Kaunas University of Technology

Faculty of Electrical and Electronics Engineering

Personalized Deep Learning Models for Pain Level Classification from Photoplethysmography Signals

Master's Final Degree Project

Povilas Piartli

Project author

Prof. Vaidotas Marozas

Supervisor

Kaunas, 2021

Kaunas University of Technology

Faculty of Electrical and Electronics Engineering

Personalized Deep Learning Models for Pain Level Classification from Photoplethysmography Signals

Master's Final Degree Project

Biomedical Engineering (6211EX002)

Povilas Piartli

Project author

Prof. Vaidotas Marozas

Supervisor

Prof. Žilvinas Nakutis

Reviewer

Kaunas, 2021

Kaunas University of Technology

Faculty of Electrical and Electronics Engineering
Povilas Piartli

Personalized Deep Learning Models for Pain Level Classification from Photoplethysmography Signals

Declaration of Academic Integrity

I confirm the following:

1. I have prepared the final degree project independently and honestly without any violations of the copyrights or other rights of others, following the provisions of the Law on Copyrights and Related Rights of the Republic of Lithuania, the Regulations on the Management and Transfer of Intellectual Property of Kaunas University of Technology (hereinafter – University) and the ethical requirements stipulated by the Code of Academic Ethics of the University;

2. All the data and research results provided in the final degree project are correct and obtained legally; none of the parts of this project are plagiarised from any printed or electronic sources; all the quotations and references provided in the text of the final degree project are indicated in the list of references;

3. I have not paid anyone any monetary funds for the final degree project or the parts thereof unless required by the law;

4. I understand that in the case of any discovery of the fact of dishonesty or violation of any rights of others, the academic penalties will be imposed on me under the procedure applied at the University; I will be expelled from the University and my final degree project can be submitted to the Office of the Ombudsperson for Academic Ethics and Procedures in the examination of a possible violation of academic ethics.

Povilas Piartli

Confirmed electronically

Piartli, Povilas. Personalized Deep Learning Models for Pain Level Classification from Photoplethysmography Signals. Master's Final Degree Project / supervisor Prof. Vaidotas Marozas; Faculty of Electrical and Electronics Engineering, Kaunas University of Technology.

Study field and area (study field group):): Bioengineering, Engineering Sciences.

Keywords: deep learning, machine learning, pain, PPG, photoplethysmography.

Kaunas, 2021. 56 pages.

Summary

The aim of this work is to develop a system which detects and classifies pain into 4 categories, based on finger photoplethysmographic signal. To accomplish the task first medical literature was analyzed to find how to experimentally stimulate pain and how the body reacts to pain, what are expected physiological changes and how currently pain is measured, it was found that with pain there is an increased activation of sympathetic nervous system, and with it changes in cardiovascular system. Currently the golden standard for pain estimation is patient self-report, in cases where this is not possible there are scales based on patient's movement, body position, sound, etc. Then technical literature review was performed to find state of the art in automatic pain recognition and classification., it was found that almost all research is focused on using facial video recordings and very few methods used physiological signals, and only one used photoplethysmogram signals, this is due to limited available databases and open databases having only ECG, EDA, and EMG signals. Other methods relied heavily on EDA signal due to its large initial reaction to pain. Methods were constructed to extract and process heartbeat pulses from PPG signals, features were then extracted to describe the pulse morphology. The extracted pulses were then passed through a quality control algorithm which determined is the PPG pulse artefact free. The extracted pulses were then separated into datasets by person. The task of 4 classes was split into 2 neural networks, an initial binary classifier for pain/no pain detection, and a secondary trinary classifier for pain class detection. Three different neural network architectures were designed with 10 different types of input. Results showed that all neural networks performed well in binary classification with feature based networks performing better than signal based networks. With highest achieved accuracy of 0.92 in testing dataset and 1.00 in training dataset. In trinary classifier all networks performed poorly, with signal based networks performing better than feature based. With highest achieved accuracy of 0.61 in testing dataset and 1.00 in training dataset.

Piartli, Povilas. Personalizuoti giliojo mokymo modeliai skausmo klasifikacijai iš fotopletizmogramos signalo. Magistro baigiamasis projektas / vadovas prof. Vaidotas Marozas; Kauno technologijos universitetas, Elektros ir Elektronikos fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Bioinžinerija, inžinerijos mokslai.

Reikšminiai žodžiai: gilusis mokymas, mašininis mokymas, skausmas, FPG, fotopletizmograma.

Kaunas, 2021. 56 p.

Santrauka

Šio darbo tikslas - sukurti sistemą, kuri atpažintų ir klasifikuotų skausmą į klases, naudojant piršto fotopletizmogramos signalą. Tam pirma buvo atlikta medicininės literatūros analizė siekiant išsiaiškinti kaip eksperimentiškai stimuliuoti skausmą, kaip kūnas reaguoja į skausmą ir kaip skausmas yra matuojamas. Buvo rasta, kad skausmo metu suaktyvėja simpatinė nervinė sistema ir dėl to yra pokyčių širdies veikloje ir kraujotakos sistemoje. Dabartinis skausmo vertinimo etalonas yra subjektyvios skausmo vertinimo skalės, pagal kurias pacientai patys įvertina patiriamo skausmo lygį. Tuo atveju kai pacientas negali apibrėžti patiriamo skausmo lygį, gydytojas, pasinaudodamas ilgamete praktika, vertina skausmą subjektyviai pagal paciento veido išraišką, judesius ir kitus intuityviai suprantamus požymius. Mokslinės techninės literatūros analizė atskleidė, kad beveik visi metodai remiasi veido vaizdo įrašais, tik keli tyrimai yra paremti fiziologinių signalų analize ir tik viename naudojama fotopletizmograma (FPG). Literatūros analizė taip pat parodė laisvai prieinamų duomenų bazių trūkumą. Siūlomas skausmo objektyvaus vertinimo algoritmas remiasi FPG signalo kokybės vertinimu, segmentavimu į širdies dūžius, signalo atkarpų aprašymu morfologiniais požymiais ir požymių klasifikavimu dirbtiniais neuroniniais tinklais (DNT) . FPG signalo segmentai buvo padalinti į 2 rinkinius, DNT apmokymui ir testavimui. Pirmajame DNT apmokymo etape buvo siekiama ištirti skausmo atpažinimo galimybes (binarinis klasifikavimas), o antrajame– skausmo lygio įvertinimo galimybes (klasifikavimas į 3 klases). Taip pat buvo įvertintas giliojo mokymo DNT efektyvumas klasifikuoti skausmo lygį remiantis FPG signalo morfologija nenaudojant jokio signalo parametrizavimo. Tyrimo rezultatai parodė, kad neuroniniai tinklai gerai pasirodė dvinarėje klasifikacijoje, savybių pagrindu veikiantys tinklai pasirodė geriau negu gryno signalo pagrindu veikiantys. Bendrai geriausi pasiekti rezultatai yra 0,92 tikslumas testavimo rinkinyje ir 1,00 apmokymo rinkinyje. Trinarėje klasifikacijoje visi tinklai pasirodė prastai, gryno signalo tinklai pasirodė geriau negu savybių pagrindo. Bendrai geriausi rezultatai pasiekti yra 0,61 tikslumas testavimo rinkinyje ir 1,00 apmokymo rinkinyje.

Table of contents

List of figures

List of tables

List of abbreviations

Abbreviations:

PPG – Photoplethysmogram
VPG – Velocity plethysmogram
APG – Acceleration plethysmogram
JPG – Jerk plethysmogram
SPG – Snap plethysmogram
CPG – Crackle plethysmogram
NN- Neural network
CNN – Convolutional neural network
RNN – Recurrent neural network
LSTM – Long-short term memory neural network
STD – standard deviation
ECG – Electrocardiogram
EDA – Electrodermal activity
EMG – Electromyogram

Introduction

Pain is a mechanism for the body to indicate that there is a threat or damage, however it is not always needed and can have undesirable effects, as such there is a need for pain reduction or elimination. Currently pain is reduced using various analgesics, however, pain is also often mismanaged, even for cases with severe pain[1]. This mismanagement stems from subjective factor in pain evaluation from the medical staff. Currently pain is recognized and classified in two ways, by patient self-report which is later evaluated by the medical staff or by direct evaluation by medical staff in cases with non-verbal patients. However, these methods are not always suitable, in cases such as heavily sedated, non-verbal adults or small children it is hard to ask or evaluate their pain properly. A solution for this problem is pain detection and classification from physiological signals, currently this area is very new but quickly developing due to application of machine learning techniques and miniaturization of sensors, including wearable sensors for continuous monitoring[2]. Particularly photoplethysmography is interesting due to its devices being small, energy efficient, and already present in clinical setting as blood oxygen saturation meters. Currently there are no commercial solutions for pain detection and classification which use physiological signals. In this research finger photoplethysmography signals will be analyzed, their features will be extracted and used in developing pain classification algorithms. Both parametrization of the signal and signal waveforms will be tested with neural networks to determine which approach provides better accuracy with basic neural networks. Additionally, different types will be tested to determine if recurrency or convolution provides additional advantage. Different time windows were tested to determine of longer recordings help with pain classification.

The aim of this work is to develop a system which detects and classifies pain into 4 categories, based on finger photoplethysmographic signal.
The objectives are the following:

1. To analyze pain effects on the body to determine how to measure it;
2. To analyze currently applied pain recognition and classification methods;
3. To propose features describing PPG signal morphology;
4. To propose PPG signal quality control algorithm;
5. To test and evaluate different types of neural networks with both features and raw signals.

1. State of the art and previous research

1.1. Clinical background

Clinical background overview is split into 2 parts, pain stimulation methods and pain estimation methods. Pain stimulation methods were overviewed in order so select the most suitable one for testing on volunteers. Pain estimation methods were overviewed to analyze

1.1.1. Pain stimulation methods

There are several different pain stimulation methods and can be classified according to how pain stimulation is performed: mechanical, thermal, electrical, and ischemic (Fig. 1). Besides, each method can be classified as invasive or non-invasive:

- Invasive methods affect more nerve fibers and provide a higher-level stimulation compared to non-invasive methods. Invasive methods activate sensory cells which are responsible for sensation and damage and always leave lasting effects on the organism.
- Non-invasive methods provide lower-level stimulation compared to invasive methods. However, they are more often used because they carry lower risk than invasive methods; also, they are easily repeatable and leave less lasting effects.

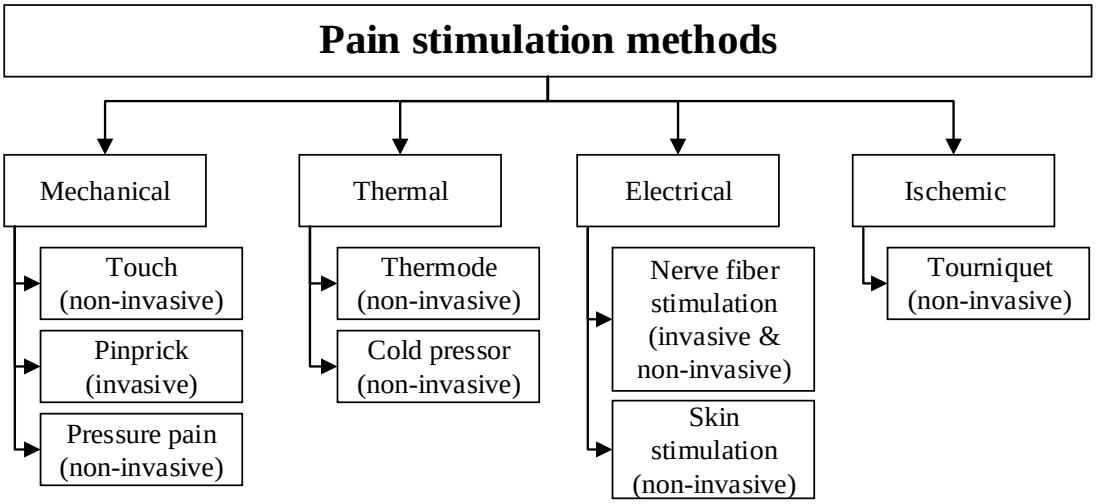

Fig. 1. Classification of pain stimulation methods

Mechanical pain stimulation methods

Mechanical pain stimulation methods are performed using a controlled pressure which affects mechanoreceptors and nociceptors. Mechanoreceptors sense a wide variety of modalities (e.g., touch, pressure, vibration). Meanwhile, nociceptors sense all previously mentioned modalities and one additional (damage of cells). Nociceptors trigger from higher values of those modalities. Most mechanical pain stimulation methods affect both receptors. In this case, when both types of receptors are affected, the sensation of pain is reduced [3]. It is hypothesized that this pain reduction is an adaptation mechanism due to mundane tasks. Mechanical pain stimulation methods can be categorized into three groups [4]:

- **Touch method** uses additional light pressure which can be performed by fingers or Von Frey hair (Fig. 2 a). Von Frey hair is calibrated filament that bends when a certain pressure is reached [4]. This method usually used when there is increased skin sensitivity.
- **Pinprick method** is invasive, and stimulation is applied to a very concentrated area using a needle or safety pin [4] (Fig. 2 b). It leaves lasting damage to the skin and is unsuitable for repeatability studies.
- **Pressure method** is performed by applying controlled pressure (controlled pain) to specific locations on the body (e.g. earlobe, finger, toe) [4]. Pressure pain method is usually performed using pressure algometer (Fig. 2.c).

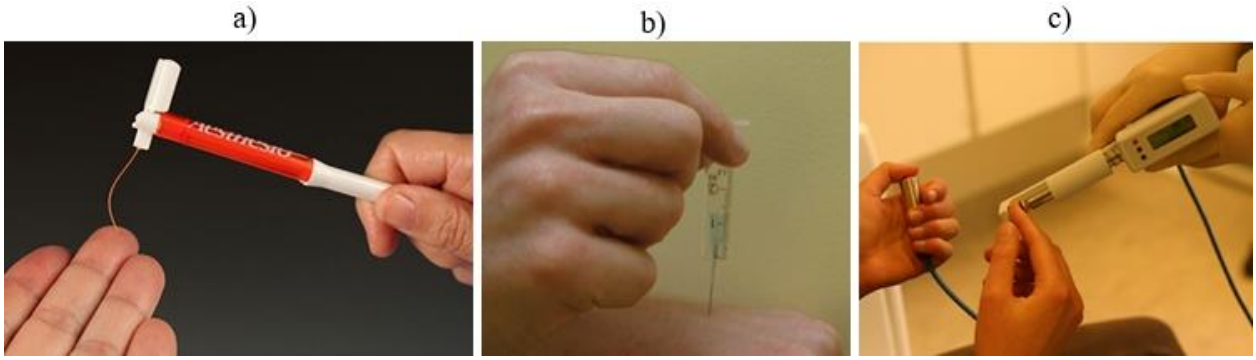

Fig. 2. Examples of mechanical pain stimulation methods: a) touch method – Von Frey hair, b) Pinprick method – needle (adapted from [5]), c) pressure method – algometer [6]

Advantages and disadvantages of mechanical pain stimulation methods are presented in Table 1. These methods can be quite easily performed but can be used only in specific situations (e.g. pinprick method only can be used by a trained specialist). In many cases, the disadvantages of mechanical methods overcome their advantages.

Table 1. Advantages and disadvantages of mechanical pain stimulation methods

	Advantages	Disadvantages
Touch method	A concentrated effect area; Non-invasive.	Stimulates mechanoreceptors, could be avoided with a slow application; Low maximum stimulation; Often used in situations with increased sensitivity on the skin from other factors.
Pinprick method	A concentrated effect area; Strong stimulation.	Invasive; Difficult to control the strength of stimulation; Cannot be repeated on the same location with repeatable results (causes hyperalgesia [4]).
Pressure method	A high degree of control (area, pressure, location); Can be applied slowly by limiting the effect of mechanoreceptor suppression [7].	Slow; A large area of minimal stimulation.

Thermal pain stimulation methods

Thermal pain stimulation methods can be easily applied by using controlled hot or cold stimulus on the subject's skin and affects skin's thermoreceptors. There are two main thermal pain stimulation methods:

- **Thermode method (TM)** uses thermodes for a fast and controlled thermal stimulation of a small area of the body. Applied temperature varies from 0 °C to +51 °C [8].
- **Cold pressor (CP) test (CPT**) uses immersion of hand, forearm or foot into the cold water of 0-15 °C [9]. The CP test is famous in the studies due to its simplicity and lack of danger to the participant. The advantages and disadvantages of thermal pain stimulation methods are summarized in Table 2.

Table 2. Advantages and disadvantages of thermal pain stimulation methods

	Advantages	Disadvantages
Thermode method	Easy location control; Limitation of maximum temperature to avoid burns; Potential for experiment automation; Could be used for both invasive and non-invasive tests [10]	Equipment is bulky (liquid exchange type thermodes) and expensive;
Cold pressor test	Cheap and simple equipment (bucket and cold water); Simple testing protocols; Easy to control (the test can be easily stopped by the participant if the pain is too high).	Very varied protocols (stirred/not stirred, temperature, depth of immersion, time of immersion); Experiment imposable to automate; Increase vasoconstriction.

Both thermal pain stimulation methods are widely applied in practice. Comparative analysis of specific protocols, times of stimulation, and other information is summarized in Table 3.

Table 3. Comparative analysis of cold pressor test (sorted by CPT water temperature)

Protocol	Baseline time	CPT water temperature, °C	CPT duration, min	Psychological evaluation	Post CPT evaluation
Stone [11]	25 min	0-1	3	-	-
Dixon [12]	-	0-2	5	VAS-I, VAS-U	-
Ghiasi [13]	3 min	0-4	3	-	4 min
Imai [14]	-	0-4	TT	VAS	-
Malarvili [15]	30 min	0-5	5	-	5 min
Patryla [16]	2 min 37 °C water	1	-	0-100 NRS	-
Mitchell [17]	1 min 32 °C water	1,3,5,7	5	VAS, PRI	1 min 32°C
Wirch [18]	5 min	4-5	6	-	15 min
Campbell [19]	-	5	5	0-20 scale	-

Table 3. Continuation

Protocol	Baseline time	CPT water temperature, °C	CPT duration, min	Psychological evaluation	Post CPT evaluation
Chalaye [20]	-	7	5	VAS-I, VAS-U (every 15 s)	-
Kaushik [21]	15 min	10	1	-	-
Gehling [22]	-	10	1	0-100 NRS	-
Mizeva [23]	10 min	-	3	-	10 min
Geisser [24]	-	-	5	0-10 (every 10 s), McGill scale	-

VAS-I – visual analogue pain intensity scale , VAS – visual analogue scale, PRI – Pain Rating Index scale, NRS – numeric rating scale; TT – till threshold.

Electrical pain stimulation methods

Electrical pain stimulation can be easily applied and controlled; however, it activates nerves unnaturally (in a very synchronized manner), excites full spectrum of peripheral nerve fibers, and bypasses sensory nerve endings. An alternative is to apply a smaller stimulus which produces a stinging or burning sensation on the skin. There are two types of electrical pain stimulation methods (summarized in Table 4):

- **Nerve fiber stimulation –** electrodes are placed in a way for the electric current to activate the nerve fibers directly, without activating the appropriate nociceptors. The stimulation can be performed invasively (needle electrodes, lower voltages, provide minimal stimulation to the surrounding areas) and non-invasively (skin electrodes, higher voltages, affect and stimulate the surrounding tissues) [4].
- **Skin mechanoreceptors stimulation** tries to imitate natural pain by using skin electrodes and limited voltage. Limited voltage reduces stimulation effect on nociceptors and in that way only mechanoreceptors are activated. Therefore, pain is more natural [7].

Table 4. Advantages and disadvantages of electrical pain stimulation methods

	Advantages	Disadvantages
Nerve fiber stimulation	Fast and simple control; Can be non-invasive.	Bypass of receptors, effects could be different from natural stimulation; May require abrasion or damage to the skin to reduce resistance.
Skin mechanoreceptors stimulation	Imitate natural pain; Easy to control; Non-invasive.	A large area of stimulation; Slow and weak stimulation.

Ischemic pain stimulation methods

Ischemic pain stimulation is a specific stimulation applied by using a tourniquet. Ischemic pain felt when blood flow is reduced to some parts of the body. The stimulation is usually performed on

limbs (e.g. hand, leg); however, it is also possible to use it for other parts of the body. There are two types of ischemic pain stimulation methods (summarized in Table 5):

- **Stationary method –** ischemic pain is caused by a tourniquet applied to a limb in a stationary position. A tourniquet is held until the pain becomes unbearable but not longer than 2 hours [4].
- **Exercise method –** the same as the tourniquet method except a tourniquet is applied not in a stationary position but before or just after the end of physical activity [25].

Table 5. Advantages and disadvantages of ischemic pain stimulation methods

	Advantages	Disadvantages
Stationary method	Simple control and equipment; Pain fades away quickly after the test.	Potentially low test-retest reliability [26]; Subjects to experience fatigue before pain [26]; Slow application.
Exercise method	Simple and fast; Indifferent to the fitness level of the participant.	Sex differences; Protocols vary across studies.

1.1.2. Body reaction to pain

Pain induces reaction in physiological systems, which in turn results in changes in specific parameters which can be recorded and evaluated. More specifically pain induces increased reactivity of the autonomic nervous, cardiovascular and respiratory systems. Pain level is also associated with physiological parameters(e.g. heartrate, respiration rate, blood pressure, skin conductance, etc.)
This section mainly analyses the changes caused only by thermal pain. When cold environments expose skin's thermoreceptors, they detect temperature drop and pass this information through the afferent nerves to the anterior hypothalamus, which controls the heat balance. The control center emits control signals to other areas of the hypothalamus that regulate the rate of heat release. When physiological thermoregulatory measures become ineffective in maintaining thermal homeostasis, the heat generation center begins to stimulate efferent nerves of the sympathetic nervous system [27]. Activation of the sympathetic nervous system tries to counter heat loss and maintain homeothermy. Increased activity of the sympathetic nervous system increases sweating (electrodermal activity), respiratory rate, heart rate, blood pressure, and vasoconstriction and dilation of the arteries.

Pain and changes in heart activity, arterial blood pressure, and respiration

The heart rate is a well-known indicator of pain and activity of the sympathetic(SNS) and parasympathetic(PNS) nervous system[28]. The relative increase in SNS activity is associated with increased HR and increase in PNS activity is associated with decreased HR. Therefore, the PNS influences are the ones capable of producing rapid changes in the measurement of the interbeat intervals (IBI). IBI sequences are described as the time interval between R peaks in ECG signal. Different time intervals are used for HRV evaluation. These are three main groups (domains) used for the analysis of IBI: 1) time-domain, 2) frequency-domain, and 3) nonlinear-domain. The time-domain analysis measures HRV parameters of normal-to-normal (N-N) intervals: 1) a standard deviation of all N-N intervals (SDNN) and 2) the square root of the

mean of the sum of squares of differences between adjacent N-N intervals (RMSSD) [29]. The frequency-domain analysis measures HRV parameters of N-N intervals by using fast Fourier transformation and calculates the power spectrum of specific frequency ranges. The pain studies focus on three main HRV parameters: low-frequency (LF), high frequency (HF), and the ratio of LF and HF (LF/HF ratio) [29]. In a study[30] of 101 subject it was found that during cold pressor test LF component increases, HF component decreases and LF/HF ration is also increasing. This indicates that during CPT SNS is dominant, similar results were found in [31] and [32]. The disadvantage of HRV analysis is that long periods (5 minute recordings being standard[33] however shorter recording based estimations exist[34]) of monitoring are required for accurate evaluation of LF component, this reduced the usability of HRV for pain evaluation into only being applicable for long term pain.

Mean arterial blood pressure (MAP) increases during cold pressure test due to increased peripheral vascular resistance [26]. The increased peripheral vascular tone of peripheral muscular arteries/arterioles has a significant effect; it fasters velocity of reflected arterial pressure waves [35]. The early return of arterial pressure waves augments the amplitude of central systolic and pulse BP, resulting in elevated wave reflection intensity. However, clinically approved continuous MAP measurement tools are invasive which limits its usage to patients which would have MAP measured for other reasons.

The amplitude of the photoplethysmogram (PPG) signal is related to volume changes in the peripheral circulation. Therefore, the amplitude of the PPG signal decreases due to the narrowing of the arterial blood vessels caused by the activated sympathetic nervous system and pain. A study [36] showed that the intensity of relatively mild pain caused by two heat stimuli (43 and 48 °C) has no relationship to changes in PPG signal amplitude. In contrast, the cold pressure test caused more severe pain and reduced the amplitude of the PPG signal. It can be concluded that the amplitude of PPG signal changes significantly in response to even mild pain stimuli, but has limited specificity in the assignment of pain intensity [36]. In another study [37] it was found that PPG AC amplitude decreases during CPT and increases with hand immersed in hot water(55°C), with IR PPG AC changing more than red PPG AC. PPG is a good candidate for pain evaluation due to its ease of measurement and quick reaction to pain.

Another parameter combining ECG and PPG signals is pulse arrival time (PAT). PAT is the time delay in between R wave peak of ECG and onset of corresponding PPG signal. A significant rise in arterial blood pressure increases the vascular tone, and the arterial wall becomes stiffer, causing the PAT to shorten. In opposite, a decrease in vascular tone due to a fall in BP causes PAT to lengthen. The study [38] showed that PAT was slightly lower during CPT compared to rest. PAT recovers within in 1 min while amplitude and slope do not recover entirely by 5 min. However, PAT requires measurement of both PPG and ECG and for good synchronization between devices, this is possible and not difficult compared to MAP, but it makes the equipment more bulky and harder to apply.

Breathing and pain model studies showed that sudden skin pain increases inspiratory flow by reducing inspiratory time or by increasing inspiratory volume, or by combining both. The inspiratory flow rate has a correlation with inspiratory intensity [8]. The studies have found that acute pain increases respiratory rate and volume [39]. However, respiration is inconvenient to measure, requiring either a respiration belt or a sensor in face area.

Pain and skin and muscular responses

Increased sympathetic tone leads to increased sweating. Sweating occurs because of sudomotor (sweat gland) activation; it has two effects – decreased skin resistance and a small electrical potential is developed during sudomotor activation. Increased skin conductance has been associated with pain response [40], [41], [42]. Studies identified time points with moderate or severe postoperative pain (with 90 % sensitivity, 64 % specificity and 89 % sensitivity, 74 % specificity respectively) [40], [42]. However, skin conductance is significantly impacted by external conditions, which can change evaporation speed. Another method to investigate sympathetic skin responses is to register sudomotor activation, during the activation a small electrical potential is developed. The advantage of this method is that the signal directly shows when the sweating is happening and not after it happened enough to have a change in skin conductance. Sudomotor activity has been shown to have a correlation with pain response [43]. However, sudomotor activity measurement is inconvenient as it shares the frequency range with EMG and requires specific electrode placement for low noise and high signal level(low muscular activity with high sweating activity).

It is generally assumed that stress induces muscular tension [44] and pain as a stress-inducing agent should have this effect also. It has been shown that electromyography (EMG) activity in trapezius muscle is significantly increased during a painful experience [44]. However, EMG monitoring is inconvenient due to high noise levels from movement and inconvenient application of electrodes.

1.1.3. Patient reported pain level

The current golden standard for pain evaluation is patient self-report. Self-reporting is done with the help of standardized pain scales. Examples of the most popular pain scales can be seen in Table 6. The table is divided into 2 main groups of scales, by the type – qualitative and quantitative. Qualitative scales serve the main purpose to find the type of pain, its location, while Quantitative scales are used for pain value, how strong is the pain.

Table 6. Comparison table of verbal patient pain estimation methods

Name	Type	Range	Method of expression
McGill pain questionnaire (MPQ)	Qualitative	-	Verbal
Coping Strategies Questionnaire (CSQ)	Qualitative	-	Verbal
Kohn Reactivity Scale (KRS)	Qualitative	-	Verbal
Pennebaker Inventory of Limbic Languidness (PILL)	Qualitative	-	Verbal
Visual analogue scale (VAS)	Quantitative	0-100%	Verbal or manipulatory
Verbal descriptor scale (VDS)	Quantitative	1-3 or 1-5 Predefined meanings 1 – none, 2 – mild, 3 – moderate, 4 – severe, 5 – unbearable	Verbal
Numerical pain rating scale (NPRS)	Quantitative	0-10 or 0-100	Verbal

Psychophysical methods are based on the participant's individual (personal, subjective) assessment of the pain. Commonly used psychophysical methods [45]:

VAS, VDS, and NPRS methods are quantitative and serve the purpose to evaluate pain level. Meanwhile, MPQ method is the multidimensional instrument and describes the pain in a qualitative way with the purpose to identify location and type of pain. VAS and NPRS methods are often used in daily practice, while, VDS and MPQ methods more commonly used in clinical research [46]. The major disadvantage of subjective pain evaluation methods is that the intensity of measured pain highly dependents on various factors such as the previous experience of pain, emotional status, physiology, location, type of pain, etc.

Patient reported scales have an additional limitation that they require a patient capable of cognitive functions and communication.

1.1.4. Medical staff reported pain level

There are various scales which can be used by medical staff, however, all of them require a responding patient. This means that sedated patients which do not show any expressions, whether they are facial or full body, cannot be evaluated for their pain levels. The scales seen in Table 7 show various characteristics that are used for pain evaluation, such as body and facial movements, sounds, body position. Particularly facial and body movements are used as they are present in all pain scales. However, body movements are harder to capture and as such are suitable for automated methods. Facial movements are easier to capture as such have been popular in machine learning community. However, they still have the mentioned issue of not being present in some patients.

Table 7. Comparison table of non verbal patient pain estimation methods[47]

Name	**Range**	**Domain**					
		Facial expression	**Body movement**	**Body posture**	**Verbal response**	**Ventilator compliance**	**Physiologic dimension**
Behavioral Pain Rating Scale [48]	0-3 per domain 0-12 total	None to constant frowning or grimacing	Quiet to very restless	Relaxed to extreme tenseness	Normal speech to crying	-	-
PAIN Algorithm [49]	0-1	None to frown/ grimace/ wince	None/slow movement to restless	None to rigid/ splinting/ tenseness	None to crying/ moaning	-	Changes in heartrate, blood pressure, respiration, pallor
Behavioral Pain Scale [50]	1-4 per domain 3-12 total	Relaxed to grimacing	None to limbs retracted permanently	-	-	Tolerating movement to unable to control ventilation	-
Nonverbal Pain Scale (NVPS) [51]	0-2 per domain 0-10 total	None/smile to frequent grimace, frown	Lying quietly to restless, excess activity or withdrawal from stimulus	Lying quietly to rigid/ stiff	-	-	Vital sign changes in past 4 hours. Changes in skin moisture/color or pupil size

Table 7. Continuation

Name	Range	Domain					
		Facial expression	**Body movement**	**Body posture**	**Verbal response**	**Ventilator compliance**	**Physiologic dimension**
Pain Behavior Assessment Tool [52]	0-1	10 descriptors	15 descriptors	-	7 descriptors	-	-
Critical Care Pain Observation Tool (CPOT) [53]	0-2 per domain 0-8 total	Relaxed to grimacing	None to restless	Relaxed to very tense/rigid	For non-intubated patients: Normal speech or no vocalization to crying	For intubated patients: tolerate ventilator or movement to fighting ventilator	-

1.2. Objective pain assessment methods

Current research on automatic pain recognition or classification is very dependent on available databases and researcher specializations. Due to these factors most of databases and research currently is based on image or audio processing and not on physiological signals. The few databases that have physiological signals have a very limited variety – ECG, EDA, EMG signals. This results in research focused on using those signals. The issue is how to describe these signals, how to extract their features and here a limitation is encountered – most researchers use statistical descriptors for specified time windows.

Additionally there is only a single commercial project for automatic pain recognition[54], however, it is not currently sold, pending FDA approval. The PainQx operates using EEG data and aims to map brain activity for pain detection. This kind of approach is suitable for longer term monitoring, but EEG equipment is comparably heave and hard to set up, especially compared to PPG readers, making this product a bad choice for quick measurement applications.

1.2.1. Automatic pain recognition

There has been a very large review in 2019 about automatic pain recognition and classification methods[55]. The methods can be separated into 3 categories, video recording analysis, physiological signal analysis and combined. In this research the goal is to classify pain from physiological signals, as such, methods using video recordings will be ignored. It is also noticed how research is very centered around the available databases, this resulted in most of the research using BioVid database. In Table 8 you can see some examples of research on pain classification using automated methods which use physiological signals, however, signals are centered around EDA, ECG, facial and back EMG. ECG is used to extract blood volume pulse it should be similar to photoplethysmography, but it is not exactly the same and that it does not have pulse wave morphology. This means that the blood volume pulse does not have information of how the vascular system is changing. This research will be new in that it will include the information from the vascular system.

Most of the research which used BioVid database mentions that the most informative signal is skin level conductance, this is an expression of sweating. However, skin level conductance has a sharp peak with the beginning of pain and then decreases to normal levels, as such, it is not suitable for longer term pain as it only reacts to the increase in pain and not presence of pain.

Table 8. Comparison table of machine learning based automatic pain recognition algorithms using physiological signals [55]

Author	**Model**	**Physiological signals**	**Accuracy achieved**	**Dataset**
Kächele ’16[56]	Random forest and k-NN	ECG, EDA, EMG (trapezius)	-	BioVid
Amirian ’16[57]	RBF Neural Network	ECG, EDA, EMG (trapezius)	80% Binary 32.1% 5 class	BioVid
Lopez-M. ’17[58]	Multi-task neural network	ECG, EDA, EMG (trapezius)	-	BioVid
Lopez-M. ’18[59]	Recurrent Neural Networks	ECG, EDA	-	BioVid
Jiang ’17[60]	Neural network	EDA, facial EMG, HR, RSP	70.6% 3 class	-
Hinduja ’20[61]	Random forest	DBP, MBP, SBP, EDA, RBP, Pulse, RSP rate, RSP volts	77.7% Binary	BP4D+
Chu ’17[62]	Genetic algorithm and principal component analysis	ECG, EDA, PPG	75% 4 class	-

DBP – diastolic blood pressure, MBP – mean blood pressure, SBP – systolic blood pressure, EDA – Electrical dermal activity, skin conductance, RBP – Raw blood pressure, RSP – respiration, HR – heart rate.

The research performed by Chu et al[62] is the closest to what is aimed at this research. The distinction from other research is in the usage of PPG signals. While others estimated blood volume pulse from ECG signal, Chu used PPG signal. However, the same as research based on BioVid database, pain intensity is registered by the applied stimulus intensity and not volunteers evaluation. While in our research pain intensity will be coded by volunteers evaluation, this makes accuracy comparisons hard.

1.2.2. Available Databases

In Table 9 publicly available databases and their available modalities can be seen. It can see that very few databases have physiological signals. For any groups attempting research without their own data acquisition this is a big limiting factor. Because of this, research is focused on video and audio methods. Only BioVid, BP4D+, EmoPain, SenseEmotion, and X-ITE databases have physiological signals and those are mostly limited to ECG, EDA, EMG, with BP4D+ being an exception with heartrate, respiration rate and blood pressure as additional measurements. As such any research which collects and examines other physiological signals will be very new and beneficial.

Table 9. Publicly available databases[55]

Database	**Subjects**	**Pain stimulation method**	**Video/Audio**	**Physiological signals**
UNBC-McMaster	25 adults	Exercise for chronic shoulder pain affected limb	Facial video	
BioVid	90 healthy adults	Heat pain, thermode	Facial video	EDA, ECG, EMG(trapezius, corrugator, zygomaticus muscles)
BP4D	41 healthy adults	Cold pressor test	Facial video	
BP4D+	140 healthy adults	Cold pressor test	Facial video	heart rate, respiration rate, blood pressure, EDA
MIntPain	20 healthy adults	Electrical pain	Facial video	
EmoPain	22 adults	Exercise for chronic lower back pain patients	Video, audio, motion capture	EMG(trapezius, lumbar, paraspinal muscles)
SenseEmotion	45 healthy adults	Heat pain	Facial video, audio	EDA, ECG, EMG (trapezius muscle), RSP
X-ITE	134 healthy adults	Heat and electrical pain	Facial video, body video, audio	EDA, ECG, EMG(trapezius, corrugator, zygomaticus muscles)

2. Methodology

2.1. Data acquisition

Data was collected in-house for a SzeleSTIM project. Data collection took place indoors at the Biomedical Engineering Institute (Kaunas, Lithuania) in a quiet, temperature-controlled (24 °C ± 1 °C) room at the same time of the day (08:00-13:00), protocol can be seen in Fig. 3. Participants were instructed to put their arm up to the middle of the forearm into water during warm and cold water phases, indicate their pain as often as they can distinguish differences and to pull out their hand if the pain feels too uncomfortable to continue.

Fig. 3. Protocol of the study

PPG was registered on a finger using Nautilus II data acquisition system (1000Hz, Kaunas University of Technology Biomedical Engineering Institute, Kaunas, Lithuania The data is labeled by using NPRS pain scale data collected during experiment. Non-uniformity of NPRS data is corrected using linear interpolation. Fifty-one healthy volunteers (26 women), 36.25 ± 10.34 years old (range 22 to 64 years), with a height of 1.76 ± 0.09 m, weight of 74.68 ± 14.89 kg, and body mass index of 24.11 ± 3.70 kg/m2 participated in the study. The subjects were fully informed about the investigation and any possible related risks and discomfort. The study was conducted by following the ethical principles of the Declaration of Helsinki and with approval from the Kaunas Region Biomedical Research Ethics Committee (No. BE-2-24). Identifiable information was removed from the collected data to ensure participant anonymity.

Sensor and electrode placement can be seen in Fig. 4 and Fig. 5.

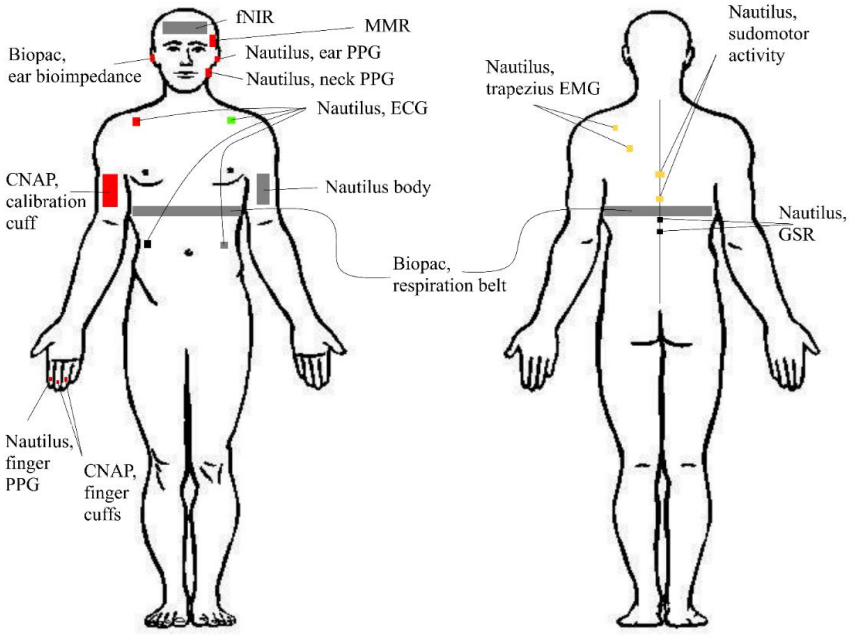

Fig. 4. Placement of sensors and electrodes. Body outline source [63]

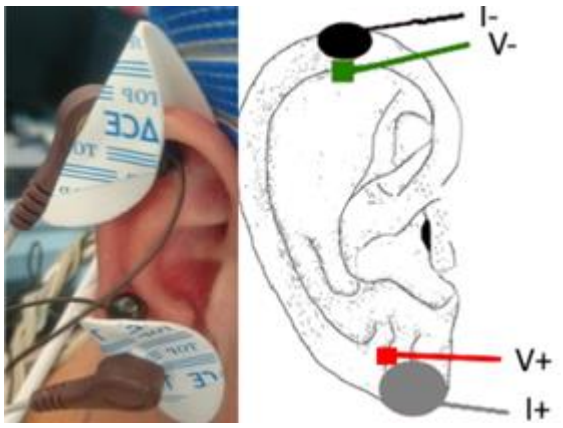

Fig. 5. Electrode placement for ear bioimpedance measurement

Below are data examples from a single volunteer. Finger photoplethysmogram (PPG), seen in Fig. 6. Changes in peak-to-peak level can be seen (2% at CPT2 to 4% and Rest1 phase), however there are no clear morphology changes.

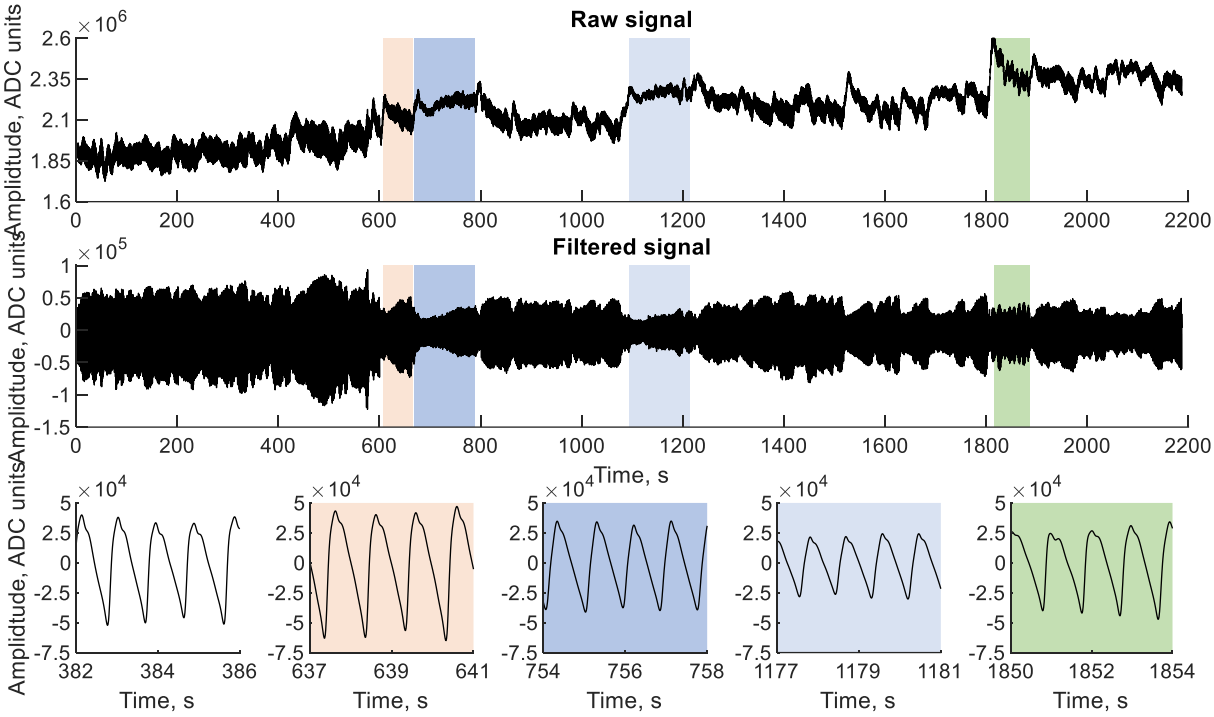

Fig. 6. Red color PPG signal recorded from finger (white – baseline, light brown – 32 °C warm water, blue – 7 °C cold water, light blue – 10 °C cold water, green – deep breathing)

2.2. Algorithm

The proposed pain classification algorithm is split into 2 main parts – data preparation and processing with neural networks. The first part can be seen in Fig. 7, while the second part is seen in Fig. 8. Data preparation algorithm is used to convert data from signals into segments of standard length and to extract morphology features. The second part of the solution is to use neural networks for pain classification, the task was split into two neural networks, first for binary classification pain/ no pain, second for 3 class classification into light/moderate/heavy pain.

This splitting was initiated after initial trials showed that a direct 4 class classifier performed very poorly.

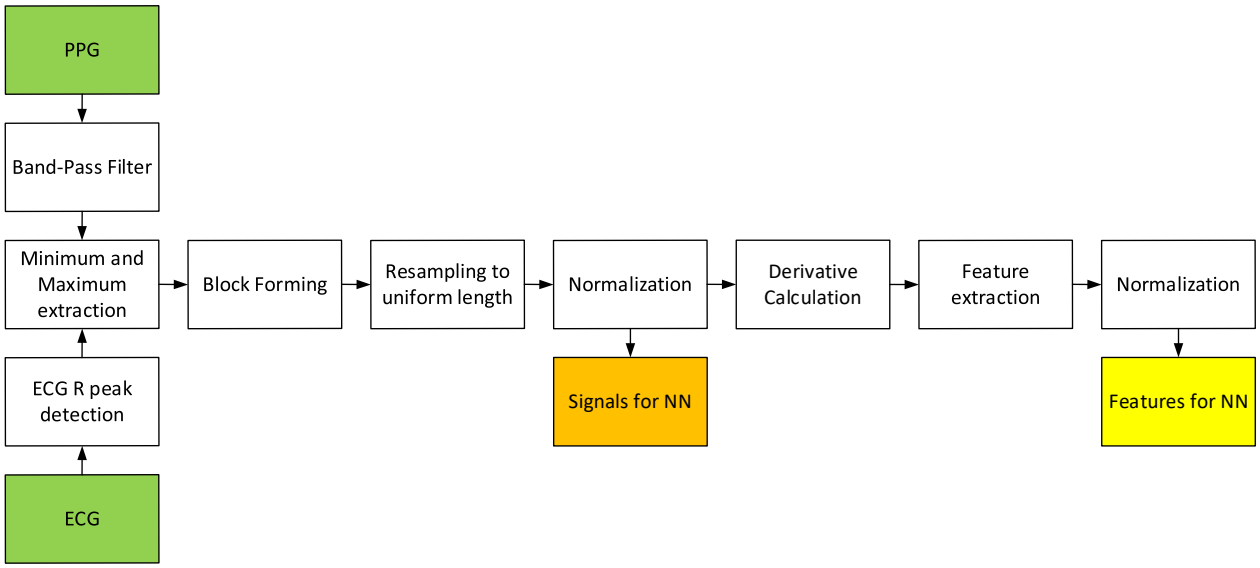

Fig. 7. Data preparation algorithm, from raw data to signals and features for usage in neural networks

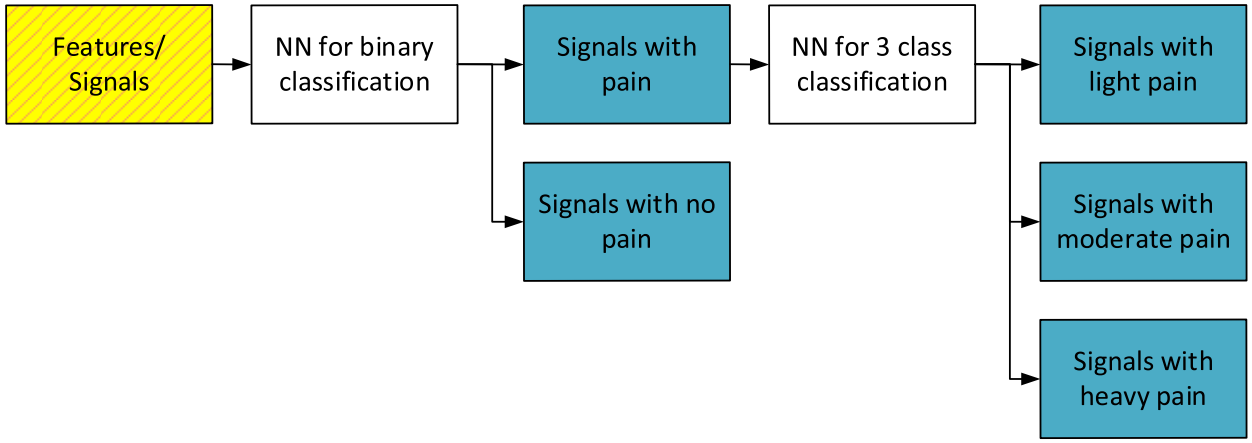

Fig. 8. Pain classification algorithm, input of features/ signals, classification into 4 classes of no/ light/ moderate/ heavy pain

2.3. Data preparation

The process is separated into 7 steps:

1. Heartbeat pulse detection
2. Baseline removal
3. Heartbeat pulse resampling to uniform length
4. Heartbeat pulse amplitude normalization
5. Derivative heartbeat pulse calculation
6. Feature extraction
7. Feature normalization

2.3.1. Data preprocessing and normalization

First the signal is filtered by high-pass (Butterworth, order of 4, 0.5hz cut-off) and low-pass (Butterworth, order of 2, 10hz cut-off) filters, using zero-phase filtering algorithm. The next step is to segment the signal heartbeat by heartbeat. For simplification this was done with the help of ECG signal, R peaks were extracted and used as guides for PPG heartbeat detection. Intervals between 2 heartbeats are searched for minimum value. These minimum values are used to describe a single heartbeat segment as a segment from one minimum to another.

Baseline is additionally removed by calculating a straight line from beginning to end and subtracting it from the segment, example can be seen in Fig. 9. Heartbeats are then additionally normalized. It is assumed that a previous 2-minute recording without pain would be available and as such its maximum is considered as 1. Additionally, derivatives up to fifth order are calculated for the entire signal, segmented based on normal signal segmentation points and normalized and unified with the same method with the exception of additional baseline removal.

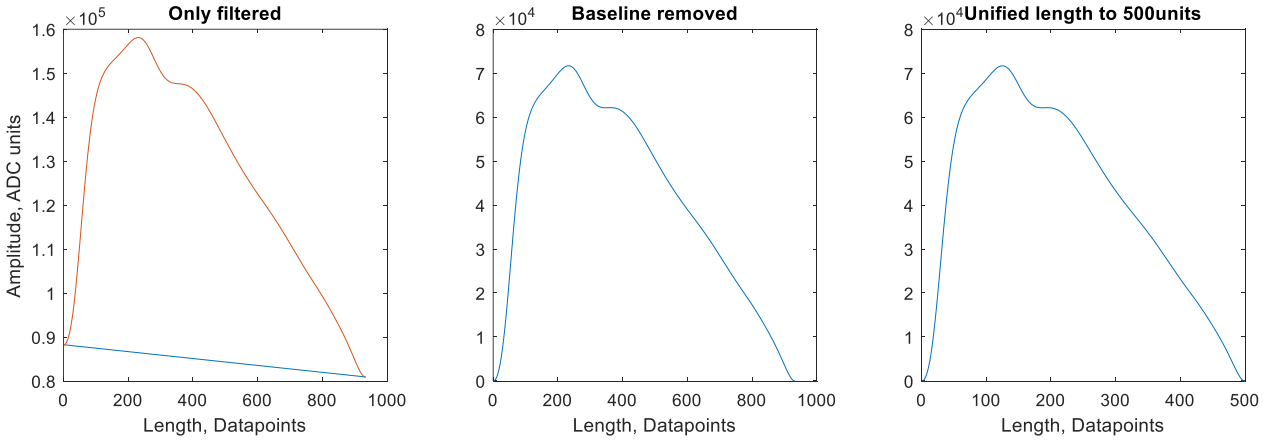

Fig. 9. Length unification algorithm

2.3.2. Quality control

Quality control is performed in 2 steps.

First step is manual median signal shapes are calculated for the first rest period, these are evaluated manually if they represent PPG signals, good example in Fig. 10, bad signal example in Fig. 11.

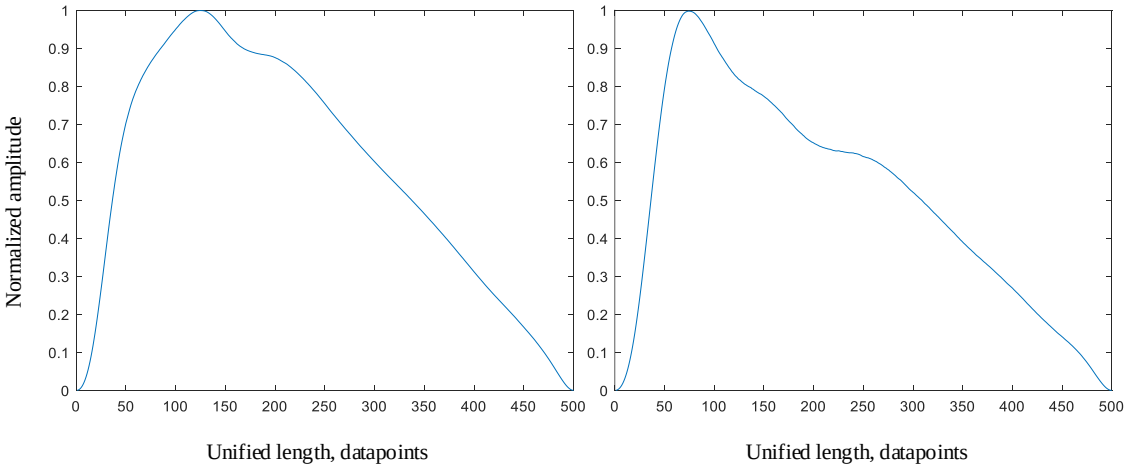

Fig. 10. Examples of good PPG signal shapes

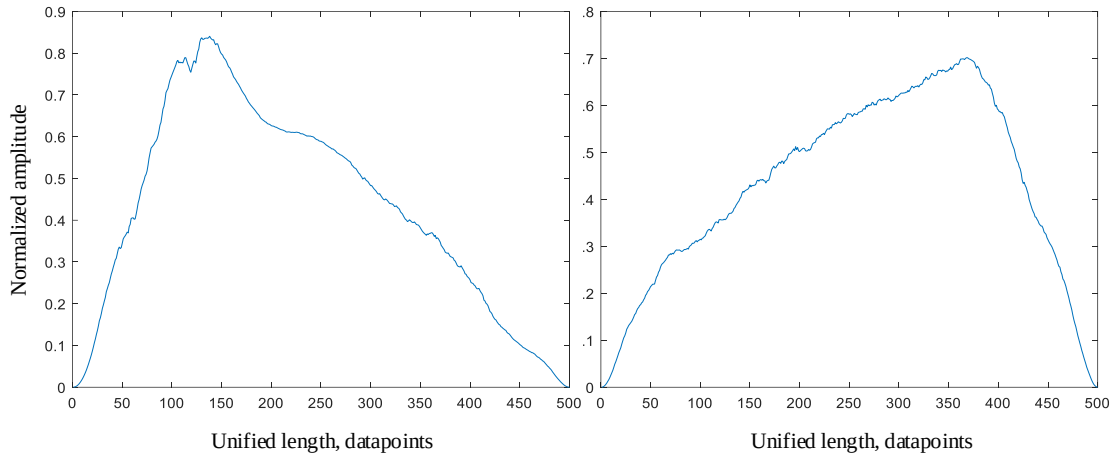

Fig. 11. PPG Finger median signals of first rest period which were rejected due to their median signal not having sufficient quality or not representing PPG signal

The second step is automatic, each segment is compared to previously generated median signal of the first rest period. Comparison is performed by calculating root square error for each datapoint in the segment and summing it into a single value. A threshold was set as 90% quantile of all segments of all signals left after step 1, which equals to 0.1327 for the tested database. Segments which have error value above the threshold were removed.

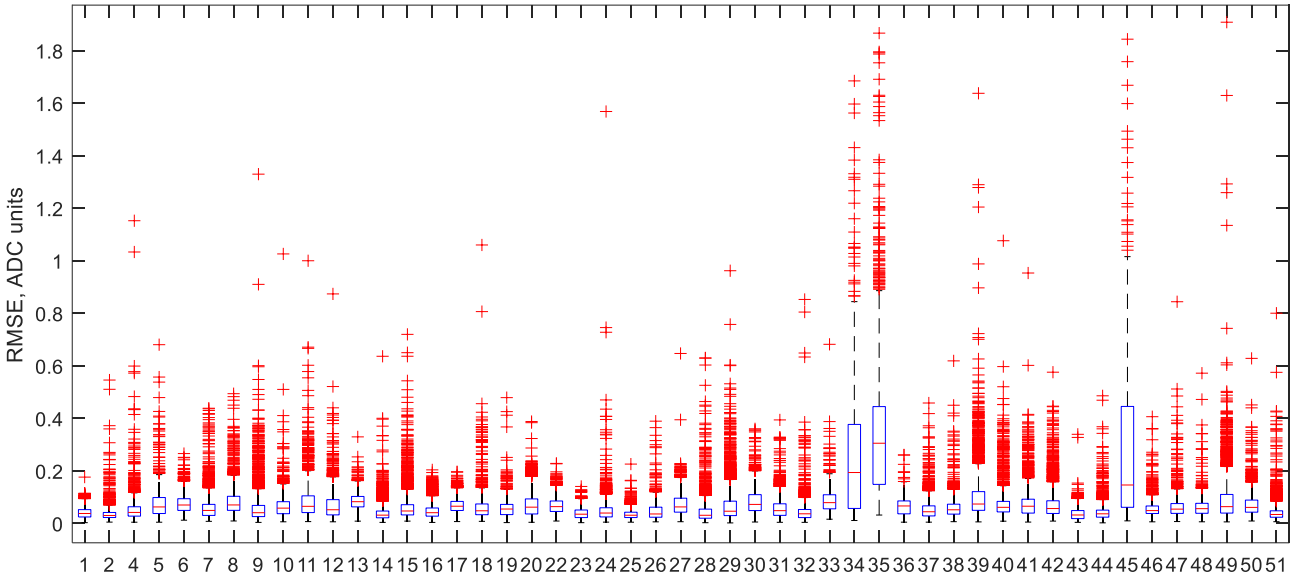

Fig. 12. RMSE values for each subject

2.3.3. Feature extraction

For PPG, VPG and APG parameters are defined as in [64]. Extraction is based on zero crossing in signals derivative. Additionally JPG (jerk plethysmogram), SPG(snap plethysmogram), CPG(crackle plethysmogram) are also used. These higher derivatives have not been investigated due to them being out of scope for most fields, however as observed in the collected data, they still have sufficient signal-noise level for feature extraction and as such could provide some new parameters. JPG and SPG parameters are extracted in the same way as PPG, VPG and APG. While SPG is only used for zero crossing detection for JPG intensity features. CPG does not have sufficient signal to noise level and as such will be rejected.

VPG, APG, JPG signals have a high enough signal to noise ratio after the initial filtering while JPG, SPG, CPG where additionally filtered with a low-pass filter (Butterworth, order of 2, 25hz cut-off, assuming 500hz sampling ratio after signal length unification), using zero-phase filtering algorithm.
Judicial points are extracted from derivates by detecting zero crossings. To have equal numbers of features minimum number of fiducial points for entire database have been calculated. This results in not all fiducial points being used, but the methods are applicable to a larger part of population.
Another distinction in features are intensity and time features. Intensity coded features are problematic for PPG signal, this is due to non-uniform signal levels for applications of the device. Currently used database has a continuous signal and as such amplitude features will be used, further deployment would require hardware or software solutions to uniform signal acquisition. Temporal features are amplitude independent and as such are a lot more resilient and more applicable for PPG.
In my research I will use both features. Usage of both features will also reveal which feature set is more useful and if amplitude features are required for pain classification. Zero crossings will be considered as temporal features.
Features are then additionally normalized. It is assumed that a previous 2-minute recording without pain would be available and as such its feature median is considered as 1. Sign is additionally corrected to have all features in positive range, this is needed for later machine learning application.

Table 10. Signal and derivatives properties and preprocessing

Signal	**Minimum number of fiducial points**	**Number of features**	**Filtering**
PPG(plethysmogram):	1	2 area+ 1 amplitude + 2 angle = 5 Amplitude features	LP (10Hz) +HP (0.5Hz)
VPG(velocity plethysmogram):	4	5 area + 4 amplitude + 5 angle = 14 Amplitude features 1 time feature	LP (10Hz) +HP (0.5Hz)
APG(acceleration plethysmogram):	4	5 area + 4 amplitude + 5 angle = 14 Amplitude features 4 time feature	LP (10Hz) +HP (0.5Hz)
JPG(jerk plethysmogram):	6	7 area + 6 amplitude + 7 angle = 20 Amplitude features 4 time feature	LP (10Hz) +HP (0.5Hz) + LP (25Hz)
SPG(snap plethysmogram):	52	6 time feature	LP (10Hz) + HP(0.5Hz) + LP (25Hz)
CPG(crackle plethysmogram):	144		LP (10Hz) +HP (0.5Hz) + LP (25Hz)

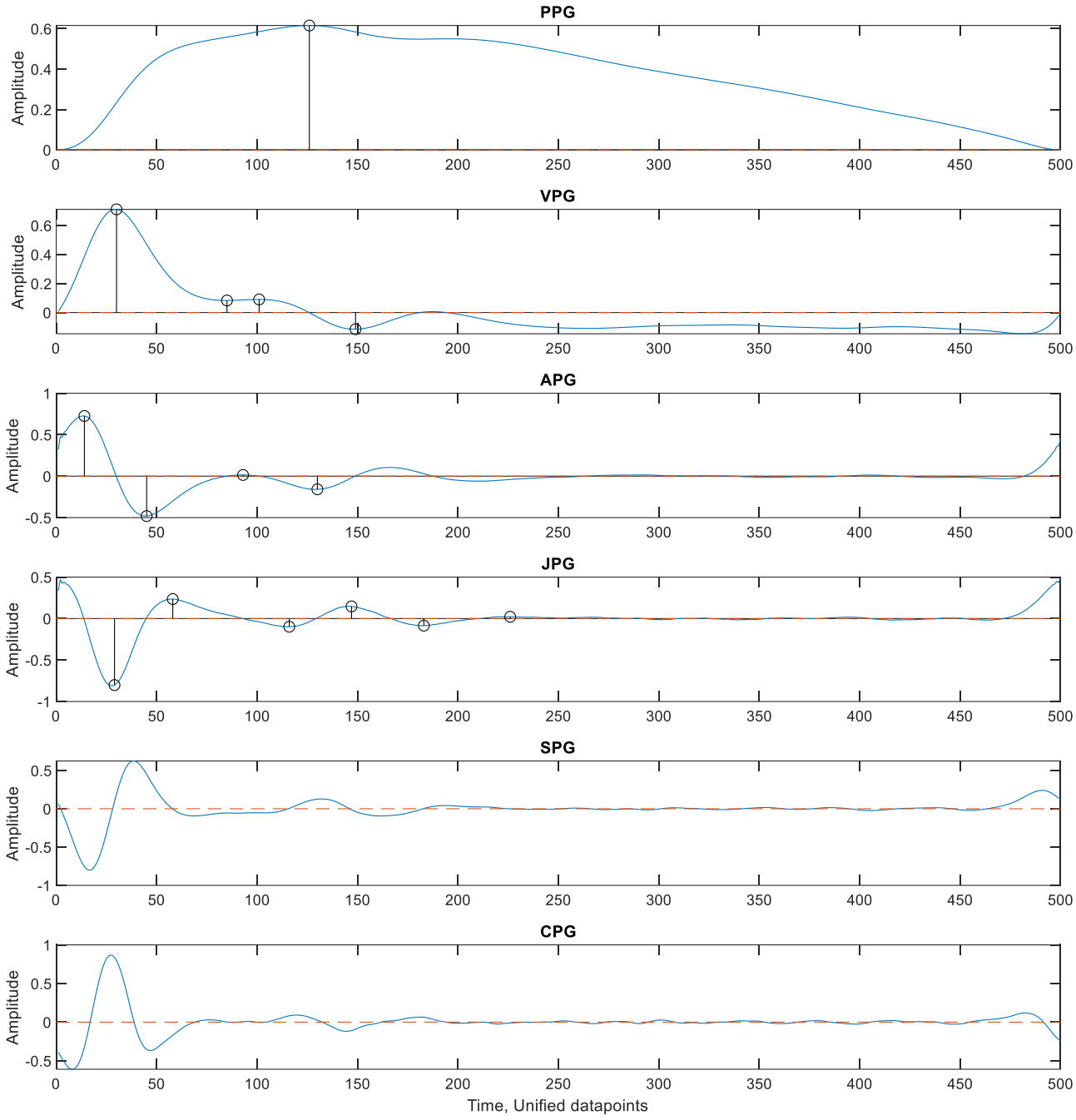

Fig. 13. Detected fiducial points which were later used for feature extraction

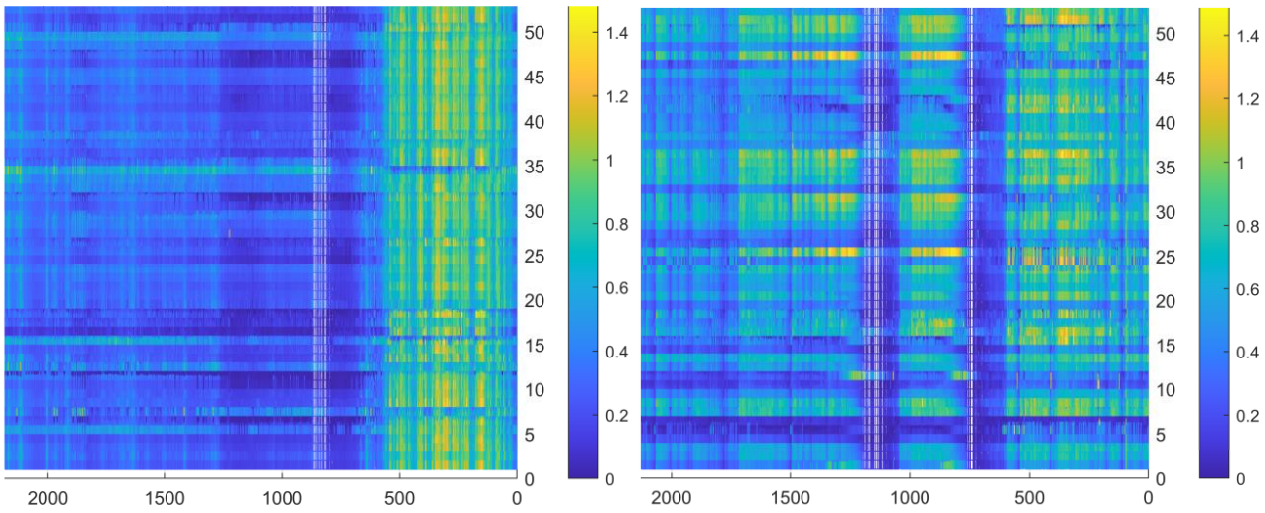

Fig. 14. Intensity feature maps of volunteers 6 and 7, empty zones – bad quality data. Volunteer 7 has very clear areas where most feature have lower amplitude, these areas match the time with cold period. X axis- time, y axis- feature no. and color-value

2.3.4. Training and testing datasets

Datasets are split by age and gender groups, by selecting 1 person from each age (20-29, 30-39, 40-49, 50-65)/gender (male, female) group for testing and validation.
This results in validation and testing dataset of 8 people. Training dataset contains 41 people. 2 people were removed due to insufficient signal quality.
Pain values are reduced from 0-100 range to 4 categories, this is based on clinical practice [65–68].Which are 0- no pain, 1 – slight pain, non-opioids, 2- mild pain -weak opioids ,3 – strong pain, strong opioids. Cut-off points and ranges seen in Table 11, cut off points were chosen based on previous research[66, 67].

Table 11. Pain classes and their respective ranges in NPRS scales

Pain class	Range
0	0-9
1	10-49
2	50-69
3	70-100

Training dataset was additionally oversampled by duplicating blocks of signals with pain classes of 1, 2 and 3. This oversampling was used in order to balance the datasets, which helps with model training. Real quantities and quantities in equalized datasets can be seen in Table 12, Table 13.

Table 12. Single-beat training dataset size table

Pain class	Real Quantity	Quantity in trinary oversampled dataset	Quantity in binary oversampled dataset
0	73171		73171
1	7444	7443	76063
2	2149	6444	
3	1294	7764	

Table 13. Single-beat testing dataset size table

Pain class	Real Quantity
0	17000
1	1562
2	505
3	366

Additionally, 5 and 12 beat sets were created. 5 beat set was created to test if multiple beats provide additional information for the neural networks. 12 beat set was created to later average it into single beat and eliminate respiratory component. Multi-beat datasets were created in the same way with addition that all consecutive beats have to pass quality control, the pain value is assigned from the last beat. Multi-beat dataset quantities can be seen in Table 14, Table 15, Table 16, Table 17.

Table 14. 5-beat training dataset size table

Pain class	Real Quantity	Quantity in trinary oversampled dataset	Quantity in binary oversampled dataset
0	60031		60031
1	5105	5104	57985
2	1396	5580	
3	708	4956	

Table 15. 5-beat testing dataset size table

Pain class	Real Quantity
0	16075
1	1467
2	471
3	371

Table 16. 12-beat training dataset size table

Pain class	Real Quantity	Quantity trinary in oversampled dataset	Quantity binary in oversampled dataset
0	51459		51459
1	3410	3409	51964
2	953	3808	
3	393	3537	

Table 17. 12-beat testing dataset size table

Pain class	Real Quantity
0	15223
1	1254
2	408
3	364

2.4. Neural networks

Neural networks were used for pain classification. Two types of input data were tested – signals and features. Feature approach is more traditional and is expected to give better results, however, raw data approach requires less computational power and, as such, is of interest in future applications. Additionally, signals may have additional information which is not available in the selected features. Tree different neural network architectures were tested – Multi Layer Perceptron (MLP) as a baseline indicating accuracy of a simple network, Long-Short Term Memory (LSTM) as an indicator if recurrent neural networks have an advantage in pain classification, Convolutional Neural Networks (CNN) as an indicator if convolution has advantage in pain classification.

MLP neural network was tested as a baseline to see if a simple network can extract the information and baseline accuracy. A simple MLP(Fig. 15) was created with 4 layers with Relu activation functions, layer sizes were changed according to which dataset was used, with the keyword multiplier indicating how many beat features there are – 1, 5, 12 and keyword input_length indicating if its feature or signal neural network and keyword n denoting whether it's a binary or trinary classificator A Softmax layer is used in the end to normalize the results for classification.

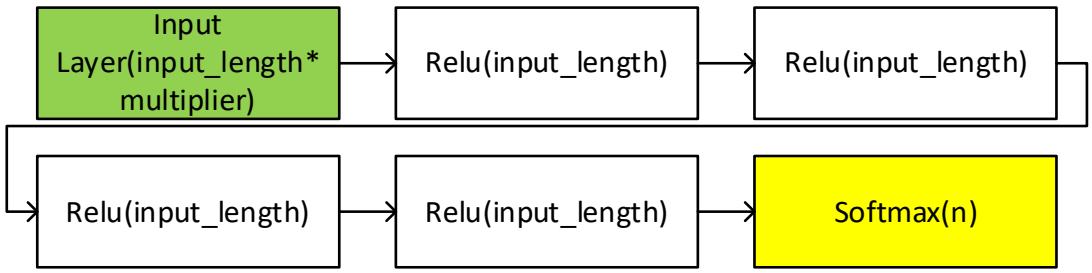

Fig. 15. MLP architecture, where multiplier denotes number of heartbeats(1, 5, 12) and n denotes number of classes(2 for first neural network and 3 for second)

The second tested architecture is based on convolutional neural networks, size is variable in the same way as MLP network with the keyword multiplier indicating how many beat features there are – 1, 5, 12 and keyword input_length indicating if its feature or signal neural network and keyword n denoting whether it is a binary or trinary classificator.
The CNN (seen in Fig. 16) is based on a 3 convolution-maxpooling layers for feature extraction and a 3 layer MLP network for decisions.
At last a LSTM neural network (Fig. 17) was tested to see if recurrent neural networks would have any advantages. For multi-beat datasets the beats were input as timesteps. LSTM length was also variable with multiplier corresponding to beat number in dataset – 1, 5, 12 and keyword input_length indicating if its feature or signal neural network and keyword n denoting whether it's a binary or trinary classificator.

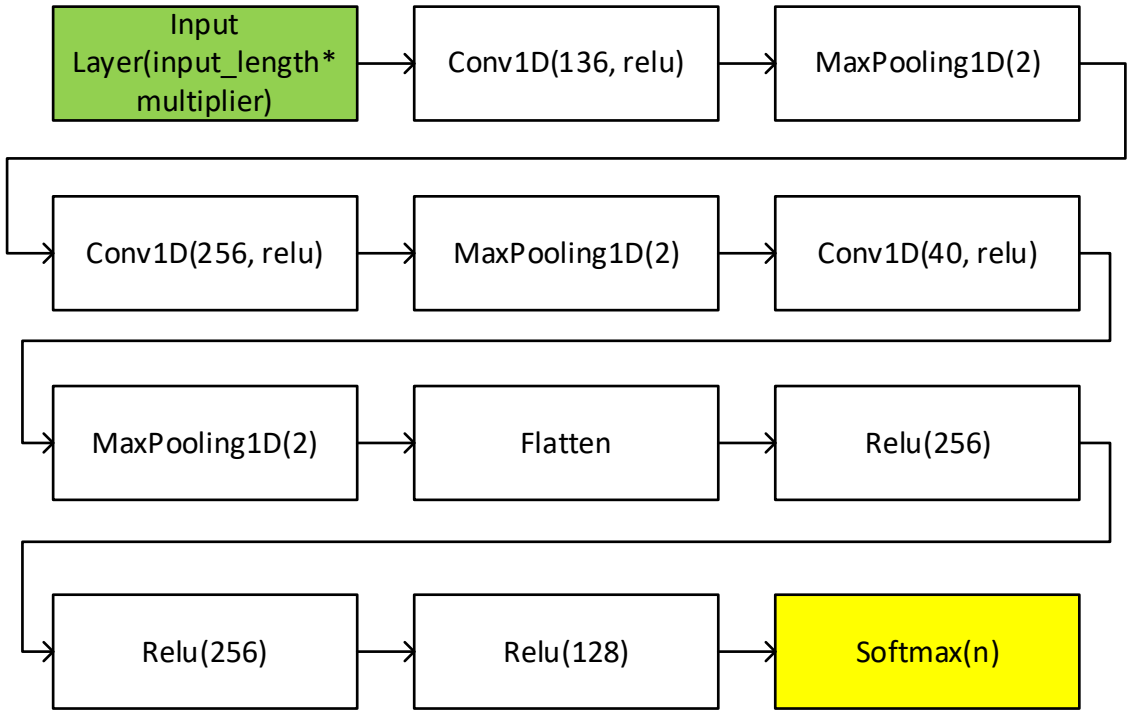

Fig. 16. CNN architecture

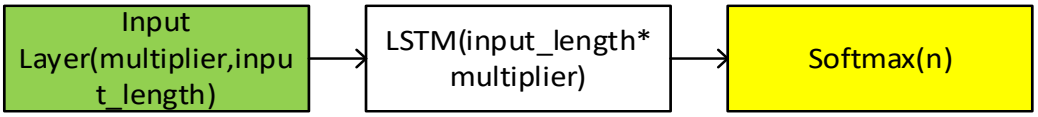

Fig. 17. LSTM architecture

2.5. Accuracy evaluation

Accuracy is evaluated separately for both classifiers by using testing datasets composed of signals with values suitable for the datasets. For binary classification pain classes from 1 to 3 are combined into a single group, locally named class 1, and tested in comparison to pain class 0.
While for trinary classification only pain classes 1 to 3 are used, while pain class 0 is excluded.

Neural networks were split in this way because third party created classifiers can be created later or based on the case an assumption can be made that the patient is in pain, but it is unknown how painful the experience is.
Binary classifier is analyzed with a standard confusion matrix and calculating its accuracy, sensitivity, specificity.
Accuracy is calculated according to formula (1), accuracy shows how much of the classes are predicted correctly.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

If: TP, TN, FP, FN can be seen in Table 18.
Sensitivity is calculated according to formula (2), sensitivity shows how much of the predicted class 0 is predicted correctly.

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

If: TP, FN can be seen in Table 18.

Specificity is calculated according to formula (3), specificity shows how much of the predicted class 1 is predicted correctly.

$$Specificity = \frac{TN}{TN+FP} \quad (3)$$

If: TP, FN can be seen in Table 18.

Trinary classifier is analyzed by forming a confusion matrix and calculating precision and recall.

Precision is calculated according to formula (4), precision shows show much of the class was correctly predicted

$$Precision = \frac{TP}{FN1+FN2+TP} \quad (4)$$

If: TP, FN1, FN2 for class 2 can be seen in Table 18.

Recall is calculated according to formula (5), recall shows how much of the predicted class 0 is predicted correctly.

$$Recall = \frac{TP}{FP1+FP2+TP} \quad (5)$$

If: TP, FP1, FP2 for class 2 can be seen in Table 18.

Table 18. Confusion matrices of binary and trinary classifiers. In trinary classifier variables are name for class 2 example calculations

Binary classification

	Predicted value	
	class 0	class
Real value class 0	TP	FP
class 1	FN	TN

Trinary classification

	Predicted value		
	class 1	class 2	class 3
Real Value class 1		FP1	
class 2	FN1	TP	FN2
class 3		FP2	

2.6. Personalization of the method

The issue with the currently proposed method is that it is partially personalized. Personalization is performed in the normalization part of the algorithm. It is assumed that previous recordings of 2 minutes in resting state would be available, the assumption is that if the method would be deployed in real life the resting state would be recorded during annual family doctor visits, the signals would then be uploaded to a patient database and associated with the persons personal ID number and could be pulled by the device if pain measurement is required.

3. Results

Result section is separated into 2 main categories – morphology and machine learning results.

Morphological changes are then separated into intensity and temporal changes and their corresponding features. Intensity changes refer to amplitude related changes and temporal changes refer to time related changes.

3.1. Analysis of signal morphology changes

The main changing component in PPG signals is amplitude. Main factors changing it are – outside light, fixing pressure, tissue composition and blood flow. All factors except blood flow are slow changing and as such are easily removable by a high pass filter and further removed by additional baseline removal. Below PPG signal and five of its derivatives are displayed. Median, 25% and 75% quantiles are displayed to show variations in signal level. From Fig. 18 it can be seen that PPG signal level highly decreases during CPT and partially recovers after it. Partial recovery could be associated with limited time available.

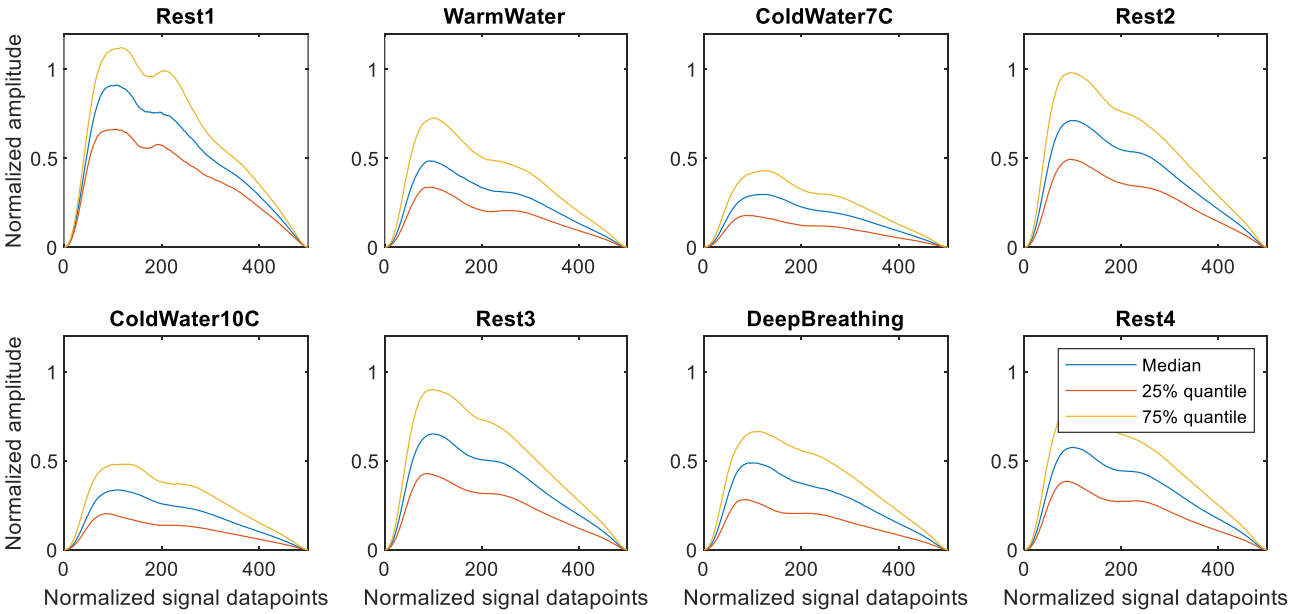

Fig. 18. Signal level changes in PPG signal of all volunteers combined

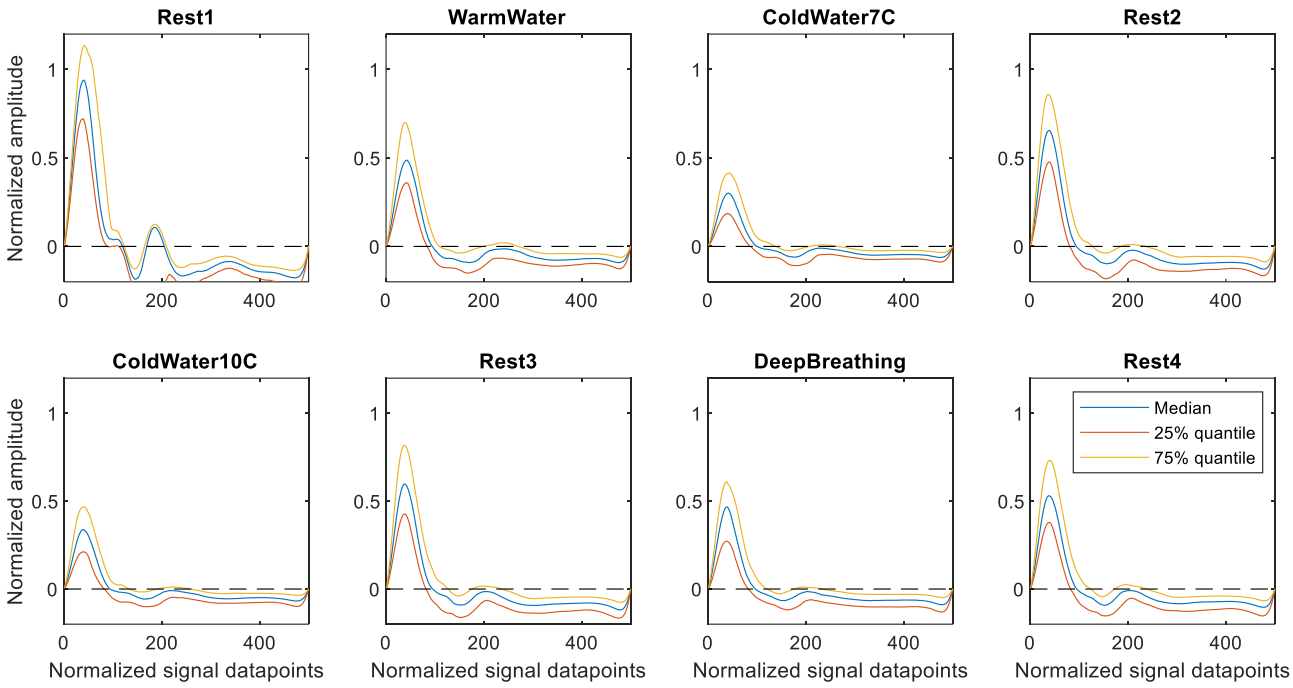

Fig. 19. Signal level changes in first derivative of PPG signal of all volunteers combined

Fig. 19 shows the first derivative of PPG – velocity plethysmogram, in all phases big changes from Rest1 phase can be seen, with the only exception being Rest4, which could be expected due to longer period from CPT to Rest4. The clearest difference can be seen in the second peak, in other phases it is suppressed and is under the 0 line.

Fig. 20 shows the second derivative signal, and again it shows similar tendency, however the distinction from the first derivative is faster recovery. During Rest2, Rest3, Rest4 a partial recovery is seen. However, stimulating phases are not only CPT, which could provide false positives if it is used for pain recognition.

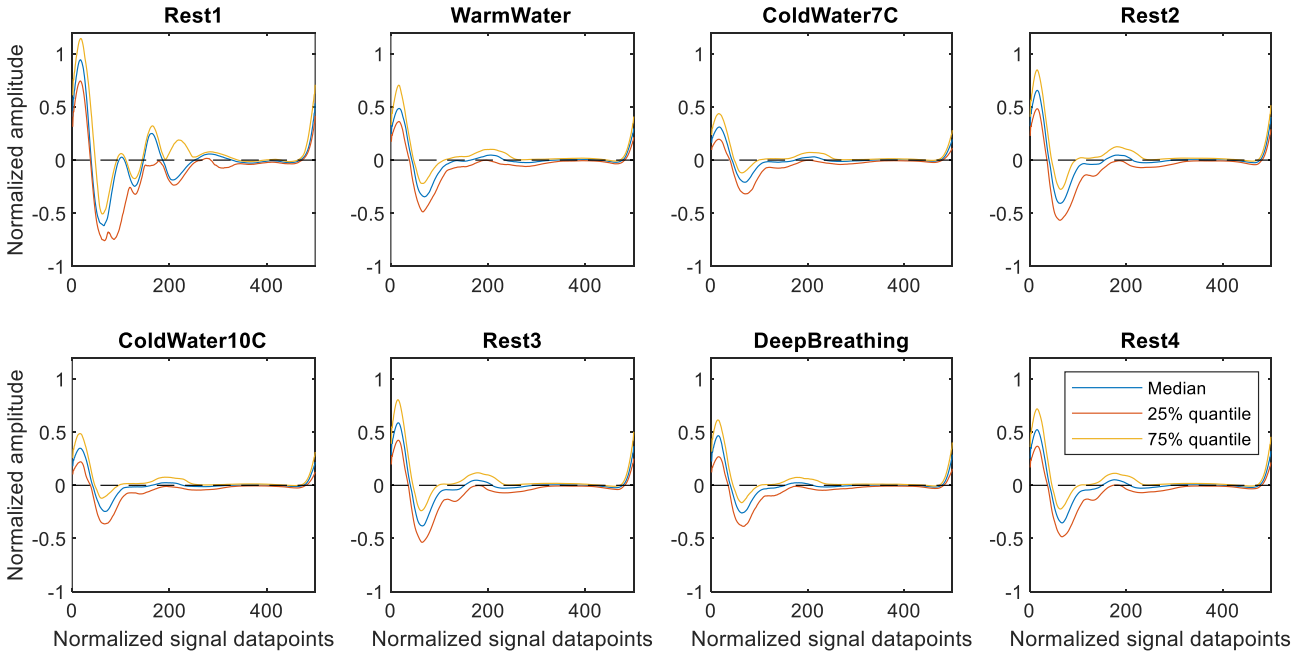

Fig. 20. Signal level changes in second derivative of PPG signal of all volunteers combined

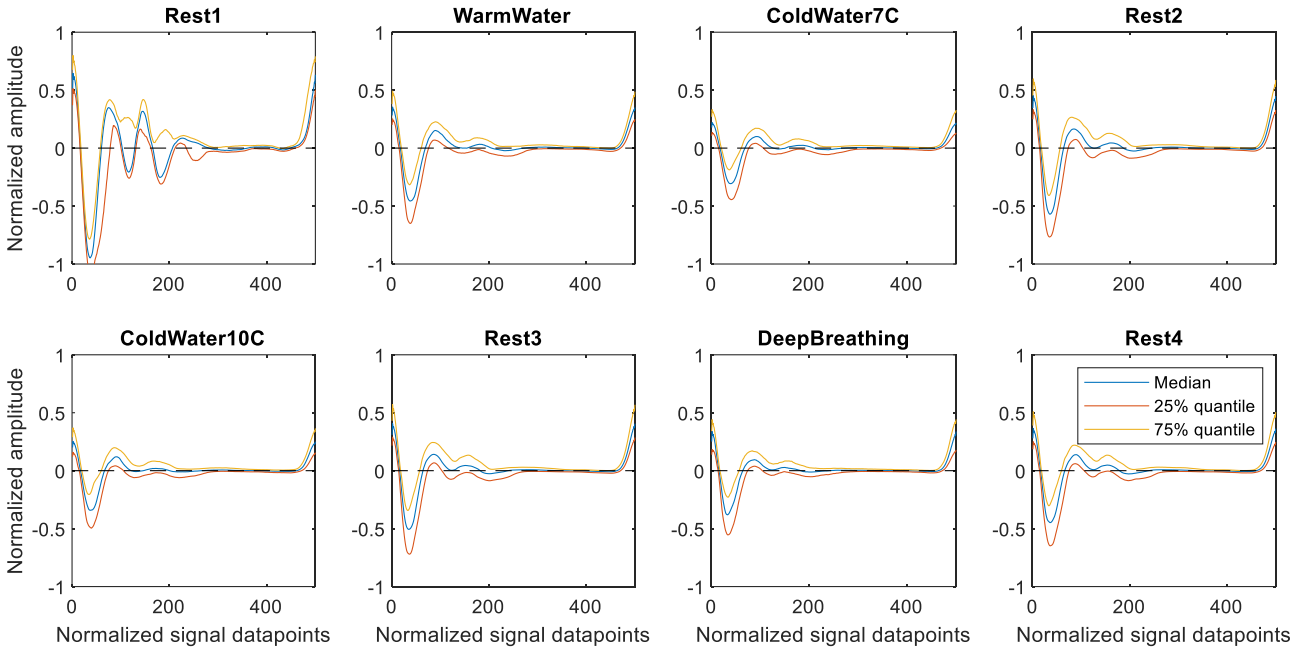

Fig. 21. Signal level changes in third derivative of PPG signal of all volunteers combined

Fig. 21 shows the third derivative. The third derivative behaves similarly to the second, however, in terms of zero crossings the third derivative stays above zero in second and third peak, while the second derivatives second peak dips below zero.

Fig. 22 shows the fourth derivative. The fourth derivative behaves similarly to the third, with the difference mostly seen in the third peak. In the fourth derivative the third peak is just reaching the zero line for the median signal in signals with stimulation and is starting to recover in Rest4, the recovery can be seen in 75% quantile signal.

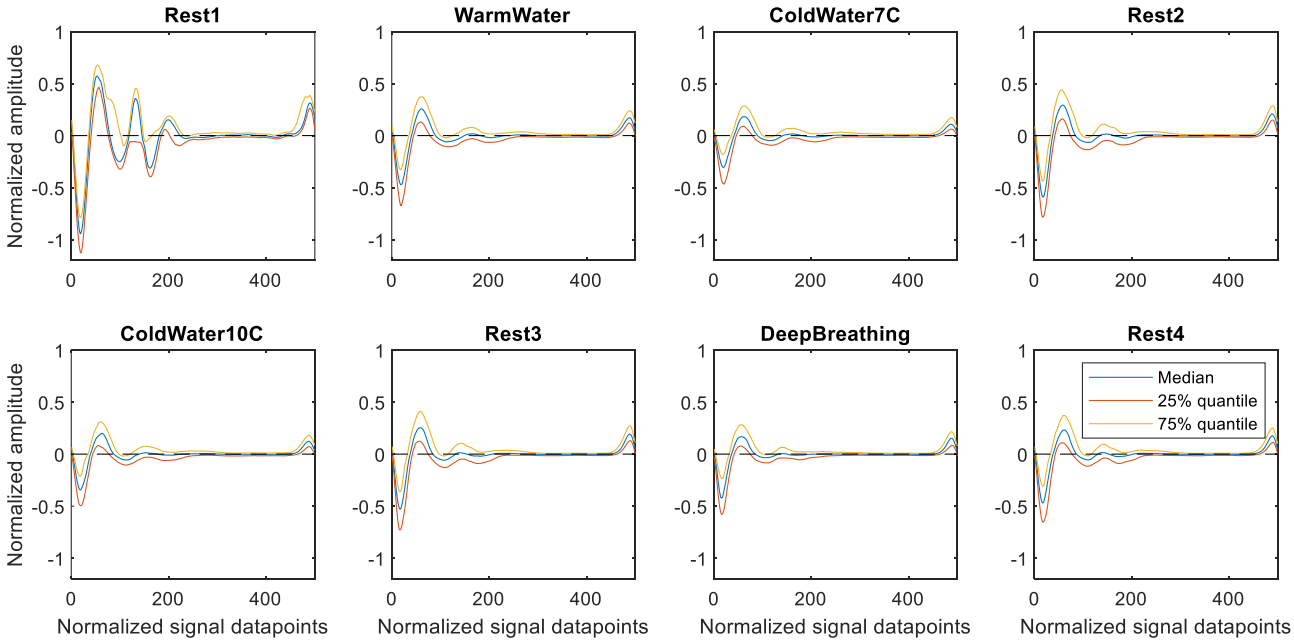

Fig. 22. Signal level changes in fourth derivative of PPG signal of all volunteers combined

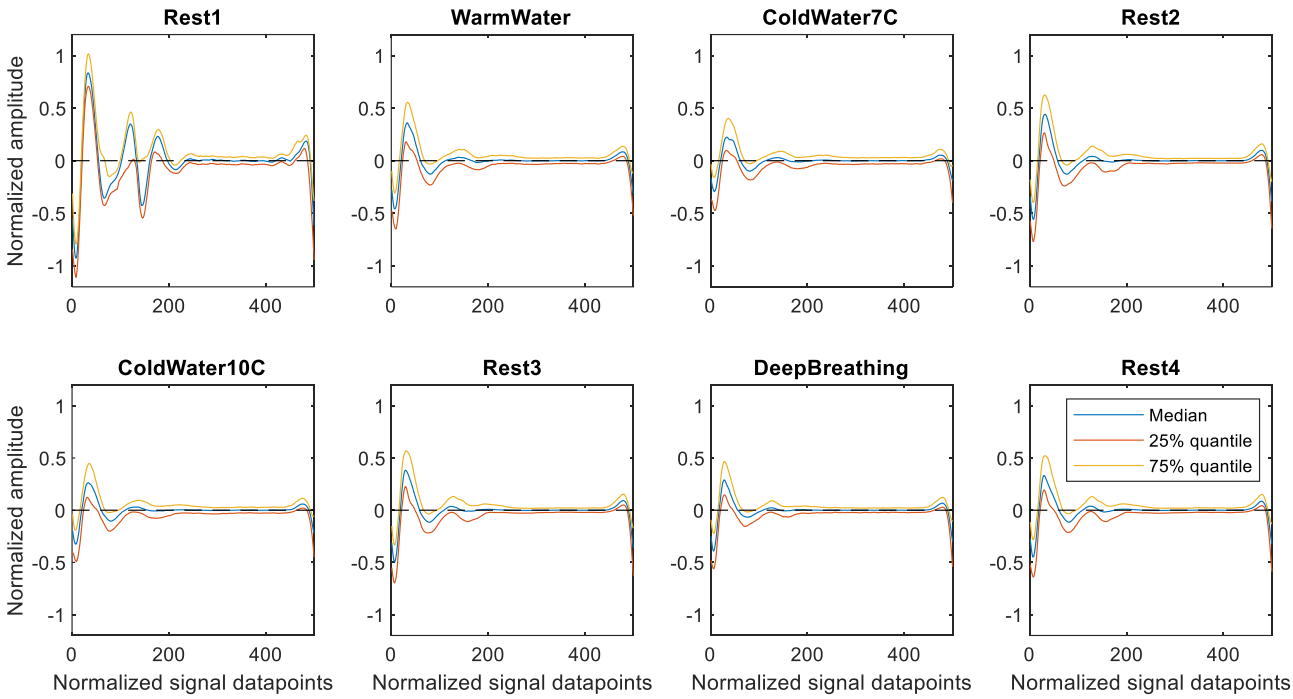

Fig. 23. Signal level changes in fifth derivative of PPG signal of all volunteers combined

Fig. 23 shows the fifth derivative. The fifth derivative behaves the most similarly to the fourth derivative, the second positive peak is balancing around the zero line and decreasing in stimulating phases.

In a short conclusion the overlook on derivative signal shows that the second, third and fifth derivatives react to all stimulating phases, not just CPT. In all derivatives the Rest2 and Rest3 periods are not long enough to fully recover, while the Rest 4 shows partial recover. The main issue is that signals tend to react to all stimulating phases, not only CPT and as such it can be hard to differentiate for pain recognition algorithms.

3.2. Analysis of feature changes

The signals are then further broken down into area (seen in Fig. 24), angle (seen in Fig. 26), amplitude (seen in Fig. 28) and time features (seen in Fig. 30) features.

Area features can be seen in Fig. 24 and Fig. 25. Fig. 24 shows a collection of all volunteers and it shows population tendencies. While Fig. 25 is an example of an individual, showing that for some features, such as Area 5, do not follow the trend of the study population. This differences between population and individual shows that there are some additional differences. A prospect for future studies would be to find identifiers for these differences to separate population into groups with similar behavior of features.

Area features 5,6, 11 do not show any clear distinction between pain levels, while area 5 and 6 also does not show distinction with different pain classes.

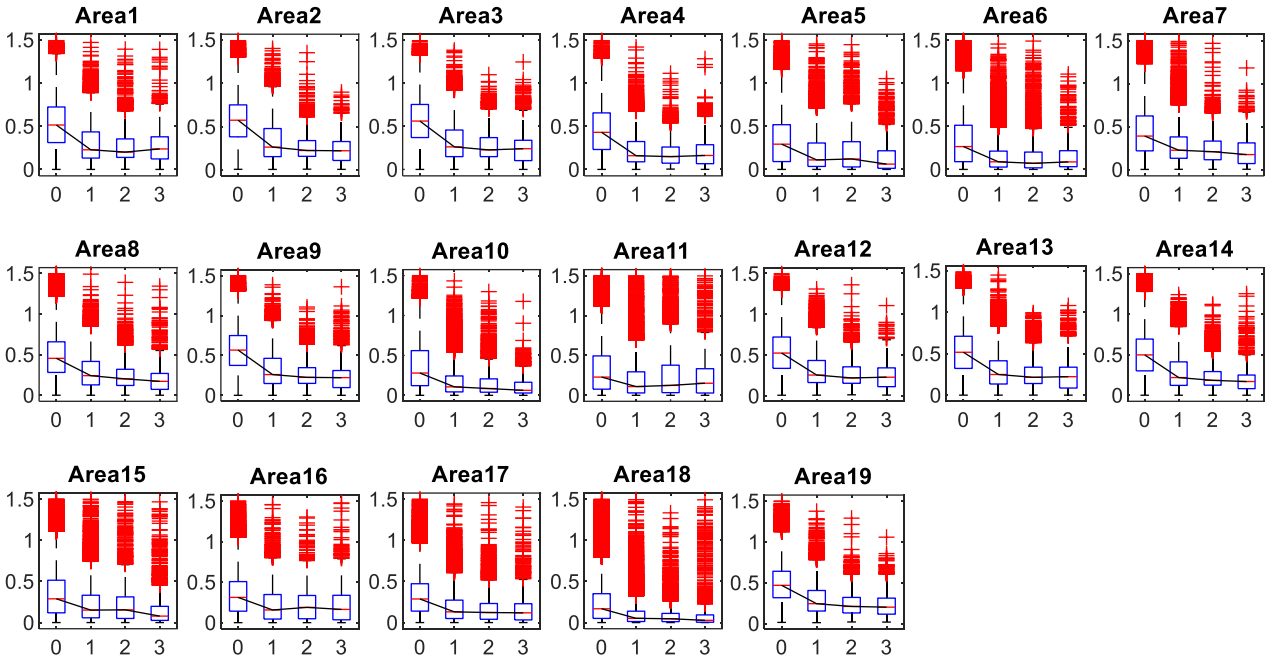

Fig. 24. Area features of all volunteers, y axis- value, x axis – pain level. Where Area 1-2 is from PPG signal, Area 3-8 is from VPG, Area 9-14 is from APG, Area 15-19 is from JPG

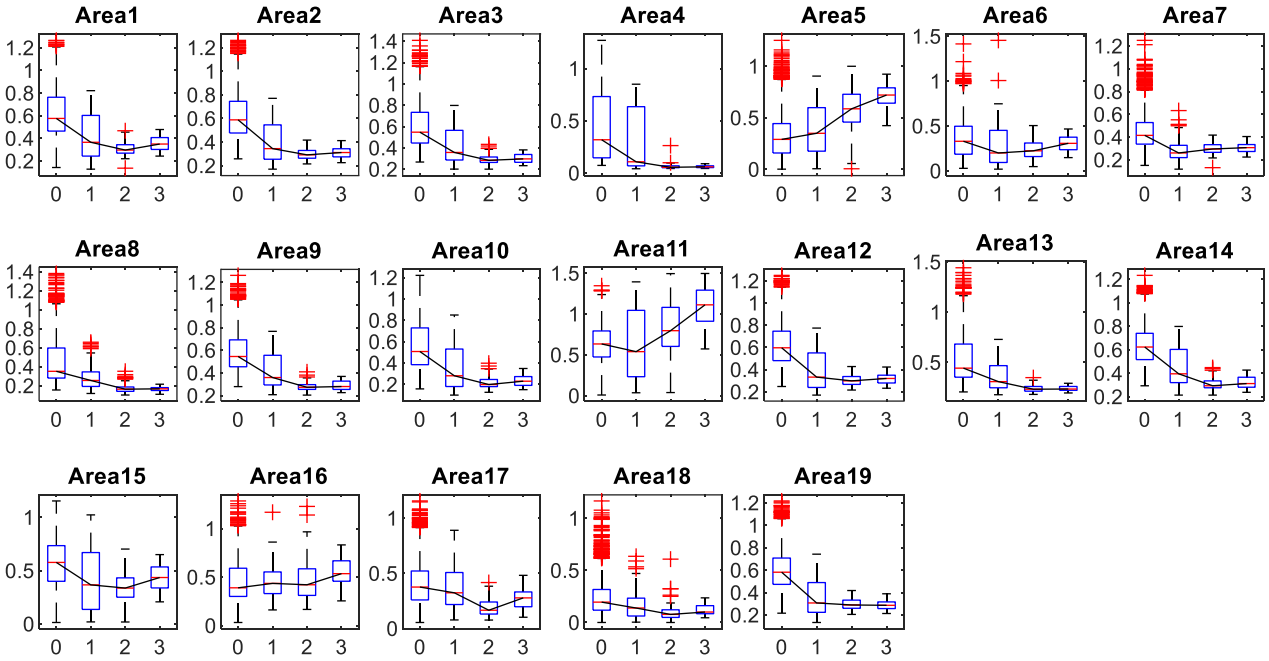

Fig. 25. Area features of volunteer no. 1, y axis- value, x axis – pain level. Where Area 1-2 is from PPG signal, Area 3-8 is from VPG, Area 9-14 is from APG, Area 15-19 is from JPG

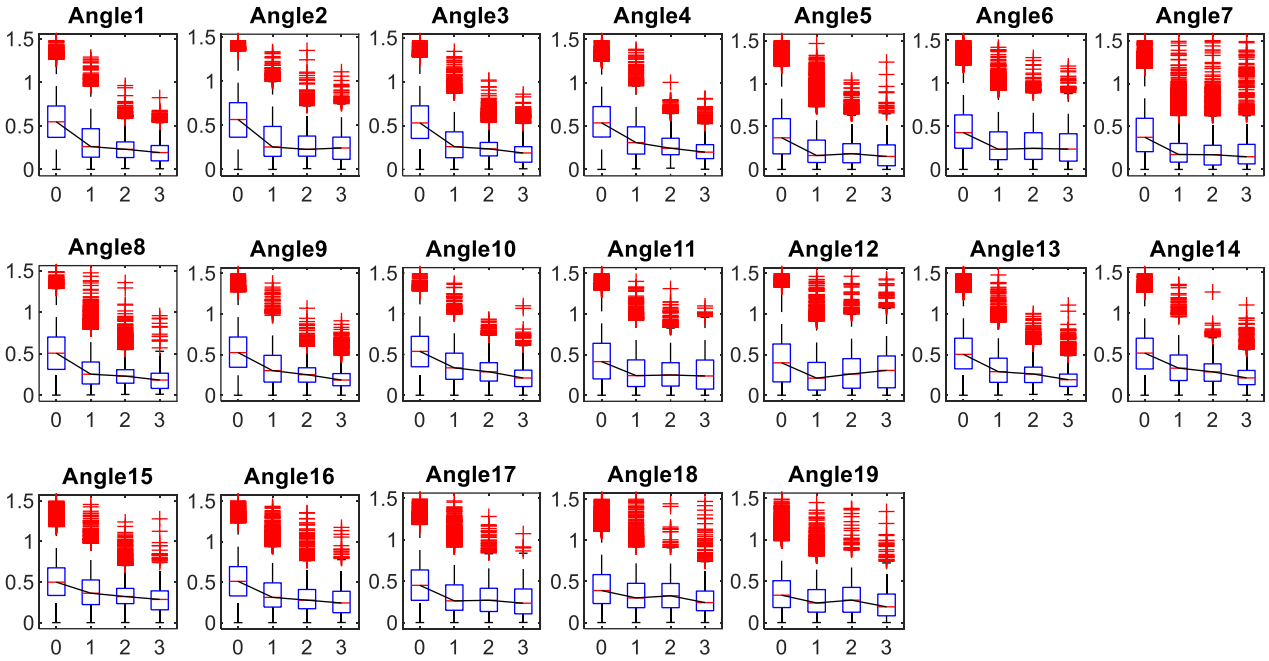

Fig. 26. Angle features of all volunteers, y axis- value, x axis – pain level. Where Angle 1-2 is from PPG signal, Angle 3-8 is from VPG, Angle 9-14 is from APG, Angle 15-19 is from JPG

Angle features are seen in Fig. 26 and Fig. 27. Angle features show the same differences as in area features, the single volunteer features are not always following study population trends. Examples of this can be seen in features Angle18 and Angle19.

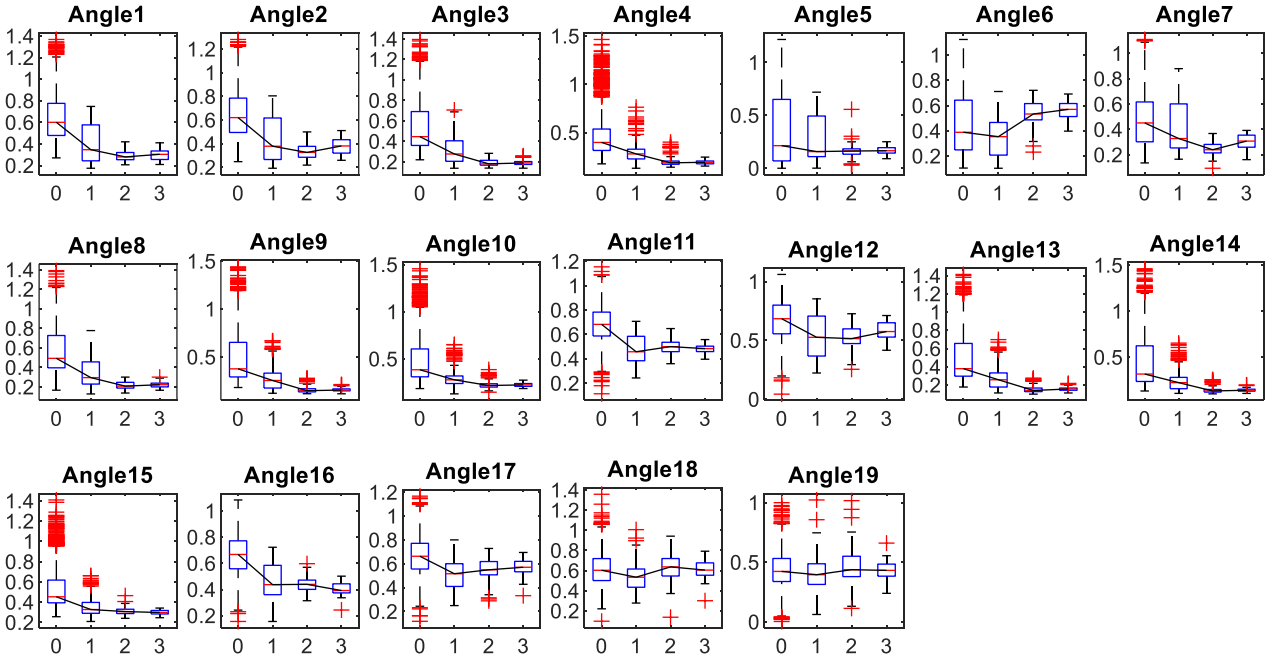

Fig. 27. Angle features of volunteer no. 1, y axis- value, x axis – pain level. Where Angle 1-2 is from PPG signal, Angle 3-8 is from VPG, Angle 9-14 is from APG, Angle 15-19 is from JPG

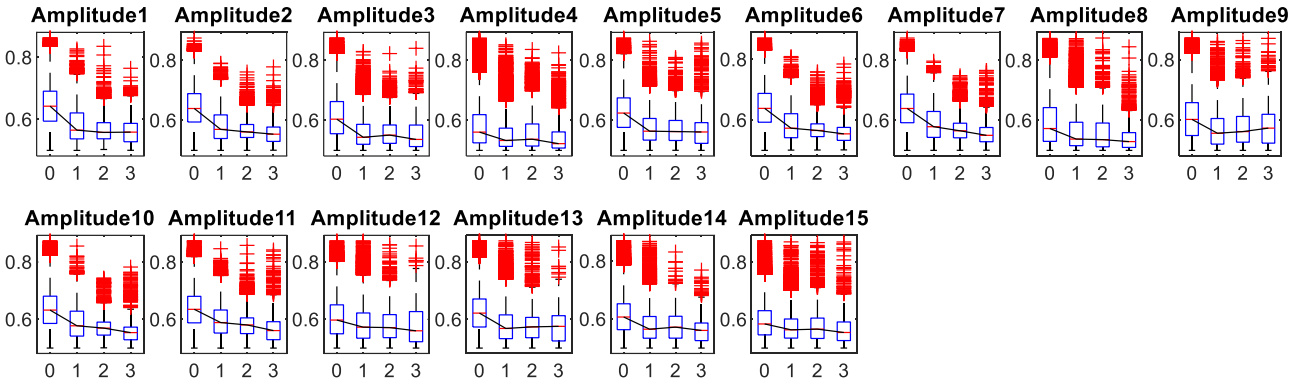

Fig. 28. Amplitude features of all volunteers, y axis- value, x axis – pain level. Where Amplitude 1 is from PPG, Amplitude 2-5 is from VPG, Amplitude 6-9 is from APG, Amplitude 9 -15 is from JPG

Amplitude features are seen in Fig. 28 and Fig. 29. Amplitude features show the most difference between individual and study population. Amplitude 3, 4, 7, 8, 9, 12, 13, 14, 15 features do not follow study population feature trends. While features like Amplitude 1, 2, 7, 10, 11 follow study population trends

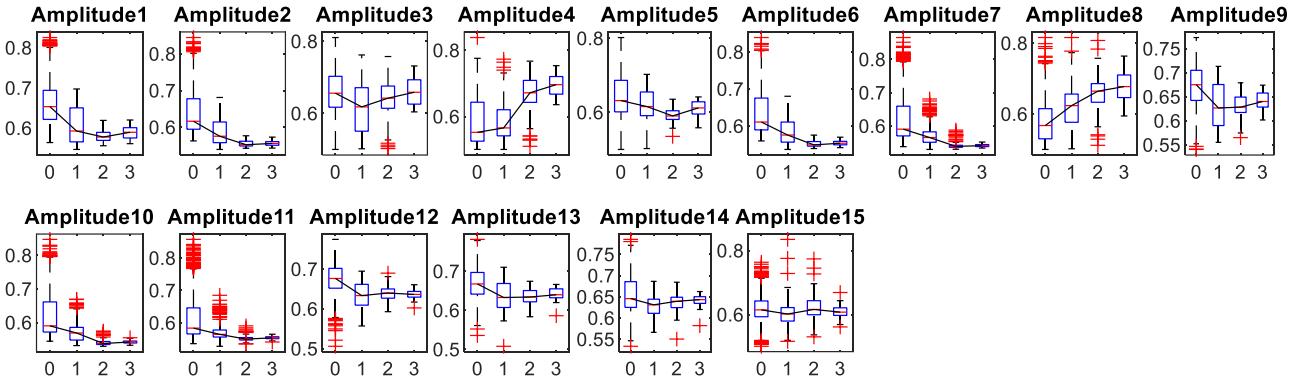

Fig. 29. Amplitude features of volunteer no. 1, y axis- value, x axis – pain level. Where Amplitude 1 is from PPG, Amplitude 2-5 is from VPG, Amplitude 6-9 is from APG, Amplitude 9 -15 is from JPG

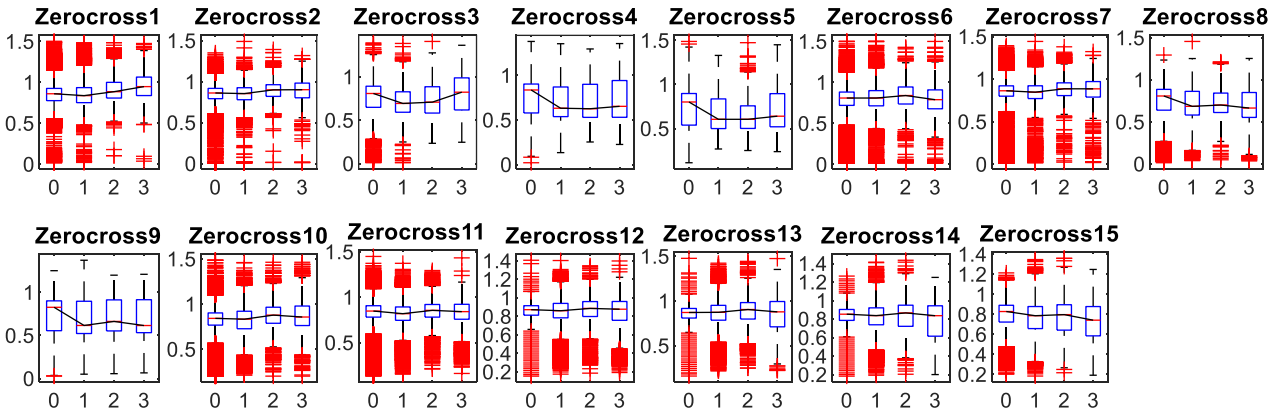

Fig. 30. Time based features- zero crossing location in the pulse, of all volunteers, y axis- value, x axis – pain level. Where Zerocross1 is from VPG, Zerocross 2-5 are from APG, Zerocross 6-9 are from JPG and Zerocross 10-15 are from SPG

Temporal features consist of signal zero crossing locations. Temporal features can be seen in Fig. 30 and Fig. 32. Just like in intensity features, differences between individual and population can be seen. We also see a very large number of outliers on both sides in Fig. 30. This indicates that the population is likely split into 2 groups- with ascending and descending feature sets. Zerocross 3, 4,

5 features in Fig. 32 show another interesting behavior, in pain classes 2 and 3 the variability is highly reduced compared to classes of 0 and 1., however this could partially be caused by limited number of samples. Fig. 31 shows histogram of Zerocross9 feature, the figure clearly shows a bimodal distribution, based on box plot figures it can expected that similar bimodal distributions are present in other zerocross features.

Zerocross 10-15 are from JPG.

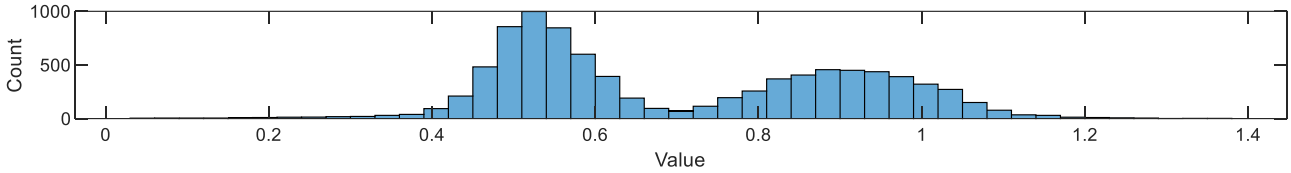

Fig. 31. Histogram of all volunteer Zerocross9 feature – showing bimodal distribution

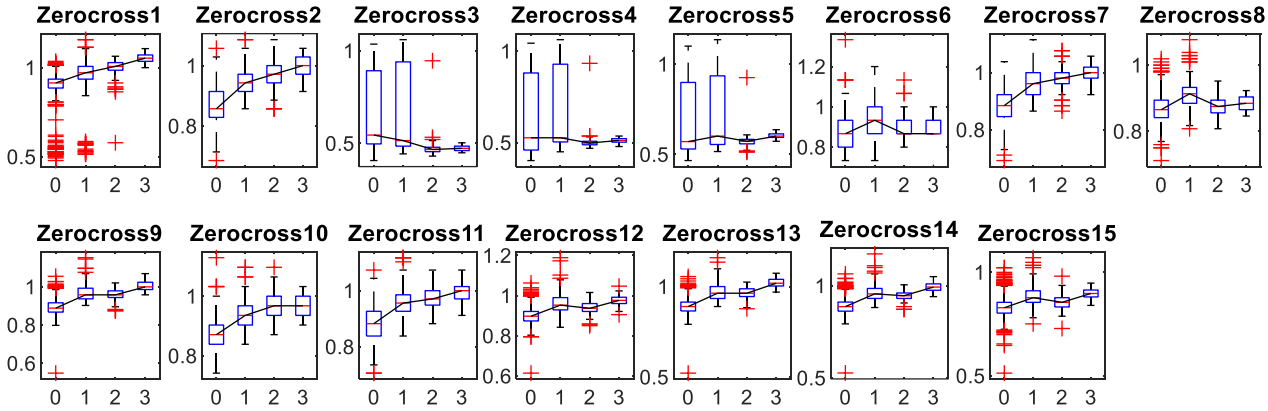

Fig. 32. Time based features- zero crossing location in the pulse, of volunteer no. 1, y axis- value, x axis – pain level. Where Zerocross1 is from VPG, Zerocross 2-5 are from APG, Zerocross 6-9 are from JPG and Zerocross 10-15 are from SPG

3.3. Binary classification results

Table 19 shows results of feature based binary classifier training. In all cases the accuracy of training dataset reached above 0.89, with some reaching 1. Those that did not reach 1, stabilized around their values. For training data the accuracy measurement is sufficient due to oversampled datasets. Additionally, sensitivity and specificity are also provided. We can see that the lowest prediction accuracy is 0.7822, however, metric of accuracy is not very suitable in this case due to imbalance in the dataset, as such measure of specificity is more suitable as it shows which models are better at assigning painful class to the painful signals. While overviewing specificity value of various models it can be seen that most models still have difficulty assigning painful signals to the painful class. While class 0 signals are mostly correctly assigned as being class 0(seen by sensitivity values).

In MLP models performed better when they had access to more heart, and even better when a sequence of heartbeats was averaged into a single heartbeat. LSTM models showed similar tendency with exception of 5 averaged heartbeat being an exception with reduced values, however it has increased sensitivity compared to non-averaged model. CNN models however showed different tendencies, with non-averaged multi-beat models being better than averaged of single beat model, however a more surprising part is that 5 beat model performed better than 12.

The best performer in terms of both accuracy and specificity is an MLP network using 12 averaged heartbeat signal features. This model achieved accuracy of 0.9 and specificity of 0.69. The models were not additionally tuned to perform in their best configuration and show only general tendencies in which methods are more suitable for application and on which it is better to focus in future research. In general, all models were good at picking out class 0 signals, but failed correctly assigning painful signals. This could be partially explained by the dataset imbalance, there were a lot more signals in class 0, which also allowed for the model to learn better.

Table 19. Feature based binary classifier results

S	Dataset set	Binary classification accuracy		Binary classification specificity	Binary classification sensitivity
		Training dataset	Testing dataset	Testing dataset	Testing dataset
MLP	Single heartbeat	0.90	0.78	0.32	0.95
	5 heartbeats	1.00	0.81	0.34	0.93
	12 heartbeats	1.00	0.83	0.37	0.94
	5 heartbeats averaged	0.99	0.84	0.39	0.93
	12 heartbeats averaged	0.96	0.92	0.69	0.94
LSTM	Single heartbeat	0.91	0.85	0.42	0.95
	5 heartbeats	1.00	0.85	0.41	0.92
	12 heartbeats	1.00	0.85	0.40	0.93
LSTM	5 heartbeats averaged	0.97	0.79	0.31	0.93
	12 heartbeats averaged	0.99	0.89	0.54	0.94
CNN	Single heartbeat	1.00	0.84	0.38	0.92
	5 heartbeats	1.00	0.88	0.52	0.93
	12 heartbeats	1.00	0.86	0.42	0.93
	5 heartbeats averaged	1.00	0.86	0.43	0.92
	12 heartbeats averaged	1.00	0.87	0.43	0.92

Table 20 shows results of signals based binary classifier training. In most cases the accuracy was above 0.89 but some failed earlier, with examples being MLP 5 and 12 heartbeat networks with only 0.69 and 0.74. For training data the accuracy measurement is sufficient due to oversampled datasets.

Contrary to feature based classifiers in signal based classifiers averaged heartbeats performed worse than non averaged. The best networks in terms of accuracy are all single heartbeat, with MLP model performing the best at 0.85, following it is LSTM and CNN models. CNN model has an interesting exception in that 12 heartbeat non-averaged model performed the best, in accuracy but suffered in specificity.

Overall feature based methods have higher accuracy, specificity and sensitivity, indicating that feature extraction and application in binary classification is beneficial. Future improvements are possible in better feature extraction, more research in possible features, model tuning and approach using decision tree type models instead of NN.

Table 20. Raw data based binary classifier results

Method	Dataset set	Binary classification accuracy		Binary classification specificity	Binary classification sensitivity
		Training dataset	Testing dataset	Testing dataset	Testing dataset
MLP	Single heartbeat	0.99	0.85	0.43	0.93
	5 heartbeats	0.69	0.77	0.33	0.97
	12 heartbeats	0.74	0.64	0.23	0.97
	5 heartbeats averaged	0.65	0.61	0.23	0.97
	12 heartbeats averaged	0.98	0.82	0.27	0.91
LSTM	Single heartbeat	0.78	0.84	0.43	0.98
	5 heartbeats	0.98	0.82	0.33	0.92
	5 heartbeats averaged	0.90	0.61	0.17	0.90
	12 heartbeats averaged	0.96	0.68	0.17	0.91
CNN	Single heartbeat	0.89	0.83	0.39	0.95
	5 heartbeats	1.00	0.80	0.26	0.90
	12 heartbeats	1.00	0.83	0.31	0.91
	5 heartbeats averaged	0.99	0.80	0.27	0.90
	12 heartbeats averaged	0.99	0.82	0.30	0.92

Both feature and signal based binary classifiers were tested. In both cases it was observed that for training datasets (made of 41 people) in feature based models the minimum accuracy was 0.89 while in signal based models the minimum accuracy was much lower at 0.65, it seems that some of the models reach their capability ceiling much earlier and as such the simple architecture is not enough. Additionally, the CNN models seem the most consistent as they all have good accuracies (above 0.98 with a single model exception) in both feature based and signal based types. While in the testing dataset(made of 8 people) accuracy is much lower. This shows that people on which the models were trained do not completely match the people it was tested on, and as such a larger dataset is needed for successful application in real world environment. 12 heartbeat LSTM network was eliminated due to excessive training time (3 days of continuous calculation).

Fig. 33 shows result of single heartbeat CNN model being applied to a signal of a volunteer from the training group. It can be seen that the network correctly predicted the range in which the volunteer is in pain, with the exception of the end of painful periods.
Fig. 34 shows the same model applied to a signal from a volunteer in the test dataset. The performance is much worse, with many false positives. However, it is also seen that most of the

false positives arise only after pain, this could indicate of a slower than noticeable physiological response. Additional post processing (10^{th} order moving average filter), like in Fig. 35 could also help with final prediction.

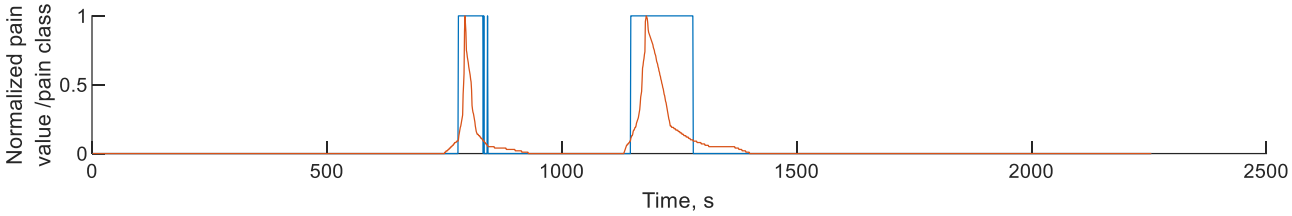

Fig. 33. Volunteer no. 4(training dataset) signal processed by single heartbeat feature binary CNN model. Blue- predicted painful period, Orange – subjective evaluation

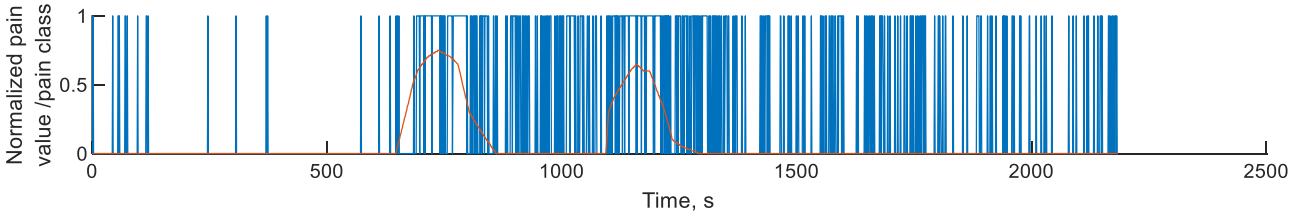

Fig. 34. Volunteer no. 1(testing dataset) signal processed by single heartbeat feature binary CNN model. Blue- predicted painful period, Orange – subjective evaluation

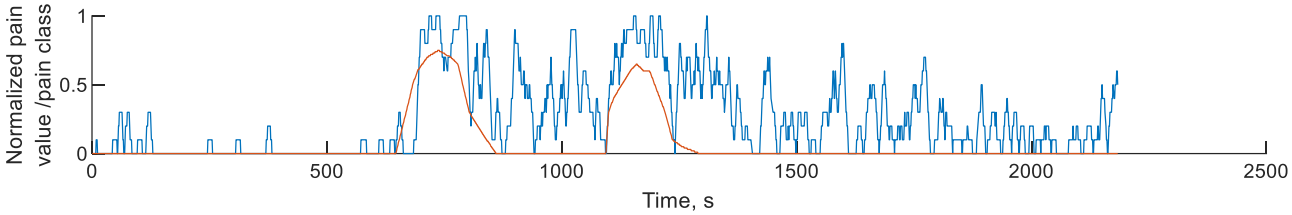

Fig. 35. Volunteer no. 1(testing dataset) signal processed by single heartbeat feature CNN model, additionally filtered with an averaging filter of order 10. Blue- predicted painful period, Orange – subjective evaluation

3.4. Trinary classification results

Comparing NN in trinary classification is difficult, especially with different dataset sizes and different effects of misclassification (assigning class 1 instead of 2 will have much smaller consequences compared to assigning class 1 instead of class 3). Comparison is performed using precision and recall metrics.

Table 21 shows precision values of all tested feature based NN. It can be seen from the accuracy column that the trained models reached the ceiling of their capability as they stabilized around the presented values, all models except 1 reached accuracy higher than 0.95. In the case of training dataset accuracy metric was suitable because training was performed with an oversampled dataset. We can see that in testing dataset class 1 has the highest precision, meaning that out of the performed predictions class 1 is the most assigned correctly, however classes 2 and 3 suffer with very low precision. This tendency is seen across all types of signals and all types of NN. An interesting example can be seen in 12 averaged heartbeat LSTM network, it has the highest precision in class 2, but fails completely in class 3, many other networks perform similarly, having good class 1 precision, but failing in classes 2 and 3. Class 3 shows especially bad results with no network correctly identifying it in the testing dataset.

Table 21. Feature based trinary classification precision

Method	Dataset set	Accuracy		Trinary classification precision, class 1	Trinary classification precision, class 2	Trinary classification precision, class 3
		Training dataset	Testing dataset	Testing dataset	Testing dataset	Testing dataset
MLP	Single heartbeat	0.95	0.56	0.79	0.22	0.04
	5 heartbeats	1.00	0.59	0.86	0.13	0.08
	12 heartbeats	1.00	0.58	0.80	0.41	0.00
	5 heartbeats averaged	0.99	0.61	0.84	0.33	0.08
	12 heartbeats averaged	0.99	0.56	0.80	0.32	0.00
LSTM	Single heartbeat	0.76	0.54	0.69	0.36	0.15
	5 heartbeats	1.00	0.54	0.82	0.09	0.00
	12 heartbeats	1.00	0.52	0.76	0.25	0.00
	5 heartbeats averaged	0.95	0.58	0.83	0.23	0.00
	12 heartbeats averaged	1.00	0.61	0.84	0.43	0.03
CNN	Single heartbeat	1.00	0.56	0.80	0.23	0.04
	5 heartbeats	1.00	0.60	0.94	0.00	0.08
	12 heartbeats	1.00	0.58	0.83	0.31	0.00
	5 heartbeats averaged	0.99	0.58	0.78	0.33	0.09
	12 heartbeats averaged	0.97	0.46	0.62	0.39	0.00

Table 22 shows recall values for feature based trinary classification. Recall shows how much of the class was assigned correctly. We can see that class 1 is has the highest rates of recall indicating that it is most assigned correctly, while classes 2 and 3 have very low values and even 0 in some cases, showing that these classifiers failed, and a different approach is required. The only exception being the previously mentioned 12 averaged heartbeat LSTM, it has a class 3 recall of 0.79, and not outstanding recall values in class 1(0.67) and class 2(0.40), but its class 3 precision is still very low (0.03). This shows that the network is skewed towards class 3 and is likely to falsely predict other class signals as class 3 signal.

Table 22. Feature based trinary classification recall

Method	Dataset set	Trinary classification recall, class 1	Trinary classification recall, class 2	Trinary classification recall, class 3
		Testing dataset	Testing dataset	Testing dataset
MLP	Single heartbeat	0.66	0.28	0.09
	5 heartbeats	0.68	0.23	0.16
	12 heartbeats	0.66	0.33	0.00

Table 22. Continuation

Method	Dataset set	Trinary classification recall, class 1	Trinary classification recall, class 2	Trinary classification recall, class 3
		Testing dataset	Testing dataset	Testing dataset
MLP	5 heartbeats averaged	0.69	0.40	0.21
	12 heartbeats averaged	0.66	0.29	0.00
LSTM	Single heartbeat	0.72	0.36	0.08
	5 heartbeats	0.61	0.16	0.15
	12 heartbeats	0.61	0.24	0.00
	5 heartbeats averaged	0.65	0.29	0.00
	12 heartbeats averaged	0.67	0.40	0.79
CNN	Single heartbeat	0.67	0.29	0.13
	5 heartbeats	0.63	0.03	0.01
	12 heartbeats	0.64	0.32	0.00
	5 heartbeats averaged	0.67	0.34	0.23
	12 heartbeats averaged	0.59	0.24	0.00

Table 23 shows precision values for signal based trinary classifiers. It can be seen from the accuracy column that the trained models reached the ceiling of their capability as they stabilized around the presented values, all models except 1 reached accuracy higher than 0.87. Overall in terms of training dataset the feature based models performed better. Signal based trinary classifier shows lower class 1 precision values compared to the feature-based ones. However, class 3 precision is higher than compared to featured based alternative, while class 2 shows just slightly higher precision, this shows that these networks are more skewed towards the higher 2 classes.

Table 23. Signal based trinary classification precision

Method	Dataset set	Accuracy		Trinary classification precision, class 1	Trinary classification precision, class 2	Trinary classification precision, class 3
		Training dataset	Testing dataset	Testing dataset	Testing dataset	Testing dataset
MLP	Single heartbeat	0.99	0.60	0.72	0.24	0.12
	5 heartbeats	0.66	0.60	0.43	0.38	0.30
	12 heartbeats	0.94	0.60	0.58	0.38	0.18
	5 heartbeats averaged	0.94	0.58	0.76	0.20	0.16
	12 heartbeats averaged	0.91	0.56	0.71	0.15	0.30

Table 23. Continuation

Method	Dataset set	Accuracy		Trinary classification precision, class 1	Trinary classification precision, class 2	Trinary classification precision, class 3
		Training dataset	Testing dataset	Testing dataset	Testing dataset	Testing dataset
LSTM	Single heartbeat	0.88	0.59	0.71	0.44	0.34
	5 heartbeats	0.95	0.56	0.74	0.25	0.20
	5 heartbeats averaged	0.89	0.51	0.70	0.17	0.18
	12 heartbeats averaged	0.91	0.51	0.55	0.54	0.36
CNN	Single heartbeat	0.98	0.53	0.81	0.28	0.16
	5 heartbeats	0.95	0.40	0.81	0.34	0.09
	12 heartbeats	0.96	0.47	0.78	0.48	0.11
	5 heartbeats averaged	0.96	0.55	0.78	0.58	0.09
	12 heartbeats averaged	0.96	0.52	0.72	0.40	0.17

Table 24 shows recall values of the signal based trinary classifier. Here the same results as in precision scores can be seen, the signal based classifiers show higher recall values indicating that the network is more evenly distributed in terms of class recognition performance.

The best performing model seems to be LSTM 12 averaged heartbeat based, it shows good results in both feature and signal based variants. However, it has difficulties with class 2.

Table 24. Signal based trinary classification recall

Method	Dataset set	Trinary classification recall, class 1	Trinary classification recall, class 2	Trinary classification recall, class 3
		Testing dataset	Testing dataset	Testing dataset
MLP	Single heartbeat	0.66	0.26	0.17
	5 heartbeats	0.76	0.17	0.28
	12 heartbeats	0.73	0.19	0.29
	5 heartbeats averaged	0.73	0.19	0.20
	12 heartbeats averaged	0.72	0.13	0.34
LSTM	Single heartbeat	0.77	0.35	0.33
	5 heartbeats	0.71	0.22	0.32
	5 heartbeats averaged	0.68	0.17	0.21
	12 heartbeats averaged	0.79	0.23	0.60

Table 24. Continuation

Method	Dataset set	Trinary classification recall, class 1	Trinary classification recall, class 2	Trinary classification recall, class 3
		Testing dataset	Testing dataset	Testing dataset
CNN	Single heartbeat	0.69	0.38	0.26
	5 heartbeats	0.71	0.32	0.22
	12 heartbeats	0.72	0.33	0.45
	5 heartbeats averaged	0.71	0.33	0.19
	12 heartbeats averaged	0.74	0.29	0.27

Fig. 36 shows a result of trinary classifier applied to the signal of the fourth volunteer(training dataset), additionally results from binary classifier, seen in Fig. 33, were used as a mask. It can be seen that the pattern of predicted pain matches the pattern of the subjective evaluation. Fig. 37 and Fig. 38 shows results of the same model applied in the same way on the first volunteer(testing dataset), here the results look much worse, especially in figure Fig. 37. However, by adding an additional moving average filter (10^{th} order) a more expressed result can be seen. The predictions in the painful regions are more frequent and consistent resulting in more continuous signals.

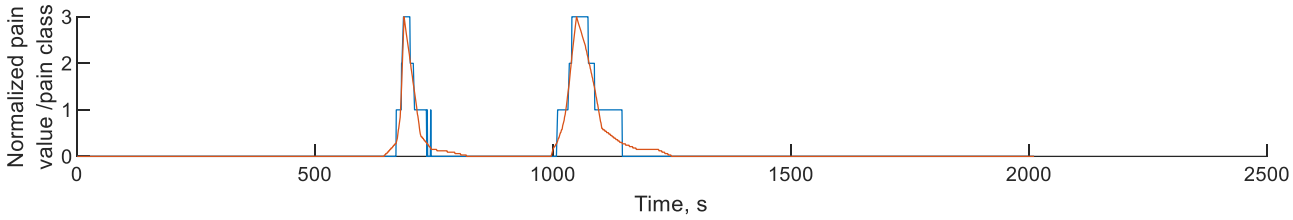

Fig. 36. Volunteer no. 4(training dataset) signal processed by single heartbeat feature trinary CNN model. Blue- predicted painful period, Orange – subjective evaluation

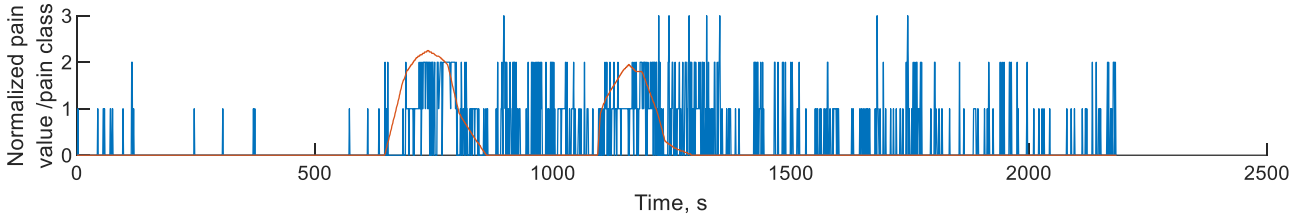

Fig. 37. Volunteer no. 1(testing dataset) signal processed by single heartbeat feature trinary CNN model. Blue- predicted painful period, Orange – subjective evaluation

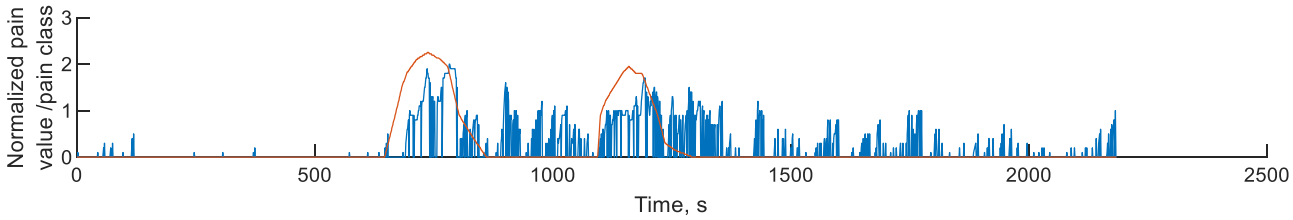

Fig. 38. Volunteer no. 1(testing dataset) signal processed by single heartbeat feature trinary CNN model, additionally filtered with an averaging filter of order 10. Blue- predicted painful period, Orange – subjective evaluation

Overall, it was found that for binary classification feature based models showed better results compared to signal based models and in trinary classification signal based models performed better than feature based. This shows an interesting possible solution to use mixed models for final pain class classification, however first the current models should be tuned to see what final accuracy could be extracted as the current model only indicated which networks are more capable in pain classification. Additionally, there is space for improvement in feature extraction, it has been noticed that PPG waveforms can differ a lot and as such an approach where the waveforms are first classified into different types and then have different features and different classifiers could provide better accuracy. Another possible approach is combining model results by using 3 or more models and deciding on the final value based on the common output of the models. Overall, There is still a lot of space for improvement in this field as all of the research is very new, currently the main limitation is data, there is not enough to properly train models for entire population, as a temporary method it would be possible to employ leave one out technique, it works by using all people except one for training, this can be repeated for all people, however if this is performed for all people computational power issues arise, depending on used NN type and architecture. Other researchers have used decision tree type classifiers these could also be interesting to test on features tested in this research.

Conclusions

1. Literature analysis was performed to analyze how pain affects the human body. It was found that during pain, sympathetic nervous system is stimulated, which subsequently modulates cardiovascular regulation. Therefore, it can be hypothesized that finger photoplethysmogram signal might be useful in an objective pain assessment.

2. State-of-the-art analysis of pain recognition algorithms revealed very few studies using physiological signals, most are using face recordings. Most algorithms use statistical methods for limited time windows, however this can also be attributed to the nature of the signals, as EDA and EMG do not provide a constant signal with morphology to analyze. A single research using PPG signal was found investigating a very small sample size.

3. PPG signal was parametrized with 53 intensity and 15 time features based on zero crossing in signal derivatives up till fourth order. Higher order derivatives were not recommended due to increasing noise level.

4. PPG signal quality control algorithm was developed to distinguish which heartbeat signals are suitable for pain recognition and classification. The algorithm operates by comparing suspected heartbeat with a median heartbeat from a baseline period.

5. In total 60 artificial neural networks were investigated. It was found that artificial neural networks perform adequately well in binary classification when classifying signals into “pain” and “no pain” episode, with highest achieved accuracy of 0.92 in testing dataset and accuracy of 1.00 in training dataset. However, artificial neural networks were significantly less accurate in trinary pain level classification, especially in identifying classes with higher pain levels with accuracies 0.61 and 1.00 for testing and training datasets, respectively.

Acknowledgements

I would like to thank the company SzeleSTIM GmbH (Vienna, Austria) for sharing the data for analysis. I would also like to Prof. Vaidotas Marozas and Dr. Andrius Rapalis for their guidance and comments.This project has partially been funded by the Research and Innovation Fund of Kaunas University of Technology (project grant No. PP2021/5), the Research Fund of Lithuanian University of Health Sciences (2021-JV-00006), and the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 880603 (SzeleSTIM GmbH).

List of references

1. MCNEILL, Jeanette A., SHERWOOD, Gwen D. and STARCK, Patricia L. The hidden error of mismanaged pain: A systems approach. *Journal of Pain and Symptom Management*. 2004. Vol. 28, no. 1, p. 47–58. DOI 10.1016/j.jpainsymman.2003.11.005.
2. LEROUX, Andrew, RZASA-LYNN, Rachael, CRAINICEANU, Ciprian and SHARMA, Tushar. Wearable Devices: Current Status and Opportunities in Pain Assessment and Management. *Digital Biomarkers*. 2021. P. 89–102. DOI 10.1159/000515576.
3. KOLTZENBURG, Martin and HANDWERKER, Hermann O. Differential ability of human cutaneous nociceptors to signal mechanical pain and to produce vasodilatation. *Journal of Neuroscience*. 1994. Vol. 14, no. 3 II, p. 1756–1765. DOI 10.1523/jneurosci.14-03-01756.1994.
4. CURATOLO, Michele, PETERSEN-FELIX, Steen, ARENDT-NIELSEN, Lars and FISHER, Dennis M. Sensory assessment of regional analgesia in humans. *Anesthesiology*. 2000. Vol. 93, no. 6, p. 1517–1530. DOI 10.1097/00000542-200012000-00025.
5. CALLIN, Sarah and BENNETT, Michael I. Assessment of neuropathic pain. *Continuing Education in Anaesthesia, Critical Care and Pain*. 1 December 2008. Vol. 8, no. 6, p. 210–213. DOI 10.1093/bjaceaccp/mkn037.
6. FELDREICH, Anna, ERNBERG, Malin, LUND, Bodil and ROSÉN, Annika. Increased β-endorphin levels and generalized decreased pain thresholds in patients with limited jaw opening and movement-evoked pain from the temporomandibular joint. *Journal of Oral and Maxillofacial Surgery*. 16 December 2012. Vol. 70, no. 3, p. 547–556. DOI 10.1016/j.joms.2011.09.013.
7. HANDWERKER, H. O. and KOBAL, G. Psychophysiology of experimentally induced pain. *Physiological Reviews*. 1993. Vol. 73, no. 3, p. 639–671. DOI 10.1152/physrev.1993.73.3.639.
8. JAFARI, Hassan, GHOLAMREZAEI, Ali, FRANSSEN, Mathijs, VAN OUDENHOVE, Lukas, AZIZ, Qasim, VAN DEN BERGH, Omer, VLAEYEN, Johan W.S. and VAN DIEST, Ilse. Can Slow Deep Breathing Reduce Pain? An Experimental Study Exploring Mechanisms. *Journal of Pain*. 2020. Vol. 21, no. 9–10, p. 1018–1030. DOI 10.1016/j.jpain.2019.12.010.
9. BIRNIE, Kathryn A., PETTER, Mark, BOERNER, Katelynn E., NOEL, Melanie and CHAMBERS, Christine T. Contemporary use of the cold pressor task in pediatric pain research: A systematic review of methods. *Journal of Pain*. 2012. Vol. 13, no. 9, p. 817–826. DOI 10.1016/j.jpain.2012.06.005.
10. PEDERSEN, Juri L. and KEHLET, Henrik. Secondary hyperalgesia to heat stimuli after burn injury in man. *Pain*. 1998. Vol. 76, no. 3, p. 377–384. DOI 10.1016/S0304-3959(98)00070-0.
11. STONE, Rachel M., AINSLIE, Philip N., KERSTENS, Thijs P., WILDFONG, Kevin W. and TYMKO, Michael M. Sex differences in the circulatory responses to an isocapnic cold pressor test. *Experimental Physiology*. 1 March 2019. Vol. 104, no. 3, p. 295–305. DOI 10.1113/EP087232.
12. DIXON, Kim E., THORN, Beverly E. and WARD, L. Charles. An evaluation of sex differences in psychological and physiological responses to experimentally-induced pain: A path analytic description. *Pain*. 2004. Vol. 112, no. 1–2, p. 188–196. DOI 10.1016/j.pain.2004.08.017.
13. GHIASI, Shadi, GRECO, Alberto, NARDELLI, Mimma, CATRAMBONE, Vincenzo, BARBIERI, Riccardo, SCILINGO, Enzo Pasquale and VALENZA, Gaetano. Investigating Phasic Activity of Time-Varying High-Order Spectra: A Heartbeat Dynamics Study during Cold-Pressor Test. In : *Computing in Cardiology*. IEEE Computer Society, 1 September 2018. ISBN 9781728109589.
14. IMAI, Y., PETERSEN, K. K., MØRCH, C. D. and ARENDT NIELSEN, L. Comparing test–retest reliability and magnitude of conditioned pain modulation using different combinations

of test and conditioning stimuli. *Somatosensory and Motor Research*. October 2016. Vol. 33, no. 3–4, p. 169–177. DOI 10.1080/08990220.2016.1229178.
15. AIMIE-SALLEH, Noor and MALARVILI, M. B. *Study of relationship between heart rate variability and autonomic function using cold pressor test for Malaysian population*. Penang : IEEE, 2011. ISBN 9781467300193.
16. PARTYLA, Tomasz, HACKER, Henriette, EDINGER, Hardy, LEUTZOW, Bianca, LANGE, Joern and USICHENKO, Taras. Remote Effects of Electromagnetic Millimeter Waves on Experimentally Induced Cold Pain: A Double-Blinded Crossover Investigation in Healthy Volunteers. In : *Anesthesia and Analgesia*. 2017. p. 980–985. ISBN 9789663353708.
17. MITCHELL, Laura A., MACDONALD, Raymond A.R. and BRODIE, Eric E. Temperature and the cold pressor test. *Journal of Pain*. May 2004. Vol. 5, no. 4, p. 233–237. DOI 10.1016/j.jpain.2004.03.004.
18. WIRCH, Jennifer L., WOLFE, Larry A., WEISSGERBER, Tracey L. and DAVIES, Gregory A.L. Cold presser test protocol to evaluate cardiac autonomic function. *Applied Physiology, Nutrition and Metabolism*. June 2006. Vol. 31, no. 3, p. 235–243. DOI 10.1139/H05-018.
19. CAMPBELL, Claudia M., EDWARDS, Robert R. and FILLINGIM, Roger B. Ethnic differences in responses to multiple experimental pain stimuli. *Pain*. January 2005. Vol. 113, no. 1–2, p. 20–26. DOI 10.1016/j.pain.2004.08.013.
20. CHALAYE, Philippe, DEVOIZE, Laurent, LAFRENAYE, Sylvie, DALLEL, Radhouane and MARCHAND, Serge. Cardiovascular influences on conditioned pain modulation. *Pain*. 2013. Vol. 154, no. 8, p. 1377–1382. DOI 10.1016/j.pain.2013.04.027.
21. KAUSHIK, Reena, HAQ, Mohd, DESAI, Hetal and PATEL, Mansi. Depressive symptoms contribute to increased response during cold pressor test in young adult persons. *National Journal of Physiology, Pharmacy and Pharmacology*. 2019. No. 0, p. 1. DOI 10.5455/njppp.2019.9.12363201804032019.
22. GEHLING, Julia, MAINKA, Tina, VOLLERT, Jan, POGATZKI-ZAHN, Esther M., MAIER, Christoph and ENAX-KRUMOVA, Elena K. Short-term test-retest-reliability of conditioned pain modulation using the cold-heat-pain method in healthy subjects and its correlation to parameters of standardized quantitative sensory testing. *BMC Neurology*. 5 August 2016. Vol. 16, no. 1. DOI 10.1186/s12883-016-0650-z.
23. MIZEVA, Irina, FRICK, Peter and PODTAEV, Sergey. Skin blood flow and temperature oscillations during cold pressor test. In : *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. Milan : IEEE, 2015. p. 7382–7385. ISBN 9781424492718.
24. GEISSER, Michael E., ROBINSON, Michael E. and PICKREN, Wade E. *Differences in cognitive coping strategies among pain-sensitive and pain-tolerant individuals on the cold-pressor test*. 1992.
25. MAURSET, A., SKOGLUND, L. A., HUSTVEIT, O., KLEPSTAD, P. and OYE, I. A new version of the ischemic tourniquet pain test. *Methods and Findings in Experimental and Clinical Pharmacology*. 1991. Vol. 13, no. 9, p. 643–647.
26. EDENS, Jennifer L. and GIL, Karen M. Experimental induction of pain: Utility in the study of clinical pain. *Behavior Therapy*. 1995. Vol. 26, no. 2, p. 197–216. DOI 10.1016/S0005-7894(05)80102-9.
27. WEHRWEIN, Erica A., ORER, Hakan S. and BARMAN, Susan M. *Overview of the Anatomy, Physiology, and Pharmacology of the Autonomic Nervous System*. 13 June 2016.
28. LAVIGNE, Gilles J., ZUCCONI, Marco, CASTRONOVO, Vincenzia, MANZINI, Christiane, VEGLIA, Fabrizio, SMIRNE, Salvatore and FERINI-STRAMBI, Luigi. Heart rate changes during sleep in response to experimental thermal (nociceptive) stimulations in healthy subjects. *Clinical Neurophysiology*. 2001. Vol. 112, no. 3, p. 532–535. DOI 10.1016/S1388-2457(00)00558-7.
29. KOENIG, J., JARCZOK, M. N., ELLIS, R. J., HILLECKE, T. K. and THAYER, J. F. Heart

rate variability and experimentally induced pain in healthy adults: A systematic review. *European Journal of Pain (United Kingdom)*. 2014. Vol. 18, no. 3, p. 301–314. DOI 10.1002/j.1532-2149.2013.00379.x.

30. PENG, Rong Chao, YAN, Wen Rong, ZHOU, Xiao Lin, ZHANG, Ning Ling, LIN, Wan Hua and ZHANG, Yuan Ting. Time-frequency analysis of heart rate variability during the cold pressor test using a time-varying autoregressive model. *Physiological Measurement*. 2015. Vol. 36, no. 3, p. 441–452. DOI 10.1088/0967-3334/36/3/441.
31. MOUROT, L., BOUHADDI, M. and REGNARD, J. Effects of the cold pressor test on cardiac autonomic control in normal subjects. *Physiological Research*. 2009. Vol. 58, no. 1, p. 83–91. DOI 10.33549/physiolres.931360.
32. JAUREGUI-RENAUD, Kathrine, HERMOSILLO, Antonio G., MÁRQUEZ, Manlio F., RAMOS-AGUILAR, Fernando, HERNÁNDEZ-GORIBAR, Mariano and CÁRDENAS, Manuel. Repeatability of heart rate variability during simple cardiovascular reflex tests on healthy subjects. *Archives of Medical Research*. 2001. Vol. 32, no. 1, p. 21–26. DOI 10.1016/S0188-4409(00)00255-1.
33. HAYANO, Junichiro. Introduction to Heart Rate Variability. In : IWASE, Satoshi, HAYANO, Junichiro and ORIMO, Satoshi (eds.), *Clinical Assessment of the Autonomic Nervous System*. Tokyo : Springer Japan, 2017. p. 109–127. ISBN 978-4-431-56012-8.
34. MUNOZ, M. Loretto, VAN ROON, Arie, RIESE, Harriëtte, THIO, Chris, OOSTENBROEK, Emma, WESTRIK, Iris, DE GEUS, Eco J.C., GANSEVOORT, Ron, LEFRANDT, Joop, NOLTE, Ilja M. and SNIEDER, Harold. Validity of (Ultra-)Short recordings for heart rate variability measurements. *PLoS ONE*. 2015. Vol. 10, no. 9, p. 1–15. DOI 10.1371/journal.pone.0138921.
35. NICHOLS, Wilmer W. Clinical measurement of arterial stiffness obtained from noninvasive pressure waveforms. *American Journal of Hypertension*. 2005. Vol. 18, no. 1 SUPPL., p. 3–10. DOI 10.1016/j.amjhyper.2004.10.009.
36. HAMUNEN, K., KONTINEN, V., HAKALA, E., TALKE, P., PALOHEIMO, M. and KALSO, E. Effect of pain on autonomic nervous system indices derived from photoplethysmography in healthy volunteers. *British Journal of Anaesthesia*. 2012. Vol. 108, no. 5, p. 838–844. DOI 10.1093/bja/aes001.
37. KHAN, Musabbir, PRETTY, Christopher G., AMIES, Alexander C., ELLIOTT, Rodney, CHIEW, Yeong Shiong, SHAW, Geoffrey M. and CHASE, J. Geoffrey. Analysing the effects of cold, normal, and warm digits on transmittance pulse oximetry. *Biomedical Signal Processing and Control* [online]. 2016. Vol. 26, p. 34–41. DOI 10.1016/j.bspc.2015.12.006. Available from: http://dx.doi.org/10.1016/j.bspc.2015.12.006
38. SELVARAJ, Nandakumar, JARYAL, Ashok, SANTHOSH, Jayashree, DEEPAK, Kishore K. and ANAND, Sneh. Monitoring of cardiovascular reactivity during cold pressor test using photoplethysmography. *Proceedings of ICSCN 2008 - International Conference on Signal Processing Communications and Networking*. 2008. P. 363–367. DOI 10.1109/ICSCN.2008.4447220.
39. JAFARI, Hassan, COURTOIS, Imke, VAN DEN BERGH, Omer, VLAEYEN, Johan W.S. and VAN DIEST, Ilse. Pain and respiration: A systematic review. *Pain*. 2017. Vol. 158, no. 6, p. 995–1006. DOI 10.1097/j.pain.0000000000000865.
40. BURTON, A. R., BIRZNIEKS, I., BOLTON, P. S., HENDERSON, L. A. and MACEFIELD, V. G. Effects of deep and superficial experimentally induced acute pain on muscle sympathetic nerve activity in human subjects. *Journal of Physiology*. 2009. Vol. 587, no. 1, p. 183–193. DOI 10.1113/jphysiol.2008.162230.
41. HULLETT, B., CHAMBERS, N., PREUSS, J, ZAMUDIO, I., LANGE, J., PASCOE, E. and LEDOWSKI, T. Monitoring electrical skin conductance: a tool for the assessment of postoperative pain in children? *Pediatric Anesthesia*. 2009. Vol. 19, no. 5, p. 556–556. DOI 10.1111/j.1460-9592.2009.02993_4.x.
42. LEDOWSKI, Thomas, ANG, B., SCHMARBECK, T. and RHODES, J. Monitoring of

sympathetic tone to assess postoperative pain: Skin conductance vs surgical stress index. *Anaesthesia*. 2009. Vol. 64, no. 7, p. 727–731. DOI 10.1111/j.1365-2044.2008.05834.x.
43. SCHESTATSKY, Pedro, VALLS-SOLÉ, Josep, COSTA, João, LEÓN, Lucia, VECIANA, Misericordia and CHAVES, Márcia L. Skin autonomic reactivity to thermoalgesic stimuli. *Clinical Autonomic Research*. 2007. Vol. 17, no. 6, p. 349–355. DOI 10.1007/s10286-007-0446-8.
44. LUNDBERG, Ulf, KADEFORS, Roland, MELIN, Bo, PALMERUD, Gunnar, HASSMÉN, Peter, ENGSTRÖM, Margareta and ELFSBERG DOHNS, Ingela. Psychophysiological stress and emg activity of the trapezius muscle. *International Journal of Behavioral Medicine*. 1994. Vol. 1, no. 4, p. 354–370. DOI 10.1207/s15327558ijbm0104_5.
45. DY, Sydney. Pain measurement. *Biobehavioral Approaches to Pain*. 2009. Vol. 43, p. 321–335. DOI 10.1007/978-0-387-78323-9_13.
46. FLAHERTY, S. A. Pain measurement tools for clinical practice and research. *AANA journal*. 1996. Vol. 64, no. 2, p. 133–140.
47. LI, Denise, PUNTILLO, Kathleen and MIASKOWSKI, Christine. A Review of Objective Pain Measures for Use With Critical Care Adult Patients Unable to Self-Report. *Journal of Pain*. 2008. Vol. 9, no. 1, p. 2–10. DOI 10.1016/j.jpain.2007.08.009.
48. MATEO, Ofelia M. and KRENZISCHEK, Dina A. A pilot study to assess the relationship between behavioral manifestations and self-report of pain in postanesthesia care unit patients. *Journal of Post Anesthesia Nursing*. February 1992. Vol. 7, no. 1, p. 15–21.
49. PUNTILLO, Kathleen A., MIASKOWSKI, Christine, KEHRLE, Karen, STANNARD, Daphne, GLEESON, Sheila and NYE, Pamela. Relationship between behavioral and physiological indicators of pain, critical care patients' self-reports of pain, and opioid administration. *Critical Care Medicine*. 1997. Vol. 25, no. 7, p. 1159–1166. DOI 10.1097/00003246-199707000-00017.
50. PAYEN, J. F., BRU, O., BOSSON, J. L., LAGRASTA, A., NOVEL, E., DESCHAUX, I., LAVAGNE, P. and JACQUOT, C. Assessing pain in critically ill sedated patients by using a behavioral pain scale. *Critical Care Medicine*. December 2001. Vol. 29, no. 12, p. 2258–2263. DOI 10.1097/00003246-200112000-00004.
51. ODHNER, Margaret, WEGMAN, Deborah, FREELAND, Nancy, STEINMETZ, Ann and INGERSOLL, Gail L. Assessing pain control in nonverbal critically ill adults. *Dimensions of Critical Care Nursing*. 2003. Vol. 22, no. 6, p. 1325. DOI 10.1097/00003465-200311000-00010.
52. PUNTILLO, Kathleen A., MORRIS, Ann B., THOMPSON, Carol L., STANIK-HUTT, Julie, WHITE, Cheri A. and WILD, Lorie R. Pain behaviors observed during six common procedures: Results from Thunder Project II. *Critical Care Medicine*. 2004. Vol. 32, no. 2, p. 421–427. DOI 10.1097/01.CCM.0000108875.35298.D2.
53. GÉLINAS, Céline, FILLION, Lise, PUNTILLO, Kathleen A., VIENS, Chantal and FORTIER, Martine. Validation of the critical-care pain observation tool in adult patients. *American Journal of Critical Care*. July 2006. Vol. 15, no. 4, p. 420–427. DOI 10.4037/ajcc2006.15.4.420.
54. Home - PainQx. [online]. [Accessed 12 May 2021]. Available from: https://painqx.com/
55. WERNER, Philipp, LOPEZ-MARTINEZ, Daniel, WALTER, Steffen, AL-HAMADI, Ayoub, GRUSS, Sascha and PICARD, Rosalind. Automatic Recognition Methods Supporting Pain Assessment: A Survey. *IEEE Transactions on Affective Computing*. 2019. Vol. X, no. July. DOI 10.1109/TAFFC.2019.2946774.
56. KÄCHELE, Markus, THIAM, Patrick, AMIRIAN, Mohammadreza, SCHWENKER, Friedhelm and PALM, Gunther. Methods for Person-Centered Continuous Pain Intensity Assessment from Bio-Physiological Channels. *IEEE Journal on Selected Topics in Signal Processing*. 2016. Vol. 10, no. 5, p. 854–864. DOI 10.1109/JSTSP.2016.2535962.
57. AMIRIAN, Mohammadreza, KÄCHELE, Markus and SCHWENKER, Friedhelm. Using radial basis function neural networks for continuous and discrete pain estimation from bio-

physiological signals. In : SCHWENKER, Friedhelm, ABBAS, Hazem M, EL GAYAR, Neamat and TRENTIN, Edmondo (eds.), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Cham : Springer International Publishing, 2016. p. 269–284. ISBN 9783319461816.

58. LOPEZ-MARTINEZ, Daniel and PICARD, Rosalind. Multi-task neural networks for personalized pain recognition from physiological signals. *arXiv*. 2017. P. 181–184.
59. LOPEZ-MARTINEZ, Daniel and PICARD, Rosalind. Continuous Pain Intensity Estimation from Autonomic Signals with Recurrent Neural Networks. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*. 2018. Vol. 2018, p. 5624–5627. DOI 10.1109/EMBC.2018.8513575.
60. JIANG, Mingzhe, MIERONKOSKI, Riitta, SYRJÄLÄ, Elise, ANZANPOUR, Arman, TERÄVÄ, Virpi, RAHMANI, Amir M., SALANTERÄ, Sanna, AANTAA, Riku, HAGELBERG, Nora and LILJEBERG, Pasi. Acute pain intensity monitoring with the classification of multiple physiological parameters. *Journal of Clinical Monitoring and Computing*. 2019. Vol. 33, no. 3, p. 493–507. DOI 10.1007/s10877-018-0174-8.
61. HINDUJA, Saurabh, CANAVAN, Shaun and KAUR, Gurmeet. Multimodal Fusion of Physiological Signals and Facial Action Units for Pain Recognition. *Proceedings - 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2020*. 2020. P. 577–581. DOI 10.1109/FG47880.2020.00060.
62. CHU, Yaqi, ZHAO, Xingang, HAN, Jianda and SU, Yang. Physiological signal-based method for measurement of pain intensity. *Frontiers in Neuroscience*. 2017. Vol. 11, no. MAY, p. 1–13. DOI 10.3389/fnins.2017.00279.
63. human body in anatomical position - Clip Art Library. [online]. [Accessed 30 June 2020]. Available from: http://clipart-library.com/clipart/6cpokXzdi.htm
64. LIANG, Yongbo, CHEN, Zhencheng, WARD, Rabab and ELGENDI, Mohamed. Hypertension Assessment via ECG and PPG Signals: An Evaluation Using MIMIC Database. *Diagnostics*. 2018. Vol. 8, no. 3, p. 65. DOI 10.3390/diagnostics8030065.
65. ON CANCER PAIN RELIEF, W H O Expert Committee, CARE, Active Supportive and ORGANIZATION, World Health. *Cancer pain relief and palliative care. Report of a WHO Expert Committee.* 1990. World Health Organization. World Health Organization technical report series ; no. 804.
66. SERLIN, Ronald C., MENDOZA, Tito R., NAKAMURA, Yoshio, EDWARDS, Katherine R. and CLEELAND, Charles S. When is cancer pain mild, moderate or severe? Grading pain severity by its interference with function. *Pain*. 1995. Vol. 61, no. 2, p. 277–284. DOI 10.1016/0304-3959(94)00178-H.
67. JENSEN, Mark P., SMITH, Douglas G., EHDE, Dawn M. and ROBINSIN, Lawrence R. Pain site and the effects of amputation pain: Further clarification of the meaning of mild, moderate, and severe pain. *Pain*. 2001. Vol. 91, no. 3, p. 317–322. DOI 10.1016/S0304-3959(00)00459-0.
68. VARGAS-SCHAFFER, Grisell. Is the WHO analgesic ladder still valid? Twenty-four years of experience. *Canadian family physician Medecin de famille canadien*. 2010. Vol. 56, no. 6, p. 514–7, e202-5.