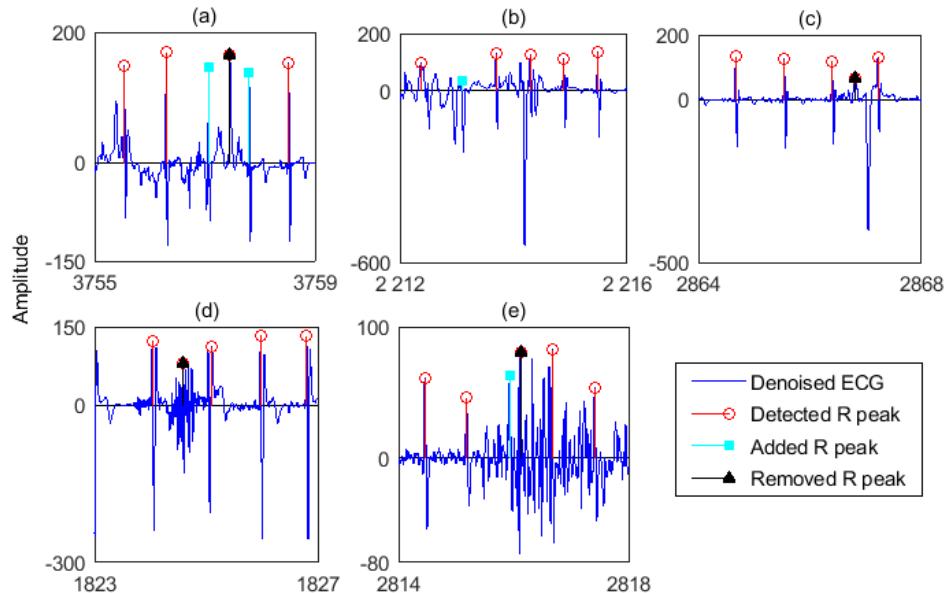


Figure 5.5: Some manual corrections for the wrong or missed R peak detections

The denoised ECG from the previous step has a high SNR in general with only a few interferences, where the amplitude of the R peak is noticeably higher than those of the other waves and the noise level in the baseline. The famous Pan-Tompkins QRS detection algorithm [128] was first implemented with the denoised ECG signal, but redundant R peak detection occurred frequently and incorrect detections happened when a sudden noise occurred.

Therefore, a simpler but more effective algorithm was proposed and implemented in this section. In the algorithm, signal peaks are detected within a time window at 10 seconds with a moving step of 8 seconds. The amplitude threshold of peak detection is defined as 0.4 times that of the peak value within the window. The second threshold is the minimum distance between two adjacent detected peaks. It is set to be 0.4 second, which is the RR interval at 150 bpm. The overlap between two data segments is to avoid missed detection on the window edge. However, the overlap area may cause redundant detection at the same location or a nearby location when the detected peak is on the edge. To solve this problem the first detected peak in each window is abandoned. In addition to this, repetitive detection is removed in the final detection result. Table 5.2 presents the error rate of the peak detection. After a visual check, it was found that most of the peaks in the recordings (99.96%) were detected correctly. In the final check phase, the incorrectly detected R peaks were removed and the missed ones were added manually. Figure 5.5 shows some incorrect or missed detected R peaks and their corrections. It can be seen that incorrect or missed detection occurs when noise with high peaks exists within the RR interval limits (e.g., (a) and (c-d)) and when the R peak is distorted and is below the amplitude threshold (e.g., (b)). The above-described steps are summarized in the flowchart in the left half of Figure 5.6.

5.2.2 ECG derived respiration

As introduced in Chapter 3, it is feasible to extract respiratory signals from ECG, PPG and ABP waveforms due to the respiratory modulation on the baseline wander, amplitude, and frequency. The respiratory signal extraction methods are classified into two categories in the review [130], the filtered-based techniques where non-respiratory frequency components are attenuated and the feature-based techniques where a beat-by-beat feature is extracted before further extraction. A feature-based method using the R peaks extracted in the previous step was implemented in this section, based on the R peak amplitude modulation. One motivation for implementing the EDR extraction is that the ECG is more reachable than wearable capacitive strip sensor, and another motivation is the overestimated breathing rates extracted from the capacitive pressure sensor waveform and provided by the Bioharness algorithm. The majority of the estimated breathing rates in one case were larger than 25, which caused a suspicion of inaccurate detection because the normal rate for an adult at rest is between 12 and 20. The respiratory signal extraction in the implementation follows the commonly used procedures and respiratory cycles counting follows the "advanced counting method" that has been proven to provide more stable results especially when data is corrupted by artifacts or irregular beats [131].

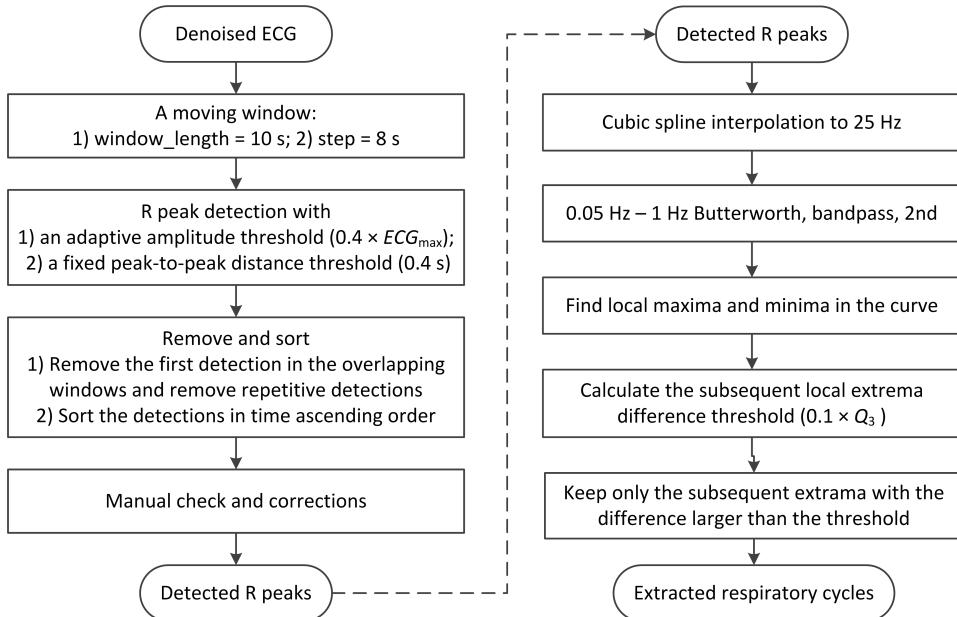


Figure 5.6: Signal processing flow in 1) R peak detection and 2) respiratory cycles extraction from the detected R peaks

The right half of Figure 5.6 presents the signal processing flow when extracting the respiratory cycles from the detected R peaks. The R-peak time series is first interpolated with a cubic spline so that the time series is evenly resampled for the following steps. Then the interpolated time series is filtered to keep only the frequencies of interest [130, 131]. Finally, respiration is detected using a threshold to eliminate the influence of spurious breaths where the pairs of max-min are conditionally counted (Q_3 is the third quartile in a distribution). The principle for making the "keep or remove" decision is that distinct respiratory cycles are categorized by the amplitude of the corresponding oscillation rather than the position of the maxima. The waveforms after each main step are presented in Figure 5.7(a)-(c). The signal segment in the figure is from a case where the breathing rate was overestimated. The estimated heart rate in (f) shows high consistency between the Bioharness reported values and the extracted values. While the reported breathing rates in this segment (e) are much higher than the extracted ones may be due to the failure in eliminating the spurious breathing cycles as shown in (d).

The breathing rates and heart rates in Figure 5.7 were calculated from R locations and peaks, and therefore the data points distributed unevenly on the timeline. The extracted values were linearly interpolated and the mean difference between them was calculated to compare the extracted values in

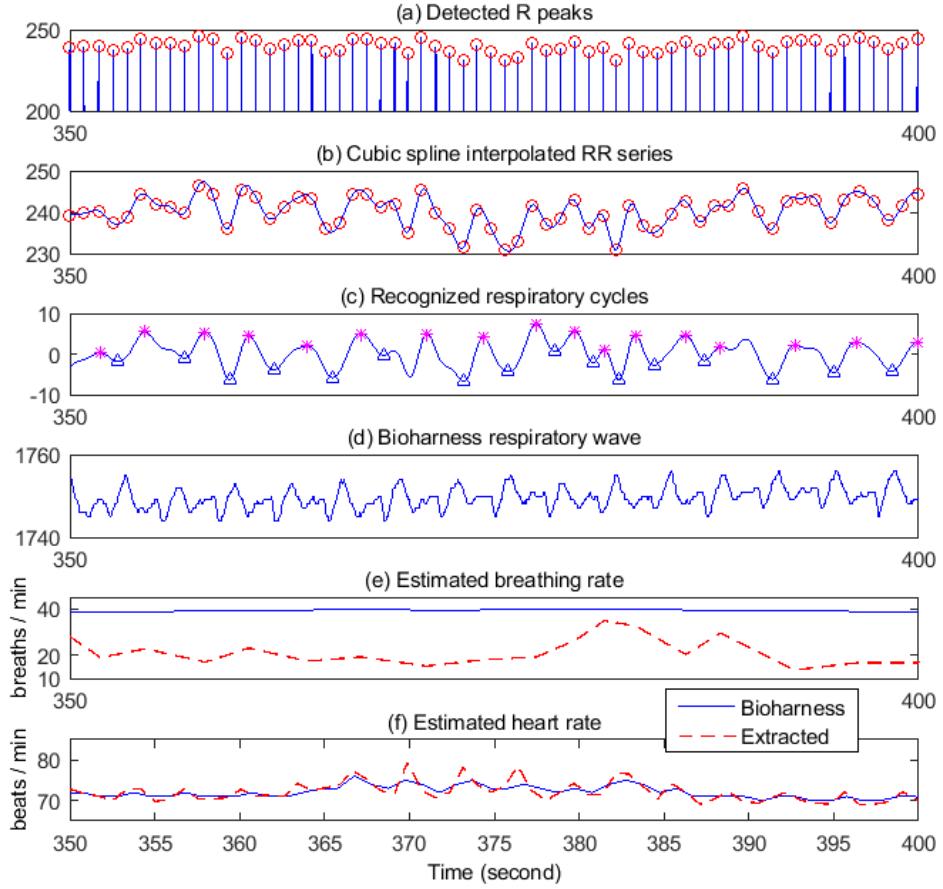


Figure 5.7: Visualization of the key steps in EDR extraction (a-c). Capacitive pressure sensor waveform (d). The comparison of the estimated breathing rate and heart rate between the reported values by Bioharness and the extracted values in this study (e-f)

this study and the reported ones by Bioharness. Figure 5.8 shows the distribution of the mean difference between the extracted values and the reported values in all the 31 study subjects with an interval of 0.5. It shows that that on average there is a small difference lies in the majority of the two heart rate results on average (± 1 beat/min, 29 out of 31; (-2.75,1.75) beat/min, 2 out of 31). However, the breathing rate results show a considerable difference and great variance in the difference (± 4 breaths/min, 24 out of 31; $\pm (4.25,8.75)$ breaths/min, 7 out of 31).

It was reported that the breathing rate (or breathing frequency) provided by Bioharness is less reliable compared to the other Bioharness parameters in a reproducibility test [82]. On the other hand, it was also noted that the performance of breathing rate extraction algorithms using ECG and PPG

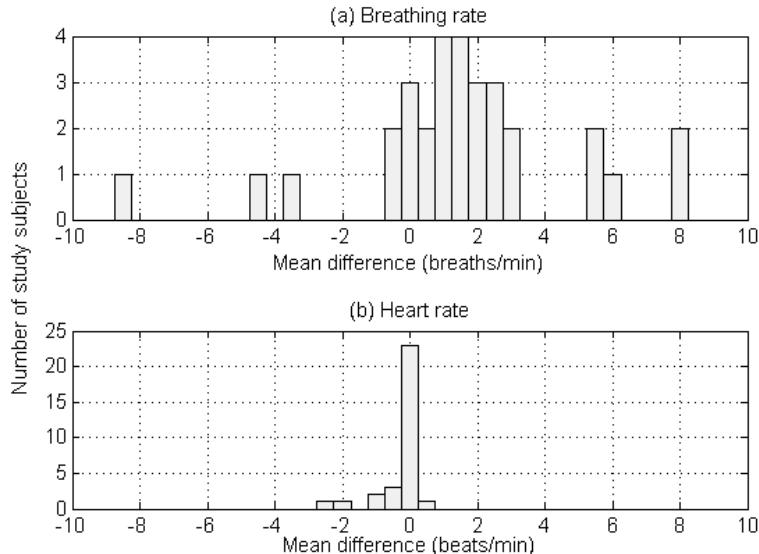


Figure 5.8: The distribution of the mean difference between the reported values and the extracted values in all the study subjects

may be gradually reduced as the true breathing rate increases because of the decrease in signal quality [132]. Since a certain degree of uncertainty lies in both signals and processing algorithms, the question of how this uncertainty impacts the next classification performance and the reproducibility of the classification results needs to be investigated. Regarding EDR itself, more algorithms (e.g., the ones that were implemented in [132]) could be implemented to find the one that suits this database best. Furthermore, a quality assessment of either signal or the extracted parameter [130] may provide such an indicator before the classification phase. The same also applies to other signals and parameters.

5.2.3 Heart rate variability

HRV is used to describe the phenomenon of oscillations between consecutive instantaneous heartbeats. As introduced before, HRV can reflect the activities of the sympathetic and parasympathetic nerves. Meanwhile, it should also be kept in mind that HRV can be affected by other factors such as cardiovascular disease states, respiration, and age [133]. HRV is measured by the variations in the RR intervals, which are usually denoted as NN intervals in contrast to the intervals calculated when erroneous, missing or ectopic beats exist. In addition to the manual heartbeat correction, automatic techniques are sporadically brought into development [134].

HRV analysis and its methods roughly fall into three categories - time domain analysis (including statistical measures and geometrical patterns), frequency domain analysis and nonlinear methods. Summaries of HRV analysis methods can be found in a comprehensive guideline in [135] and in a more recent review [133]. Table 5.3 lists most of the commonly used features in short-term (and ultra-short-term) HRV analysis. Short-term is standardized as 5 min, which is defined in contrast to the 24 h long-term analysis. Geometric methods in the time domain are not considered in this study because at least 20 min recording of NN intervals is needed to construct the geometric pattern [135].

Table 5.3: HRV features and their definitions

Feature	Unit	Description	Estimate
<i>Time domain, statistical measures</i>			
SDNN	ms	Standard Deviation of all NN intervals	Overall HRV
RMSSD	ms	Root Mean Square of the Successive Differences	Short-term variation of HRV
NNx		The number of pairs of successive NN intervals that differ by more than x intervals	estimate, high frequency variations in heart rate
pNNx	%	The proportion of NNx divided by the total number of NN intervals	
<i>Frequency domain measures</i>			
Total power	ms ²	Variance of all NN intervals	
VLF	ms ²	Power in very low frequency range	0.0033 - 0.04 Hz
LF	ms ²	Power in low frequency range	0.04 - 0.15 Hz
LF norm	n.u.	LF power in normalized units: LF/(Total power - VLF) × 100	
HF	ms ²	Power in high frequency range	0.15 - 0.4 Hz
HF norm	n.u.	HF power in normalized units: HF/(Total Power - VLF) × 100	
LF/HF		Ratio of LF [ms ²]/HF [ms ²]	
LF _{peak} ,HF _{peak}	Hz	LF and HF band peak frequency	
<i>Nonlinear methods</i>			
ApEn/SampEn		Approximate/Sample entropy, measures of quantifying the regularity or predictability of time series data; Higher values indicate the normal heart signals and smaller values indicate the cardiac abnormal signals.	

Among the statistical measures in the time domain, the measures derived from interval differences including RMSSD, NNx, and pNNx are highly correlated because they all estimated high-frequency variations in heart rate and RMSSD is preferred over the other two due to its better statistical properties [135]. In a long-term analysis of 24 h, the time domain features are strongly correlated to the frequency domain features (e.g., SDNN - total power, RMSSD - HF, and pNN50 - HF). These theoretical high correlations among the time domain measures and with HF was also observed in our pre-

vious principal component analysis in [110], where ultra-short-term analysis less than 1 min was applied. The time windows smaller than 5 min, which were mostly equal to and smaller than 1 min, are considered as ultra-short-term in HRV analysis. Since slower regulatory mechanisms contribute to longer-term analysis, the three terms of HRV analysis values are not interchangeable [136]. The choice of HRV analysis window and NN time series segmentation will be discussed later.

The frequency domain methods were preferred due to providing more easily interpretable results in terms of physiological regulations compared to time domain methods [135]. Several methods are frequently found to estimate power spectral density in HRV analysis, including the parametric method, the autoregressive model (AR) estimation, and non-parametric ones including Welch's method, fast Fourier transform (FFT) with resampling and the Lomb-Scargle periodogram without resampling. The advantage of using the AR model is a better spectrum resolution when short data frames are used, although choosing an optimum model order is a difficult issue. The difference ranges of the model order were recommended to be between 8 and 25 in different studies [135, 137, 138] and normalized HRV frequency features may not be significantly affected when choosing different orders within the range [138]. The implementation of an HRV feature extraction relies on an open-source Matlab toolbox PhysioZoo [137], which enables the standardization and reproducibility of HRV analysis in mammalian models.

5.2.4 sEMG

Many features extracted from the sEMG signal have been studied thoroughly due to the boom of myoelectric control applications. The time domain of the sEMG features fall into four categories according to their mathematical properties [139]: 1) calculations based on amplitude values of the sEMG signal, which include the features that contain signal energy information (e.g., RMS) and the features that contain signal complexity information (e.g., WL¹); 2) features contain frequency information (e.g., ZC and SSC) where a threshold parameter needs to be pre-defined; 3) coefficients of the prediction model (e.g., AR); and 4) features extracted from two adjacent or several segments of an sEMG signal (e.g., MAVS). In addition to these, there are also frequency domain features extracted from power spectral density. Time domain features are more commonly found in sEMG pattern recognition studies and the frequency domain features did not show better performance compared to the time domain ones [141–144].

Among all the proposed sEMG features, no single one is unanimously recommended as the best representative of sEMG signal across applications

¹In BioPatRec platform [140], WL is calculated from FFT spectrum and therefore is a frequency feature.

Table 5.4: sEMG features and their descriptions

Feature	Mathematical description	Description
MAV	$= \frac{1}{N} \sum_{i=1}^N x_i $	mean absolute value
MAVS	$= \text{MAV}_{k+1} - \text{MAV}_k$	mean absolute value slope
RMS	$= \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$	root mean square
MPV	$= \max x_i $	maximum absolute value, absolute peak
LD	$= e^{\frac{1}{N} \sum_{i=1}^N \log(x_i)}$	log detector
WL	$= \sum_{i=1}^{N-1} x_{i+1} - x_i $	wave length
ZC	$= \sum_{i=1}^{N-1} [\text{sgn}(x_i \times x_{i+1}) \cap x_i - x_{i+1} \geq \text{threshold}]$ $\text{sgn}(x) = 1, if x \geq \text{threshold}; = 0, otherwise$	zero crossing, the number of times that the signal changes sign
SSC	$= \sum_{i=2}^{N-1} [f[(x_i - x_{i-1}) \times (x_i - x_{i+1})]]$ $f(x) = 1, if x \geq \text{threshold}; = 0, otherwise$	slope sign change, the number of times that slope of the signal changes sign
AR	$x(n) = -\sum_{k=1}^p a_k x(n-k) + e(n)$	the AR coefficients (a_k) in the signal's AR model of order p
DAMV	$= \frac{1}{N} \sum_{i=1}^{N-1} x_{i+1} - x_i $	difference absolute mean value
FMN	$= \sum_{j=1}^M f_j P_j / \sum_{j=1}^M P_j$	mean frequency
FMD	$\sum_{j=1}^M P_j = \sum_{j=\text{MDF}}^M P_j = \frac{1}{2} \sum_{j=1}^M P_j$	median frequency
FPK	$f @ \max(P_j)$	peak frequency
where x_i represents the signal in a segment i , N represents segment length		
k represents the sequence number of the feature		
f_j represents the frequency of the spectrum at frequency bin j		
P_j represents the power spectrum at frequency bin j		

and scenarios. For example, MAV and WL were recommended by some hand gesture studies [139, 141], but WL was found to have a weak performance in recognizing facial gestures [142, 143]. Moreover, other factors such as the sampling rate of the signal could influence the classification performance of the same feature so that a new feature is introduced to compensate [142, 144]. This study includes the sEMG features in all the four categories in the time domain and some in the frequency domain. Table 5.4 [139, 145] lists the names and descriptions of the involved sEMG features. Each of the features is extracted from each sEMG signal segment within a sliding time window and there are N signal samples. The Matlab implementation is based on the definitions in [139] and an open-source toolbox BioPatRec [146].

5.2.5 GSR

As previously introduced, the GSR signal is composed of the tonic level, SCL, and the faster phasic change, SCR. The SCRs that can be attributed to a specific eliciting stimulus are defined as event-related SCRs, while those without an identifiable eliciting stimulus are defined as non-specific SCRs. For the event-related SCRs, the latency time between the stimulus and the response is one parameter. The other parameters are the descriptions of the SCR or SCL's amplitude, frequency, shape (e.g., rise-time and recovery-time), and area under the curve [106]. The typical values of the GSR parameters provided by Biopac guide [64] are:

- The latency period between stimulus onset and the first significant deviation in the signal: 1-3 seconds (event-related SCR)
- The minimum deflection in the signal that is considered as an SCR: 0.01-0.05 μ S, or the amplitude resolution value of the device
- SCR amplitude range: minimum deflection to 2-3 μ S
- SCL amplitude range: 2-16 μ S
- SCR frequency: 1-3 (and is possible to be around 10) per minute during rest, 20-25 during high arousal (non-specific SCR)

Figure 5.9 shows the SCRs detection using a biosignal processing tool provided by [147]. The start and peak of an SCR in (b) are located by the two zero-crossing points in (a). In each time window, the features extracted from GSR are listed in Table 5.5. In addition to the classic trough-to-peak methods in GSR signal decomposition, mathematical deconvolution models were built to better treat the superposed SCRs [148, 149] or be more robust to signal noise due to motion and measurement [150].

Table 5.5: GSR features and their descriptions

Feature	Description
GSR mean	The average of the GSR values
GSR standard deviation	The standard deviation of the GSR values
GSR AUC	The area under the GSR signal curve
Number of SCRs	The number of detected SCRs in the time window
SCR rise-time mean	The average of the SCRs rise-time (from start to peak in GSR signal)
SCR amplitude mean	The average of SCR amplitude (amplitude difference between start and peak in GSR signal)

5.2.6 Window length for feature extraction

Feature extraction from time series is on the basis of the signal segmentation. To provide a continuous estimation of pain intensity, a sliding time window with a fixed length is usually used in time series segmentation. However, the features cannot be synchronously updated from all the signals because

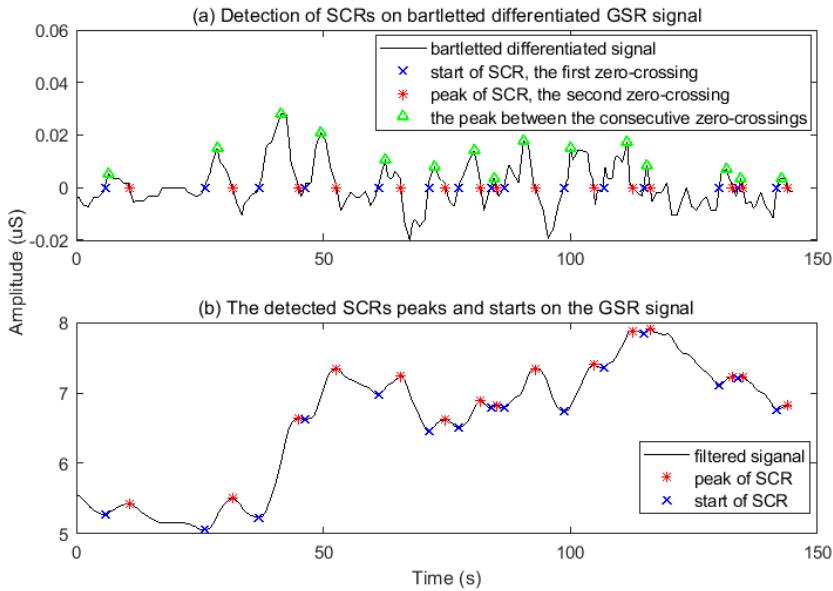


Figure 5.9: Decompose GSR into tonic SCL and phasic SCR

each feature or signal has a different requirement on the window length. The minimum required window length of the features are summarized in Table 5.6 from literature. The features mentioned in the previous parts are denoted in the rest of the thesis with the format `<signalName>_<featureName>`.

The choice of the time window for HRV features in automatic pain assessment studies presents a dilemma. On the one hand, the pain stimulation period is usually short when causing phasic pain which equal or less than 5 s [28, 158]. On the other hand, the ultra-short-term HRV features, other than the nonlinear ones [159], have hardly been validated in terms of validity reliability as being good surrogates for the short-term ones [152]. The spectral analysis especially, is recommended to be performed in a recording lasting at least 10 times that of the lower frequency so as to ensure the stability of the signal [152, 153], and therefore the window length for `ecg_vlf`, `ecg_lf`, and `ecg_hf` are recommended to be 50 min, 250 s, and 1 min, respectively. Comparatively, two time-domain features were validated with a much shorter length, where 30 s or the average of three 10 s interval features is recommended regarding `ecg_rmssd` and `ecg_sdnn` due to their strong correlation with the 5 min measurement [151, 160].

In continuous GSR analysis, no clear principle can be found as regards choosing the window length. A typical value could be 10 s or 15 s [30, 66]. The stimulus duration length also needs to be considered when it is known. The average duration of SCRs is reported to be less than 2 s (when sampling

Table 5.6: The recommended minimum time window length in feature extraction

Feature	Principle	Min length
ecg_sdnn	strong correlation with 5 min measurement [151]	30 s
ecg_rmssd		10 s
ecg_nn50	-	2 min [136]
ecg_vlf		50 min
ecg_lf	at least 10 times of the lower frequency [152, 153]	250 s
ecg_hf		1 min
ecg_apen	sensitivity to the input parameters [154]	data length
ecg_sampen		$N \geq 200$
Feature	Typical length in previous studies	
emg_mav	- 256 ms during beginning of the movement or the steady-start part [139, 143]	
emg_mavs		
emg_rms	- non-overlapped 256 ms during an active facial expression [142]	
emg_mpv	- classification error decreased with longer window lengths	
emg_ld	from 50 to 550 ms, optimal to be between 150 and 250 ms with	
emg_wl	with acceptable controller delays [155]	
emg_zc	100 ms bins [156]	
emg_ssc		
emg_ar		
emg_damv		
emg_fmn		
emg_fmd		
gsr_mean	- 10 s [30], the mean of the signal amplitude is considered as SCL	
gsr_std	- 6 s [157]	
gsr_auc	- 15 s window updated each second [66]	
gsr_nscr		
gsr_scr_rt		
gsr_scr_apavg		

rate is 32 Hz) [149], which can be considered as the extreme minimum window length to extract the number of SCRs. This could explain the results in [66] where the average number of skin conductance peaks per second is smaller than 0.4. A similar definition, the number of skin conductance fluctuations (NSCF) is used in some studies [31, 66] in the unit of number-per-minute. The reported SCRs duration, on the other hand, indicates that the sampling rate of GSR in this study (1 Hz) may not be sufficient to capture the subtle phasic component in GSR.

Regarding sEMG signal analysis, the conventional choice of window length is around 250 ms (200, 250, or 256 ms [139, 142, 143]), or as short as 100 ms [156]. In continuous sEMG analysis in myoelectric control especially, or in similar human-computer interaction applications, the same window length is chosen by taking into consideration both classification error and controller delay [155, 161]. However, it was also pointed out that sEMG features could have different sensitivity to changes in window length, where manipulating

window length has a greater effect on reliability and inter-individual variability for peak EMG than mean EMG [162].

The window length differences in the feature extraction summarized in Table 5.6 indicate that the features would update asynchronously if several physiological signals are involved in the multimodal pain assessment and when mimicking a continuous assessment.

5.2.7 Inter-subject and intra-subject variability, and signal or feature normalization

The variability is multidimensional in this study. First, the variability lies in the pain sensitivity and pain tolerance regarding pain perception, which is to say the subjectivity of pain. In the electrical tests, where pain tolerance was reached most of the time, the self-report VAS score varied among the subjects, as shown in Figure 4.4(a), which indicates the inter-subject variability in pain tolerance. In the electrical tests, the difference in the VAS score from the same person was 0.59 on average, and the difference was 0.62 for the thermal tests.

Another variability dimension is in signal level, which could be understood as the difference in signal amplitude range or parameter value range. In the GSR signal, the range of SCL is different across subjects and tests due to individual differences such as age, gender and culture, and environmental variables such as temperature and humidity. Thus, it is recommended to use the relative response measures after transformation based on an individual's signal range or using z-score standardization [64, 163]. Similar to the GSR signal, the sEMG signal is also required to be normalized when comparing across time, between individuals, or between muscles. There are multiple methods for sEMG normalization using difference normalization reference values, however there is no wide agreement on which is the best one [164]. Some examples are being normalized with respect to a defined baseline [156, 165], to the individual mean amplitude [166], and using the min-max normalization on signal [167] or features [139].

In HRV analysis, it was pointed out that R-R interval or ECG analyses should be normalized to eliminate its mathematical dependence on HR [168, 169] so that the variation of an HRV feature would not just represent the variation of HR (both within-subject-wise and between-subject-wise). Hallstrom A. *et al.* [168] proposed to rescale the R-R interval to an HR of 75 (or equivalently an average interval of 800 ms) and the correlations between HRV features were halved due to the normalization. Similarly, normalizing the R-R intervals was suggested by dividing by the average value [169]. In this study, the HRV features that are normalized to an HR of 75 are denoted with a prefix 'n'(e.g., *ecg_nsdnn*). The frequency features in the HRV analysis have their own defined normalization, where *ecg_lf* and

ecg_hf are divided by their sum. The normalized frequency features are denoted as *ecg_lf_nu* and *ecg_hf_nu*. They are prone to providing more exchangeable results across many implementation differences such as the interpolation density and the transformation algorithm from time domain to frequency domain [170]. However, these two normalized features are predictable from each other as well as from *ecg_lf/hf* due to their mathematical definitions. In other words, once any one of these three features is specified, the other two are completely determined. There is another commonly found HRV feature transformation, which is calculating the feature's natural logarithm. This is more commonly used in statistical analysis to modify the skewed distribution in the HRV features (e.g., [171]). The logically transformed features are usually denoted with an 'ln' prefix (e.g., *ecg_lnrmssd*).

The two variability dimensions discussed above are manageable respectively through study design (i.e., all the subjects are supposed to reach their self-defined pain tolerance) and signal/data processing (i.e., signal/feature normalization). The third variability dimension, which is the variability or consistency of the physiological signals' response to pain, will be observed and be further recognized using machine learning techniques in later analysis.

5.3 Summary and future work

This chapter presents the rationales, methods, and implementation results in processing and feature extraction of the collected signals. Some of the chosen methods are basic rather than advanced due to a compromise between research width and depth. More profound research or implementation with more sophisticated techniques can be conducted when addressing each sub research question mentioned in this chapter. Future work may target on robust processing methods for more complicated cases (e.g., irregular ECG and unexpected noises). Taking ECG processing and analysis as an example:

- The impact of the ECG sampling rate on HRV analysis and further on pain analysis can be investigated to see whether upsampling is needed [172].
- Manual or any R peak correction is unfeasible in long recordings and low-quality signals. A signal index (e.g., [173]) may be introduced to help choose analyzable segments.
- For extracting respiratory rate from ECG signals in complex but real situations (e.g., with artifacts or atrial fibrillation), more robust extraction techniques [174–176] can be considered.

Chapter 6

Features and Visualization

Based on the discussion in the previous parts, the chosen physiological signal features are summarized in Table 6.1 together with analysis window length and feature update time step.

Table 6.1: A summary of features of interest

Feature	Window length	Update time step
ecg_nsdnn	30 s	1 s
ecg_nrmssd	10 s	1 s
ecg_nhf	60 s	1 s
ecg_nlf	250 s	1 s
ecg_lf/hf	250 s	1 s
ecg_sampen	250 s	1 s
ecg_hr	5 s	1 s
respr_<source>	20 s	1 s
emg_<feature>	250 ms	250 ms
gsr_scl	1 s	1 s
gsr_t_<tonic_feature>	15 s	1 s
gsr_p_<phasic_feature>	15 s	1 s

The feature observation in this part is to provide hints to these questions:

- How does each feature react to the stimulation since t_0 ?
- How does each feature react to the pain sensation induced by the stimulation since t_1 ?
- Does each feature have the potential to differentiate pain intensities across subjects?

The tests were divided into three groups based on the final self-reported VAS/NRS; the three groups were VAS[4,6), VAS[6,8) and VAS[8,10]. The number of the tests that fall into each pain category were 25, 61 and 36, respectively. The average curves in the three categories were then extracted and compared. One attribute of this database is that the length of each period (i.e., from t_0 to t_1 and from t_1 to t_2) was loosely controlled. However, the *pain threshold* was uniformly defined and the time of reaching the *pain*

threshold was reported as t_1 in all the tests. Moreover, the primary interest of this study is the pain sensation induced by pain stimulation rather than other responses to the stimulation. Therefore, the time axis of all the tests is aligned at t_1 in the observational plots, which means the new time axis shows the relative time to t_1 in seconds (relative time = $t - t_1$). Then, the median time offset of t_0 and t_2 to t_1 is calculated and marked in the plots to further observe how the feature changes in reaction to the stimulation.

6.1 GSR features

The GSR values were missing when the measurement electrode pair was moved from one side of the hand to the other, and thus were denoted as NaNs. The signal gaps formed by the missing values naturally separate one recording into four segments and each contains one test as presented in Figure 4.1, from before t_0 and until after t_2 . A decomposition algorithm, cvxEDA [148] is used for deriving the tonic component and phasic component from the GSR signal, which is based on Bayesian statistics, mathematical convex optimization, and sparsity. In the model, the GSR signal is considered as a sum of three parts: tonic component (t), phasic component (r) and an additive Gaussian noise term (ϵ). Meanwhile, the GSR signal is also considered as a convolution process between SudoMotor Nerve Activity (SMNA) and the Impulse Response Function (IRF), where SMNA is a sum of Driver_{tonic} and Driver_{phasic} (p). The modeling structure is crystallized with Equation 6.1.

$$\text{GSR} = r + t + \epsilon = \text{SMNA} * \text{IRF} + \epsilon = (\text{Driver}_{\text{tonic}} + \text{Driver}_{\text{phasic}}) * \text{IRF} + \epsilon \quad (6.1)$$

The adjustable parameters in the model are set as $\tau_0 = 2$, $\tau_1 = 0.7$, $\delta = 10$, $\alpha = 0.016$, $\gamma = 0.01$ and solver = quadprog, where most are defined as the same as [148] except for α . A larger α value is chosen in this study to yield a sparser spikes estimation due to the sparser signal sampling. One subject with two heat tests was excluded in the visualization of this part due to too many missing values. The number of tests in each pain category therefore changed into 25, 58 and 35.

The GSR features are listed in Figure 6.1 and the linear correlation between every feature pair are presented with Pearson's correlation coefficients. The features are defined based on Table 5.5. The very strong linearly correlated features (Pearson's $r > 0.8$) are not repetitively visualized. The tonic features to observe includes 1) *scl*, 2) *t_auc*, and 3) *t_std*; and the phasic features to observe are 1) *p_rise_time_avg*, 2) *p_num_pk*s, 3) *p_pk_amp_max* and 4) *p_auc*. All the features are extracted from amplitude standardized GSR with z-score within test, except for *scl*.

p_rise_time_avg	1	0.33	0.31	0.13	0.087	0.15	0.23	0.12	-0.051	0.35	0.36	0.35
p_num_pk	0.33	1	0.51	0.24	0.17	0.18	0.33	0.21	-0.0041	0.43	0.48	0.45
p_pk_amp_max	0.31	0.51	1	0.1	0.011	0.29	0.46	0.067	0.073	0.88	0.87	0.89
t_max	0.13	0.24	0.1	1	0.98	-0.0031	0.23	0.98	-0.051	0.1	0.13	0.11
t_avg	0.087	0.17	0.011	0.98	1	-0.14	0.043	0.97	-0.064	0.0047	0.034	0.019
t_auc	0.15	0.18	0.29	-0.0031	-0.14	1	0.63	0.073	0.0059	0.3	0.3	0.29
t_std	0.23	0.33	0.46	0.23	0.043	0.63	1	0.16	0.013	0.5	0.49	0.5
scl	0.12	0.21	0.067	0.98	0.97	0.073	0.16	1	-0.065	0.062	0.094	0.075
p_auc	-0.051	-0.0041	0.073	-0.051	-0.064	0.0059	0.013	-0.065	1	0.046	-0.092	0.028
p_std	0.35	0.43	0.88	0.1	0.0047	0.3	0.5	0.062	0.046	1	0.91	0.98
p_avg	0.36	0.48	0.87	0.13	0.034	0.3	0.49	0.094	-0.092	0.91	1	0.95
p_max	0.35	0.45	0.89	0.11	0.019	0.29	0.5	0.075	0.028	0.98	0.96	1

Figure 6.1: The GSR feature list and the linear correlation (Pearson's r) between every two features

6.1.1 Tonic component

The decomposed tonic component (gsr_scl) in three pain categories are visualized in Figure 6.2 where the amplitude of the curves are normalized by subtracting the amplitude value at t_1 so that the amplitude value at time zero of all the curves are zero. The three average curves are compared in Figure 6.2(d) with median t_0 and median t_2 , which all nearly increase on a monotone from the start of the stimulation. The highest pain category has the greatest slope before the *pain threshold* and greater than VAS[4,6] after that. Moreover, interestingly, the peak amplitude of the three pain categories between t_1 and t_2 is in direct proportion to the pain intensity. The amplitude peaks are located at 13th s, 16th s and 20th s, which are all before reaching the median t_2 . This may have occurred due to the longer stimulation duration between t_1 and t_2 when higher pain intensity is reported (median: 21th s, 28th s and 40th s, respectively). It also indicates that the tonic component amplitude could stop increasing and drop even though the

stimulation or pain sensation continues. These observations are consistent with the result in [31].

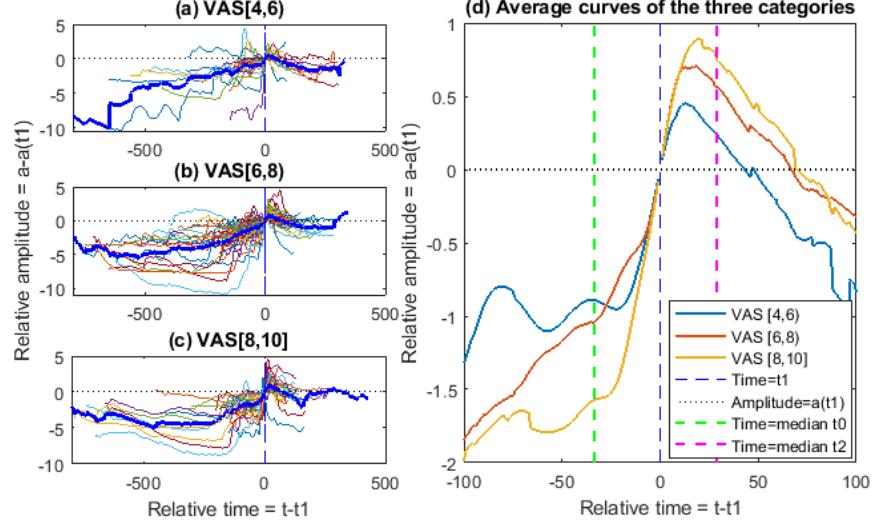


Figure 6.2: The tonic component of all the tests in each pain category normalized to t_1 and their average curve (a-c); a comparison of the average curves (d).

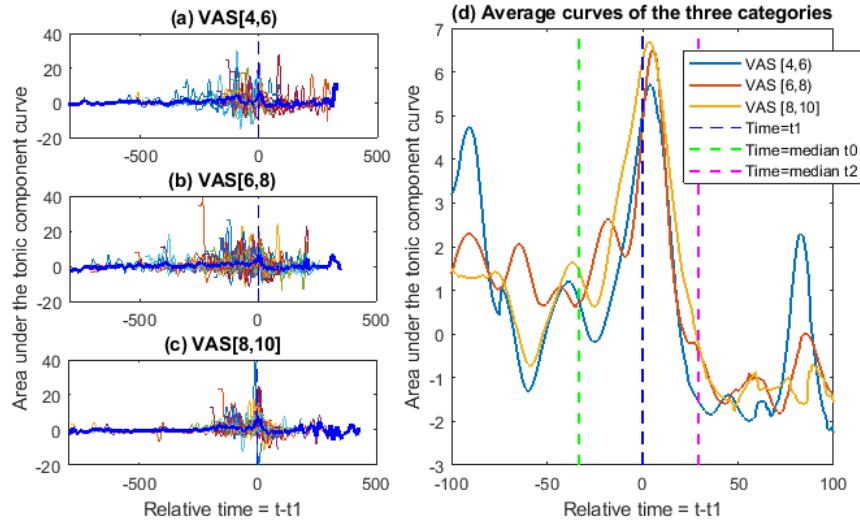


Figure 6.3: Area under the tonic component (t_auc)

The two tonic features, t_auc and t_std , extracted from standardized t with a sliding 15 s time window are plotted in Figure 6.3 and Figure 6.4, respectively. The average curves indicate the start of the pain sensation

better than scl . Similar to scl , the peak amplitude of these two tonic features may also help differentiate pain intensities.

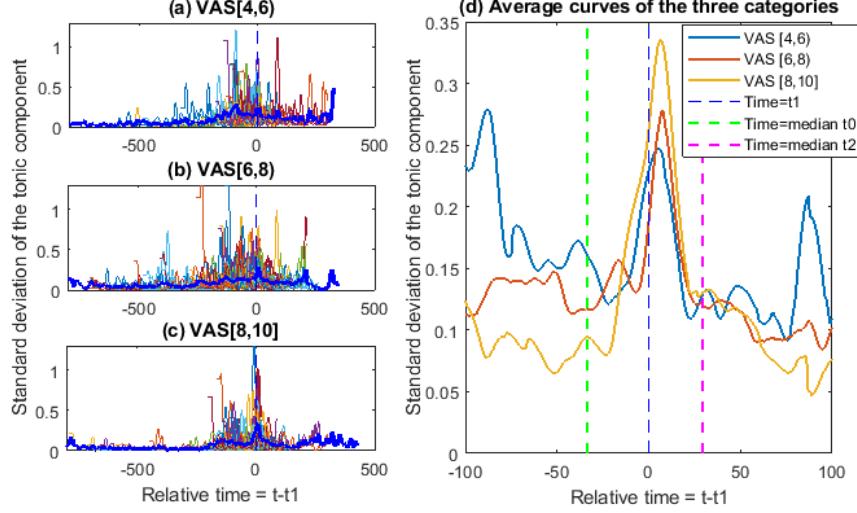


Figure 6.4: Standard deviation of the tonic component (t_std) in each pain category and a comparison of their averages

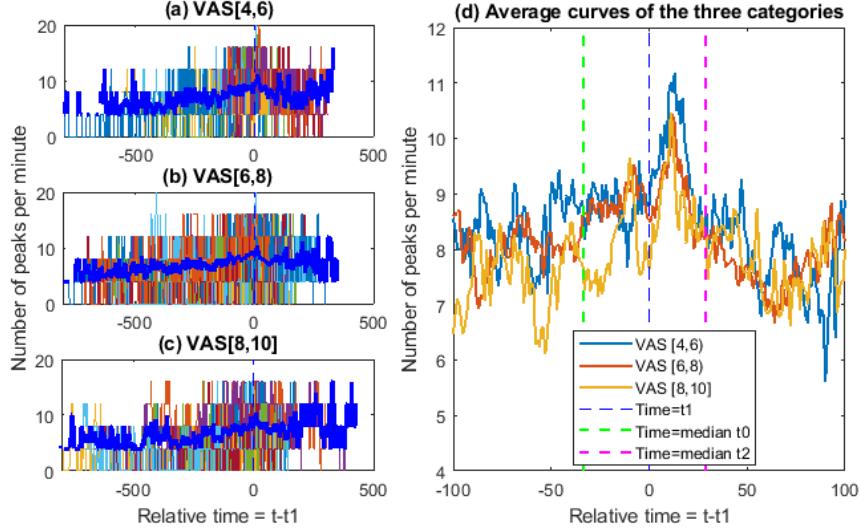


Figure 6.5: Number of phasic driver peaks per minute (p_num_pk)

6.1.2 Phasic component

One phasic feature, p_num_pk , was observed in other studies as well. For example, in the model that applies to infants and postoperative patients,

peaks per second was concluded to increase along with pain intensity ([66]: NRS 0, 1-3, 4-5, 6-10, and [177]: NRS \leq 3, NRS $>$ 3, or NRS \leq 5, NRS $>$ 5) . In the study with tonic experimental heat stimulus [31], the average NSCF per minute was observed to increase as pain stimulation continued, and most of the pain categories could be differentiated (no pain, low-NRS 30, medium-NRS 60, high-NRS 90) except for differentiating between low and medium.

The average trend of p_num_pks in the three pain categories are presented in Figure 6.5. The figure shows that the number of peaks slightly increased on average after t_1 . However, the difference between the average curves is not clear enough to distinguish pain categories. Some other phasic features were checked as well, including $p_rise_time_avg$, $p_pks_amp_max$, p_auc and p_std . However, they are not presented with the details as they could not differentiate either pain from no pain or pain intensities.

One reason for the inconsistency with the literature could be the sparse signal sampling, which may lead to missing or inaccurately decomposed phasic peaks. The 1 Hz sampling rate is able to capture the SCRs that last 6 s or longer but is unable to capture the small fluctuations that last for 2 s or even shorter [149]. In a postoperative pain study with between-group comparison [178], it was concluded that patients having moderate pain (NRS 4-5) have 0.28 NSCF per second on average (16.8/min) and those with severe pain (NRS 6-10) have 0.33 NSCF per second on average (19.8/min)); these are higher than the average p_num_pks in this study as shown in Figure 6.5(d). On the other hand, the p_num_pks before t_0 in this study is higher than 4.2/min in the no pain group in [178].

6.2 ECG features

The ECG features observed are listed in Table 6.1. All the 122 tests are included in the visualization. However, the feature values during the beginning period of a recording are represented as "NaN" due to the wide analyzing time window. The feature extraction was implemented using PhysioZoo toolbox [137]. Lomb-Scargle periodogram is used to estimate power spectral density in extracting frequency domain features.

The average value of ecg_hr in each subject was normalized to 75 bpm to diminish the influence of the individual difference on average heart rate. Similarly, the average breathing rate was normalized to 20 breaths per minute in each subject. A scalar λ was introduced for the normalization purpose [168] where the values of the feature, X , are rescaled to Y by $Y = \lambda \times X$. Four scalar values were defined, and they are:

$$\lambda_1 = \frac{75}{\text{subject_mean_hr}} \quad (6.2)$$

$$\lambda_2 = \frac{800}{\text{subject_mean_nn}} \quad (6.3)$$

$$\lambda_3 = \left(\frac{800}{\text{subject_mean_nn}} \right)^2 \quad (6.4)$$

$$\lambda_4 = \frac{20}{\text{subject_mean_respr}} \quad (6.5)$$

where λ_1 was used in heart rate; λ_2 was used in features with the unit "ms" (e.g., $sdnn$); λ_3 was used in features with the unit " ms^2 "; λ_4 was used in respiration rate.

	hr	1	-0.087	-0.21	-0.0057	-0.12	0.074	-0.2	0.2
sdnn	-0.087	1	0.65	0.59	0.66	-0.12	-0.15	0.015	
rmssd	-0.21	0.65	1	0.44	0.52	-0.22	0.052	0.0013	
lf	-0.0057	0.59	0.44	1	0.6	0.097	-0.13	-0.032	
hf	-0.12	0.66	0.52	0.6	1	-0.19	-0.0029	-0.057	
lf/hf	0.074	-0.12	-0.22	0.097	-0.19	1	-0.27	0.032	
sampen	-0.2	-0.15	0.052	-0.13	-0.0029	-0.27	1	-0.098	
respr	0.2	0.015	0.0013	-0.032	-0.057	0.032	-0.098	1	
	hr	sdnn	rmssd	lf	hf	lf/hf	sampen	respr	

Figure 6.6: The ECG feature list and the linear correlation (Pearson's r) between every two features

Pearson's correlation coefficients between the features are presented in Figure 6.6. Most of the extracted features show weak ($r=0.2-0.39$) or very weak ($r=0-0.19$) linear correlation to the other features except for the ones between $sdnn$, $rmssd$, lf and hf where the linear correlation is either strong ($r=0.6-0.79$) or moderate ($r=0.4-0.59$).

6.2.1 Heart rate

The average curves in Figure 6.7(d) show the heart rate responses to pain stimulation in general, and the pattern in reaction to pain stimulation is quite similar in all the three pain categories. The heart rate first rises since the pain stimulation is present, then stays near the peak value when the pain sensation continues, and starts to drop when the stimulation is removed. The

peak heart rate value in the average curves and the recovering speed after the stimulation was removed show it was inversely proportional to pain intensity with subtle differences.

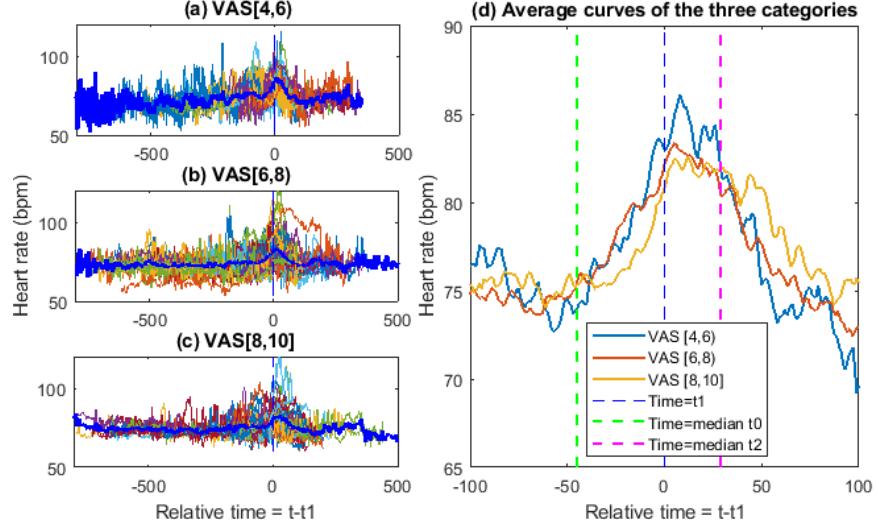


Figure 6.7: Heart rate (ecg_hr) rescaled with λ_1

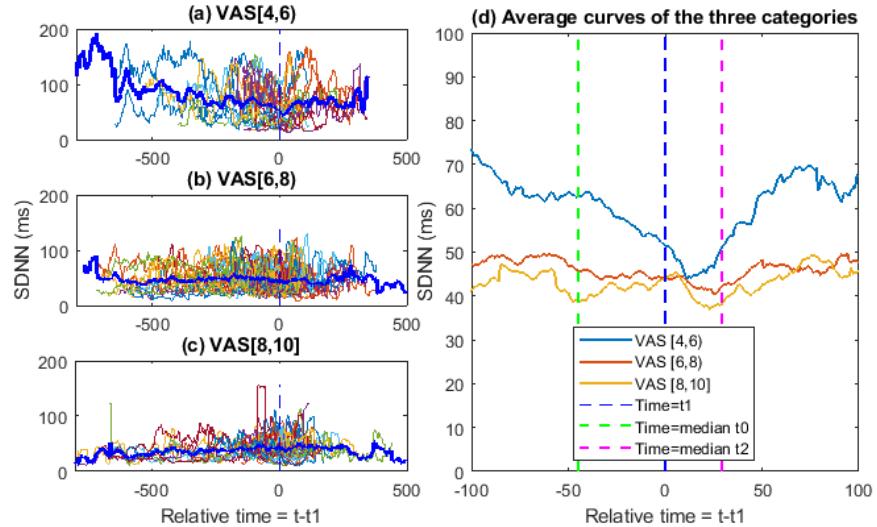


Figure 6.8: SDNN (ecg_nsdnn) rescaled with λ_2

6.2.2 Heart rate variability

The four moderate to strong linearly correlated features, ecg_nsdnn , ecg_nrmssd , ecg_nlf , and ecg_nhf are plotted from Figure 6.8 to Figure 6.11. The

response to the pain stimulation can hardly be observed from the average curves of the pain categories VAS(6,8] and VAS[8,10]. By contrast, the category VAS(4,6] responded in all the four features (decreased in ecg_nsdnn , ecg_rmssd and ecg_nhf ; increased in ecg_nlf). The observed ecg_lf/hf in Figure 6.12 and ecg_sampen in Figure 6.13 show a slight decrease on average to pain. However, the change is minor compared to the feature value range.

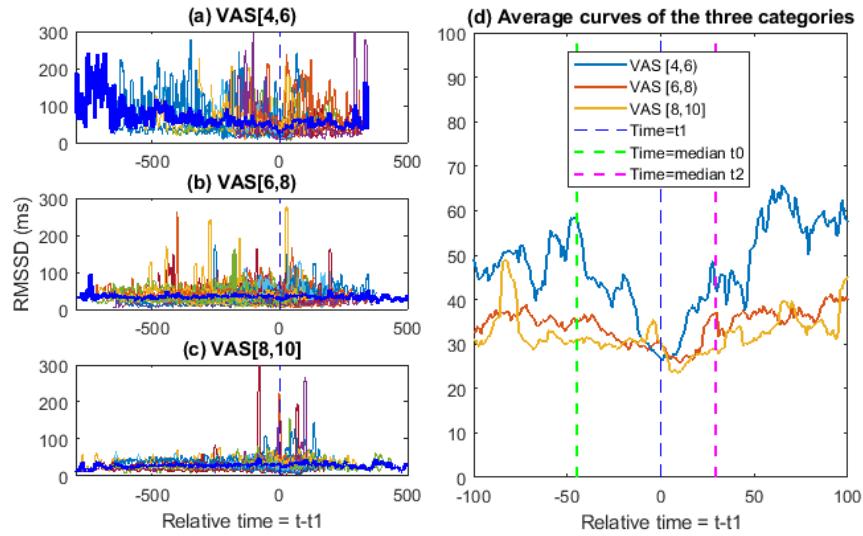


Figure 6.9: RMSSD (ecg_nrmssd) rescaled with λ_2

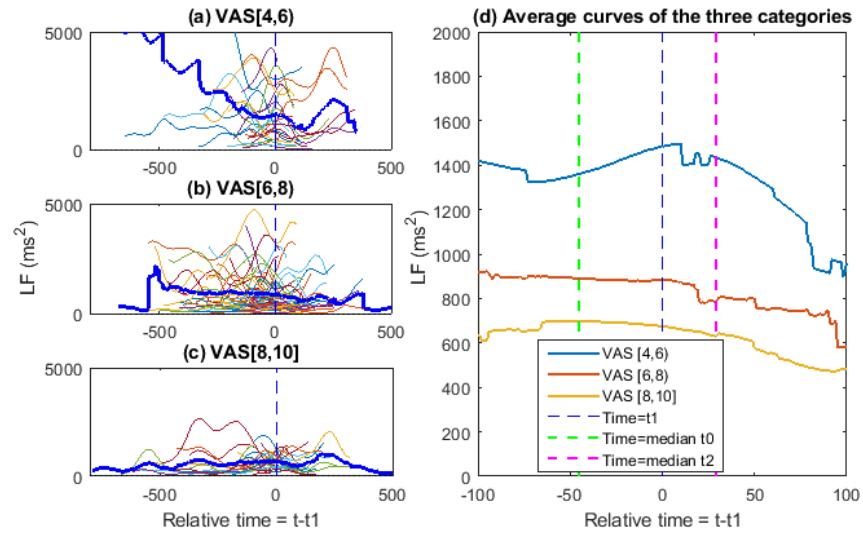


Figure 6.10: LF (ecg_nlf) rescaled with λ_3

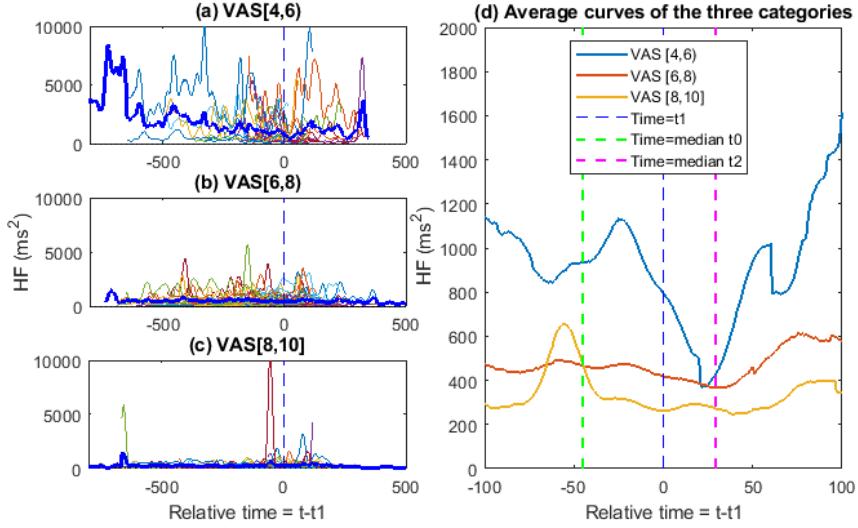


Figure 6.11: HF (ecg_nhf) rescaled with λ_3

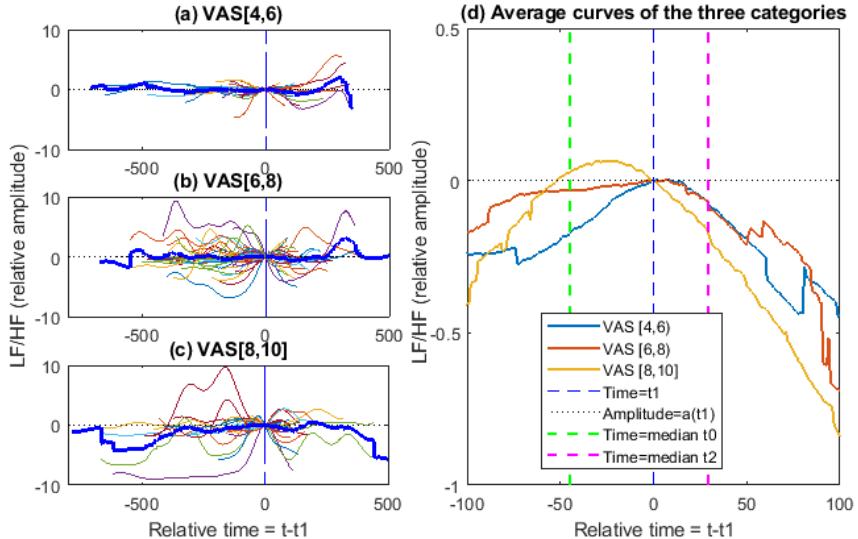


Figure 6.12: LF/HF (ecg_lf/hf) normalized to t_1

Thus far, little clear evidence has been found that any of the HRV features have a linear correlation with pain intensity that is more than weak (either VAS or other self-report tools) in previous studies [179, 180]. So did the observations in this section from Figure 6.8 to Figure 6.13. On the other hand, HF has been found to react to pain (or pain stimulus) in the literature [30, 31, 181–183], and thus HF is considered as a potential NAN balance in-

dicator [182, 183]. Therefore, it is commonly used in pain intensity modeling with the multi-parameter approach [30, 31].

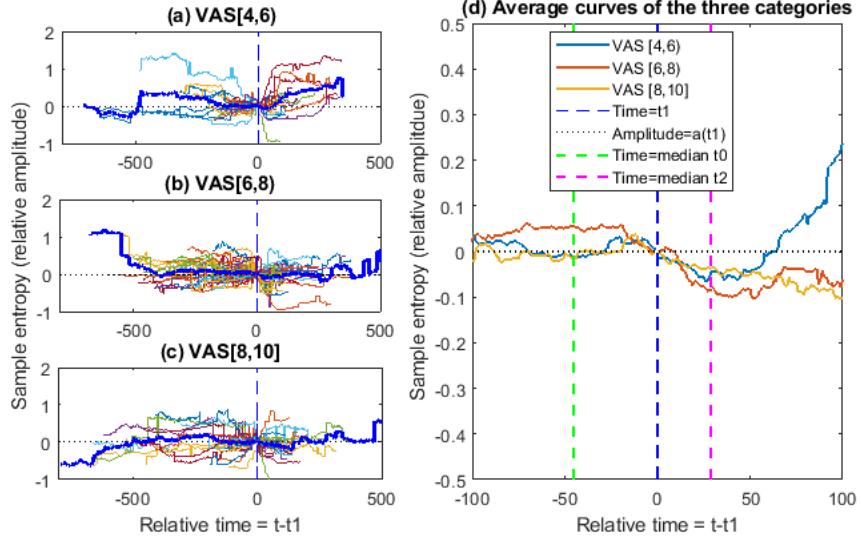


Figure 6.13: Sample entropy (*ecg_sampen*) normalized to t_1

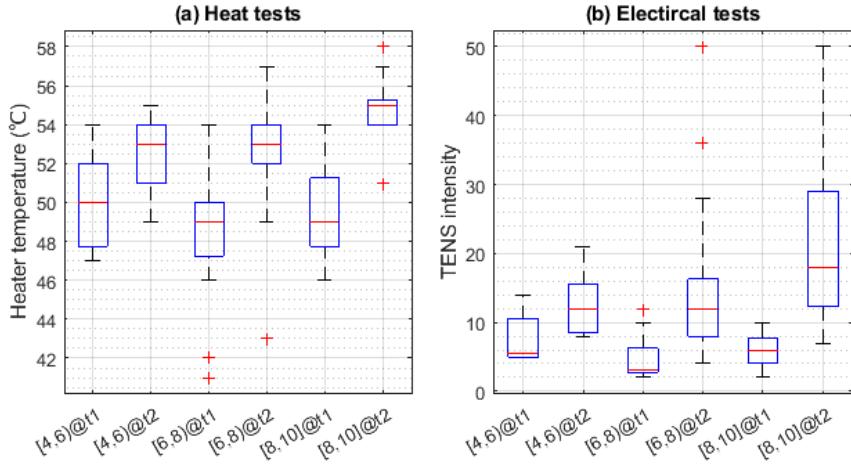


Figure 6.14: Stimulus intensity distribution in each pain category

In addition to the response to pain stimulation or pain, one interesting observation from Figure 6.8 to Figure 6.11 is: the higher the reported pain intensity, the lower the average feature value. The difference is especially clear within LF. The observed LF value difference in this study setting is independent of pain stimulus, which may result from the inherent difference among people. By looking at the literature, LF (and HF) at rest do have

the potential to predict pain sensitivity, pain threshold, and possibly pain intensity when pain is evoked by an experimental stimulus [184–186]. Thus, the distributions of the stimulus intensity reported at t_1 and t_2 in each pain category are further checked in Figure 6.14. The figure shows that the stimulus intensity of VAS[4,6] at t_1 is higher than that of VAS[6,8) in general under both stimuli. This is in line with the finding in [184] that a greater LF is associated with a higher *pain threshold*. However, the highest pain category is inconsistent with the finding.

6.2.3 Respiration rate

Figure 6.15 presents the respiration rates calculated from the ECG derived respiration. The average respiration for each subject was normalized to 20 breaths per minute. In all the three pain categories, a local trough can be observed in the average curve either near median t_0 or t_1 . However, the fluctuation is small (2-3 breaths/min) and is not unique across the time axis.

As mentioned before, there is another breathing signal source from the Bioharness’ capacitive pressure sensor, and the respiration rate extracted from this signal is denoted as *respr_bha*. The *respr_bha* extraction method is the same as *respr_edr* except for the interpolation step as shown in Figure 5.6. A third data source (*respr_dev*) was calculated on-device from the same capacitive pressure sensor signal, provided by the Bioharness device. Table 6.2 shows the average root mean square error (RMSE) of all the subjects and the correlation (Spearman’s ρ) between respiration rate sources. It shows that both the signal (between *respr_edr* and *respr_bha*) and the feature extraction algorithm (between *respr_bha* and *respr_dev*) contribute to a different outcome. This is further illustrated by comparing Figure 6.16(d) to Figure 6.15(d). The average curve fluctuation in Figure 6.16 is comparatively higher, and the pattern is slightly changed. How such uncertainties originating from the measurements, signals, and processing can influence the final estimation result would be a question to be looked further.

Table 6.2: The RMSE and Spearman’s ρ between the respiration rate calculated from difference sources, mean (std). *One extremely large value is excluded in the averaging.

	<i>respr_edr</i>	<i>respr_bha</i>	<i>respr_dev</i>
<i>respr_edr</i>	-	$\rho=0.32(0.20)$	$\rho=0.29(0.18)$
<i>respr_bha</i>	RMSE=6.31(1.64)	-	$\rho=0.35(0.24)$
<i>respr_dev</i>	RMSE=4.79(1.42)	RMSE=6.86*(1.62*)	-

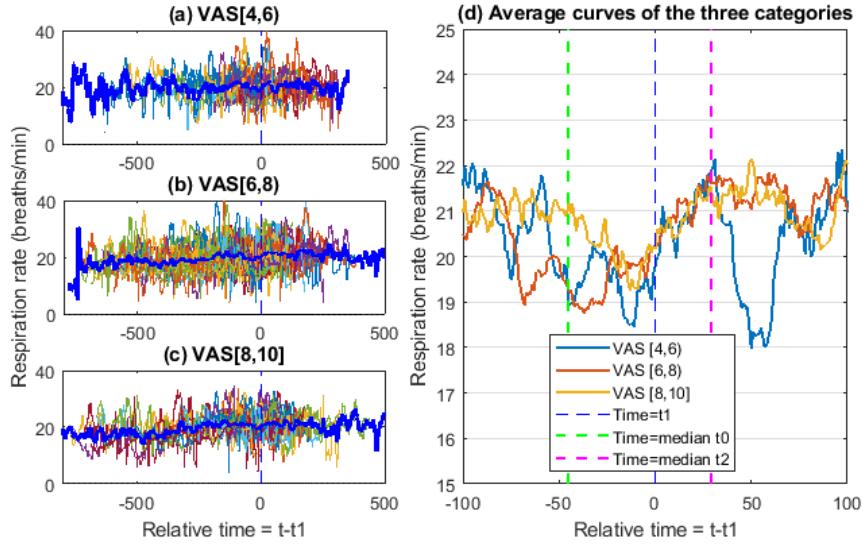


Figure 6.15: ECG derived respiration rate (*respr_edr*) rescaled with λ_4

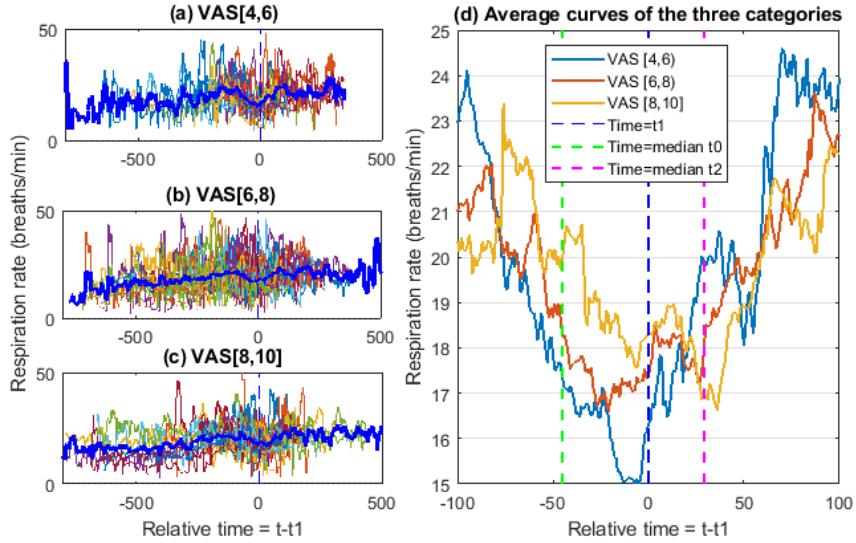


Figure 6.16: Bioharness respiration rate (*respr_bha*) rescaled with λ_4

6.3 sEMG features

The five electrodes placed on one side of the face are close to each other whilst the facial muscles are usually activated jointly. The selectivity of the signals might be weak for this reason. Therefore, the correlation between the five filtered sEMG channels is first observed with Spearman's ρ . The

Spearman's ρ is calculated within each subject and then the mean absolute values are presented in Figure 6.17. Two subjects were excluded in the averaging because the absolute value of ρ is nearly 1 in all channels. The results show that *cor* on the upper face has the weakest correlation with the other muscles (avg. $\rho = 0.34$ v.s. avg. $\rho = 0.63$) on the mid face and the lower face, especially with *zyg*. For *orb* on the mid face, the correlation is weaker with distal facial muscles than the proximal ones.

	<i>cor</i>	<i>orb</i>	<i>lev</i>	<i>zyg</i>	<i>ris</i>
<i>cor</i>	1	0.42	0.4	0.27	0.34
<i>orb</i>	0.42	1	0.71	0.68	0.61
<i>lev</i>	0.4	0.71	1	0.54	0.57
<i>zyg</i>	0.27	0.68	0.54	1	0.65
<i>ris</i>	0.34	0.61	0.57	0.65	1

Figure 6.17: Mean absolute Spearman's ρ of the subjects between adaptively filtered sEMG signals, $\text{std} \in (0.13, 0.22)$.

Based on the correlation result above, another observation was made regarding the correlation between the sEMG features. The 13 features listed in Table 5.4 were extracted from *cor* and *zyg*, respectively, with related Matlab functions in the BioPatRec toolbox. As the order of the AR model was defined as 4, the *ar* feature has four values and in total there are 32 rows/columns in the correlation matrix in Figure 6.18.

For the two signals that have a weak or negligible correlation, the correlation of the features in one to another (between *c_<feature>* and *z_<feature>*) are also weak or very weak. The features in the separate signal show a similar correlation coefficient pattern. Among those, the features in the first group described in Section 5.2.4 (*ld*, *damv*, *mav*, *mpv*, *rms*, and *wl*) show very strong correlation ($\rho > 0.8$ or 0.9) due to a similar mathematical definition. Features *wl* and *damv* are especially similar because the only difference is a scalar. With a similar definition, the two frequency features *fmn* and *fmd* show a very strong correlation ($\rho > 0.9$) as well. The *ar* feature values show moderate to strong ($\rho > 0.4$) correlation with each other and strong correlation ($\rho > 0.6$) with *fmd/fmn* as well as *ssc*. This is consistent with the theoretical definition where *ssc* can represent frequency information of the sEMG signal [139]. The relatively representative features are *ar_1*, *rms*, *mavs*, and *fpk*. However, as the average *mavs* and *fpk* were not found to have a clear response to pain stimulation or sensation, only the other two features are presented below.

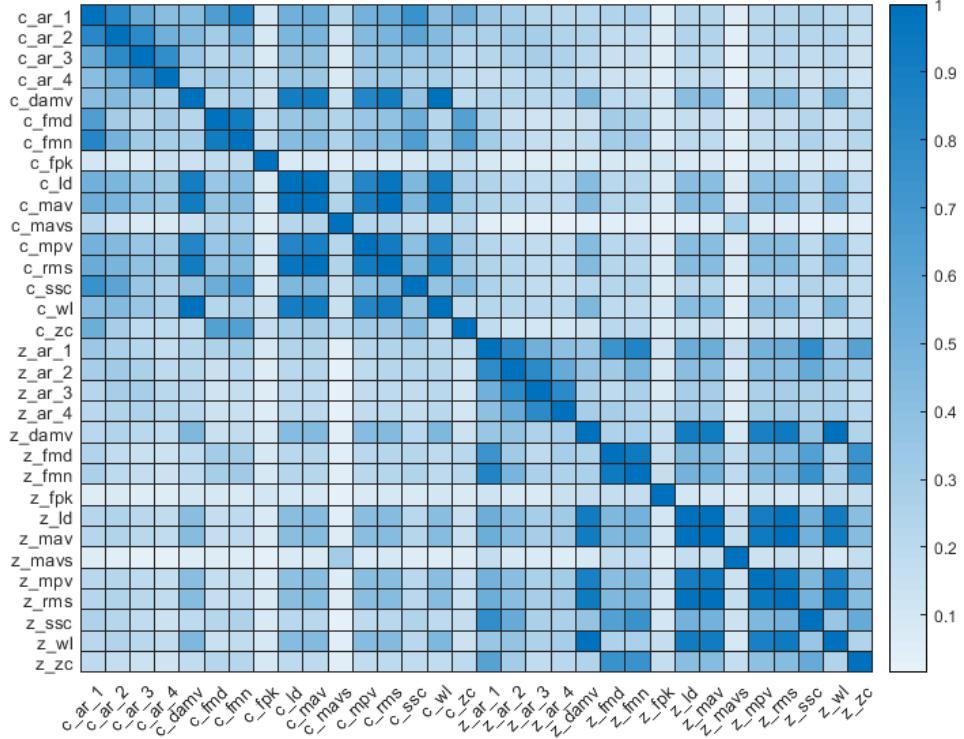


Figure 6.18: Mean absolute Spearman's ρ of the subjects between fifteen sEMG features of muscle corrugator and muscle zygomaticus major, $\text{std} \in (0, 0.25)$.

In addition to the facial expression of pain that we are interested in, other facial expressions and facial movements could also be reflected in the sEMG signals, such as talking and interaction with the guidance personnel. In the tests, talking was not allowed during the pain stimulation period but was not controlled before or after these periods. Such influence can be revealed in the *rms* feature indicating signal amplitude (Figure 6.19) to some extent. The baseline and variation of the average curves stay around zero from t_0 to t_1 while they both increase from t_1 to t_2 and remain or continue increasing after t_2 . The change is more clearly observed in *cor* compared to *zyg* (similar to *cor* and thus is not presented). In both facial muscles, the *rms* feature does not show a clear difference between the three pain categories.

The first coefficients of the 4th order AR model are presented in Figure 6.20 (*cor*) and Figure 6.21 (*zyg*). In both facial muscles, the average curves of VAS[8,10] show a slightly different response after t_1 than the other two pain categories. In *zyg*, the average curves in all the pain categories rise slightly from t_0 to t_1 and decline from t_1 except for VAS[8,10].

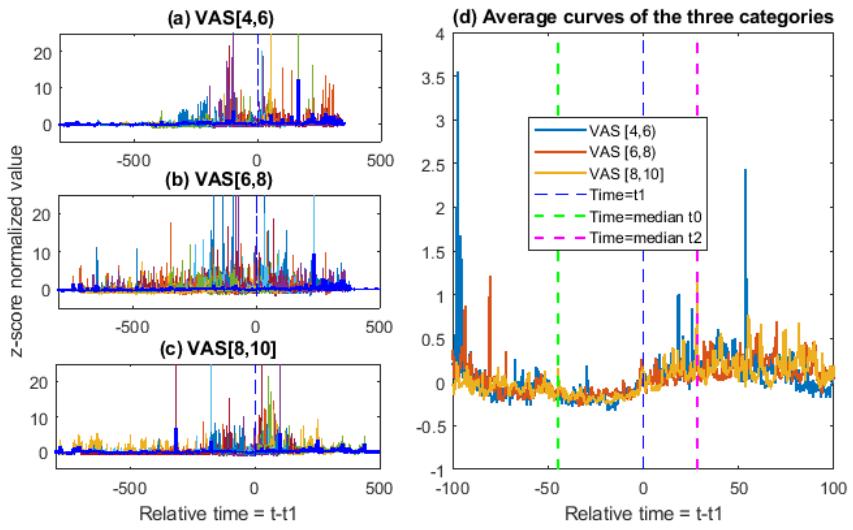


Figure 6.19: Root mean square of corrugator sEMG (*emg_cor_rms*)

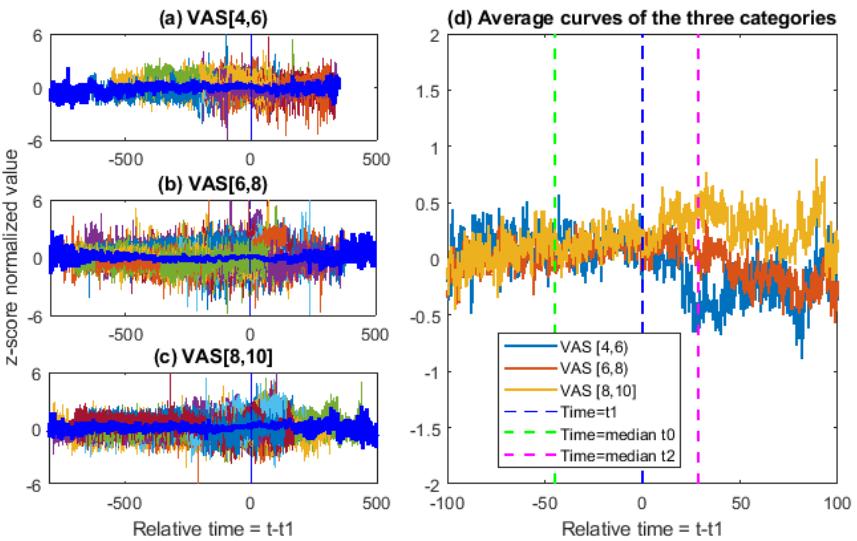


Figure 6.20: The first coefficient in the 4th autoregressive model of corrugator sEMG (*emg_cor_ar_1*)

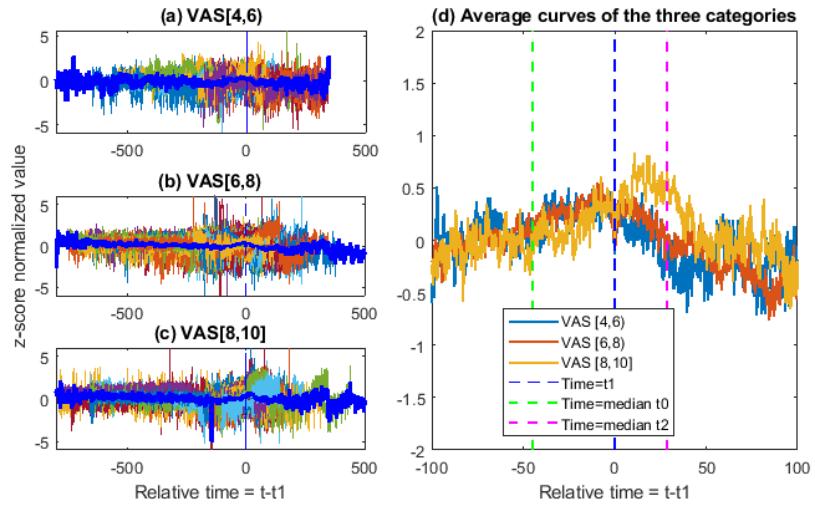


Figure 6.21: The first coefficient in the 4th autoregressive model of zygomaticus major sEMG (*emg_zyg_ar_1*)

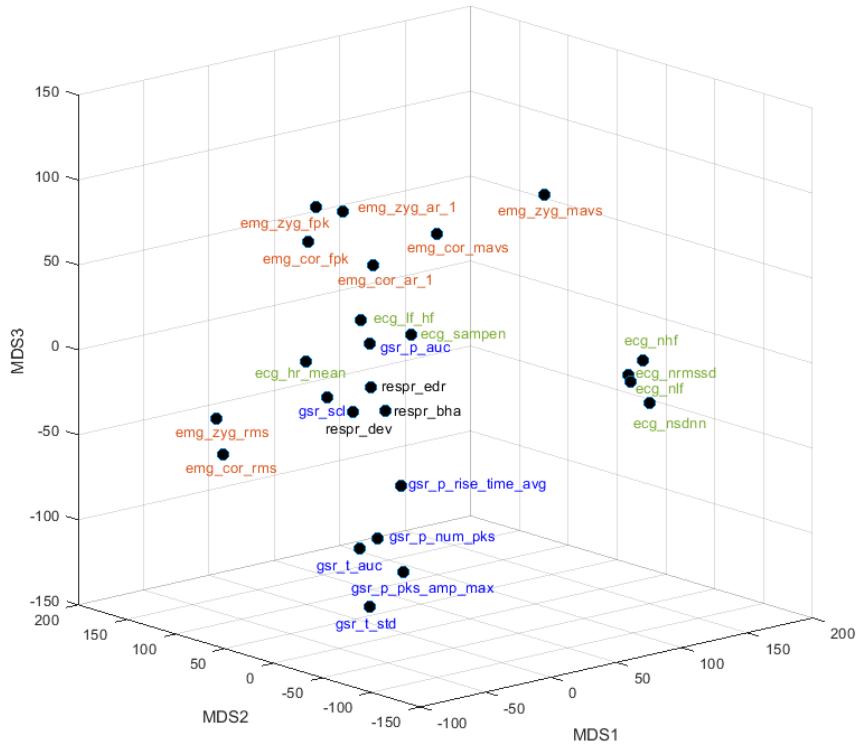


Figure 6.22: Multidimensional scaling plot of the Euclidean distance between the representative features

6.4 Distance in representative features

The features without strong linear correlations to other features from the same signal are considered as representative features. The similarities between them across signals were examined with multidimensional scaling (MDS) plot of the Euclidean distance between the standardized feature values in Figure 6.22. It further validated the dissimilarity between the chosen features as well as the signals.

6.5 Summary and discussion

This chapter presents the implementation, correlation, and visualization results with the physiological features. The findings and discussion are summarized below in signal wise.

6.5.1 GSR features

The phasic features show a weak response to both stimulation and pain sensation from the visualization (represented by p_num_pks). They could not directly indicate pain intensity either. However, as NSCF in previous studies has shown its potential in indicating pain intensity, the phasic features are still worth attention in further studies. Consistent with previous studies, gsr_scl showed a good response to pain in general and may help index pain intensity. Nevertheless, it also responded to the stimulation when no pain had yet been caused. In contrast, the tonic features (t_auc and t_std) may indicate the presence of the pain when it occurs.

6.5.2 ECG features

The heart rate increases in reaction to stimulation on average. However, the responses of the heart rate, the HRV features, and the respiration rate to pain sensation or self-reported pain intensity are much less directly perceived from the plots. Instead of being a direct pain index, the HRV features may help index the subjectivity of pain.

The conducted HRV analysis failed in finding the dynamic changes of HRV features in reaction to pain stimulation, although we tried to attenuate the dependence of HRV feature values on mean HR value in each person [187, 188]. The observations made in this chapter are mainly in the time domain or frequency domain alone. Observations and feature extraction for classification can be made, for example, with time-frequency or time-varying approaches [189] in future work.

Regarding the reproduction of results, one of the challenges with physiological signals is the large inter-subject variability, as discussed in Section

5.2.7. In ECG signal, HR and HRV features were normalized by the reference to the average RR interval of the whole recording from the same subject in this chapter due to the fact that the baseline at rest was not recorded among several subjects. This normalization choice may influence results reproduction on new subjects or when study design alters. One solution in future data collection could be recording the basal state of each person or having access to the physiological data in similar cases. Alternatively, if individual-level patterns are of interest rather than a generalized one, data of the same person in different scenarios and during different time periods would be required.

6.5.3 sEMG features

Some of the collected sEMG signals, as well as the investigated features, show weak selectivity between each other. The amplitude features, represented by *rms*, responded to pain sensation. There was also a common failing among the other parameters as well, as they did not respond specifically to pain sensation. The features representing frequency information may help differentiate high pain intensity from others.

Chapter 7

Pain Estimation from Physiological Features

The study design in Figure 4.1 was described mainly from the perspective of the stimulus. The design is further interpreted from pain stimulus to pain sensation as shown in Figure 7.1. Based on the time marks (t_0 , t_1 , and t_2) throughout one subject's data collection time, the timeline is divided into three types of periods - P_0 , P_1 , and P_2 . The sensation during P_1 is considered as "no pain" if using the definition of the *pain threshold* at t_1 or could be considered as having "adequate pain control" from a clinical point of view. However, the value of the physiological parameters could be altered due to factors other than pain, such as emotional states (anxiety, fear, and stress) [190]. Such negative emotional states may also accompany and interact with pain during P_2 , but this is not further differentiated in the interpretation. In P_2 , the pain sensation intensity is assumed to be positively correlated to time due to the increase of both pain stimulus intensity and duration.

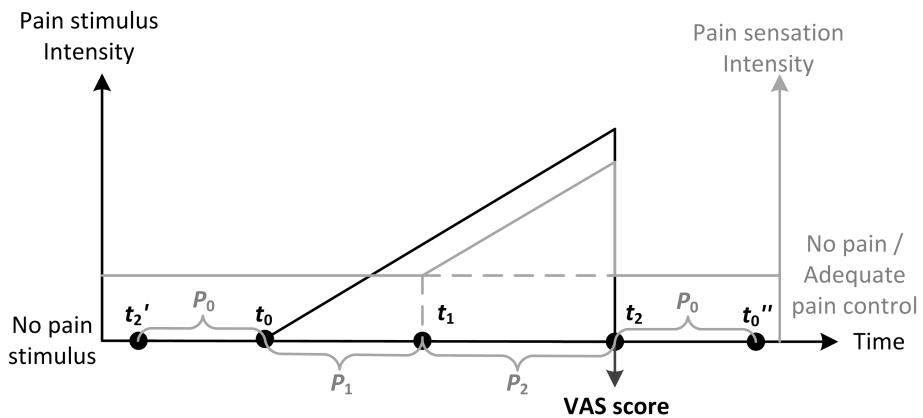


Figure 7.1: Study design interpretation

In our previous work [33, 110], the design was interpreted slightly differently, where P_1 was interpreted as *Mild pain* and P_2 was interpreted as *Moderate/Severe pain*. This interpretation was based on the defined VAS score at t_1 and the correspondence between the four-point VRS and the NRS presented in [37]. Based on the updated interpretation in Figure 7.1 and the feature observation results in Chapter 6, an attempt has been made to solve the pain estimation problem with a two-step classification, where the first step (S1) is to recognize the presence of pain or inadequate pain control in the timeline (i.e., differentiating P_2 from P_0 and P_1), and the second step (S2) is to further recognize the pain intensity after the pain period is recognized (i.e., differentiating between P_2 periods).

7.1 Methods

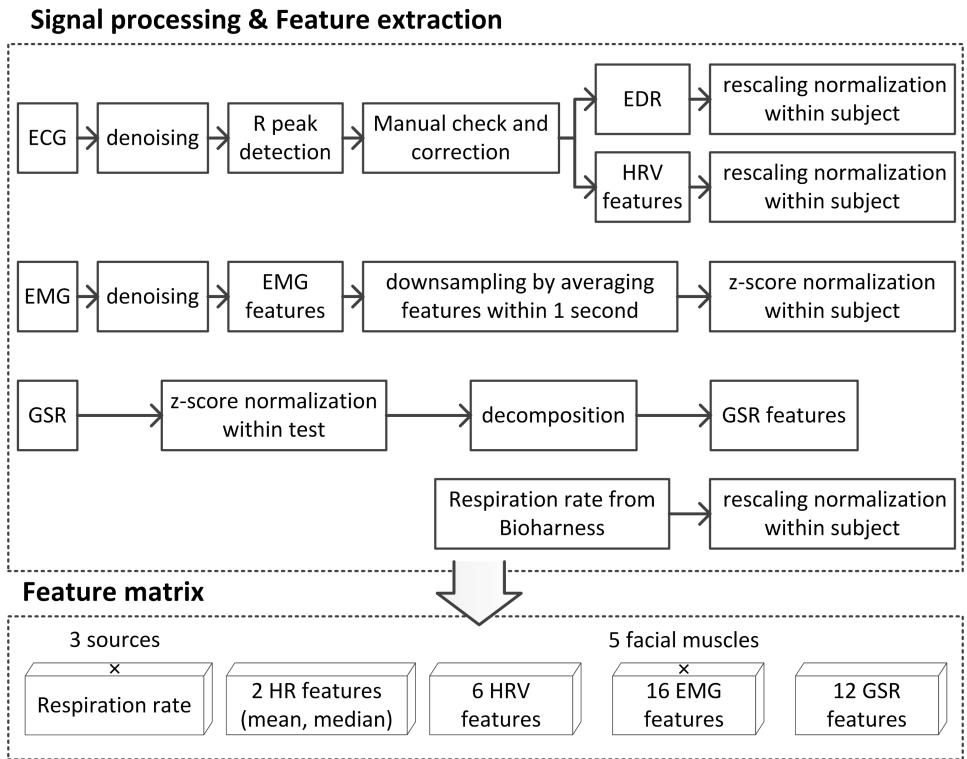


Figure 7.2: Signal processing & feature extraction flow summary and the composition of the feature matrix

The composition of the feature matrix used in the classification is presented in Figure 7.2. In total 103 columns of features constitute the 103-dimensional

feature matrix. The length of the feature matrix is approximately the total duration of the recordings measured in seconds from 30 subjects (one subject was excluded because of incomplete GSR recording). However, missing values in some of the features exist intermittently in the matrix due to signal loss (i.e., when changing GSR measuring side) or the long feature extraction window in the beginning of a recording (e.g., the first 249 seconds features of the first 250-s window when extracting `ecg_nlf`). The total feature matrix length in P_0 , P_1 , and P_2 are 39615, 5054, and 4613, respectively. While the lengths with valid values in all the dimensions are 27408, 4522, and 4167.

The matrix lengths show a between-class imbalance in S1 classification, and thus neither accuracy nor error rate is a proper performance assessment metric because they are sensitive to data distributions [191, 192]. Imbalanced data is a common case in datasets. Two of the solutions to this problem are sampling methods which modify the imbalanced distributions into balanced ones and cost-sensitive learning where the costs associated with misclassifying examples are considered. The details of the classification methods are further presented in the subsections below.

7.1.1 Leave-subject-out cross-validation

Cross-validation is a typical method to test the performance of the developed classification model. The cross-validation loop includes a training phase and a testing phase in each iteration. The test dataset is independent of the training dataset and thus provides an unbiased evaluation of the trained classifier. In some cases, a third dataset, a validation dataset is used in the training phase to optimize the model (e.g., tuning the hyperparameters). Usually, the cross-validation is conducted in 5/10-fold [193] or leave-one-out in an extreme case. However, the time series from the same subject is non-independent and thus leave-subject-out cross-validation is recommended [194], where a single subject is iteratively left out as the test dataset. This is to avoid attaining good but biased results where the data from the same person is involved in both training and testing. The non-independence especially lies between the four tests with the same subject where the reactions could be similar. In the implementation, leave-subject-out cross-validation is conducted in the outer-loop for performance evaluation, while 5-fold cross-validation is conducted in the inner-loop when hyperparameters are optimized in the training phase.

7.1.2 Performance metrics

As introduced before, the overall classification accuracy is dominated by the accuracy of the majority class and insensitive to the minority class accuracy. Therefore, the accuracy of each class is reported when comparing different classifiers. Particularly in the S1 binary classification, the class accuracy

(ACC_x) is either a true positive rate (TPR) or a true negative rate (TNR) as shown in Figure 7.3. In the S1 hyperparameter optimization part, the models are tuned based on AUC (area under the ROC curve).

		True Class	
		Pain or Inadequate pain control	No pain or Adequate pain control
Estimation	Pain or Inadequate pain control	True Positive (TP)	False Positive (FP)
	No pain or Adequate pain control	False Negative (FN)	True Negative (TN)
		True Positive Rate (TPR) = $\frac{TP}{TP+FN}$ (Sensitivity)	True Negative Rate (TNR) = $\frac{TN}{TN+FP}$ (Specificity)

Figure 7.3: Confusion matrix in S1 binary classification

7.1.3 Hyperparameter or Model tuning

Hyperparameter tuning or model tuning takes place in the training phase as introduced in the cross-validation. The cross-validation with hyperparameter tuning is also called nested cross-validation [195] which has an inner cross-validation loop in each training phase. The best hyperparameter set in each outer loop fold is derived after iterating the inner loops and the unbiased generalization performance is calculated in each fold. It should be made clear that hyperparameter tuning in the nested cross-validation is not to choose the best hyperparameters, but to check the performance, which is consistent with the purpose of using cross-validation. Some commonly used tuning methods minimize the objective (e.g., classification error) using grid search, random search, Bayesian optimization, etc. In the implementation, Bayesian optimization is used.

7.1.4 Misclassification cost matrix in cost-sensitive learning

In the binary classification, a misclassification cost matrix C has the structure shown in Table 7.1. In the matrix, c_{00} and c_{11} are correct predictions and thus are negated costs. The inequality of the misclassification costs is adjustable by changing the relative value between c_{01} and c_{10} . Given a cost matrix C , an example x is classified into the class having the minimum

expected cost. $R(i|x)$ is the expected cost of x being classified into class i :

$$R(x, i) = \sum_j P(j|x)C(i, j) \quad (7.1)$$

where $P(j|x)$ is the probability estimation of classifying an instance into class j . However, the implementation does not apply to all and is case-specific [191]. For example, in Matlab, C is integrated into a classification either in the prediction phase by minimizing the expected classification cost (e.g., decision tree) or in the training phase by adjusting prior probabilities (e.g., ensemble learning) [196].

Table 7.1: Cost matrix for binary classification

	actual negative	actual positive
predict negative	c_{00} , TN	c_{01} , FN
predict positive	c_{10} , FP	c_{11} , TP

7.1.5 Feature importance

Feature selection or feature ranking for classification is to reduce the dimensionality and noise in the feature matrix. It may help improve accuracy and efficiency in the classification because some irrelevant, redundant and noisy features may exist. On the other hand, models such as tree-based classifiers and linear discriminant analysis classifiers could indicate the relative level of importance among the features after training.

7.1.6 Classifiers

Table 7.2 lists five types of classifiers. In addition to the listed methods, a shallow neural network classifier with 1 hidden layer of 10 neurons is tested as well. The sub-types of each classifier are specified in the table, and their performance is first evaluated without hyperparameter optimization. Discriminant analysis classification is a parametric method, where the parameters of a Gaussian distribution is estimated for each class. The sub-types of the discriminant analysis classifier are defined under different modeling assumptions regarding the means and covariances. For the k-nearest neighbor (knn) classifier, the sub-types are defined according to distance metrics. The hyperparameter *NumNeighbors* is defined as 5 when no optimization is applied. The features are standardized in all the knn classifiers. Ensemble learning is to enhance the performance of multiple "weak" learners by combining them. Many ensemble learning methods are based on decision trees. The chosen tree-based methods are AdaBoost, Bag (i.e., Random forest), RUSBoost which was developed to deal with the class imbalance problem

Table 7.2: Specifications of the classifiers

Classifier	Multicass	Cost-sensitive learning	Hyperparameters	Sub-types	Feature importance	Solution to missing values
Decision tree	Yes	Minimizing the expected classification cost	-MaxNumSplits -MinLeafSize -SplitCriterion	None	Yes	-surrogate decision splits
Discriminant analysis	Yes	Minimizing the expected classification cost	-Delta -DiscrimType -Gamma	-linear -quadratic -diaglinear -diagquadratic -pseudolinear -pseudoquadratic	Yes	
k-Nearest neighbor	Yes	Minimizing the expected classification cost	-Distance -DistanceWeight -Exponent -NumNeighbors -Standardize	-cityblock -chebychev -correlation -cosine -euclidean -hamming -jaccard -minkowski -seuclidean -spearman		
Ensemble tree	Yes Yes No	Prior probabilities	-Method -NumLearningCycles -LearnRate	-AdaBoost -Bag -RUSBoost -RobustBoost	Yes	
Support vector machine	Yes	Prior probabilities where C describe penalties	-BoxConstraint -KernelScale -KernalFunction -PolynomialOrder -Standardize	-gaussian -rbf -linear -polynomial		

[197], and RobustBoost which is more robust to label noise [198]. Finally, the sub-types of the support vector machine classifier are defined based on the used kernel function used.

7.2 Results

In this section, the results of the classification in the two steps are separately presented separately. In each subsection, the non-optimized models are built in the first round. The relative feature importance and classification performance are checked. After comparing the performance, a selective group of models are tuned in the second round by hyperparameter optimization, adding misclassification matrix, or using both.

7.2.1 Step 1 - Estimating the presence of the pain or inadequate pain control

This part presents the results of the binary classification between pain (or inadequate pain control) and no pain (or adequate pain control). The samples at period P_2 were labelled as pain and the remaining samples were labelled as no pain.

Feature importance

The relative importance of the features in all the training folds of S1 classification is visualized in Figure 7.4. On the x-axis, from 1 to 80 are sEMG features; from 81 to 92 are GSR features; from 93 to 95 are respiration rates; from 96 to 103 are ECG features. Consistent with the visualized features in Chapter 6, GSR features, respiration rate and ECG features show higher importance than sEMG features in most of the models. Comparatively, sEMG features played a more important role in linear discriminant analysis classifiers than in tree-based classifiers. After averaging the curves (the values in each curve are normalized between 0 and 1), the five most important features with the highest average values are: (feature number – feature name – average importance value out of 1)

- feature 88 – *gsr_scl* – .99/1
- feature 96 – *ecg_hr* – .54/1
- feature 85 – *gsr_t_avg* – .43/1
- feature 84 – *gsr_t_max* – .41/1
- feature 94 – *bha_respr* – .31/1

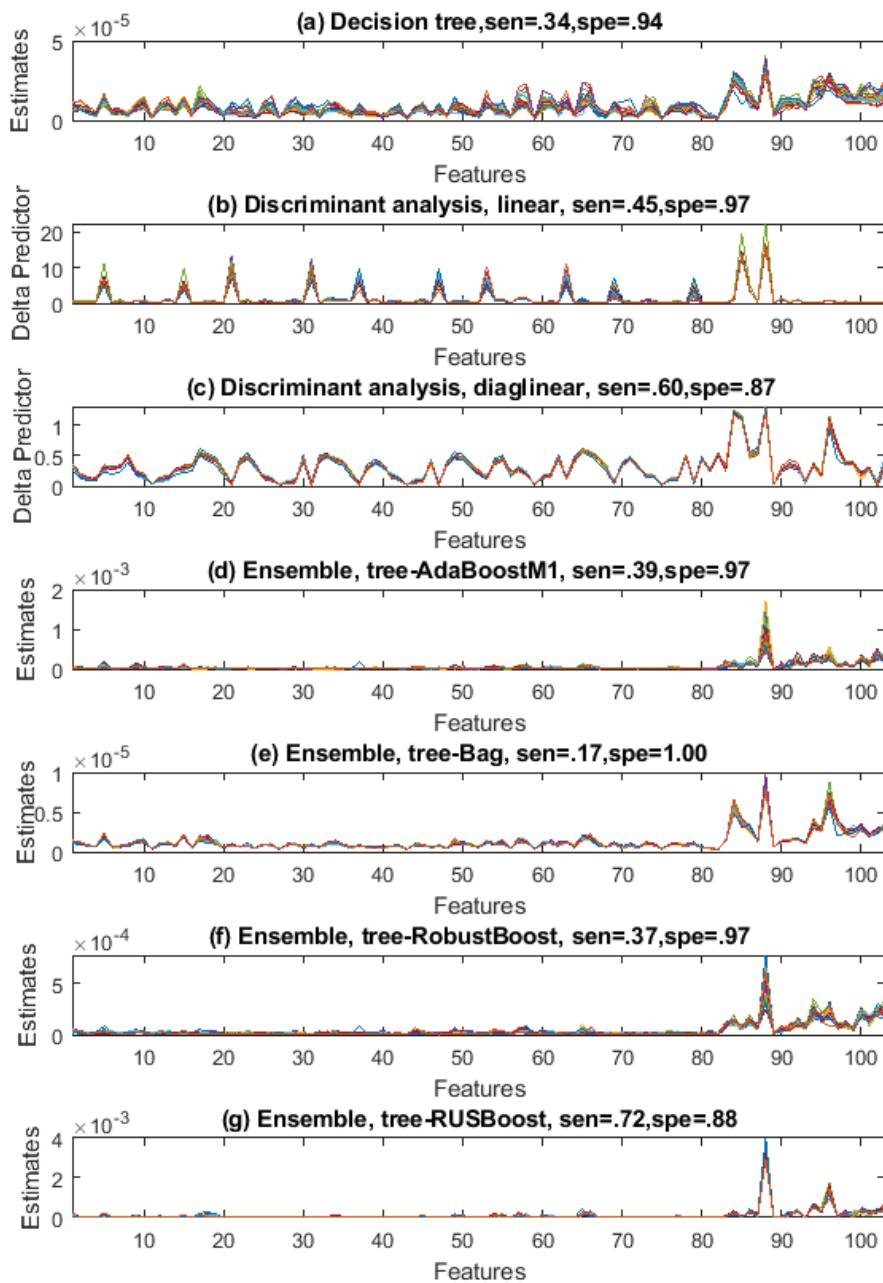


Figure 7.4: S1 - relative importance of the features in tree-based classifiers and linear discriminant analysis classifiers, with average test sensitivity and specificity

Performance of the models without tuning

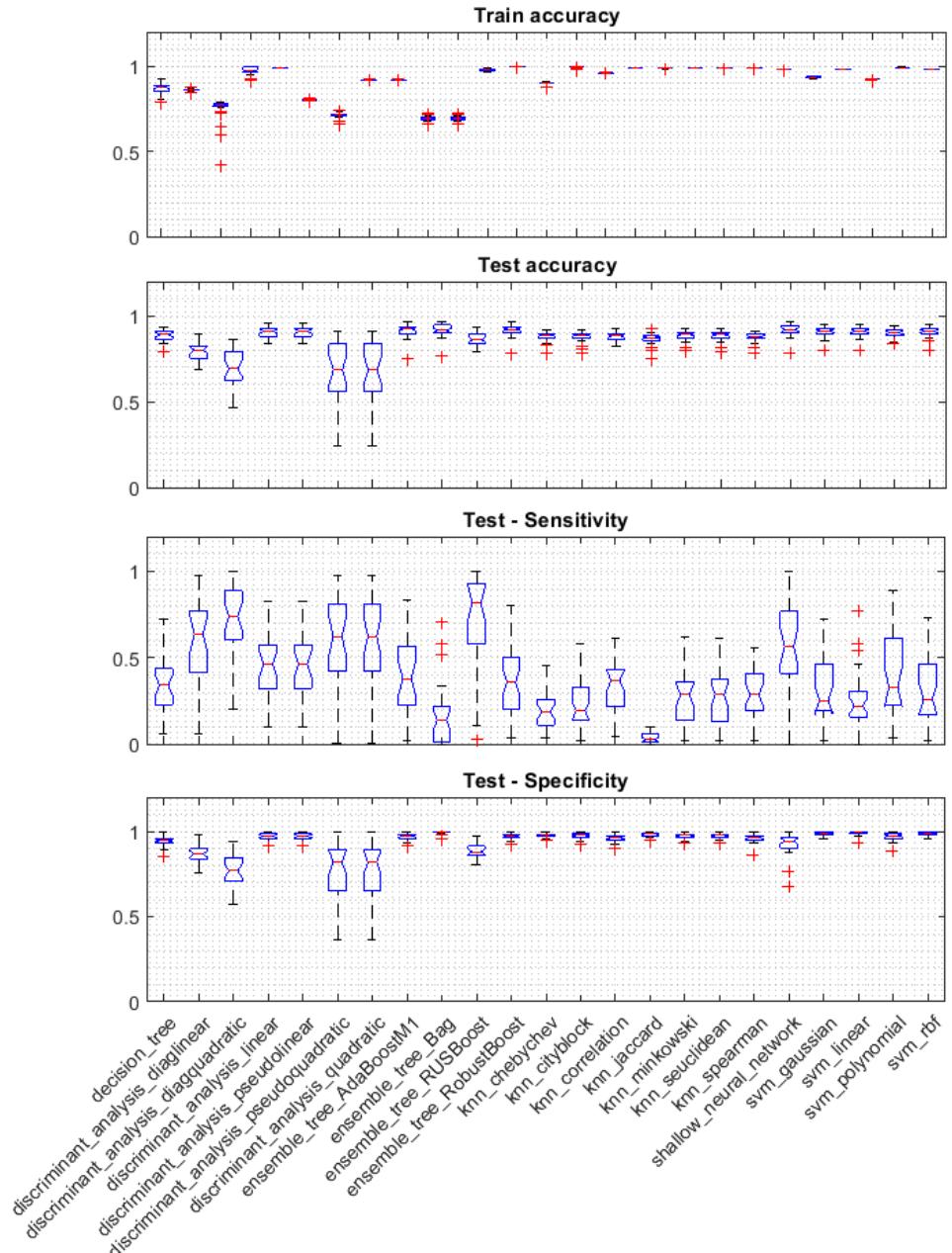


Figure 7.5: S1 - Model performance without hyperparameter optimization

The performances of the models without tuning are presented in Figure 7.5 with the distribution of training accuracy, test accuracy, as well as the sensitivity and specificity in the test phase of the leave-subject-out cross-

validation. The distribution of the test accuracy across different models is close to the distribution of training accuracy in the same model but with a larger variance. As expected, the test accuracy is dominated by the test specificity where the majority class lies.

Table 7.3: S1 - Four selected models

Model	Sensitivity median (min, max)	Specificity median (min, max)
discriminant_analysis_diaglinear	.64 (.06, .98)	.87 (.76, .98)
discriminant_analysis_diagquadratic	.74 (.20, .99)	.77 (.57, .94)
ensemble_tree_RUSBoost	.81 (.03, 1)	.88 (.80, .97)
shallow_neural_network	.56 (0, 1)	.94 (.68, 1)

Four models are selected for further comparison at an individual-level to consider the overall sensitivity, and specificity in the performance. The four models are listed in Table 7.3, and the detailed performances from each test fold are compared in Figure 7.6. The test subject was sorted in a descending order of the test sensitivity in the ensemble_tree_RUSBoost model, whose top two important features are gsr_scl and ecg_hr as shown in Figure 7.4(g). A comprehensive look into Figure 7.6 and Figure 7.4 indicates that the test sensitivity could vary a lot within the same train-test fold across different models. For example, the test sensitivity in subject No.28 is .72 in discriminant_analysis_diaglinear while it is .19 in ensemble_tree_RUSBoost, but the case is reversed in subject No.15.

Model tuning and the tuned performance

As introduced in Table 7.2, the hyperparameters to be tuned are case dependent, and the room for hyperparameter tuning varies. For example, among the eligible parameters to tune in discriminant analysis classifiers, both linear coefficient threshold δ and the amount of regularization γ are tunable in real values for linear discriminant analysis, whereas δ must be 0 and γ could only be either 0 or 1 in quadratic sub-types. Another example is the neural network where far more options and possibilities exist. However, the aim of implementing model tuning in this part is to discuss the effect of it on classification performance rather than exploring the best machine learning model in solving the pain recognition problem. Thus, two models with default parameter settings, discriminant_analysis_diaglinear, and ensemble_tree_RUSBoost, are optimized and the changes in performance in each test fold are presented in Figure 7.7.

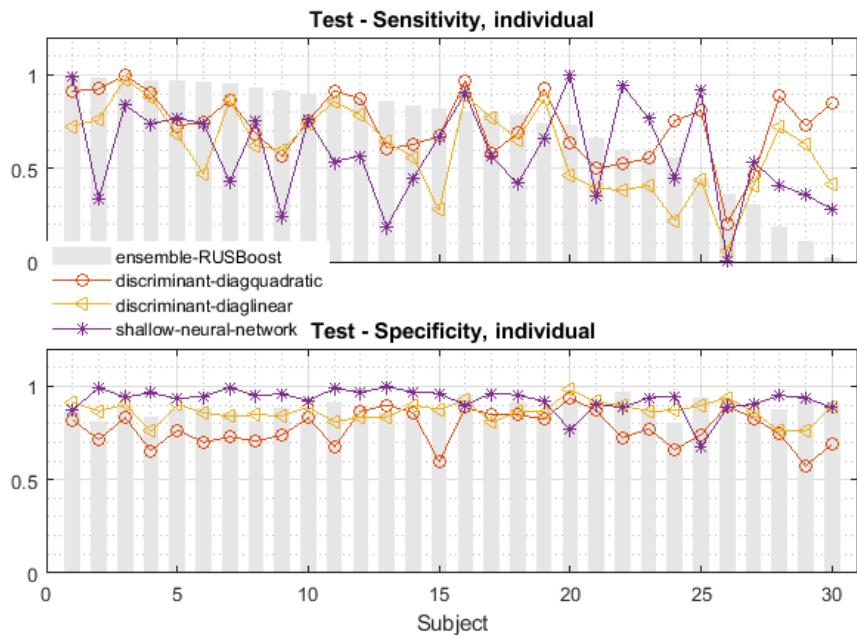


Figure 7.6: S1 - The performance in each leave-subject-out cross-validation test fold (subjects in ensemble_RUSBoost sensitivity descending order)

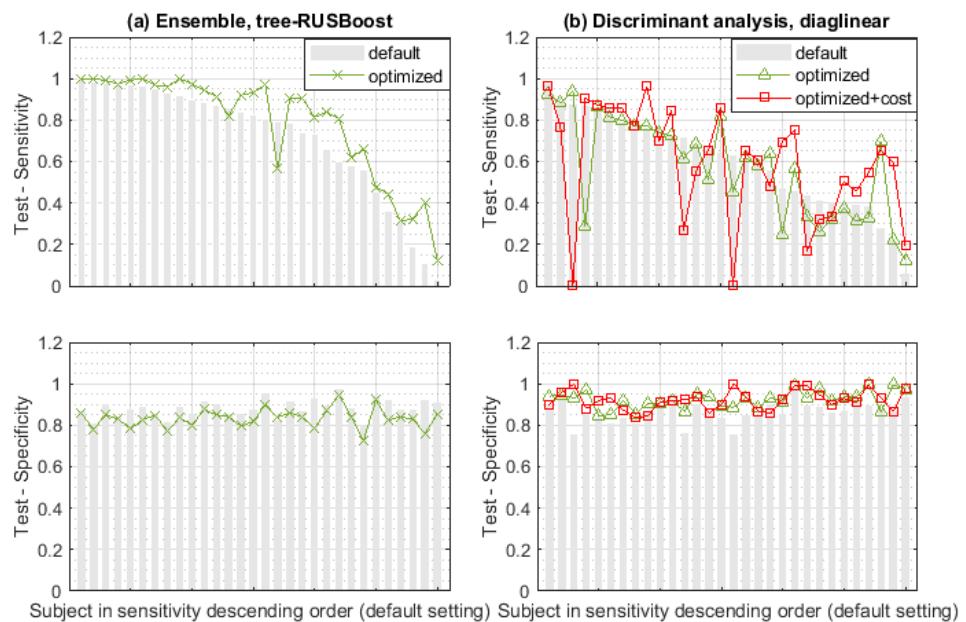


Figure 7.7: S1 - Classification performance with default model settings and optimized settings

RUSBoost is the abbreviation for random undersampling boosting. It solves the imbalanced class problem by undersampling the majority classes to the smallest class (*RatioToSmallest*) with a pre-defined ratio vector. As an ensemble learning method, learning rate and number of learning cycles are tunable where a lower learning rate yields to a larger number of learning cycles and a longer training time. Meanwhile, as a decision tree method, the tree parameters are also tunable. The tuned performance shown in Figure 7.7(a) is the result after tuning learning rate at 100 learning cycles and the other hyperparameters are the same as the default setting. It shows that the sensitivity improved (median from .81 to .91) with a sacrifice of specificity (median from .88 to .84), which results in an overall test accuracy decrease (median from .87 to .83). The misclassification cost was not added as undersampling already solved the imbalanced class issue.

In diaglinear discriminant analysis optimization, both δ and γ were optimized automatically with Bayesian optimization. When adding the misclassification cost matrix, the cost of misclassifying pain was set to be five times the cost on misclassifying no pain. The performance change in Figure 7.7(b) shows improved specificity after optimization (median from .87 to .93) while sensitivity decreased (median from .64 to .62) and thus the overall accuracy increased from median .80 to .87. After adding the misclassification cost to the optimization, many test folds had a better sensitivity without harming the specificity. However, the sensitivity of several other test folds meanwhile dropped significantly.

Discussion

With the aim of recognizing pain, the simply optimized RUSBoost was the one that reached the best median sensitivity. However, the sensitivity variance among subjects is not limited to the RUSBoost method but also occurs with any other method. This may indicate that the variability exists in nature in the pattern of the physiological features in pain, which cannot be tackled by the implemented methods yet or the information other than the extracted features is needed.

When differentiating pain from no pain, it is natural to hypothesize that pain with higher intensity is more recognizable from physiological parameters than a lower intensity pain as with a stronger reaction. In the optimized RUSBoost model, the self-reported pain intensity from the best estimated 15 subjects is 7.2 on average ($SD = 1.2$), while the average pain intensity is 6.6 ($SD = 1.4$) among the other 15 subjects. In another study, where the pain threshold and pain tolerance were reported instead of using the self-report pain scale, a higher classification accuracy was reached between baseline and pain with a higher intensity [199]. This may be explained by the fact that SCL dominated the classification performance in RUSBoost from the clues

of feature importance (Figure 7.4), SCL visualization (Figure 6.2), and when comparing the skin conductance level and the linear combination cuves in [31].

Table 7.4: S1 - The best sensitivity and the corresponding model for each test subject fold and the sensitivity standard deviation among all the 26 models (subjects in best sensitivity descending order)

Subject	Best sensitivity	Best model	Sensitivity SD
1	1	optimized_ensemble_tree_RUSBoost	.260
7	1	optimized_ensemble_tree_RUSBoost	.238
18	1	optimized_ensemble_tree_RUSBoost	.271
23	1	shallow_neural_network	.243
29	1	optimized_ensemble_tree_RUSBoost	.242
30	.997	discriminant_analysis_diagquadratic	.294
4	.993	optimized_ensemble_tree_RUSBoost	.275
27	.974	optimized_ensemble_tree_RUSBoost	.237
19	.974	discriminant_analysis_pseudoquadratic	.270
21	.974	optimized_ensemble_tree_RUSBoost	.291
24	.971	optimized_ensemble_tree_RUSBoost	.300
25	.968	optimized_ensemble_tree_RUSBoost	.241
6	.961	optimized_ensemble_tree_RUSBoost	.257
15	.944	shallow_neural_network	.239
13	.943	optimized_ensemble_tree_RUSBoost	.241
8	.938	discriminant_analysis_pseudoquadratic	.292
10	.933	optimized_ensemble_tree_RUSBoost	.273
28	.924	discriminant_analysis_diagquadratic	.272
9	.920	optimized_ensemble_tree_RUSBoost	.207
16	.914	optimized_ensemble_tree_RUSBoost	.247
22	.905	optimized_ensemble_tree_RUSBoost	.226
20	.885	discriminant_analysis_diagquadratic	.266
17	.857	optimized_discriminant_analysis_diaglinear	.255
26	.856	discriminant_analysis_pseudoquadratic	.246
11	.839	optimized_ensemble_tree_RUSBoost	.190
14	.795	ensemble_tree_RUSBoost	.193
3	.766	shallow_neural_network	.183
2	.727	discriminant_analysis_diagquadratic	.180
12	.680	discriminant_analysis_pseudoquadratic	.163
5	.443	optimized_ensemble_tree_RUSBoost	.138

The factor that limits the S1 classification performance could be the imperfection of the chosen model or its hyperparameters. However, on the other hand, the performance could be restricted by the fact that the data patterns in some of the test folds may not fundamentally follow the patterns in the corresponding training folds. Although the classifier models are not exhaustively explored, the best performance for each subject and the performance variations among the models may give a hint as to the pattern difference in different test folds. Table 7.4 presents the best sensitivity that

the models could reach for each test subject and the sensitivity variations among the models. The subjects in the table are sorted in descending order starting from the best sensitivity. The table shows that two thirds (21/30) of the best sensitivity are larger than .90 and most of them (15/21) are from the optimized RUSBoost model. Among the first 24 subjects, the standard deviations of the model sensitivities are relatively steady between .20 and .30, while the variations for the subjects with a lower sensitivity are on the contrary smaller than .20. This indicates that a relatively low upper performance limit might exist within some people when using this method.

In the RUSBoost model optimization, the current solution optimized the learning rate by improving the AUC. The optimization could be further customized to reach a trade-off between the two types of error. Moreover, the results of the models are mainly evaluated by test sensitivity because the S1 classification aims to recognize pain. Similar to the performance metrics in the model optimization, the performance metrics in the automatic pain assessment field is worth further discussion.

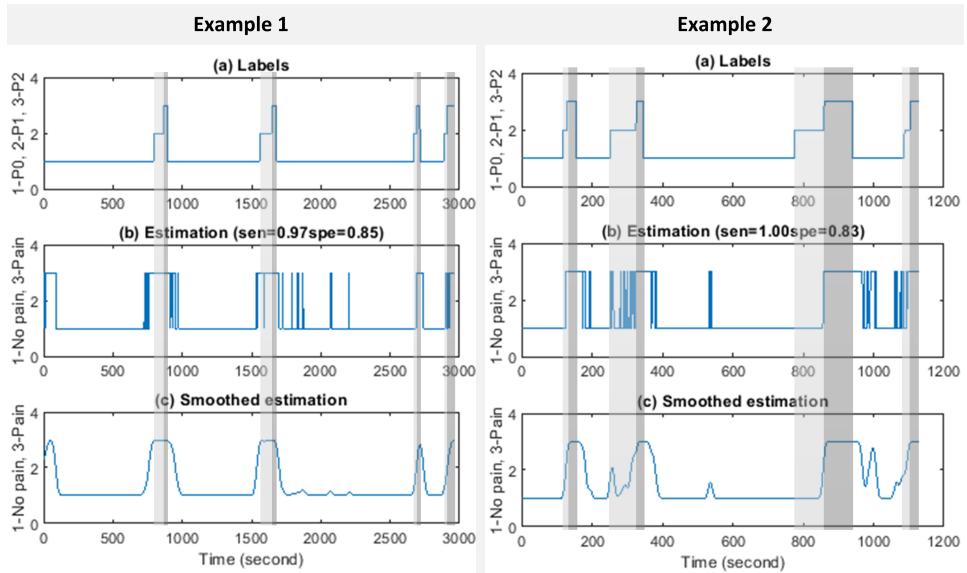


Figure 7.8: Two S1 RUSBoost estimation examples

Ultimately, as the current good predictions rely very highly on the ANS-based signals, the main limitation when bringing the model to real life is that the reactions of the signals could be non-specific to pain. Figure 7.8 shows two good pain prediction examples with nearly perfect sensitivity. The smoothed estimation could eliminate many prediction noises. However, some theoretical no pain periods were steadily recognized as pain (e.g., the beginning period and some P1 periods in Example 1). On the one hand, ANS-based signals are widely studied in emotion recognition studies such

as differentiating the positive arousal and negative arousal of emotion (e.g., [200]), and stress detection (e.g., [201]). On the other hand, pain itself is an emotion which is measured with the level of pain unpleasantness or pain distress [22, 202] in addition to the sensory intensity, and could occur with other negative emotions such as stress and fear. It may not be necessary to differentiate it from other negative emotions as they interact with each other within the whole body system [203]. Nevertheless, this limitation would probably restrict the application within the scenarios where the pain is more likely to occur such as in a recovery room or ICU.

7.2.2 Step 2 - Estimating pain level when pain is present or pain control is inadequate

The results in the S2 classification are provided assuming the presence of the pain/inadequate pain control has been correctly estimated in S1. In this part, only P_2 periods are labeled and classified. Three pain intensity classes are defined according to the VAS/NRS score, where intensity-1 is VAS/NRS [4,6), intensity-2 is VAS/NRS [6,8), and intensity-3 is VAS/NRS [8,10]. As it is known that P_2 always starts with VAS/NRS 3 or 4, the beginning of all the P_2 periods are labeled as intensity-1, while the end of each P_2 period is labeled based on the final self-report score. During each P_2 period, the pain stimulus intensity continues increasing, and thus the pain sensation intensity is considered as increasing (as visualized in Figure 7.1), and the samples in the middle are labeled linearly based on the start label and the end label. For example, if the final self-report VAS was 8, the samples in P_2 are labelled intensity-1 at the beginning, intensity-2 in the middle part and intensity-3 in the end part.

As S2 is a three-class classification problem instead of a binary classification one, the models used in this step are slightly different from the ones used in S1. The performance in this part is evaluated with accuracy. The same leave-subject-out cross-validation was conducted in S2.

Feature importance

The relative importance of the features in all the training folds of S2 classification is visualized in Figure 7.9. Generally, the average test accuracies in all these tree-based classifiers are poor even though the training performances are good. Among them, the ensemble_tree_Bag method gives the best training and test performance. Similar to the feature importance in S1, GSR features, and ECG features show an importance higher than the other ones. After averaging the curves (the values in each curve are normalized between 0 and 1), the top five features in S2 with the highest average values are: (feature number – feature name – average importance value out of 1)

- feature 85 – gsr_t_avg – .81/1

- feature 100 – ecg_nlf – .63/1
- feature 102 – ecg_lf_nf – .58/1
- feature 103 – ecg_sampen – .48/1
- feature 97 – ecg_hr_median – .44/1

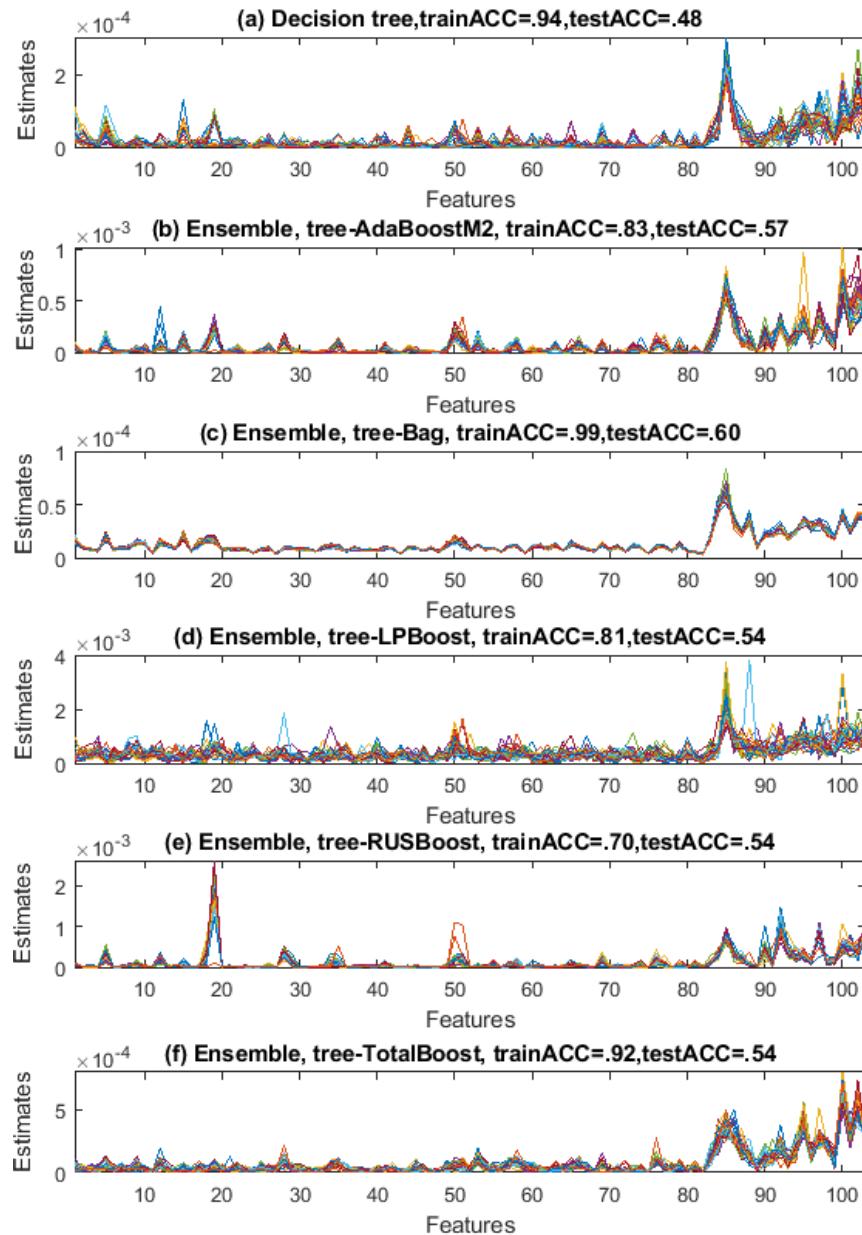


Figure 7.9: S2 - relative importance of the features in tree-based classifiers with average train accuracy and test accuracy

Performance of the models without tuning

The performance of all the implemented models are presented in Figure 7.10. Similar to the test sensitivity results in S1 (Figure 7.5), the test accuracy in all the models shows considerable variance in the 30 test folds. The median test accuracy is always around .5 meaning random guessing, which indicates that half of the test folds are unpredictable. A majority of the models show overfitting, where the test accuracy is much lower than the training accuracy. Among all the tested methods, the best four methods with the highest three average test accuracy are `svm_linear` (.61), `ensemble_tree_bag` (.60), `knn_mahalanobis` (.59), and `ensemble_tree_AdaBoostM2` (.57). These four models are further optimized in the following part.

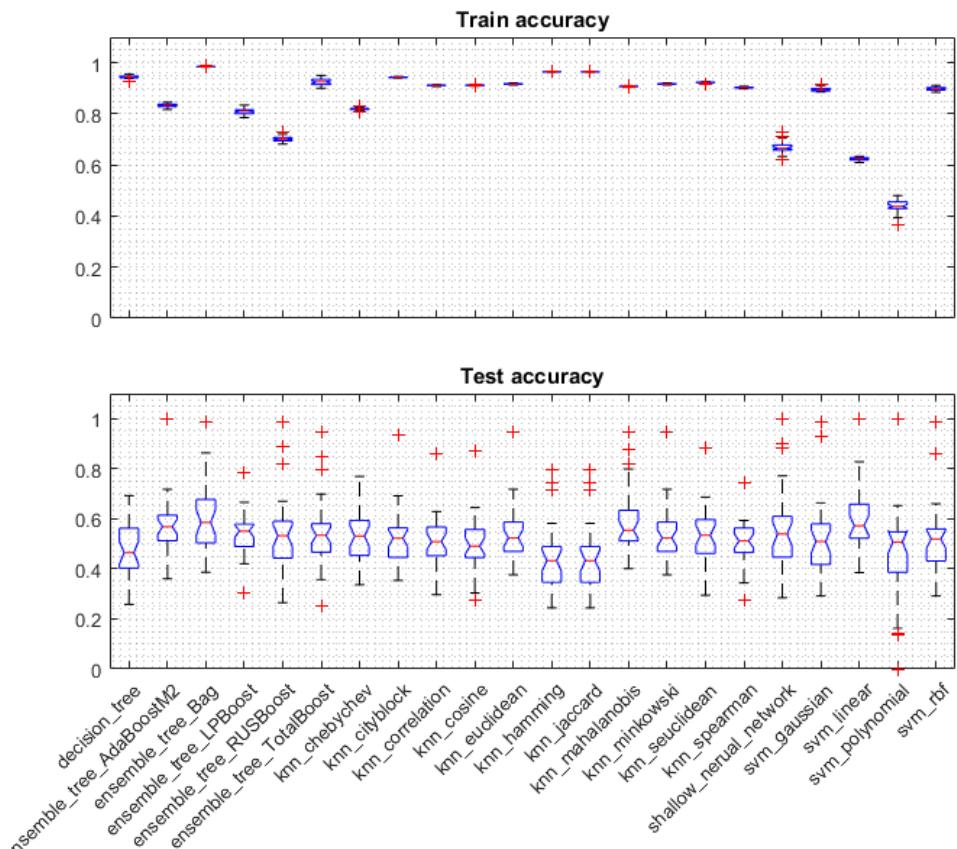


Figure 7.10: S2 - Model performance without hyperparameter optimization

Model tuning and the tuned performance

In contrast to the optimization in S1, the performance metrics used in S2 optimization was accuracy. In the linear SVM multiclass models, the Box-

Constraint, KernelScale, and Standardize parameters were optimized; in the ensemble_tree_Bag model, the number of learning cycles was optimized; in the ensemble_tree_AdaBoostM2 model, both learning rate and number of learning cycles were optimized; and in the knn_mahalanobis model, distance weight, number of neighbors, and standardize were optimized.

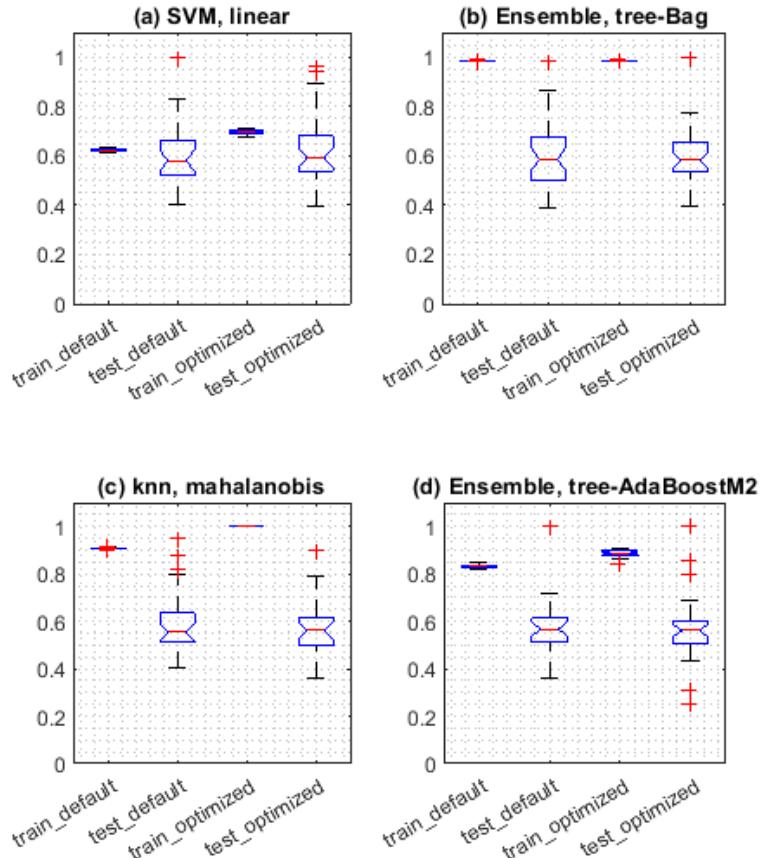


Figure 7.11: S2 - Training accuracy and test accuracy change after hyperparameter optimization

The accuracy comparison between before and after optimization are visualized in Figure 7.11 as boxplots. The training accuracy was improved notably in most of the models (except for ensemble_tree_Bag, which was already nearly 1 before optimization). However, the average test accuracy was improved only in optimized_svm_linear from .606 to .625. Thus, all the pre-defined pain intensity labels and the predictions by the optimized_svm_linear model are presented in Figure 7.12. The predictions in Figure 7.12(b) show that none of the samples labeled as intensity-3 were correctly recognized, and thus the subjects without rapid and large inten-

sity oscillations were prone to show a better test performance (e.g., subject 12 and 29). In some test folds, some relative pain intensity changes were reflected in the predictions, although the absolute pain intensity was not correctly estimated (e.g., subject 1, 4, 15, 25 and 27).

Table 7.5: S2 - The best test accuracy and the corresponding model for each test subject fold and the test accuracy standard deviation among the implemented 26 models (subjects in best test accuracy descending order)

Subject	S1 best sensitivity	S2 best testACC	S2 best model	TestACC SD
14	.795	1	ensemble_tree_AdaBoostM2	.264
17	.857	1	svm_linear	.217
18	1	1	svm_polynomial	.167
8	.938	.942	optimized_svm_linear	.152
12	.680	.900	shallow_neural_network	.151
13	.943	.854	optim_ensemble_tree_AdaBoostM2	.097
6	.961	.799	knn_mahalanobis	.125
5	.443	.797	knn_hamming	.141
20	.885	.760	shallow_neural_network	.080
11	.839	.744	knn_mahalanobis	.108
30	.997	.720	optimized_ensemble_tree_Bag	.107
16	.914	.716	optimized_svm_linear	.105
2	.727	.701	svm_linear	.076
23	1	.699	ensemble_TotalBoost	.080
28	.924	.694	svm_linear	.104
7	1	.691	optimized_ensemble_tree_Bag	.100
25	.968	.683	optimized_svm_linear	.097
22	.905	.679	shallow_neural_network	.125
21	.974	.673	svm_linear	.106
29	1	.664	svm_linear	.099
3	.766	.661	knn_chebychev	.079
4	.993	.655	optimized_ensemble_tree_Bag	.079
24	.971	.653	ensemble_tree_Bag	.061
27	.974	.618	optimized_ensemble_tree_Bag	.067
10	.933	.615	ensemble_AdaBoostM2	.064
26	.856	.606	knn_mahalanobis	.077
15	.944	.591	svm_linear	.072
19	.974	.587	optimized_ensemble_tree_Bag	.063
1	1	.550	knn_spearman	.079
9	.920	.524	knn_hamming	.075

Discussion

The best accuracies and models for each test subject fold are listed in Table 7.5. Similar to the S1 classification, a test fold with a higher best performance is in general prone to have a larger performance variation among all the models in general in S2. Unlike to S1 classification, no model gave a best performance to the majority of the test folds (optimized_ensemble_tree_Bag in

5 test folds and svm_linear in 6 test folds).

The defined labels changed rapidly in some cases, which may limit the prediction performance. An extreme example is that the labels changed from intensity-1 to intensity-3 in 11 seconds in subject 1. First, the predictions can hardly be smoothed in post-classification processing to remove the noise in output (as done in Figure 7.8). Second, in cases where the GSR features and ECG features dominate the predictions (presented in Figure 7.9), their responses may not catch up with the transient change in pain.

7.3 Discussion of automatic pain assessment algorithms

Pain has been studied within clinical decision support systems for use in diagnosis, treatment, screening, and risk assessment purpose since the middle 1990s [204], where the input data were the descriptions of, for example, demographics, symptoms and clinical history. With the development of the sensing and data acquisition technology, an increasing number of signals can now be streamed, recorded and analyzed, which makes continuous monitoring of pain feasible. In the current context, automatic pain assessment means predicting the presence of pain or pain intensity from any reachable data, especially in a continuous manner. Like most of the cases where machine learning is used to solve real-life problems, building machine learning models is highly reliant on domain knowledge. The models built/constructed can also help update the domain knowledge in turn.

The basic machine learning principles are the same across application domains, but how to generalize the learning methods and results across different but similar pain studies has not yet been sufficiently discussed. In this part, the methods and results from automatic pain assessment studies are concisely reviewed. The representative studies are listed in Table 7.6 in terms of pain scene, group size, learning variables, variable labeling in supervised learning, signal processing if there are any, learning algorithms, and performances.

Among the 11 studies, 7 pain databases are involved covering categorical/descriptive variables [15], video frames [26, 28, 205], RGB color photographs [27], and physiological signals [28, 206]. The subject groups and pain types covered were: postoperative pain in adults (aged 21 years and over) [15], postoperative pain in children (aged 5 to 18 years) [205], sensation after stress-inducing stimuli in infants (aged 18 hours to 3 days) [207], self-identified shoulder pain participants underwent active and passive motion tests [208, 209], blood pressure cuff induced experimental pain in adults [210], and experimental heat pain in adults [157, 199, 210–212]. Other similarities and differences on algorithm and performance are summarized below.

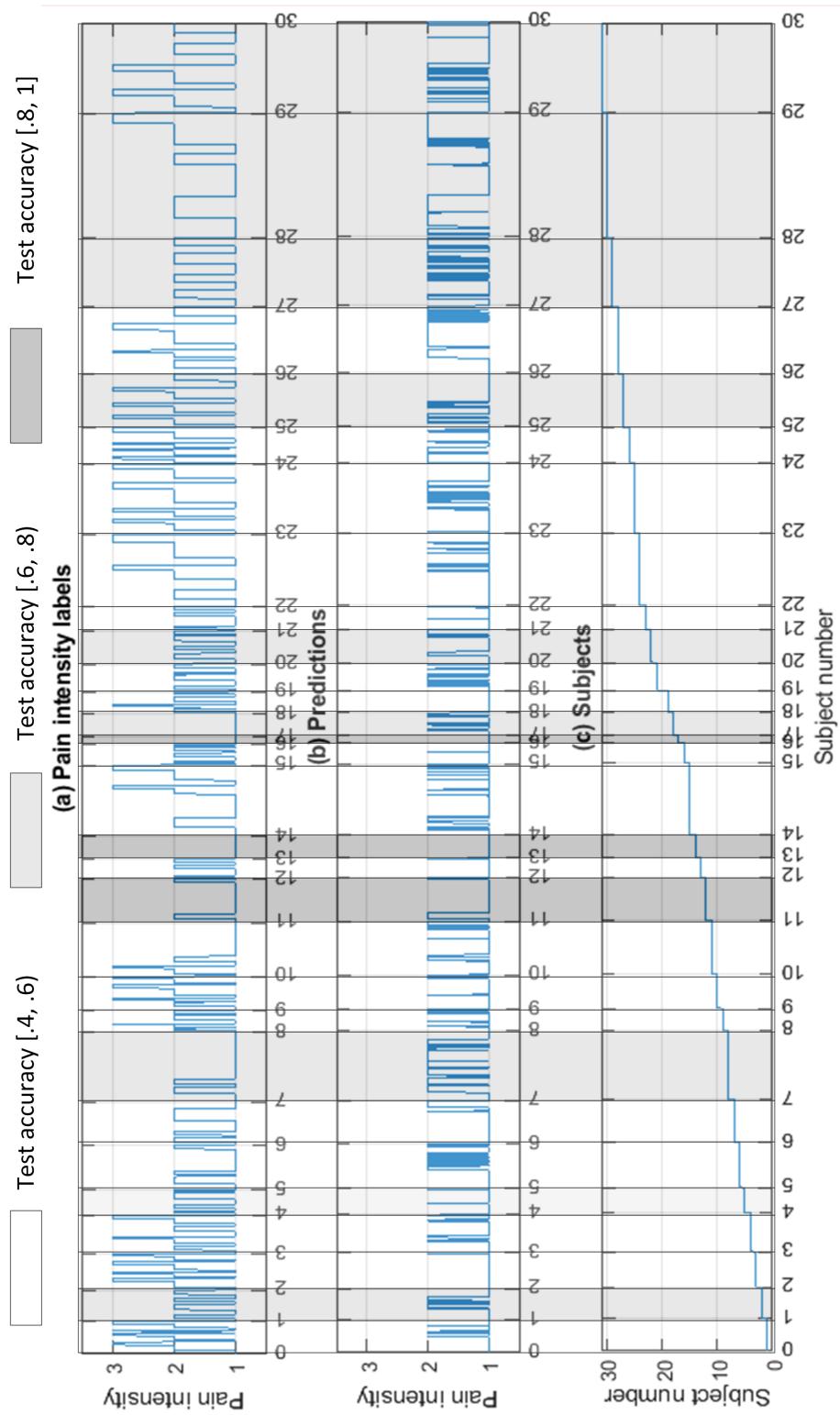


Figure 7.12: S2 - Pain intensity labels and predictions in optimized_svm_linear 3-class classification

Table 7.6: A list of representative studies on automatic pain assessment using machine learning

Study	Pain scene	Group size	Learning variables	Labels	ML algorithm	Performance
Tighe <i>et al</i> (2015) [15]	postoperative pain	8071	796 variables including demographics, home medications, comorbidities, surgical procedure and surgical circumstances	presence or absence of a moderate to severe pain score (NRS 6, 7-10), pain score recorded every 4h	LASSO, gradient-boosted decision tree, SVM, neural network, k-NN, logistic regression; 40% training, 30% validation, and 30% testing	AUC from .5 to .727, accuracy from 5 to 68
Sikka <i>et al</i> (2015) [205]	postoperative pain	50	Video face (analyzed with CERT), demographics, elapsed time since surgery	1. presence or absence of clinically significant pain (NRS \geq 4); 2. subjective pain-intensity (NRS from 0 to 10); 3. number of days since surgery as objective ground truth for pain	ongoing/transient pain, 1. binary classification, linear regression in pain-intensity estimation; 10-fold (leave-5-subject-out) cross-validation	1. average AUC from .84 to .91; 2. average Pearson's correlation from .65 to .90 within-subject, from .45 to .68 overall
Misra <i>et al</i> (2016) [206]	experimental pain (thermal)	30	128-electrode EEG (analyzed with EEGLAB Toolbox)	low pain (VAS=1) and high pain (VAS=5), each trial lasted for 4 s	logistic regression and SVM (linear and Gaussian kernels); leave-one-out cross-validation	maximum accuracy from .81 to .90
Jang <i>et al</i> (2015) [210]	experimental pain (blood pressure cuff), boredom and surprise	217	HR, LF, HF, SCL, mean temperature	pain, boredom and surprise	discriminant function analysis, linear discriminant analysis, classification and regression trees, self-organizing map, Naïve Bayes and SVM; 10-fold cross-validation	accuracy from .62 to .85

Gholami <i>et al</i> (2010) [207]	stress-inducing stimuli to neonates	16 out of 21	181 photographs (Infant COPE Database)	RGB color	10 pain intensities out of pain score between 0 and 100	relevance vector machine (RVM, a Bayesian extension of SVM); leave-one-out cross-validation (within subject)	weighted kappa efficient .47 between human examiners and RVM; accuracy .91 in binary classification (a score above 50 is pain)
Gruss <i>et al</i> (2015) [199]	experimental pain (thermal)	85	42 features from EMG, EDA and ECG (BioVid heat pain database)	from old, pain tolerance and two pain intensities in the middle	no pain, pain threshold	SVM; 75% training and 25% testing	accuracy of 5-class classification .43, accuracy of binary classification between no pain and one pain intensity is from .79 to .91
Lopez-Martinez <i>et al</i> (2018) [157]	experimental pain (thermal)	87	EDA and HRV features (BioVid heat pain database)	no pain and four pain levels	binary classification with logistic regression, SVM linear and SVM RBF; regression with linear regression, SVR linear, SVR RBF, non-recurrent neural networks, recurrent neural networks and long short-term memory networks (LSTM-NN); leave-subject-out cross-validation	best accuracy of binary classification between no pain and pain tolerance: .74 with logistic regression and SC, best mean absolute error with LSTM-NN and SC: 1.05	
Werner <i>et al</i> (2016) [211]	experimental pain (thermal)	87	head pose and facial expression features from 8700 video samples (BioVid heat pain database)	no pain and four pain levels	random forest, SVM linear and SVM RBF; leave-subject-out cross-validation	best accuracy of binary classification between no pain and pain tolerance: .72, best accuracy of 5-class classification: .31	

Kächele <i>et al</i> (2015) [22]	experimental pain (ther- mal)	86	features from EMG, ECG, EDA and video (BioVid heat pain database)	no pain and four pain levels or four pain lev- els in regression	random forest, binary classification with early or late fusion; leave-subject-out validation	best accuracy of bi- nary classification be- tween no pain and pain tolerance: .83 with late fusion, best root mean square error: .98
Kaltwang <i>et al</i> (2012) [208]	shoulder pain	25	200 sequences video, in total 48398 frames (UNBC-McMaster Shoulder Pain Ex- pression Archive Database), ana- lyzed with active appearance model	6 levels AU intensity (0-5) and 16 levels pain intensity (0-16)	relevance gression with kernel; leave-subject-out cross-validation	best mean squared er- ror when pain is cal- culated from the esti- mated AU intensities: 1.368
Zhou <i>et al</i> (2016) [209]	shoulder pain	25	UNBC-McMaster Shoulder Pain Ex- pressions Database	6 levels AU intensity (0-5) and 16 levels pain intensity (0-16)	recurrent works leave-subject-out validation	mean squared error: 1.54

7.3.1 Signal processing time window and the pain assessment frequency

The first differences are the signal processing time window and the pain assessment frequency, which are closely associated with the study design and the chosen signal(s). The data labeling and pain assessment frequency are summarized in three types in [211]: frame-level, window-level, sequence-level (e.g., video sequences, or sample-level); and day-level was found in [15].

The frame-level classification is specific to image or video methods, where the pain intensity and even the AU intensities in each video frame are manually labeled and classified. The concept could also be generalized to the facial sEMG method due to the same facial expressions information being collected. Although the AU intensity can be recognized automatically (with the expression recognition toolbox, e.g., [213]), it still requires expensive manpower work to label pain intensity frame by frame. Among all the assessment frequencies, a frame-level assessment may give the best assessment resolution (e.g., 30 frames per second) and could form a continuous monitoring method [208]. Analyzing each frame could be easily be alleged to lose dynamic facial expression information (e.g., an eye blink or eye enclosure). The dynamics of facial expressions are taken into consideration in the frame-level analysis in [209] where a sequence of 30 consecutive frames was the input to recurrent recurrent convolutional neural network (RNN) regressions. The RNN regression performance (mean squared error) in [209] shows its performance was comparable but not superior to the best regression performance with feature fusion in [208]. Compared to the window-level and sequence-level classifications, it is easier to generalize the algorithms across databases in frame-level classifications where no preset window-length parameter is needed.

The window-level classification is another solution to continuous monitoring, where a time window length and a time step or offset are defined for feature extraction or/and classification. The pain state or intensity could change as the time window slides. This method is mainly found in experimental pain studies (BioVid and this study), and values are restricted by the study design, especially the pain stimulus duration. For example, a pain stimulus duration of 4 s, which is followed by an 8-12 s recovery/baseline, and therefore a 5.5 s time window was defined in the sample-level analysis in [199]. Correspondingly, the sliding time window was defined with a width of 5 s (5.5 s) and a step of 0.5 s (0.1 s) in [157] ([212]). The choice of the feature extraction window length could be more flexible and adapted to the needs of a feature when pain is continuous (e.g., SCL in 10 s window and HRV-HF in 60 s window during surgery pain [30]).

The sequence-level or sample-level classification is commonly used [157, 199, 205, 206, 210–212]. Usually, a fixed time window with a known start

and end is used in the analysis. The pain or pain stimulus is controllable most of the time in this case. For example, manual pressure was exerted at the surgical site for a 2 or 10-second periods which is a typical clinical examination and could cause transient postoperative pain [205]. Some other examples can be found in experimental pain studies where both the stimulus duration and intensity are controllable. Analysis time windows of different lengths (5 min and 10 s in [205], 5.5 s in [199, 206, 211, 212], 6 s and 8 s in [157], and 30 s in [210]) were selected based on the study design and/or a visual inspection of the signal's temporal dynamics which makes the methods and results less interpretable.

How frequent the pain assessment/reassessment is needed depends on the circumstances, such as, the type of pain (acute or chronic) and care setting (inpatient or outpatient) so as to recognize the changing needs of the patients [14]. For example, the nociception/antinociception balance index or depth of anesthesia is nearly a real-time monitoring during surgery [214]. The pain intensity record interval is a quality indicator in acute pain management, which is recorded as the number of times recorded every 24 hours [14]. In practice, the nurses may assess postoperative pain every 15 min in the recovery room [16] and may also assess postoperative pain every 4 h with a repeat query within 1 h after administration of analgesic medications from postoperative day 1 to day 3 [15]. Therefore, from the postoperative pain point of view, either frame-level or window-level classification can provide a much better time resolution in continuous monitoring than manually requesting a pain self-report, and thus, improving the assessment frequency is not yet prior to the aim of performance improvement. However, on the other hand, the labels could be sparse during continuous data (e.g. DATA@1s and label@10 min) which is a challenge when using supervised learning methods. For the critically ill patients in an ICU, pain is often assessed before and during painful procedures such as endotracheal suctioning and mobilization [22, 215]. Therefore, the near real-time attribute in continuous pain monitoring may be more emphasized in surgical pain and the pain induced by procedures. Moreover, pain in medical procedures may be classified at a sequence-level as each procedure is an event with a short duration.

7.3.2 Labels and proxy/objective pain assessment

In clinical practice, NRS or VAS are usually used to evaluate pain intensity when self-report can be obtained. In the cases where patients are unable to self-report, validated pain behavior tools and proxy reporting from family members, parents, and caregivers are the alternatives to assess pain [24]. Similarly, numeric scales, objective and proxy assessments are found in automatic pain assessment studies but with some differences in, e.g., scale range and description as introduced in Section 2.2.

In some studies, more than one pain measurement was collected. In UNBC-McMaster database, in addition to the 16 intensity objective PSPI scores labeled in the frame-level, the sequence-level labels include a self-report 10-cm VAS, and an objective pain intensity (0-5) assessed from facial expressions by two trained observers (two objective ratings Pearson's correlation .8). Another two 15-item pain descriptor scales were also collected in the same study as described in [216]. One reflected the sensory intensity of pain starting at "extremely weak" and finishing at "extremely intense", and the other an affective-motivational scale starting at "bearable" and finishing at "excruciating". However, neither of these subjective two-verbal pain descriptors were found in use, and the VAS scores were only occasionally found being used [217] as labels in classification. Moreover, most of the subsequent studies estimated pain from the AU intensity or PSPI formula only. This may be explained by the motivation of "estimating the pain intensity like an expert observer could" [216]. Similar to the UNBC-McMaster database, the sensory dimension (i.e., pain intensity) and affective dimension (i.e., pain distress or unpleasantness) of pain was also collected from ICU patients and analyzed in [22].

In another face video study [205], both parents and inpatient nurses gave proxy 11-point NRS reports in addition to the self-report NRS reports from the children. With a slightly different goal, self-reporting was taken as the subjective ground truth. The second ground truth was an objective one which was defined as the elapsed time since surgery. This objective ground truth was defined based on the study design where the data collection occurred within 24 hours, after 1 day, and 21 days after the surgery, and the pain intensity logically decreased over time. Interestingly, the performance results show that the machine estimation with reference to the objective ground truth was superior to those of the subjective ground truth, and slightly inferior to parents measuring subjective pain because the parents knew the amount of time that had elapsed. The machine estimates were better than the parent measures when the elapsed time was added as a model input in pain intensity estimation.

The occasionally unreachable self-reports were treated as missing values in [15]. In the cases where a self-report was impossible to obtain (e.g., during a surgery), the pain or pain intensity references were: 1) some physiological signs such as increased blood pressure or heart rate, movement, and coughing (e.g., [51]); 2) a comprehensive evaluation on nociceptive stimuli and analgesic drug (e.g., opioid) concentration [30, 59] or 3) human rating from pain facial expression images [207]. Other than these measures, the level of sedation and its interaction with the time of assessment (before or during the procedure) were collected when validating the behavior pain assessment tool [22]. These provide reference experience for the physiological signal based methods as, for example, patient arousal has been found to influence

the predictivity of postoperative pain with the surgical pleth index [218].

Supervised learning models were built in all the studies, as shown in Table 7.6 and were all highly reliant on pre-defined labels. To the best of my knowledge, there have been no knowledge discovery or data mining methods applied to any of the pain assessment databases. The classifications mainly fall into three types: binary classification, multi-class pain intensity classification, and multi-state classification where the pain was one of the emotional states [210]. The binary classes were consistently defined across the postoperative pain studies in the table, and also in some other non-machine learning studies where NRS was divided at a score of 4 due to clinical significance. A variety of other choices can be found in experimental pain studies including pain v.s. no pain (e.g., [27]), one pain intensity v.s. no pain (e.g., [199]), and between two pain categories (e.g., [206]). For the multi-class pain intensity classification, the pain intensity categories could also be defined differently (e.g., [216] and [28]) in addition to the standard NRS and VAS definitions which limit the interoperability across databases. Another challenge to multi-class pain intensity classification is the limited number of samples in each pain intensity. A pain database is already hard to build even though with a small sample size, not to mention the feasibility of dividing the total sample size into 10 groups if using NRS. Some solutions currently being applied to the challenges are merging adjacent intensities and using probabilistic classification or regression instead.

7.3.3 Data exclusion

One fact observed from the studies that may hinder the generalization of the results was the additional data inclusion/exclusion rules in some studies. Besides the general inclusion/exclusion rules (e.g., patient group), the data selection was found at the subject level, trial level, and label level. The inclusion or exclusion rules were mostly defined based on pain intensity aiming at maximizing experimental variance and minimizing error variance [219]. For example, "subjects were included if they had a minimum of one trial with objective pain intensity (0-5) rating of 0 and one trial with a rating of 3, 4, or 5,..., movements with intermediate ratings of 1 or 2 were omitted." [219]; "...trials with self-ratings of 1 to 3 (NRS) were excluded from the binary classification analysis." [205]; no pain data were excluded in the training of pain level regression in continuous predictions [212]. In addition, the analyzed data segment might be a biased cut from an entire recording [210]. It could be expected that such data exclusion might have elevated the performance and in the interim limit use cases. Moreover, the performance difference with and without data exclusion was rarely presented.

7.3.4 Machine learning algorithms, cross-validation and performance evaluation

The studies reviewed so far mainly follow by the standard algorithm flow, which starts with pre-processing, followed by feature extraction and machine learning. Within an individual study, the performances are compared among features/data fusion/face descriptors [199, 205, 206, 208, 212], between different pain measures [205, 207], among machine learning algorithms [15, 157, 206, 211], with previous studies [209, 211], or covering several of these angles. Table 7.7 compares 10 of the 11 studies regarding machine learning algorithm and performance listed in Table 7.6, which focuses on pain and pain intensity estimation.

A variety of machine learning algorithms on classification and regression and deep learning methods were implemented in the studies. In these incomplete statistics, support vector machine and logistic regression most frequently appeared, especially in studies with a binary classification. Regression was found to be used in many studies with pain intensity estimation, especially in a continuous manner. In most of the reviewed studies, the influence of intra-individual similarity was eliminated by using leave-subject-out cross-validation. Two proper exceptions were when no dependent data from the same person was involved in the model [15] and when the prediction was conducted within an individual [207]. The remaining two studies achieved higher performance than the others, probably due to mixing intra-individual similarity with inter-individual similarity.

The performance metrics in the listed studies cover validity and reliability measures. In the classification models, either AUC or ACC was used to measure the model validity. The best AUC or ACC was .90; however, these performances may be overestimated due to the biased exclusion or improper cross-validation as mentioned before. In regression models, the validity measures including mean absolute error ($MAE = \frac{1}{n} \sum |y - \hat{y}|$, where y are the actual values and \hat{y} are predicted values), mean square error ($MSE = \frac{1}{n} \sum (y - \hat{y})^2$), root mean squared error ($RMSE = \sqrt{MSE}$), and R squared (R^2 is the percent of variance explained by the model). In R^2 , the higher the value, the better the prediction performance, whereas in other regression validity measures the lower the better. RMSE or MSE appeared in four of the total five regression models. The RMSE values are between .98 and 1.29 among the four models across two databases. A single validity metrics to evaluate both classification and regression models could be concordance statistic/index (C-index) as a discrimination measure [34, 220]. The C-index is always between .5 and 1 and represents AUC with binary classes.

The reliability measures are Cohen's kappa coefficient (κ) and Pearson's r , which are agreement measures between two methods (or observers). κ is supported in measuring the degree agreement with the binary classes [221].

For classification with more categories, weighted κ is used instead [207]. The neonate pain study [207] resulted in a κ .47 between the estimated pain intensity (within-subject) and human expert examiners, which is moderate but close to a fair agreement between the two methods. In the children's postoperative pain study [205], κ was measured in the binary classification. The measured values show that the estimated transient pain had a better agreement with all the other methods (i.e., subjective report, proxy estimates by nurses and parents, elapsed time since surgery) compared to the estimated ongoing pain. In general, the proxy estimates by parents agreed more with self-report than the machine estimates (.72/.5 v.s. .61/.36), and the machine estimates had equally (transient: .61) or slightly better agreement with self-reports than proxy estimates by a nurse (ongoing: .36 v.s. .15).

Pearson's r was used to check the agreement between a regression model and a second method. It was suggested that Lin's concordance correlation coefficient [222] and intraclass correlation coefficient are more appropriate measures for the agreement because a systematic effect (or bias/distance) is not considered when measuring the level of agreement between two methods using Pearson's r [221]. Nevertheless, the overall Pearson's r from the three studies were between .45 to .68 between the machine estimates and the self-report method or an objective assessment regardless of the database [205, 208, 209], which was a moderate to strong linear correlation.

7.3.5 Future work

It can be seen that many efforts on different data or signal sources have been made for automatic pain assessment. Some more recent efforts are from the perspective of feature selection [223], developing personalized machine learning models (e.g., [217]), and exploring new behaviors or clues of pain [224, 225]. Most of the studies were conducted within a single database and study design, therefore lacking a generalization of methods and results. Hence, the generalization problems such as methods migrating from one database to another and from experimental pain to pain in real-life should be to be examined further. This is feasible as a growing number of databases are accessible or under construction. It is meanwhile challenging as the varieties with data collection protocols may limit the method choices, and may alter pain perception, pain behaviors, and even pain reports.

Improving the validity and reliability of automatic pain assessment tools may not merely rely on algorithm optimization. The pattern recognition results within one database show that a step towards a better estimation may be possible, but the improvements are not yet significant. This may be constrained by the inherent patterns in data with the defined classes/labels. The commonality and difference between the pain behaviors in different groups of people (e.g., gender [226]) are not yet fully understood or considered in

automatic pain assessment studies. Moreover, the potential pattern forms in the databases have barely been explored with unsupervised learning methods such as change point detection [227] and motif discovery [228]. For the time being, the subjective and objective reports (especially when subjective reports are unavailable) may not be the most accurate and repeatable references with a near real-time resolution but are the best available ones. With unsupervised or semi-supervised learning methods, the self-report or objective measures may be less relied on, and the data collection cost may be less expensive.

Table 7.7: Algorithm and performance comparison (studies in Table 7.6)

Study	[15]	[205]	[206]	[207]	[199]	[157]	[211]	[212]	[208]	[209]
LinR	✓	✓								
DT	✓									
RF	✓									
SVM	✓									
RVM			✓							
NN	✓				✓					
kNN	✓									
LogR	✓									
RNNR										
Window	day	seq.								
LSO CV?	No	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Eval. metric.	AUC	AUC, κ / r , z-test	ACC	κ	ACC R^2	ACC / MAE, RMSE, RNNR	ACC	ACC / RMSE	MAE, r	MSE, r
Best ML	LinR	-	SVM (Gaussian)	-	-	LogR/LSTM-NN (RNNR)	RF	-	-	-
Best perf.	.70/.73	.91,.61*/ .68*, 0	.90	.47**	.91***	.74*** / 1.05, 0.24	.129, .72***	.81*** / .84, .98	.1.37, .59**	.1.54, .65**
ML mean(std)	.607(.086)/ .614(.094)	-	.85 (.06)	-	-	.72(.04) / 1.11(.04), 1.33(.03), .13(.08)	.71 (.02)	-	-	-

LinR-linear regression, DT-decision tree, RF-random forest, SVM-support vector machine, RVM-relevance vector machine, NN-neural network, kNN-k nearest neighbor, LogR-logistic regression, RNNR-recurrent neural network regression; ML-machine learning, LSO-leave subject out, CV-cross validation; AUC-area under the ROC curve, ACC-accuracy, κ -Cohen's kappa coefficient, r -Pearson correlation, MAE-mean absolute error, RMSE-root mean squared error, R^2 -R squared, MSE-mean square error,

* between estimation and subjective rating;

** between estimation and human expert coded PSPI;

*** classification between low/no pain and high pain tolerance.

Chapter 8

Conclusions

The aim of this thesis has been to develop an automatic pain assessment tool from the continuous-time signals, ECG, GSR, and EMG. The signals collected before and during the experimental pain were induced. The processed outcomes (features) were then fed separately into different classification models in the estimations in two steps - predicting the presence first and then the intensity. In this concluding chapter, the findings and resulting conclusions will first be presented, and then followed by the contributed highlights of the study and recommendations for future research.

8.1 Main findings

The main findings and resulting conclusions of this thesis are:

- Many efforts have been put into developing automatic pain assessment tools with different signs of pain, yet the expected assessment aim and outcome may be divorced from reality especially among the ones with experimental pain stimulus.

There have been a growing number of studies on automatic pain assessment since 2010. However, the study designs show many varieties. The different choices of the ground truth definitions (e.g., scale with different ranges and descriptions) may indicate the slight difference in estimation aim (e.g., expect the machine to work comparable to an objective observer or how the patient subjectively feels). One aim might be better achieved than another with the chosen signal(s) when there are multiple ground truth sources. Nevertheless, the labeling variety is mainly found in the databases for non-clinical pain. Except for the cases where self-report is impossible to obtain, 11-point NRS and 100 mm VAS are the most commonly used where 0 means no pain and the maximum value means the worst possible pain. Moreover, although $NRS \geq 4$ is a clinically significant pain it has seldom been considered in non-clinical pain studies. The input signals have covered ANS based biosignals,

facial expression, and neuroimaging signals, which overlap with the existing nociception/anti-nociception balance tools and behavioral pain assessment tools.

- A good estimation was reached with the existence of pain or inadequate pain control, but the method and result should be used with caution. Further improvement is still needed in predicting pain intensities.

In the designed signal processing flow, features were extracted from the sliding windows with the same offset and different lengths so that the features were updated in the same step for continuous prediction. The data in either step of the classification are imbalanced in nature, which could explain why in the S1 classification the random undersampling boosting model reached the best prediction sensitivity (median value: 90%) and good specificity (median value: 84%) when the model was optimized by reference to AUC. In the S2 classification, the best test accuracy for the three pain intensities was 62.5% with failure to recognize all the highest intensities. However, it should be noticed that the tonic component of GSR dominated both predictions especially in S1, which was followed by heart rate in S1 and the low-frequency part of heart rate variability in S2. The dominant features do not respond specifically to pain and thus they should be used with caution. The importance results for the feature were consistent with the feature visualization observations. It is worth noting that the low-frequency part of the heart rate variability (0.04-0.15Hz) extracted from a 250 s long time window may to some extent reflect the subjective difference in pain perception rather than a change of pain.

- Within the signal processing system developed from SpaExp, the order of importance among the processing parts is proposed to be: estimation goal \approx signal > feature > window setting \approx machine learning model.

This order was achieved by comparing the results within and across the four studies based on the SpaExp database which includes the three published works [study 1-3] and the work presented in this thesis [study 4].

(a) signal > feature: Among the two studies where the GSR signal was involved in addition to other signals, the GSR signal either showed the highest linear correlation [study 1] or contributed most to the estimation [study 4] regardless of whether a feature extraction was implemented. Moreover, when comparing the two studies using only GSR [study 3] or ECG [study 2], the ECG features could not differentiate between P_1 , P_2 , and partial P_0 ¹ better only than random guessing, but the GSR features could and reached c-index above 0.8.

(b) feature > machine learning model: With a further signal decomposition and feature extraction, study 3 showed that the effectiveness of GSR was actually reflected by the tonic component and features extracted from it among

¹partly P_0 was the 30 s period cut before t_0 for data balancing

all the 6 machine learning models. Using tonic features a better performance of around 0.25 c-index was achieved in contrast to using only the phasic features (around 0.6), while the variation among the machine learning models using the same feature set was around 0.1.

(c) feature > window setting: In the only study where time windows in multiple lengths were compared with ECG features [study 2], the AUC variation with the same features and the same binary classifier (between partly P_0 and $P_1 + P_2$ in separate pain stimulus) was no larger than 0.1 (from 0.7). However, the length choice for each feature may be subject to the study design and its definition, which were both discussed in study 4.

(d) estimation goal \approx signal: In the two studies where multiple signals were involved [study 1,4], the average prediction accuracy was improved from 68% to 83% in a continuous estimation by adjusting the goal to estimate the presence of pain/inadequate pain control (between whole $P_0 + P_1$ and P_2). The other efforts to improve many of the other aspects may contribute to the overall improvement as well. However, this may indicate a need to ask more questions from the data, especially questions about clinical concerns may lead to clearer evidence. Finally, it can be concluded that the estimation performance is probably limited by the inner pattern of the estimation goal in one or several signals, and thus finding new evidence of the truth may help develop automatic pain assessment methods, and vice versa.

- Despite a reasonable performance on average, the performance difference among subjects should be noticed. Meanwhile, the challenge of data labeling when estimating pain in continuous pain intensity should be either solved or bypassed.

The performance difference among the test folds in leave-subject-out cross-validation is the limitation of the method, which was observed in both classification steps. For example, of all the 30 subjects in the S1 optimized_ensemble_tree_RUSBoost model, the sensitivity was nearly 100% in 12 test folds but below 60% in 7 folds, even though the median value was as high as 91%. A similar performance difference was found in other studies, which reflects the level of generalization and reliability of a method across people, and thus needs to be reported. It is almost impossible to label each sample when using supervised learning methods to make a dense estimation of window-level pain intensity over time. Therefore the intensity labels during P_2 were manipulated with interpolation. The interpolation resulted in a larger data imbalance ratio and might introduce a bias to the labels. Solutions to this challenge or how to bypass it needs to be further explored.

8.2 Significance of the study

The significance of this study to the automatic pain assessment field include:

- A review of the system level of the automatic pain assessment method development studies and gives an overall picture of the field bridging the clinical and experimental points of views, a gap that has rarely been discussed previously;
- Development of a system for pain estimation with ECG, GSR, and facial EMG signals and a thorough analysis of their roles in the estimation; this could provide a benchmark value for future studies;
- Summary on the importance of each part in the classic pattern recognition processing flow that contributes to pain estimation, which is usually discussed separately.

8.3 Future research

The proposed processing flow in this study were conducted offline step by step, and the synchronization between the processes was not fully taken into consideration. Additionally, the feature normalization was based on the distribution of the data within each subject after the data collection, which is another limitation to be solved in near real-time estimations. This study was conducted with well-controlled pain stimuli and signal quality where the subject was always half-lying on an armchair with few movements. Moreover, the signal processing results were strictly checked and corrected to ensure processing accuracy. However, an application in real life may encounter different uncertainties as such. These uncertainties need to be evaluated and taken into consideration in future work. More evidence from a broader range of signals and pointcuts are needed for a better understanding and to improve automatic pain assessment validity in general. Studies across different databases or/and from the experimental to the clinical environment should be encouraged, although there still exist many challenges.

Bibliography

- [1] A. C. Metting van Rijn, A. Peper, and C. A. Grimbergen. High-quality recording of bioelectric events - Part 1 Interference reduction, theory and practice. *Medical & Biological Engineering & Computing*, 28(5):389–397, 1990.
- [2] Michael R. Neuman. Biopotential amplifiers. In *Medical Instrumentation: Application and Design, 4th Edition*, pages 241–292. 2009.
- [3] Michael J. Shea. Electrocardiography (Merck Manual, Professional Version). <https://www.merckmanuals.com/professional/cardiovascular-disorders/cardiovascular-tests-and-procedures/electrocardiography>. Accessed: 2019-June-09.
- [4] J. De Jonckheere, V. Bonhomme, M. Jeanne, E. Boselli, M. Gruenewald, R. Logier, and P. Richebé. Physiological signal processing for individualized anti-nociception management during general anesthesia: A review. *Yearbook of Medical Informatics*, 10(1):95–101, 2015.
- [5] R. Cowen, M. K. Stasiowska, H. Laycock, and C. Bantel. Assessing pain objectively: The use of physiological markers. *Anaesthesia*, 70(7):828–847, 2015.
- [6] Graham D. Fraser, Adrian D C Chan, James R. Green, and Dawn T. Macisaac. Automated biosignal quality analysis for electromyography using a one-class support vector machine. *IEEE Transactions on Instrumentation and Measurement*, 63(12):2919–2930, 2014.
- [7] D.F. Lovely. Signals and signal processing for myoelectric control. In *Powered Upper Limb Prostheses: Control, Implementation and Clinical Application*, pages 35–54. 2004.
- [8] Charles E Argoff. Recent management advances in acute postoperative pain. *Pain Practice*, 14(5):477–487, 2014.
- [9] Nancy Wells, Chris Pasero, and Margo McCaffery. Improving the quality of care through pain assessment and management. In *Patient safety and quality: An evidence-based handbook for nurses*. Agency for Healthcare Research and Quality (US), 2008.
- [10] A Vania Apkarian, M Catherine Bushnell, Rolf-Detlef Treede, and Jon-Kar Zubieta. Human brain mechanisms of pain perception and regulation in health and disease. *European Journal of Pain*, 9(4):463–463, 2005.

- [11] Joel Katz and Ronald Melzack. Measurement of pain. *Surgical Clinics of North America*, 79(2):231–252, 1999.
- [12] Jacques Simon Lee, Elisabeth Hobden, Ian G Stiell, and George A Wells. Clinically important change in the visual analog scale after adequate pain control. *Academic Emergency Medicine*, 10(10):1128–1130, 2003.
- [13] P S Myles, D B Myles, W Galagher, D Boyd, C Chew, N Macdonald, and A Dennis. Measuring acute postoperative pain using the visual analog scale : The minimal clinically important difference and patient acceptable symptom state. 118(3):424–429, 2017.
- [14] Debra B. Gordon, June L. Dahl, Christine Miaskowski, Bill McCarberg, Knox H. Todd, Judith A. Paice, Arthur G. Lipman, Marilyn Bookbinder, Steve H Sanders, Dennis C Turk, and Daniel B Carr. American pain society recommendations for improving the quality of acute and cancer pain management. 165:1574–1580, 2005.
- [15] Patrick J. Tighe, Christopher A. Harle, Robert W. Hurley, Haldun Aytug, Andre P. Boezaart, and Roger B Fillingim. Teaching a machine to feel postoperative pain: Combining high-dimensional clinical data with machine learning algorithms to forecast acute postoperative pain. *Pain Medicine*, 16(7):1386–1401, 2015.
- [16] Katja Heikkinen, Sanna Salanterä, Marjaana Kettu, and Markku Taittonen. Prostatectomy patients' postoperative pain assessment in the recovery room. *Journal of Advanced Nursing*, 52(6):592–600, 2005.
- [17] J. F. Payen, O. Bru, J. L. Bosson, A. Lagrasta, E. Novel, I. Deschaux, P. Lavagne, and C. Jacquot. Assessing pain in critically ill sedated patients by using a behavioral pain scale. *Critical Care Medicine*, 29(12):2258–2263, 2001.
- [18] Terri Voepel-Lewis, Jennifer Zanotti, Jennifer A. Dammeyer, and Sandra Merkel. Reliability and validity of the face, legs, activity, cry, consolability behavioral tool in assessing acute pain in critically ill patients. *American Journal of Critical Care*, 19(1):55–61, 2010.
- [19] Kathleen A. Puntillo, Ann B. Morris, Carol L. Thompson, Julie Stanik-Hutt, Cheri A. White, and Lorie R. Wild. Pain behaviors observed during six common procedures: Results from Thunder Project II. *Critical Care Medicine*, 32(2):421–427, 2004.
- [20] Deborah G. Klein, Michelle Dumpe, Ethan Katz, and James Bena. Pain assessment in the intensive care unit: Development and psychometric testing of the nonverbal pain assessment tool. *Heart and Lung*, 39(6):521–528, 2010.
- [21] Anne Marie Kabes, Janet K. Graves, and Joan Norris. Further validation of the nonverbal pain scale in intensive care patients. *Critical Care Nurse*, 29(1):59–66, 2009.

- [22] Céline Gélinas, Kathleen A. Puntillo, Pavel Levin, and Elie Azoulay. The behavior pain assessment tool for critically ill adults: A validation study in 28 countries. *Pain*, 158(5):811–821, 2017.
- [23] Laurie Kelly McCorry. Physiology of the autonomic nervous system. *American Journal of Pharmaceutical Education*, 71(4):78, 2007.
- [24] Keela Herr, Patrick J. Coyne, Margo McCaffery, Renee Manworren, and Sandra Merkel. Pain assessment in the patient unable to self-report: Position statement with clinical practice recommendations. *Pain Management Nursing*, 12(4):230–250, 2011.
- [25] Min SH Aung, Sebastian Kaltwang, Bernardino Romera-Paredes, Brais Martinez, Aneesha Singh, Matteo Celli, Michel Valstar, Hongying Meng, Andrew Kemp, Moshen Shafizadeh, et al. The automatic detection of chronic pain-related expression: Requirements, challenges and the multimodal EmoPain dataset. *IEEE Transactions on Affective Computing*, 7(4):435–451, 2015.
- [26] Patrick Lucey, Jeffrey F Cohn, Kenneth M Prkachin, Patricia E Solomon, and Iain Matthews. Painful data: The UNBC-McMaster shoulder pain expression archive database. In *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops*, pages 57–64. IEEE, 2011.
- [27] Sheryl Brahnam, Loris Nanni, and Randall Sexton. Introduction to neonatal facial pain detection using common and advanced face classification techniques. In *Advanced Computational Intelligence Paradigms in Healthcare-1*, pages 225–253. Springer, 2007.
- [28] Steffen Walter, Sascha Gruss, Hagen Ehleiter, Junwen Tan, Harald C. Traue, Stephen Crawcour, Philipp Werner, Ayoub Al-Hamadi, Adriano O. Andrade, and Gustavo Moreira Da Silva. The BioVid heat pain database. *IEEE International Conference on Cybernetics*, pages 128–131, 2013.
- [29] Sascha Gruss, Mattis Geiger, Philipp Werner, Oliver Wilhelm, Harald C Traue, Ayoub Al-Hamadi, and Steffen Walter. Multi-modal signals for analyzing pain responses to thermal and electrical stimuli. *JoVE (Journal of Visualized Experiments)*, (146):e59057, 2019.
- [30] Nir Ben-Israel, Mark Kliger, Galit Zuckerman, Yeshayahu Katz, and Ruth Edry. Monitoring the nociception level: A multi-parameter approach. *Journal of Clinical Monitoring and Computing*, 27(6):659–668, 2013.
- [31] Roi Treister, Mark Kliger, Galit Zuckerman, Itay Goor Aryeh, and Elon Eisenberg. Differentiating between heat pain intensities: The combined effect of multiple autonomic parameters. *Pain*, 153(9):1807–1814, 2012.
- [32] Mingzhe Jiang, Tuan Nguyen Gia, Arman Anzanpour, Amir Mohammad Rahmani, Tomi Westerlund, Sanna Salanterä, Pasi Liljeberg, and Hannu Tenhunen. IoT-based remote facial expression monitoring system with sEMG signal. In *IEEE Sensors Applications Symposium*, pages 1–6, 2016.

- [33] Mingzhe Jiang, Riitta Mieronkoski, Elise Syrjälä, Arman Anzanpour, Virpi Terävä, Amir M. Rahmani, Sanna Salanterä, Riku Aantaa, Nora Hagelberg, and Pasi Liljeberg. Acute pain intensity monitoring with the classification of multiple physiological parameters. *Journal of Clinical Monitoring and Computing*, 33(3):493–507, 2019.
- [34] Elise Syrjälä, Mingzhe Jiang, Tapani Pahikkala, Sanna Salanterä, and Pasi Liljeberg. Skin conductance response to gradual-increasing experimental pain. In *Proceedings of the 41th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2019.
- [35] Victor Kathan Sarker, Mingzhe Jiang, Tuan Nguyen Gia, Arman Anzanpour, Amir M. Rahmani, and Pasi Liljeberg. Portable multipurpose bio-signal acquisition and wireless streaming device for wearables. In *IEEE Sensors Applications Symposium*, 2017.
- [36] R. Melzack and K.L. Casey. Sensory , motivational , and central control determinants of pain - A new conceptual model. In *The Skin Senses: Proceedings of the First International Symposium of the Skin Senses*, pages 423–443, 1968.
- [37] H. Breivik, P. C. Borchgrevink, S. M. Allen, L. A. Rosseland, L. Romundstad, E. K. Breivik Hals, G. Kvarstein, and A. Stubhaug. Assessment of pain. *British Journal of Anaesthesia*, 101(1):17–24, 2008.
- [38] Raymond Sinatra. Causes and consequences of inadequate management of acute pain. *Pain Medicine (Malden, Mass.)*, 11(12):1859–71, 2010.
- [39] Amelia Swift. Understanding pain and the human body’s response to it. *Nursing Times*, pages 22–26, 2018.
- [40] Kenneth M Prkachin, Patricia E Solomon, and Joan Ross. Underestimation of pain by health-care providers: towards a model of the process of inferring pain in others. *The Canadian Journal of Nursing Research = Revue canadienne de recherche en sciences infirmières*, 39(2):88–106, 2007.
- [41] Jacqueline F M Van Dijk, Albert J M Van Wijck, Teus H. Kappen, Linda M. Peelen, Cor J. Kalkman, and Marieke J. Schuurmans. Postoperative pain assessment based on numeric ratings is not the same for patients and professionals: A cross-sectional study. *International Journal of Nursing Studies*, 49(1):65–71, 2012.
- [42] Polly E Bijur, Wendy Silver, and E John Gallagher. Reliability of the visual analog scale for measurement of acute pain. *Academic Emergency Medicine*, 8(12):1153–1157, 2001.
- [43] Erin E Krebs, Timothy S Carey, and Morris Weinberger. Accuracy of the pain numeric rating scale as a screening test in primary care. *Journal of General Internal Medicine*, 22(10):1453–1458, 2007.
- [44] Edgar E Ohnhaus and Rolf Adler. Methodological problems in the measurement of pain: a comparison between the verbal rating scale and the visual analogue scale. *Pain*, 1(4):379–384, 1975.

- [45] Céline Gélinas, Lise Fillion, Kathleen A Puntillo, Chantal Viens, and Martine Fortier. Validation of the critical-care pain observation tool in adult patients. *American Journal of Critical Care*, 15(4):420–427, 2006.
- [46] Maria Alexandra Ferreira-Valente, José Luís Pais-Ribeiro, and Mark P. Jensen. Validity of four pain intensity rating scales. *Pain*, 152(10):2399–2404, 2011.
- [47] Kenneth M Prkachin and Patricia E Solomon. The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. *Pain*, 139(2):267–274, 2008.
- [48] Edwin B. Liem, Teresa V. Joiner, Kentaro Tsueda, and Daniel I. Sessler. Increased sensitivity to thermal pain and reduced subcutaneous lidocaine efficacy in redheads. *Anesthesiology*, 102(3):509–514, 2005.
- [49] A. E. Olesen, T. Andresen, C. Staahl, and A. M. Drewes. Human experimental pain models for assessing the therapeutic efficacy of analgesic drugs. *Pharmacological Reviews*, 64(3):722–779, 2012.
- [50] Anders Jespersen, Kirstine Amris, Thomas Graven-Nielsen, Lars Arendt-Nielsen, Else Marie Bartels, Søren Torp-Pedersen, Henning Bliddal, and Bente Danneskiold-Samsøe. Assessment of pressure-pain thresholds and central sensitization of pain in lateral epicondylalgia. *Pain Medicine (United States)*, 14(2):297–304, 2013.
- [51] J. Wennervirta, M. Hynynen, A. M. Koivusalo, K. Uutela, M. Huiku, and A. Vakkuri. Surgical stress index as a measure of nociception/antinociception balance during general anesthesia. *Acta Anaesthesiologica Scandinavica*, 52(8):1038–1045, 2008.
- [52] Tanja Schlereth and Frank Birklein. The sympathetic nervous system and pain. *NeuroMolecular Medicine*, 10(3):141–147, 2008.
- [53] Eulália Silva Dos Santos Pinheiro, Fernanda Costa De Queirós, Pedro Montoya, Cleber Luz Santos, Marion Alves Do Nascimento, Clara Hikari Ito, Manuela Silva, David Barros Nunes Santos, Silvia Benevides, José Garcia Viñas Miranda, Katia Nunes Sá, and Abrahão Fontes Baptista. Electroencephalographic patterns in chronic pain: A systematic review of the literature. *PLoS ONE*, 11(2):1–26, 2016.
- [54] A. May. Neuroimaging: Visualising the brain in pain. *Neurological Sciences*, 28(SUPPL. 2):101–107, 2007.
- [55] Debbie L. Morton, Javin S. Sandhu, and Anthony K.P. Jones. Brain imaging of pain: State of the art. *Journal of Pain Research*, 9:613–624, 2016.
- [56] Stephan Geuter, Matthias Gamer, Selim Onat, and Christian Büchel. Parametric trial-by-trial prediction of pain by easily available physiological measures. *Pain*, 155(5):994–1001, 2014.

- [57] Mateusz Magda. EMG onset detection – Development and comparison of algorithms. Master's thesis, Blekinge Institute of Technology, 2015.
- [58] E. Boselli, M. Daniela-Ionescu, G. Bégou, L. Bouvet, R. Dabouz, C. Magnin, and B. Allaouchiche. Prospective observational study of the non-invasive assessment of immediate postoperative pain using the analgesia/nociception index (ANI). *British Journal of Anaesthesia*, 111(3):453–459, 2013.
- [59] M. Huiku, K. Uutela, M. van Gils, I. Korhonen, M. Kymalainen, P. Meriläinen, M. Paloheimo, M. Rantanen, P. Takala, H. Vierthio-Oja, and A. Yli-Hankala. Assessment of surgical stress during general anaesthesia. *British Journal of Anaesthesia*, 98(4):447–455, 2007.
- [60] V. Bonhomme, K. Uutela, G. Hans, I. Maquoi, J. D. Born, J. F. Brichant, M. Lamy, and P. Hans. Comparison of the Surgical Pleth IndexTMwith haemodynamic variables to assess nociception-anti-nociception balance during general anaesthesia. *British Journal of Anaesthesia*, 106(1):101–111, 2011.
- [61] I. Korhonen and A. Yli-Hankala. Photoplethysmography and nociception: Review article. *Acta Anaesthesiologica Scandinavica*, 53(8):975–985, 2009.
- [62] M. Rossi, A. Cividjian, M. C. Fevre, M. E. Oddoux, J. Carcey, C. Halle, M. Frost, M. Gardellin, J. F. Payen, and L. Quintin. A beat-by-beat, on-line, cardiovascular index, CARDEAN, to assess circulatory responses to surgery: A randomized clinical trial during spine surgery. *Journal of Clinical Monitoring and Computing*, 26(6):441–449, 2012.
- [63] Marcella Saccò, Michele Meschi, Giuseppe Regolisti, Simona Detrenis, Laura Bianchi, Marcello Bertorelli, Sarah Pioli, Andrea Magnano, Francesca Spagnoli, Pasquale Gianluca Giuri, Enrico Fiaccadori, and Alberto Caiazza. The relationship between blood pressure and pain. *Journal of Clinical Hypertension*, 15(8):600–605, 2013.
- [64] Jason J Braithwaite, Derric G Watson, Jones Robert, and Rowe Mickey. A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. Technical report, 2015.
- [65] IMotions. Galvanic skin respons (GSR): The complete pocket guide. <https://imotions.com/blog/galvanic-skin-response/>. Accessed: 2019-June-09.
- [66] Hanne Storm. Changes in skin conductance as a tool to monitor nociceptive stimulation and pain. *Current Opinion in Anaesthesiology*, 21(6):796–804, 2008.
- [67] Hassan Jafari, Imke Courtois, Omer Van Den Bergh, Johan W.S. Vlaeyen, and Ilse Van Diest. Pain and respiration: A systematic review. *Pain*, 158(6):995–1006, 2017.
- [68] Jose Alfredo L de Jesus, Rosana M Tristao, Hanne Storm, Adson F da Rocha, and Dioclecio Campos. Heart rate, oxygen saturation, and skin conductance:

A comparison study of acute pain in Brazilian newborns. In *Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, pages 1875–1879, 2011.

- [69] Céline Gélinas and Céleste Johnston. Pain assessment in the critically ill ventilated adult: Validation of the critical-care pain observation tool and physiologic indicators. *Clinical Journal of Pain*, 23(6):497–505, 2007.
- [70] Mónica Vázquez, Miren-Idoia Pardavila, María Lucia, Yara Aguado, MaÁngeles Margall, and Ma Carmen Asiain. Pain assessment in turning procedures for patients with invasive mechanical ventilation. *Nursing in Critical Care*, 16(4):178–185, 2011.
- [71] Hossein Alimohammadi, Alireza Baratloo, Ali Abdalvand, Alaleh Rouhipour, and Saeed Safari. Effects of pain relief on arterial blood O₂ saturation. *Trauma Monthly*, 19(1):25–28, 2014.
- [72] Layra Viviane Rodrigues Pinto Dantas, Thiago Silveira Pinto Dantas, Valter Joviniano Santana-Filho, Isabela Freire Azevedo-Santos, and Josimari Melo De Santana. Pain assessment during blood collection from sedated and mechanically ventilated children. *Revista Brasileira de Terapia Intensiva*, 28(1):49–54, 2016.
- [73] Lise Fillion, Kathleen a Puntillo, Chantal Viens, Martine Fortier, and Quebec City. Validation of the critical-care pain observation tool in adult patients. *American Journal of Critical Care*, 15(4):18–20, 2006.
- [74] Paul Ekman and Erika Rosenberg. *What the face reveals: basic and applied studies of spontaneous expression using the facial action codin system (FACS)*, 2nd edition. 2005.
- [75] Kenneth M Prkachin. Assessing pain by facial expression: Facial expression as nexus. *Pain Research and Management*, 14(1):53–58, 2009.
- [76] Loic Kessous, Ginevra Castellano, and George Caridakis. Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. *Journal on Multimodal User Interfaces*, 3(1):33–48, 2010.
- [77] Yale Song, David Demirdjian, and Randall Davis. Continuous body and hand gesture recognition for natural human-computer interaction. *ACM Transactions on Interactive Intelligent Systems*, 2(1):1–28, 2012.
- [78] Andre Marquand, Matthew Howard, Michael Brammer, Carlton Chu, Steven Coen, and Janaina Mourão-Miranda. Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. *NeuroImage*, 49(3):2178–2189, 2010.
- [79] Enrico Schulz, Andrew Zherdin, Laura Tiemann, Claudia Plant, and Markus Ploner. Decoding an individual’s sensitivity to pain from the multivariate analysis of EEG data. *Cerebral Cortex*, 22(5):1118–1123, 2012.

- [80] Maurizio Garbarino, Matteo Lai, Dan Bender, Rosalind W. Picard, and Simone Tognetti. Empatica E3 - A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *Proceedings of the 2014 4th International Conference on Wireless Mobile Communication and Healthcare - "Transforming Healthcare Through Innovations in Mobile and Wireless Technologies"*, pages 39–42, 2014.
- [81] James A Johnstone, Paul A Ford, Gerwyn Hughes, Tim Watson, and Andrew T Garrett. BioharnessTM multivariable monitoring device. Part I: Validity. *Journal of Sports Science and Medicine*, 11:400–408, 2012.
- [82] James A. Johnstone, Paul A. Ford, Gerwyn Hughes, Tim Watson, and Andrew T. Garrett. BioharnessTM multivariable monitoring device. Part II: Reliability. *Journal of Sports Science and Medicine*, 11:409–417, 2012.
- [83] Amir M Rahmani, Tuan Nguyen Gia, Behailu Negash, Arman Anzanpour, Iman Azimi, Mingzhe Jiang, and Pasi Liljeberg. Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: A fog computing approach. *Future Generation Computer Systems*, 78:641–658, 2018.
- [84] Geng Yang, Mingzhe Jiang, Wei Ouyang, Guangchao Ji, and Haibo Xie. IoT-based remote pain monitoring system : From device to cloud platform. *IEEE Journal of Biomedical and Health Informatics*, 22(6):1711–1719, 2018.
- [85] Iman Azimi, Janne Takalo-Mattila, Arman Anzanpour, Amir M. Rahmani, Juha-Pekka Soininen, and Pasi Liljeberg. Empowering healthcare IoT systems with hierarchical edge-based deep learning. *IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies*, pages 63–68, 2018.
- [86] Phillip A Laplante and Nancy Laplante. The Internet of Things in healthcare. *IT Professional*, 18(June):2–4, 2016.
- [87] Texas Instruments. *ADS1299-x low-noise, 4-, 6-, 8-channel, 24-bit, analog-to-digital converter for EEG and biopotential measurements (datasheet)*, January 2017.
- [88] Björn Gerdle, Stefan Karlsson, Scott Day, and Mats Djupsjöbacka. Acquisition, processing and analysis of the surface electromyogram. In *Modern Techniques in Neuroscience Research*, pages 705–755. Springer, 1999.
- [89] Krikor Tufenkjian. EEG instrumentation, montage, polarity, and localization. In *Epilepsy Board Review*, pages 15–32. Springer, 2017.
- [90] Venkatesh Acharya. Improving common-mode rejection using the right-leg drive amplifier. Technical Report July, 2011.
- [91] Teplan Michal. Fundamentals of EEG measurement. *Measurement Science Review*, 2:1–11, 2002.
- [92] Carlo J. De Luca. Electromyography. In Webster John, G, editor, *Encyclopedia of Medical Devices and Instrumentation*, pages 98–109. John Wiley Publisher, 2006.

- [93] Realtime Technologies Ltd. *Shimmer EMG user guide*, 2017. Rev 1.12.
- [94] Luca, Carlo J De. *Delsys surface electromyography: Detection and recording*, 2002.
- [95] Nitish V. Thakor. Biopotentials and electrophysiology measurement. In John G. Webster, editor, *The Measurement, Instrumentation, and Sensors Handbook*, chapter XI. CRC Press LLC, 1999.
- [96] ADInstruments. Essentials of data quality: Optimizing your range, sampling and filtering settings. <https://www.adinstruments.com/tips/data-quality>. Accessed: 2019-June-18.
- [97] Delsys Inc. How to improve EMG signal quality? <https://www.delsys.com/emgworks/signal-quality-monitor/improve/>. Accessed: 2019-June-18.
- [98] Diana Devine Lopez. Bispectral index monitoring basics. *Nursing made Incredibly Easy*, 14(2):15–16, 2016.
- [99] James A. Johnstone, Paul A. Ford, Gerwyn Hughes, Tim Watson, Andrew C. S. Mitchell, and Andrew T. Garrett. Field based reliability and validity of the Bioharness multivariable monitoring device. *Journal of Sports Science and Medicine*, 11:643–652, 2012.
- [100] Jeffrey Plourde, David Arney, and Julian M. Goldman. OpenICE: An open, interoperable platform for medical cyber-physical systems. *ACM/IEEE International Conference on Cyber-Physical Systems*, page 221, 2014.
- [101] Amit K Gupta. Respiration rate measurement based on impedance pneumography. Technical report, 2011.
- [102] Texas Instruments. *Low-Power, 8-Channel, 24-Bit Analog Front End for Biopotential Measurements (datasheet)*, August 2015.
- [103] George B Moody, Roger G Mark, Andrea Zoccola, and Sara Mantero. Derivation of respiratory signals from multi-lead ECGs. *Computers in Cardiology*, 12:113–116, 1985.
- [104] Laura Mason. *Signal processing methods for non-invasive respiration monitoring*. PhD thesis, 2002.
- [105] Maxime Cannesson, Mateo Aboy, Christoph K. Hofer, and Mohamed Rehman. Pulse pressure variation: Where are we today? *Journal of Clinical Monitoring and Computing*, 25(1):45–56, 2011.
- [106] Wolfram Boucsein and MJ Christie. *Electrodermal Activity*. Springer, 2012.
- [107] Ming-Zher Poh, Nicholas C. Swenson, and Rosalind W. Picard. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Transactions on Biomedical Engineering*, 57(5):1243–1252, 2010.
- [108] Giorgio Gabella. Autonomic nervous system. *eLS*, pages 1–6, 2012.

- [109] K. Sunil Kumar Reddy, M. U.R. Naidu, P. Usha Rani, and T. Ramesh Kumar Rao. Human experimental pain models: A review of standardized methods in drug development. *Journal of Research in Medical Sciences*, 17(6):587–595, 2012.
- [110] Mingzhe Jiang, Riitta Mieronski, Amir M. Rahmani, Nora Hagelberg, Sanna Salanterä, and Pasi Liljeberg. Ultra-short-term analysis of heart rate variability for real-time acute pain monitoring with wearable electronics. In *IEEE International Conference on Bioinformatics and Biomedicine*, pages 1025–1032, 2017.
- [111] Dale Purves, George J. Augustin, David Fitzpatrick, William C. Hall, Anthony-Samuel LaMantia, James O. McNamara, and S. Mark Williams, editors. *Neuroscience, 3rd edition*. Sinauer Associates, 2004.
- [112] Larry Kramer. Tap water and scald burns (part I). <https://www.burn-injury-resource-center.com/tap-water-and-scald-burns-part/>. Accessed: 2019-June-18.
- [113] Michal Granota, Elliot Sprecher, and David Yarnitsky. Psychophysics of phasic and tonic heat pain stimuli by quantitative sensory testing in healthy subjects. *European Journal of Pain*, 7(2):139–143, 2003.
- [114] Y. J. Lue, H. H. Wang, K. I. Cheng, C. H. Chen, and Y. M. Lu. Thermal pain tolerance and pain rating in normal subjects: Gender and age effects. *European Journal of Pain (United Kingdom)*, 22(6):1035–1042, 2018.
- [115] Texas Instruments. *EEG front-end performance demonstration kit—User's guide*, January 2016.
- [116] Alan J. Fridlund and John T. Cacioppo. Guidelines for human electromyographic research. *Psychophysiology*, 23(5):567–589, 1986.
- [117] Yu Mike Chi, Tzyy-Ping Jung, and Gert Cauwenberghs. Dry-contact and noncontact biopotential electrodes: Methodological review. *IEEE Reviews in Biomedical Engineering*, 3:106–119, 2010.
- [118] Carlo J. De Luca, L. Donald Gilmore, Mikhail Kuznetsov, and Serge H. Roy. Filtering the surface EMG signal: Movement artifact and baseline noise contamination. *Journal of Biomechanics*, 43(8):1573–1579, 2010.
- [119] Anton Van Boxtel. Facial EMG as a tool for inferring affective states. In *Proceedings of Measuring Behavior*, pages 104–108, 2010.
- [120] Mohammed AlMahamdy and H. Bryan Riley. Performance study of different denoising methods for ECG signals. *Procedia Computer Science*, 37:325–332, 2014.
- [121] Jiawei Xu, Srinjoy Mitra, Chris Van Hoof, Refet Yazicioglu, and Kofi A. A. Makinwa. Active electrodes for wearable EEG acquisition: Review and electronics design methodology. *IEEE Reviews in Biomedical Engineering*, 10:1–1, 2017.

- [122] Miroslav Zivanovic and Miriam González-Izal. Simultaneous powerline interference and baseline wander removal from ECG and EMG signals by sinusoidal modeling. *Medical Engineering and Physics*, 35(10):1431–1441, 2013.
- [123] Georgy Mihov. Subtraction procedure for removing powerline interference from ECG: Dynamic threshold linearity criterion for interference suppression. In *4th International Conference on Biomedical Engineering and Informatics*, volume 2, pages 858–861, 2011.
- [124] Shuang Qiu, Jing Feng, Rui Xu, Jiapeng Xu, Kun Wang, Feng He, Hongzhi Qi, Xin Zhao, Peng Zhou, Lixin Zhang, and Dong Ming. A stimulus artifact removal technique for SEMG signal processing during functional electrical stimulation. *IEEE Transactions on Biomedical Engineering*, 62(8):1959–68, 2015.
- [125] Steven W. Smith. *The Scientist and Engineer’s Guide to Digital Signal Processing*, volume 7. California Technical Publishing, 1999.
- [126] Jyh-shing Roger Jang. ANFIS : Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man and Cybernetics*, 23(3), 1993.
- [127] Isabelle Guyon, Steve Gunn, Masoud Nikravesh, and Lotfi Zadeh, editors. *Feature Extraction, Foundations and Applications*. Springer, 2006.
- [128] Jiapu Pan and Willis J. Tompkins. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3):230–236, 1985.
- [129] P. Hamilton. Open source ECG analysis. *Computers in Cardiology*, pages 101–104, 2002.
- [130] Peter H. Charlton, Drew A. Birrenkott, Timothy Bonnici, Marco A.F. Pimentel, Alistair E.W. Johnson, Jordi Alastruey, Lionel Tarassenko, Peter J. Watkinson, Richard Beale, and David A. Clifton. Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review. *IEEE Reviews in Biomedical Engineering*, 11:2–20, 2018.
- [131] Axel Schäfer and Karl W. Kratky. Estimation of breathing rate from respiratory sinus arrhythmia: Comparison of various methods. *Annals of Biomedical Engineering*, 36(3):476–485, 2008.
- [132] Peter H. Charlton, Timothy Bonnici, Lionel Tarassenko, Jordi Alastruey, David A. Clifton, Richard Beale, and Peter J. Watkinson. Extraction of respiratory signals from the electrocardiogram and photoplethysmogram: Technical and physiological determinants. *Physiological Measurement*, 38(5):669–690, 2017.
- [133] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, Choo Min Lim, and Jasjit S. Suri. Heart rate variability: A review. *Medical and Biological Engineering and Computing*, 44(12):1031–1051, 2006.

- [134] Luca Citi, Emery N. Brown, and Riccardo Barbieri. A real-time automated point-process method for the detection and correction of erroneous and ectopic heartbeats. *IEEE Transactions on Biomedical Engineering*, 59(10):2828–2837, 2012.
- [135] Task Force Members. Heart rate variability: standards of measurement, physiological interpretation and clinical use. *European Heart Journal*, 17:354–381, 1996.
- [136] Fred Shaffer and J. P. Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5(September):1–17, 2017.
- [137] Joachim A. Behar, Aviv A. Rosenberg, Ido Weiser-Bitoun, Ori Shemla, Alexandra Alexandrovich, Eugene Konyukhov, and Yael Yaniv. PhysioZoo: A novel open access platform for heart rate variability analysis of mammalian electrocardiographic data. *Frontiers in Physiology*, 9(October):1–14, 2018.
- [138] Eduardo Miranda Dantas, Marcela Lima Sant’Anna, Rodrigo Varejão Andreão, Christine Pereira Gonçalves, Elis Aguiar Morra, Marcelo Perim Baldo, Sérgio Lamêgo Rodrigues, and José Geraldo Mill. Spectral analysis of heart rate variability with the autoregressive method: What model order to choose? *Computers in Biology and Medicine*, 42(2):164–170, 2012.
- [139] Angkoon Phinyomark, Pornchai Phukpattaranont, and Chusak Limsakul. Feature reduction and selection for EMG signal classification. *Expert Systems With Applications*, 39(8):7420–7431, 2012.
- [140] Nicolas Berryman, Louis Bherer, Michel Audiffren, and Olivier Dupuy. Reliability of heart rate measures used to assess post-exercise parasympathetic reactivation. *Clinical Physiology and Functional Imaging*, pages 1–9, 2012.
- [141] Zainal Arief, Indra Adji Sulistijono, and Roby Awal Ardiansyah. Comparison of five time series EMG features extractions using Myo armband. In *International Electronics Symposium*, pages 11–14. IEEE, 2015.
- [142] Mahyar Hamed, Sh-hussain Salleh, Mehdi Astaraki, and Alias Mohd Noor. EMG-based facial gesture recognition through versatile elliptic basis function neural network. *BioMedical Engineering Online*, pages 1–22, 2013.
- [143] Mahyar Hamed, Alias Mohd, Noor Universiti, Tian Swee, Tan Universiti, Automotive Turbocharger Turbine, Ceramic Coating, Volute Casing View, and Mahyar Hamed. Comparison of different time-domain feature extraction methods on facial gestures’ EMGs. In *Progress in Electromagnetics Research Symposium Proceedings.*, pages 1897–1900, 2012.
- [144] Angkoon Phinyomark, Rami N. Khushaba, and Erik Scheme. Feature extraction and selection for myoelectric control based on wearable EMG sensors. *Sensors (Switzerland)*, 18(5):1–17, 2018.
- [145] Kang Soo Kim, Heung Ho Choi, Chang Soo Moon, and Chi Woong Mun. Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions. *Current Applied Physics*, 11(3):740–745, 2011.

- [146] Max Ortiz-Catalan, Rickard Brânemark, and Bo Håkansson. BioPatRec: A modular research platform for the control of articial limbs based on pattern recognition algorithms. *Source Code for Biology and Medicine*, pages 1–18, 2012.
- [147] Mohsen Nabian, Athena Nouhi, Yu Yin, Sarah Ostadabbas, and A Dataset. A biosignal-specific processing tool for machine learning and pattern recognition. pages 76–80, 2017.
- [148] Alberto Greco, , Gaetano Valenza, Antonio Lanata, Enzo Pasquale Scilingo, and Luca Citi. cvxEDA : A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Biomedical Engineering*, 63(4):797–804, 2016.
- [149] Mathias Benedek and Christian Kaernbach. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1-5):80–91, 2010.
- [150] Swayambhoo Jain, Urvashi Oswal, Kevin Shuai Xu, Brian Eriksson, and Jarvis Haupt. A compressed sensing based decomposition of electrodermal activity signals. *IEEE Transactions on Biomedical Engineering*, 64(9):2142–2151, 2017.
- [151] M. Loretto Munoz, Arie Van Roon, Harriëtte Riese, Chris Thio, Emma Oostbroek, Iris Westrik, Eco J.C. De Geus, Ron Gansevoort, Joop Lefrandt, Ilja M. Nolte, and Harold Snieder. Validity of (ultra-)short recordings for heart rate variability measurements. *PLoS ONE*, 10(9):1–15, 2015.
- [152] Leandro Pecchia, Rossana Castaldo, Luis Montesinos, and Paolo Melillo. Are ultra-short heart rate variability features good surrogates of short-term ones? State-of-the-art review and recommendations. *Healthcare Technology Letters*, 5(3):94–100, 2018.
- [153] Gari D Clifford. *Signal Processing Methods For Heart Rate Variability Analysis*. PhD thesis, 2002.
- [154] Jennifer M. Yentes, Nathaniel Hunt, Kendra K. Schmid, Jeffrey P. Kaipust, Denise McGrath, and Nicholas Stergiou. The appropriate use of approximate entropy and sample entropy with short data sets. *Annals of Biomedical Engineering*, 41(2):349–365, 2013.
- [155] Lauren H Smith, Levi J. Hargrove, Blair A. Lock, and Todd A. Kuiken. Determining the optimal window length for pattern recognition-based myoelectric control: Balancing the competing effects of classification error and controller delay. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 19(2):186–192, 2014.
- [156] Björn Hart, Marijn E Struiksma, Anton Van Boxtel, and Jos J A Van Berkum. Emotion in stories : Facial EMG evidence for both mental simulation and moral evaluation. *Frontiers in Psychology*, 9(April):1–13, 2018.

- [157] Daniel Lopez-Martinez and R Picard. Continuous pain intensity estimation from autonomic signals with recurrent neural networks. In *Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5624–5627, 2018.
- [158] Marie Udnesseter Lie, Dagfinn Matre, Per Hansson, Audun Stubhaug, John-Anker Zwart, and Kristian Bernhard Nilsen. A tonic heat test stimulus yields a larger and more reliable conditioned pain modulation effect compared to a phasic heat test stimulus. *Pain Reports*, 2(6):e626, 2017.
- [159] Mimma Nardelli, Alberto Greco, Juan Bolea, Gaetano Valenza, Enzo Pasquale Scilingo, and Raquel Bailon. Reliability of lagged poincaré plot parameters in ultrashort heart rate variability series: Application on affective sounds. *IEEE Journal of Biomedical and Health Informatics*, 22(3):741–749, 2018.
- [160] Tran Thong, Kehai Li, James McNames, Mateo Aboy, and Brahm Goldstein. Accuracy of ultra-short heart rate variability measures. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2424–2427, 2003.
- [161] Triwiyanto, Oyas Wahyunggoro, Hanung Adi Nugroho, and Herianto. Effect of window length on performance of the elbow-joint angle prediction based on electromyography. *Journal of Physics: Conference Series*, 853(1), 2017.
- [162] Adrian Mark Burden, Sandra Elizabeth Lewis, and Emma Willcox. The effect of manipulating root mean square window length and overlap on reliability, inter-individual variability, statistical significance and clinical relevance of electromyograms. *Manual Therapy*, 19(6):595–601, 2014.
- [163] Wolfram Boucsein, Don C. Fowles, Sverre Grimnes, Gershon Ben-shakhar, Walton T. Roth, Michael E. Dawson, and Diane L. Filion. Publication recommendations for electrodermal measurements. *Psychophysiology*, 49:1017–1034, 2012.
- [164] Mark Halaki and Karen Ginn. Normalization of EMG signals: To normalize or not to normalize and what to normalize to? *Computational Intelligence in Electromyography Analysis - A Perspective on Current Applications and Future Challenges*, page 448, 2012.
- [165] Mariateresa Sestito, Maria Alessandra Umiltà, Giancarlo De Paola, Renata Fortunati, Andrea Raballo, Emanuela Leuci, Simone Maffei, Matteo Tonna, Mario Amore, Carlo Maggini, and Vittorio Gallese. Facial reactions in response to dynamic emotional stimuli in different modalities in patients suffering from schizophrenia: a behavioral and EMG study. *Frontiers in Human Neuroscience*, 7(July):1–12, 2013.
- [166] Nikolaus Peter Schumann, Kevin Bongers, Orlando Guntinas-Lichius, and Hans Christoph Scholle. Facial muscle activation patterns in healthy male humans: A multi-channel surface EMG study. *Journal of Neuroscience Methods*, 187(1):120–128, 2010.

- [167] Marcel Riehle, Jürgen Kempkensteffen, and Tania M. Lincoln. Quantifying facial expression synchrony in face-To-face dyadic interactions: Temporal dynamics of simultaneously recorded facial EMG signals. *Journal of Nonverbal Behavior*, 41(2):85–102, 2017.
- [168] Alfred P Hallstrom, Phyllis K Stein, Raphael Schneider, Morrison Hodges, Georg Schmidt, and Kurt Ulm. Structural relationships between measures based on heart beat intervals : Potential for improved risk assessment. *IEEE Transactions on Biomedical Engineering*, 51(8):1414–1420, 2004.
- [169] Jerzy Sacha. Why should one normalize heart rate variability with respect to average heart rate. *Frontiers in Physiology*, 4(October):1–2, 2013.
- [170] Robet L. Burr. Interpretation of normalized spectral heart rate variability indices in sleep research: A critical review. *Sleep*, 30:913–919, 2007.
- [171] D. Ramaekers, H. Ector, A. E. Aubert, A. Rubens, and F. Van De Werf. Heart rate variability and heart rate in healthy volunteers: Is the female autonomic nervous system cardioprotective? *European Heart Journal*, 19(9):1334–1341, 1998.
- [172] Mario Merri, David C. Farden, Jack G. Mottley, and Edward L. Titlebaum. Sampling frequency of the electrocardiogram for spectral analysis of the heart rate variability. *IEEE Transactions on Biomedical Engineering*, 37(1):99–106, 1990.
- [173] C Satija, Barathram Ramkumar, and Sabarimalai M. Manikandan. Real-time signal quality-aware ECG telemetry system for IoT-based health care monitoring. *IEEE Internet of Things Journal*, 4(3):815–823, 2017.
- [174] Raquel Bailón, Leif Sörnmo, and Pablo Laguna. A robust method for ECG-based estimation of the respiratory frequency during stress testing. *IEEE Transactions on Biomedical Engineering*, 53(7):1273–1285, 2006.
- [175] Jesús Lázaro, Alejandro Alcaine, Daniel Romero, Eduardo Gil, Pablo Laguna, Esther Pueyo, and Raquel Bailón. Electrocardiogram derived respiratory rate from QRS slopes and R-wave angle. *Annals of Biomedical Engineering*, 42(10):2072–2083, 2014.
- [176] Spyridon Kontaxis, Jesus Lazaro, Valentina D.A. Corino, Frida Sandberg, Raquel Bailon, Pablo Laguna, and Leif Sornmo. ECG-derived Respiratory Rate in Atrial Fibrillation. *IEEE Transactions on Biomedical Engineering*, X(X):1–1, 2019.
- [177] Thomas Ledowski, B. Ang, T. Schmarbeck, and J. Rhodes. Monitoring of sympathetic tone to assess postoperative pain: Skin conductance vs surgical stress index. *Anaesthesia*, 64(7):727–731, 2009.
- [178] Thomas Ledowski, James Bromilow, Jianrong Wu, Michael James Paech, Hanne Storm, and Stephan Alexander Schug. The assessment of postoperative pain by monitoring skin conductance : results of a prospective study. *Anaesthesia*, (62):989–993, 2007.

- [179] Ling-hua Chang, Tso-chiang Ma, Shio-w-luan Tsay, and Gwo-ping Jong. Relationships between pain intensity and heart rate variability in patients after abdominal surgery: a pilot study. *Chinese Medical Journal*, 125(9804):1964–1969, 2012.
- [180] Jan J Meeuse, Marco S P Löwik, Sabine A M Löwik, Eline Aarden, Arie M Van Roon, Reinold O B Gans, Marten Van Wijhe, Joop D Lefrandt, and Anna K L Reyners. Heart rate variability parameters do not correlate with pain intensity in healthy volunteers. *Pain Medicine*, (14):1192–1201, 2013.
- [181] J Koenig, M N Jarczok, R J Ellis, T K Hillecke, and J F Thayer. Heart rate variability and experimentally induced pain in healthy adults : A systematic review. *European Journal of Pain*, 18:301–314, 2014.
- [182] M. Jeanne, R. Logier, J. De Jonckheere, and B. Tavernier. Validation of a graphic measurement of heart rate variability to assess analgesia/nociception balance during general anesthesia. In *The 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1840–1843, 2009.
- [183] Mathieu Jeanne, Régis Logier, Julien De Jonckheere, and Benoît Tavernier. Autonomic neuroscience : Basic and clinical heart rate variability during total intravenous anesthesia : Effects of nociception and analgesia. *Autonomic Neuroscience: Basic and Clinical*, 147(1-2):91–96, 2009.
- [184] Bradley M Appelhans and Linda J Luecken. Heart rate variability and pain : Associations of two interrelated homeostatic processes. *Biological Psychology*, 77:174–182, 2008.
- [185] Lincoln M Tracy. Heart rate variability and sensitivity to experimentally induced pain: A replication. *Pain Practice*, 18(5):687–689, 2018.
- [186] Lincoln M Tracy, Julian Koenig, Nellie Georgiou-karistianis, Stephen J Gibson, and Melita J Giummarrra. Heart rate variability is associated with thermal heat pain threshold in males , but not females. *International Journal of Psychophysiology*, 131(March):37–43, 2018.
- [187] Jerzy Sacha, Szymon Barabach, Gabriela Statkiewicz-Barabach, Krzysztof Sacha, Alexander Müller, Jaroslaw Piskorski, Petra Barthel, and Georg Schmidt. How to strengthen or weaken the HRV dependence on heart rate - Description of the method and its perspectives. *International Journal of Cardiology*, 168(2):1660–1663, 2013.
- [188] Jerzy Sacha. Interplay between heart rate and its variability: A prognostic game. *Frontiers in Physiology*, 5 AUG(September):1–4, 2014.
- [189] By Luca T. Mainardi. On the quantification of heart rate variability spectral parameters using time-frequency and time-varying methods. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1887):255–275, 2009.

- [190] Mark A Lumley, Jay L Cohen, George S Borscz, Annmarie Cano, M Alison, Laura S Porter, Howard Schubiner, and Francis J Keefe. Pain and emotion: A biopsychosocial review of recent research. *Journal of Clinical Psychology*, 67(9):942–968, 2011.
- [191] Haibo He and Edwardo A Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, 2009.
- [192] Mikel Galar, Alberto Fern, Edurne Barrenechea, and Humberto Bustince. A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, 42(4):1–22, 2011.
- [193] Sylvain Arlot and Alain Celisse. A survey of cross-validation procedures for model selection. *Statstics Surveys*, 4:40–79, 2010.
- [194] Michael Esterman, Benjamin J Tamber-rosenau, Yu-chin Chiu, and Steven Yantis. NeuroImage avoiding non-independence in fMRI data analysis : Leave one subject out. *NeuroImage*, 50(2):572–576, 2010.
- [195] Gavin C Cawley and Nicola L C Talbot. On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11:2079–2107, 2010.
- [196] Mathworks. Statistics and machine learning toolboxTMuser’s guide R2018b. 2018.
- [197] Chris Seiffert, Taghi M Khoshgoftaar, Jason Van Hulse, and Amri Napolitano. RUSBoost : A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: a System and Humans*, 40(1):185–197, 2010.
- [198] Yoav Freund. A more robust boosting algorithm. *arXiv preprint arXiv:0905.2138*, 2009.
- [199] Sascha Gruss, Roi Treister, Philipp Werner, Harald C. Traue, Stephen Crawcour, Adriano Andrade, and Steffen Walter. Pain intensity recognition rates via biopotential feature patterns with support vector machines. *PLoS ONE*, (October):1–15, 2015.
- [200] Jonghwa Kim and Elisabeth Andre. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12):2067–2083, 2008.
- [201] Alberto De Santos Sierra, Carmen Sánchez Ávila, Javier Guerra Casanova, and Gonzalo Bailador. A stress-detection system based on physiological signals and fuzzy logic. *IEEE Transactions on Industrial Electronics*, 58(10):4857–4865, 2011.
- [202] Chantal Villemure and M Catherine Bushnell. Cognitive modulation of pain : How do attention and emotion influence pain processing ? *Pain*, 95:195–199, 2002.

- [203] Ronald Melzack. Pain and stress: A new perspective. *Psychosocial Factors in Pain: Critical Perspectives*, pages 89–106, 1999.
- [204] Nuno Pombo, Pedro Araújo, and Joaquim Viana. Knowledge discovery in clinical decision support systems for pain management : A systematic review. *Artificial Intelligence In Medicine*, 60(1):1–11, 2014.
- [205] Karan Sikka, Alex A Ahmed, Damaris Diaz, Matthew S Goodwin, and Kenneth D Craig. Automated assessment of children’ s postoperative pain using computer vision. *Pediatrics*, 136(1), 2015.
- [206] Gaurav Misra, Wei-en Wang, Derek B Archer, Arnab Roy, Stephen A Coombes, G Misra, Wang We, Archer Db, A Roy, and Coombes S A Auto. Automated classification of pain perception using high-density electroencephalography data. *Journal of Neurophysiology*, 117(2):786–795, 2017.
- [207] Behnoor Gholami, Wassim M Haddad, and Allen R Tannenbaum. Relevance vector machine learning for neonate pain intensity assessment using digital imaging. *IEEE Transactions on Biomedical Engineering*, 57(6):1457–1466, 2010.
- [208] Sebastian Kaltwang, Ognjen Rudovic, and Maja Pantic. Continuous pain intensity estimation from facial expressions. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 7432 LNCS, pages 368–377, 2012.
- [209] Jing Zhou, Xiaopeng Hong, Fei Su, and Guoying Zhao. Recurrent convolutional neural network regression for continuous pain intensity estimation in video. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1535–1543, 2016.
- [210] Eun Hye Jang, Byoung Jun Park, Mi Sook Park, Sang Hyeob Kim, and Jin Hun Sohn. Analysis of physiological signals for recognition of boredom, pain, and surprise emotions. *Journal of Physiological Anthropology*, 34(1):1–12, 2015.
- [211] Philipp Werner, Ayoub Al-hamadi, Kerstin Limbrecht-ecklundt, Steffen Walter, Sascha Gruss, and Harald C Traue. Automatic pain assessment with facial activity descriptors. *IEEE Transactions on Affective Computing*, 8(3):286–299, 2017.
- [212] Markus Kächele, Patrick Thiam, Mohammadreza Amirian, Philipp Werner, Steffen Walter, Friedhelm Schwenker, and Günther Palm. Multimodal data fusion for person-independent, continuous estimation for pain intensity. In *Proceedings of Engineering Appcliations of Neural Networks*, volume 517, 2015.
- [213] Gwen Littlewort, Jacob Whitehill, Tingfan Wu, Ian Fasel, Mark Frank, Javier Movellan, and Marian Bartlett. The computer expression recognition toolbox (CERT). *Face and Gesture 2011*, pages 298–305.

- [214] Xinzong Chen, Carsten Thee, Matthias Gruenewald, Jan Wnent, Christoph Illies, Jan Hoecker, Robert Hanss, Markus Steinfath, and Berthold Bein. Comparison of surgical stress index-guided analgesia with standard clinical practice during routine general. *Anesthesiology*, 112(May):1175–1183, 2010.
- [215] Jean-Francois Payen, Gerald Chanques, Jean Mantz, Christiane Hercule, Igor Auriant, Jean-Luc Leguillou, Michele Binhas, Celine Centy, Carole Rolland, and Jean-Luc Bosson. Current practices in sedation and analgesia for mechanically ventilated critically ill patients - A prospective multicenter patient-based study. *Anesthesiology*, 106(4):687–695, 2007.
- [216] Patrick Lucey, Jeffrey F Cohn, Kenneth M Prkachin, Patricia E Solomon, Sien Chew, and Iain Matthews. Painful monitoring : Automatic pain monitoring using the UNBC-McMaster shoulder pain expression archive database. *Image and Vision Computing*, 30(3):197–205, 2012.
- [217] Daniel Lopez Martinez, Ognjen Rudovic, and Rosalind Picard. Personalized automatic estimation of self-reported pain intensity from facial expressions. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, volume 2017-July, pages 2318–2327, 2017.
- [218] T. Ledowski, J. Burke, and J. Hruby. Surgical pleth index: Prediction of postoperative pain and influence of arousal. *British Journal of Anaesthesia*, 117(3):371–374, 2016.
- [219] Ahmed Bilal Ashraf, Simon Lucey, Jeffrey F Cohn, Tsuhan Chen, Zara Ambadar, Kenneth M Prkachin, and Patricia E Solomon. The painful face – Pain expression recognition using active appearance models. *Image and Vision Computing*, 27(12):1788–1796, 2009.
- [220] S.J. Caetano, G. Sonpavde, and G.R. Pond. C-statistic : A brief explanation of its construction , interpretation and limitations. *European Journal of Cancer*, 90:130–132, 2018.
- [221] P F Watson and A Petrie. Method agreement analysis : A review of correct methodology. *Theriogenology*, 73(9):1167–1179, 2010.
- [222] Lawrence I-kuei Lin. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45(1):255–268, 1989.
- [223] Evan Campbell, Angkoon Phinyomark, and Erik Scheme. Feature extraction and selection for pain recognition using peripheral physiological signals. *Frontiers in Neuroscience*, 13(May):1–17, 2019.
- [224] Philipp Werner, Frerk Saxen, Ayoub Al-Hamadi, and Hui Yu. Generalizing to unseen head poses in facial expression recognition and action unit intensity estimation. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 1 2019.
- [225] Philipp Werner, Ayoub Al-Hamadi, Kerstin Limbrecht-Ecklundt, Steffen Walther, and Harald C Traue. Head movements and postures as pain behavior. *PloS ONE*, 13(2):e0192767, 2018.

- [226] E J Bartley and R B Fillingim. Sex differences in pain : A brief review of clinical and experimental findings. *British Journal of Anaesthesia*, 111(1):52–58, 2013.
- [227] Samaneh Aminikhanghahi and Diane J Cook. A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51(2):339–367, 2017.
- [228] Sahar Torkamani and Volker Lohweg. Survey on time series motif discovery. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(2):e1199, 2017.

Turku Centre for Computer Science

TUCS Dissertations

1. **Marjo Lipponen**, On Primitive Solutions of the Post Correspondence Problem
2. **Timo Käkälä**, Dual Information Systems in Hyperknowledge Organizations
3. **Ville Leppänen**, Studies on the Realization of PRAM
4. **Cunsheng Ding**, Cryptographic Counter Generators
5. **Sami Viitanen**, Some New Global Optimization Algorithms
6. **Tapio Salakoski**, Representative Classification of Protein Structures
7. **Thomas Långbacka**, An Interactive Environment Supporting the Development of Formally Correct Programs
8. **Thomas Finne**, A Decision Support System for Improving Information Security
9. **Valeria Mihalache**, Cooperation, Communication, Control. Investigations on Grammar Systems.
10. **Marina Waldén**, Formal Reasoning About Distributed Algorithms
11. **Tero Laihonen**, Estimates on the Covering Radius When the Dual Distance is Known
12. **Lucian Ilie**, Decision Problems on Orders of Words
13. **Jukkapekka Hekanaho**, An Evolutionary Approach to Concept Learning
14. **Jouni Järvinen**, Knowledge Representation and Rough Sets
15. **Tomi Pasanen**, In-Place Algorithms for Sorting Problems
16. **Mika Johnsson**, Operational and Tactical Level Optimization in Printed Circuit Board Assembly
17. **Mats Aspnäs**, Multiprocessor Architecture and Programming: The Hathi-2 System
18. **Anna Mikhajlova**, Ensuring Correctness of Object and Component Systems
19. **Vesa Torvinen**, Construction and Evaluation of the Labour Game Method
20. **Jorma Boberg**, Cluster Analysis. A Mathematical Approach with Applications to Protein Structures
21. **Leonid Mikhajlov**, Software Reuse Mechanisms and Techniques: Safety Versus Flexibility
22. **Timo Kaukoranta**, Iterative and Hierarchical Methods for Codebook Generation in Vector Quantization
23. **Gábor Magyar**, On Solution Approaches for Some Industrially Motivated Combinatorial Optimization Problems
24. **Linas Laibinis**, Mechanised Formal Reasoning About Modular Programs
25. **Shuhua Liu**, Improving Executive Support in Strategic Scanning with Software Agent Systems
26. **Jaakko Järvi**, New Techniques in Generic Programming – C++ is more Intentional than Intended
27. **Jan-Christian Lehtinen**, Reproducing Kernel Splines in the Analysis of Medical Data
28. **Martin Büchi**, Safe Language Mechanisms for Modularization and Concurrency
29. **Elena Troubitsyna**, Stepwise Development of Dependable Systems
30. **Janne Näppi**, Computer-Assisted Diagnosis of Breast Calcifications
31. **Jianming Liang**, Dynamic Chest Images Analysis
32. **Tiberiu Seceleanu**, Systematic Design of Synchronous Digital Circuits
33. **Tero Aittokallio**, Characterization and Modelling of the Cardiorespiratory System in Sleep-Disordered Breathing
34. **Ivan Porres**, Modeling and Analyzing Software Behavior in UML
35. **Mauno Rönkkö**, Stepwise Development of Hybrid Systems
36. **Jouni Smed**, Production Planning in Printed Circuit Board Assembly
37. **Vesa Halava**, The Post Correspondence Problem for Market Morphisms
38. **Ion Petre**, Commutation Problems on Sets of Words and Formal Power Series
39. **Vladimir Kvassov**, Information Technology and the Productivity of Managerial Work
40. **Frank Tétard**, Managers, Fragmentation of Working Time, and Information Systems

- 41.** **Jan Manuch**, Defect Theorems and Infinite Words
- 42.** **Kalle Ranto**, Z₄-Goethals Codes, Decoding and Designs
- 43.** **Arto Lepistö**, On Relations Between Local and Global Periodicity
- 44.** **Mika Hirvensalo**, Studies on Boolean Functions Related to Quantum Computing
- 45.** **Pentti Virtanen**, Measuring and Improving Component-Based Software Development
- 46.** **Adekunle Okunoye**, Knowledge Management and Global Diversity – A Framework to Support Organisations in Developing Countries
- 47.** **Antonina Kloptchenko**, Text Mining Based on the Prototype Matching Method
- 48.** **Juha Kivijärvi**, Optimization Methods for Clustering
- 49.** **Rimvydas Rukšėnas**, Formal Development of Concurrent Components
- 50.** **Dirk Nowotka**, Periodicity and Unbordered Factors of Words
- 51.** **Attila Gyenesei**, Discovering Frequent Fuzzy Patterns in Relations of Quantitative Attributes
- 52.** **Petteri Kaitovaara**, Packaging of IT Services – Conceptual and Empirical Studies
- 53.** **Petri Rosendahl**, Niho Type Cross-Correlation Functions and Related Equations
- 54.** **Péter Majlender**, A Normative Approach to Possibility Theory and Soft Decision Support
- 55.** **Seppo Virtanen**, A Framework for Rapid Design and Evaluation of Protocol Processors
- 56.** **Tomas Eklund**, The Self-Organizing Map in Financial Benchmarking
- 57.** **Mikael Collan**, Giga-Investments: Modelling the Valuation of Very Large Industrial Real Investments
- 58.** **Dag Björklund**, A Kernel Language for Unified Code Synthesis
- 59.** **Shengnan Han**, Understanding User Adoption of Mobile Technology: Focusing on Physicians in Finland
- 60.** **Irina Georgescu**, Rational Choice and Revealed Preference: A Fuzzy Approach
- 61.** **Ping Yan**, Limit Cycles for Generalized Liénard-Type and Lotka-Volterra Systems
- 62.** **Joonas Lehtinen**, Coding of Wavelet-Transformed Images
- 63.** **Tommi Meskanen**, On the NTRU Cryptosystem
- 64.** **Saeed Salehi**, Varieties of Tree Languages
- 65.** **Jukka Arvo**, Efficient Algorithms for Hardware-Accelerated Shadow Computation
- 66.** **Mika Hirvikorpi**, On the Tactical Level Production Planning in Flexible Manufacturing Systems
- 67.** **Adrian Costea**, Computational Intelligence Methods for Quantitative Data Mining
- 68.** **Cristina Seceleanu**, A Methodology for Constructing Correct Reactive Systems
- 69.** **Luigia Petre**, Modeling with Action Systems
- 70.** **Lu Yan**, Systematic Design of Ubiquitous Systems
- 71.** **Mehran Gomari**, On the Generalization Ability of Bayesian Neural Networks
- 72.** **Ville Harkke**, Knowledge Freedom for Medical Professionals – An Evaluation Study of a Mobile Information System for Physicians in Finland
- 73.** **Marius Cosmin Codrea**, Pattern Analysis of Chlorophyll Fluorescence Signals
- 74.** **Aiying Rong**, Cogeneration Planning Under the Deregulated Power Market and Emissions Trading Scheme
- 75.** **Chihab BenMoussa**, Supporting the Sales Force through Mobile Information and Communication Technologies: Focusing on the Pharmaceutical Sales Force
- 76.** **Jussi Salmi**, Improving Data Analysis in Proteomics
- 77.** **Orieta Celiku**, Mechanized Reasoning for Dually-Nondeterministic and Probabilistic Programs
- 78.** **Kaj-Mikael Björk**, Supply Chain Efficiency with Some Forest Industry Improvements
- 79.** **Viorel Preoteasa**, Program Variables – The Core of Mechanical Reasoning about Imperative Programs
- 80.** **Jonne Poikonen**, Absolute Value Extraction and Order Statistic Filtering for a Mixed-Mode Array Image Processor
- 81.** **Luka Milovanov**, Agile Software Development in an Academic Environment
- 82.** **Francisco Augusto Alcaraz Garcia**, Real Options, Default Risk and Soft Applications
- 83.** **Kai K. Kimppa**, Problems with the Justification of Intellectual Property Rights in Relation to Software and Other Digitally Distributable Media
- 84.** **Dragoş Trușcan**, Model Driven Development of Programmable Architectures
- 85.** **Eugen Czeizler**, The Inverse Neighborhood Problem and Applications of Welch Sets in Automata Theory

- 86.** **Sanna Ranto**, Identifying and Locating-Dominating Codes in Binary Hamming Spaces
- 87.** **Tuomas Hakkarainen**, On the Computation of the Class Numbers of Real Abelian Fields
- 88.** **Elena Czeizler**, Intricacies of Word Equations
- 89.** **Marcus Alanen**, A Metamodeling Framework for Software Engineering
- 90.** **Filip Ginter**, Towards Information Extraction in the Biomedical Domain: Methods and Resources
- 91.** **Jarkko Paavola**, Signature Ensembles and Receiver Structures for Oversaturated Synchronous DS-CDMA Systems
- 92.** **Arho Virkki**, The Human Respiratory System: Modelling, Analysis and Control
- 93.** **Olli Luoma**, Efficient Methods for Storing and Querying XML Data with Relational Databases
- 94.** **Dubravka Ilić**, Formal Reasoning about Dependability in Model-Driven Development
- 95.** **Kim Solin**, Abstract Algebra of Program Refinement
- 96.** **Tomi Westerlund**, Time Aware Modelling and Analysis of Systems-on-Chip
- 97.** **Kalle Saari**, On the Frequency and Periodicity of Infinite Words
- 98.** **Tomi Kärki**, Similarity Relations on Words: Relational Codes and Periods
- 99.** **Markus M. Mäkelä**, Essays on Software Product Development: A Strategic Management Viewpoint
- 100.** **Roope Vehkalahti**, Class Field Theoretic Methods in the Design of Lattice Signal Constellations
- 101.** **Anne-Maria Ernvall-Hytönen**, On Short Exponential Sums Involving Fourier Coefficients of Holomorphic Cusp Forms
- 102.** **Chang Li**, Parallelism and Complexity in Gene Assembly
- 103.** **Tapio Pahikkala**, New Kernel Functions and Learning Methods for Text and Data Mining
- 104.** **Denis Shestakov**, Search Interfaces on the Web: Querying and Characterizing
- 105.** **Sampo Pyysalo**, A Dependency Parsing Approach to Biomedical Text Mining
- 106.** **Anna Sell**, Mobile Digital Calendars in Knowledge Work
- 107.** **Dorina Marghescu**, Evaluating Multidimensional Visualization Techniques in Data Mining Tasks
- 108.** **Tero Säntti**, A Co-Processor Approach for Efficient Java Execution in Embedded Systems
- 109.** **Kari Salonen**, Setup Optimization in High-Mix Surface Mount PCB Assembly
- 110.** **Pontus Boström**, Formal Design and Verification of Systems Using Domain-Specific Languages
- 111.** **Camilla J. Hollanti**, Order-Theoretic Methods for Space-Time Coding: Symmetric and Asymmetric Designs
- 112.** **Heidi Himmanen**, On Transmission System Design for Wireless Broadcasting
- 113.** **Sébastien Lafond**, Simulation of Embedded Systems for Energy Consumption Estimation
- 114.** **Evgeni Tsivtsivadze**, Learning Preferences with Kernel-Based Methods
- 115.** **Petri Salmela**, On Commutation and Conjugacy of Rational Languages and the Fixed Point Method
- 116.** **Siamak Taati**, Conservation Laws in Cellular Automata
- 117.** **Vladimir Rogojin**, Gene Assembly in Stichotrichous Ciliates: Elementary Operations, Parallelism and Computation
- 118.** **Alexey Dudkov**, Chip and Signature Interleaving in DS CDMA Systems
- 119.** **Janne Savela**, Role of Selected Spectral Attributes in the Perception of Synthetic Vowels
- 120.** **Kristian Nybom**, Low-Density Parity-Check Codes for Wireless Datacast Networks
- 121.** **Johanna Tuominen**, Formal Power Analysis of Systems-on-Chip
- 122.** **Teijo Lehtonen**, On Fault Tolerance Methods for Networks-on-Chip
- 123.** **Eeva Suvitie**, On Inner Products Involving Holomorphic Cusp Forms and Maass Forms
- 124.** **Linda Mannila**, Teaching Mathematics and Programming – New Approaches with Empirical Evaluation
- 125.** **Hanna Suominen**, Machine Learning and Clinical Text: Supporting Health Information Flow
- 126.** **Tuomo Saarni**, Segmental Durations of Speech
- 127.** **Johannes Eriksson**, Tool-Supported Invariant-Based Programming

- 128.** **Tero Jokela**, Design and Analysis of Forward Error Control Coding and Signaling for Guaranteeing QoS in Wireless Broadcast Systems
- 129.** **Ville Lukkarila**, On Undecidable Dynamical Properties of Reversible One-Dimensional Cellular Automata
- 130.** **Qaisar Ahmad Malik**, Combining Model-Based Testing and Stepwise Formal Development
- 131.** **Mikko-Jussi Laakso**, Promoting Programming Learning: Engagement, Automatic Assessment with Immediate Feedback in Visualizations
- 132.** **Riikka Vuokko**, A Practice Perspective on Organizational Implementation of Information Technology
- 133.** **Jeanette Heidenberg**, Towards Increased Productivity and Quality in Software Development Using Agile, Lean and Collaborative Approaches
- 134.** **Yong Liu**, Solving the Puzzle of Mobile Learning Adoption
- 135.** **Stina Ojala**, Towards an Integrative Information Society: Studies on Individuality in Speech and Sign
- 136.** **Matteo Brunelli**, Some Advances in Mathematical Models for Preference Relations
- 137.** **Ville Junnila**, On Identifying and Locating-Dominating Codes
- 138.** **Andrzej Mizera**, Methods for Construction and Analysis of Computational Models in Systems Biology. Applications to the Modelling of the Heat Shock Response and the Self-Assembly of Intermediate Filaments.
- 139.** **Csaba Ráduly-Báka**, Algorithmic Solutions for Combinatorial Problems in Resource Management of Manufacturing Environments
- 140.** **Jari Kyngäs**, Solving Challenging Real-World Scheduling Problems
- 141.** **Arho Suominen**, Notes on Emerging Technologies
- 142.** **József Mezei**, A Quantitative View on Fuzzy Numbers
- 143.** **Marta Olszewska**, On the Impact of Rigorous Approaches on the Quality of Development
- 144.** **Antti Airola**, Kernel-Based Ranking: Methods for Learning and Performance Estimation
- 145.** **Aleksi Saarela**, Word Equations and Related Topics: Independence, Decidability and Characterizations
- 146.** **Lasse Bergroth**, Kahden merkkijonon pisimmän yhteisen alijonon ongelma ja sen ratkaiseminen
- 147.** **Thomas Canhao Xu**, Hardware/Software Co-Design for Multicore Architectures
- 148.** **Tuomas Mäkilä**, Software Development Process Modeling – Developers Perspective to Contemporary Modeling Techniques
- 149.** **Shahrokh Nikou**, Opening the Black-Box of IT Artifacts: Looking into Mobile Service Characteristics and Individual Perception
- 150.** **Alessandro Buoni**, Fraud Detection in the Banking Sector: A Multi-Agent Approach
- 151.** **Mats Neovius**, Trustworthy Context Dependency in Ubiquitous Systems
- 152.** **Fredrik Degerlund**, Scheduling of Guarded Command Based Models
- 153.** **Amir-Mohammad Rahmani-Sane**, Exploration and Design of Power-Efficient Networked Many-Core Systems
- 154.** **Ville Rantala**, On Dynamic Monitoring Methods for Networks-on-Chip
- 155.** **Mikko Peltola**, On Identifying and Locating-Dominating Codes in the Infinite King Grid
- 156.** **Anton Tarasyuk**, Formal Development and Quantitative Verification of Dependable Systems
- 157.** **Muhammad Mohsin Saleemi**, Towards Combining Interactive Mobile TV and Smart Spaces: Architectures, Tools and Application Development
- 158.** **Tommi J. M. Lehtinen**, Numbers and Languages
- 159.** **Peter Sarlin**, Mapping Financial Stability
- 160.** **Alexander Wei Yin**, On Energy Efficient Computing Platforms
- 161.** **Mikołaj Olszewski**, Scaling Up Stepwise Feature Introduction to Construction of Large Software Systems
- 162.** **Maryam Kamali**, Reusable Formal Architectures for Networked Systems
- 163.** **Zhiyuan Yao**, Visual Customer Segmentation and Behavior Analysis – A SOM-Based Approach
- 164.** **Timo Jolivet**, Combinatorics of Pisot Substitutions
- 165.** **Rajeev Kumar Kanth**, Analysis and Life Cycle Assessment of Printed Antennas for Sustainable Wireless Systems
- 166.** **Khalid Latif**, Design Space Exploration for MPSoC Architectures

- 167.** **Bo Yang**, Towards Optimal Application Mapping for Energy-Efficient Many-Core Platforms
- 168.** **Ali Hanzala Khan**, Consistency of UML Based Designs Using Ontology Reasoners
- 169.** **Sonja Leskinen**, m-Equine: IS Support for the Horse Industry
- 170.** **Fareed Ahmed Jokhio**, Video Transcoding in a Distributed Cloud Computing Environment
- 171.** **Moazzam Fareed Niazi**, A Model-Based Development and Verification Framework for Distributed System-on-Chip Architecture
- 172.** **Mari Huova**, Combinatorics on Words: New Aspects on Avoidability, Defect Effect, Equations and Palindromes
- 173.** **Ville Timonen**, Scalable Algorithms for Height Field Illumination
- 174.** **Henri Korvela**, Virtual Communities – A Virtual Treasure Trove for End-User Developers
- 175.** **Kameswar Rao Vaddina**, Thermal-Aware Networked Many-Core Systems
- 176.** **Janne Lahtiranta**, New and Emerging Challenges of the ICT-Mediated Health and Well-Being Services
- 177.** **Irum Rauf**, Design and Validation of Stateful Composite RESTful Web Services
- 178.** **Jari Björne**, Biomedical Event Extraction with Machine Learning
- 179.** **Katri Haverinen**, Natural Language Processing Resources for Finnish: Corpus Development in the General and Clinical Domains
- 180.** **Ville Salo**, Subshifts with Simple Cellular Automata
- 181.** **Johan Ersfolk**, Scheduling Dynamic Dataflow Graphs
- 182.** **Hongyan Liu**, On Advancing Business Intelligence in the Electricity Retail Market
- 183.** **Adnan Ashraf**, Cost-Efficient Virtual Machine Management: Provisioning, Admission Control, and Consolidation
- 184.** **Muhammad Nazrul Islam**, Design and Evaluation of Web Interface Signs to Improve Web Usability: A Semiotic Framework
- 185.** **Johannes Tuikkala**, Algorithmic Techniques in Gene Expression Processing: From Imputation to Visualization
- 186.** **Natalia Díaz Rodríguez**, Semantic and Fuzzy Modelling for Human Behaviour Recognition in Smart Spaces. A Case Study on Ambient Assisted Living
- 187.** **Mikko Päkkälä**, Potential and Challenges of Analog Reconfigurable Computation in Modern and Future CMOS
- 188.** **Sami Hyrynsalmi**, Letters from the War of Ecosystems – An Analysis of Independent Software Vendors in Mobile Application Marketplaces
- 189.** **Seppo Pulkkinen**, Efficient Optimization Algorithms for Nonlinear Data Analysis
- 190.** **Sami Pyöttiälä**, Optimization and Measuring Techniques for Collect-and-Place Machines in Printed Circuit Board Industry
- 191.** **Syed Mohammad Asad Hassan Jafri**, Virtual Runtime Application Partitions for Resource Management in Massively Parallel Architectures
- 192.** **Toni Ernvall**, On Distributed Storage Codes
- 193.** **Yuliya Prokhorova**, Rigorous Development of Safety-Critical Systems
- 194.** **Olli Lahdenoja**, Local Binary Patterns in Focal-Plane Processing – Analysis and Applications
- 195.** **Annika H. Holmbom**, Visual Analytics for Behavioral and Niche Market Segmentation
- 196.** **Sergey Ostroumov**, Agent-Based Management System for Many-Core Platforms: Rigorous Design and Efficient Implementation
- 197.** **Espen Suenson**, How Computer Programmers Work – Understanding Software Development in Practise
- 198.** **Tuomas Poikela**, Readout Architectures for Hybrid Pixel Detector Readout Chips
- 199.** **Bogdan Iancu**, Quantitative Refinement of Reaction-Based Biomodels
- 200.** **Ilkka Törmä**, Structural and Computational Existence Results for Multidimensional Subshifts
- 201.** **Sebastian Okser**, Scalable Feature Selection Applications for Genome-Wide Association Studies of Complex Diseases
- 202.** **Fredrik Abbors**, Model-Based Testing of Software Systems: Functionality and Performance
- 203.** **Inna Pereverzeva**, Formal Development of Resilient Distributed Systems
- 204.** **Mikhail Barash**, Defining Contexts in Context-Free Grammars
- 205.** **Sepinoud Azimi**, Computational Models for and from Biology: Simple Gene Assembly and Reaction Systems
- 206.** **Petter Sandvik**, Formal Modelling for Digital Media Distribution

- 207.** **Jongyun Moon**, Hydrogen Sensor Application of Anodic Titanium Oxide Nanostructures
- 208.** **Simon Holmbacka**, Energy Aware Software for Many-Core Systems
- 209.** **Charalampos Zinoviadis**, Hierarchy and Expansiveness in Two-Dimensional Subshifts of Finite Type
- 210.** **Mika Murtojärvi**, Efficient Algorithms for Coastal Geographic Problems
- 211.** **Sami Mäkelä**, Cohesion Metrics for Improving Software Quality
- 212.** **Eyal Eshet**, Examining Human-Centered Design Practice in the Mobile Apps Era
- 213.** **Jetro Vesti**, Rich Words and Balanced Words
- 214.** **Jarkko Peltomäki**, Privileged Words and Sturmian Words
- 215.** **Fahimeh Farahnakian**, Energy and Performance Management of Virtual Machines: Provisioning, Placement and Consolidation
- 216.** **Diana-Elena Gratie**, Refinement of Biomodels Using Petri Nets
- 217.** **Harri Merisaari**, Algorithmic Analysis Techniques for Molecular Imaging
- 218.** **Stefan Grönroos**, Efficient and Low-Cost Software Defined Radio on Commodity Hardware
- 219.** **Noora Nieminen**, Garbling Schemes and Applications
- 220.** **Ville Taajamaa**, O-CDIO: Engineering Education Framework with Embedded Design Thinking Methods
- 221.** **Johannes Holvitie**, Technical Debt in Software Development – Examining Premises and Overcoming Implementation for Efficient Management
- 222.** **Tewodros Deneke**, Proactive Management of Video Transcoding Services
- 223.** **Kashif Javed**, Model-Driven Development and Verification of Fault Tolerant Systems
- 224.** **Pekka Naula**, Sparse Predictive Modeling – A Cost-Effective Perspective
- 225.** **Antti Hakkala**, On Security and Privacy for Networked Information Society – Observations and Solutions for Security Engineering and Trust Building in Advanced Societal Processes
- 226.** **Anne-Maarit Majanoja**, Selective Outsourcing in Global IT Services – Operational Level Challenges and Opportunities
- 227.** **Samuel Rönnqvist**, Knowledge-Lean Text Mining
- 228.** **Mohammad-Hashem Hahgbayan**, Energy-Efficient and Reliable Computing in Dark Silicon Era
- 229.** **Charmi Panchal**, Qualitative Methods for Modeling Biochemical Systems and Datasets: The Logicome and the Reaction Systems Approaches
- 230.** **Erkki Kaila**, Utilizing Educational Technology in Computer Science and Programming Courses: Theory and Practice
- 231.** **Fredrik Robertsén**, The Lattice Boltzmann Method, a Petaflop and Beyond
- 232.** **Jonne Pohjankukka**, Machine Learning Approaches for Natural Resource Data
- 233.** **Paavo Nevalainen**, Geometric Data Understanding: Deriving Case-Specific Features
- 234.** **Michal Szabados**, An Algebraic Approach to Nivat's Conjecture
- 235.** **Tuan Nguyen Gia**, Design for Energy-Efficient and Reliable Fog-Assisted Healthcare IoT Systems
- 236.** **Anil Kanduri**, Adaptive Knobs for Resource Efficient Computing
- 237.** **Veronika Suni**, Computational Methods and Tools for Protein Phosphorylation Analysis
- 238.** **Behailu Negash**, Interoperating Networked Embedded Systems to Compose the Web of Things
- 239.** **Kalle Rindell**, Development of Secure Software: Rationale, Standards and Practices
- 240.** **Jurka Rahikkala**, On Top Management Support for Software Cost Estimation
- 241.** **Markus A. Whiteland**, On the k-Abelian Equivalence Relation of Finite Words
- 242.** **Mojgan Kamali**, Formal Analysis of Network Routing Protocols
- 243.** **Jesús Carabaño Bravo**, A Compiler Approach to Map Algebra for Raster Spatial Modeling
- 244.** **Amin Majd**, Distributed and Lightweight Meta-heuristic Optimization Method for Complex Problems
- 245.** **Ali Farooq**, In Quest of Information Security in Higher Education Institutions: Security Awareness, Concerns, and Behaviour of Students
- 246.** **Juho Heimonen**, Knowledge Representation and Text Mining in Biomedical, Healthcare, and Political Domains

- 247.** **Sanaz Rahimi Moosavi**, Towards End-to-End Security in Internet of Things based Healthcare
- 248.** **Mingzhe Jiang**, Automatic Pain Assessment by Learning from Multiple Biopotentials

TURKU CENTRE *for* COMPUTER SCIENCE

<http://www.tucs.fi>

tucs@abo.fi



University of Turku

Faculty of Science and Engineering

- Department of Future Technologies
 - Department of Mathematics and Statistics
- Turku School of Economics*
- Institute of Information Systems Science



Åbo Akademi University

Faculty of Science and Engineering

- Computer Engineering
- Computer Science

Faculty of Social Sciences, Business and Economics

- Information Systems

ISBN 978-952-12-3889-5
ISSN 1239-1883

Mingzhe Jiang

Automatic Pain Assessment by Learning from Multiple Biopotentials

Mingzhe Jiang

Automatic Pain Assessment by Learning from Multiple Biopotentials

Mingzhe Jiang

Automatic Pain Assessment by Learning from Multiple Biopotentials