



Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 170 (2020) 947–952

Procedia
Computer Science

www.elsevier.com/locate/procedia

The 10th International Symposium on Frontiers in Ambient and Mobile Systems (FAMS)
April 6-9, 2020, Warsaw, Poland

Objective Pain Assessment Using Vital Signs

Burak Erdoğ an^{a,*}, Hasan Oğ ul^b

^a*Department of Computer Engineering, Baskent University, Ankara 06810, Turkey*

^b*Faculty Of Computer Sciences, Østfold University College, Halden 1757, Norway*

Abstract

Pain is considered as an emotional experience and unrestful feeling associated with tissue damage. The feeling of pain occurs when the interpretation starts in brain; as a signal is sent through nerve fiber to the brain. Pain allows the body to prevent further tissue damage. Since there are different ways of expressing and feeling pain, the experience of pain is unique for everybody. In this respect, objective pain assessment is a key step and a major challenge in proper management of pain in different individuals. In this study, we offer a computational solution for objective assessment of pain using vital signs. To this end, we have reported the prediction for existence of pain by calculating the performances of several computational methods that take the sequence of vital signs acquired until pain observation as input. We claim that the use of computational intelligence methods can encourage computer-aided monitoring of pain in a hospitalized environment to a certain degree.

© 2020 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the Conference Program Chairs.

Keywords: prognosis; time-series classification; vital signs

* Corresponding author. Tel.: +90-312-246-6629; fax: +0-312-246-6630 .

E-mail address: erdoganburak93@gmail.com

1877-0509 © 2020 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the Conference Program Chairs.

10.1016/j.procs.2020.03.103

1. Introduction

Pain is an unpleasant feeling which may be caused from several physical conditions, interactions or diseases. It can be formally defined as an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage [1]. Its evolutionary function is usually described as to defend an organism against potential tissue damage from a noxious stimulus. Pain contains an intricate interchange amongst nociceptive pathways and an individuals' psychological and cognitive state. Pain is possibly dangerous and disturbing. Long-term ramifications such as chronic pain and change of physiological factors can occur due to lack of appropriate treatment.

While the treatment of pain aims at finding and eliminating its major cause, it is essential to reduce its severity during this treatment. Assessment is the groundwork of management of pain because it reveals the severity of pain and estimates the effectiveness of potential interferences. Pain assessment is usually done subjectively based the story of patients themselves. Sensing, indentifying and responding to the pain may vary remarkably between people. Currently, physicians depend on subjective pain measures like the visual analog scale (VAS). The VAS requires patients to show their pain stages on a scale from 0-10. Subjective measurement leads to an unreliable assessment of a patient's pain state at baseline and follow-up leading to uncertainty inappropriate treatment options and dosage, resulting in the over/under treatment of patients [4].

Inability of communicating is a further difficulty in pain assessment. It's essential to realize that the incapability of speaking doesn't deny the occurrence of pain and its necessary management. In situations where self-assessment tools cannot be applied, such as in individuals with neurological or cognitive impairment, other techniques are needed. This personal experience is assessed by self-reporting assessment measurements [2]. Researchers have shown that 25–40% of patients acknowledged to the hospital suffer mild to severe pain at any time. Management is frequently weighed down by weak judgment, notably in patients who are incapable to report. As a consequence, scientists and clinicians recognized the necessity to create more objective measures of pain for pain assessment.

A number of possible objective assessment instruments have been developed which are influenced by cardiovascular and respiratory factors (patterns of blood pressure, heart rate responses, and heart rate variability, interval of pulse beat and pulse wave amplitude), pupillary changes and skin sweating [13]. The best objective assessment instrument should eliminate viewer error, be trustworthy in those incapables to speak [3]. Healthcare professionals are usually untrustworthy in assessing pain severity. Presently, possible algorithms utilize variables that are only not direct measures of nociception and therefore are not reliable. Hence these algorithms are weak to the impact of other aspects, such as illness processes or prescription. It's usually complicated to decide which of the modifications in the parameter are sincerely due to pharmacological, physiological or pathological events, and which happen as a result of pain. In addition, many analgesics are sedatives, hence it makes difficult matters in the critical care settings, where tools are needed that can consistently split analgesia from sedation. Various methods are not practical. For example, the cost of MRI is an obstacle for being used in pain assessment on a daily basis. [13].

Assessment of pain is a basic task when dealing with acute pain. Assessment of intensity and location of the pain usually serves in clinical practice. Though other crucial attributes of acute pain, in addition to pain such as intensity at rest, need to be evaluated or unrelated data and false ramifications can be seen as a result. Assessment of consequences of treatment and long-lasting is more complicated, both in patients with pain as a result of cancer and in patients suffering pain caused by non-malignant causes. Various tools and algorithms have been developed for different types of chronic pain. Assessing pain in patients with whom they cannot communicate well is tough and to this date, the list of available tools implies that pain assessment remains to be a hurdle because pain is such a subjective experience [2].

In this study, we analyze the feasibility of using vital signs collected via a set of wearable sensors for objective pain assessment. We focus on the patients in intensive care units (ICUs) since their vital signs are being monitored frequently in a controlled environment. Pain assessment is usually done by well-trained nursing staff and recorded. We build a machine learning-based model that takes a short multi-modal time-series of vital signs as input and report the existence of severe pain. The prediction performances of several classifications algorithms with different feature extraction techniques are evaluated.

2. Methods

2.1. Data collection and preprocessing

To the best of our knowledge, there is no publicly available benchmark dataset for computational assessment of pain. Here, we built a dataset acquired from Medical Information Mart in Intensive Care (MIMIC)-III database using the v1.4 release [5]. The MIMIC database gathered anonymous medical data from more than 40,000 patients admitted to Beth Israel Deaconess Medical Center in Boston Between 2001 and 2012. The database includes information such as vital sign measurements, demographics made at the bedside, procedures, medications, laboratory test results, caregiver notes, mortality and imaging reports.

MIMIC supports various range of analytic studies traversing clinical decision-rule improvement, electronic tool development, and epidemiology. It is freely available to researchers worldwide, it includes a very large group of ICU patients and it contains data including electronic documentation, lab results, waveforms and bedside monitor trends. [14].

We collected the measurements of seven different physiological parameters for each of 8 hours prior to pain recording time at ICU. These variables include:

- Glasgow Coma Scale (GCS-eye opening),
- Heart Rate,
- Oxygen Saturation,
- Pupil Size (Left and Right),
- Respiration Rate,
- Skin Temperature,
- Urine Color.

To build our own dataset for pain prediction, we applied two further inclusion criteria: (1) an assessment score of pain is reported in addition to its existence, and (2) at least one measurement is available for each of the vital signs 8 hour prior to recording time. Each patient was labelled by either positive (having pain) and negative (no pain) by considering all samples which are assessed by a pain score larger than zero. This was resulted with a dataset of 10403 ICU stays comprising seven physiological parameters on 8 time points, pain recording time, pain scores and binary class labels (6163 positives and 4240 negatives) for existence of pain. To manage missing data we considered the following technique; the latest measurement is carried forward to fill subsequent empty time points to manage the imputation of missing data. The time points that precede the collection of any measurements are backfilled with the first subsequent hour with recorded value.

2.2. Problem formulation

Our main goal is to predict if the current individual has a pain or not using a computational method. This can be formulated as a time-series binary classification task, where the input is a short multi-variate signal of physiological or other vital medical measurements and the output is a value in binary format representing the occurrence of pain. We examined a set of model-based machine learning methods for time-series classification to adapt for pain assessment problem in a binary categorization setup.

2.3. Model-based methods

A model-based time-series classification method takes input samples with associated labels of class and then creates a model that can detect differences between predefined categorical labels. A crucial obstacle that should be solved is feeding the model with potential time-shifts, time-series signals of different scales, and differing lengths. A solution can be extracting several numerical features and use them as samples of input. We are guided by some recent applications [6, 7, 15] and extracted some statistical time-domain features from time-series vital signs. In our case, these features include; median, variance, mean, root mean square, entropy, root mean square, mean absolute

deviation and interquartile range.

We consider four different machine learning models: AdaBoost, Multilayer Perceptron, LogitBoost, and Random Forest Due to their popularity in computer-aided examination systems.

AdaBoost was originally titled AdaBoost.M1 by the founders of the technique Freund and Schapire. Because it is used for classification rather than regression lately, it may be referred to as discrete AdaBoost. Consequently, the most common algorithm used with AdaBoost is one level of decision trees. They are often called decision stumps because these trees are so short and only contain one decision for classification [8].

A multilayer perceptron is a class of feedforward artificial neural network. When multilayer perceptrons have a single hidden layer, they are sometimes referred to as "vanilla" neural networks [9].

Hastie et al. formulated LogitBoost [10]. One can derive the LogitBoost algorithm by considering AdaBoost as a generalized additive model and applying the cost function of logistic regression.

To train a model, Random Forest builds several multiple decision trees. A decision tree is a flowchart-like structure whereas each internal node expresses a test on a feature demonstrating the related sample [11]. By voting overbuilt decision trees, a query is classified.

3. Results

We used model-based methods to our dataset for binary classification (pain exists, pain does not exist) of pain. Firstly, model-based methods were fed directly by the sequence of measurements of vital signs. Since the same number of measurements was taken for each patient, the problem of varying length inputs did not affect the models. Subsequently, by using extracted features from time series we replaced the inputs. Table 1, Table 2 and Table 3 list the accuracy for correctly classified instances with leave-one-out cross-validation tests and predicting each method using Area Under Receiver Operating Characteristics Curve (AUROC).

We examined the need for sequential measurement data for all time points during 3, 6 and 8 hours before pain onset. The performance is poor when instant values measured 3 hours after the onset are used for correctly classified instances. For the AUROC score as the hour increases score increased as well. For both cases Random Forest performs slightly better except 3 hours on set for extracted features is performed when LogitBoost is used.

Table 1. Comparison of methods for pain prediction at 3 hours prior to pain observation.

Sequence Representation	Method	AUROC	Accuracy
Direct	Ada Boost	0.647	% 75.1
	Multilayer Perceptron	0.628	% 75.4
	LogitBoost	0.649	% 75.5
	Random Forest	0.655	% 76.1
Extracted	Ada Boost	0.648	% 75.1
	Multilayer Perceptron	0.615	% 73.8
	LogitBoost	0.656	% 75.4
	Random Forest	0.644	% 75.0

Table 2. Comparison of methods for pain prediction at 6 hours prior to pain observation.

Sequence Representation	Method	AUROC	Accuracy
Direct	Ada Boost	0.633	% 65.8
	Multilayer Perceptron	0.631	% 65.6
	LogitBoost	0.660	% 66.7
	Random Forest	0.704	% 68.9
Extracted	Ada Boost	0.640	% 66.1
	Multilayer Perceptron	0.656	% 66.5
	LogitBoost	0.662	% 66.8
	Random Forest	0.700	% 68.4

Table 3. Comparison of methods for pain prediction at 8 hours prior to pain observation.

Sequence Representation	Method	AUROC	Accuracy
Direct	Ada Boost	0.634	% 65.6
	Multilayer Perceptron	0.630	% 65.9
	LogitBoost	0.654	% 66.5
	Random Forest	0.711	% 68.9
Extracted	Ada Boost	0.651	% 63.6
	Multilayer Perceptron	0.646	% 64.2
	LogitBoost	0.667	% 65.0
	Random Forest	0.727	% 68.2

4. Discussion and Conclusion

This study assesses the accuracy of a computational system that observes the sequence of effortlessly obtainable physiological parameters to predict the existence of pain. The prediction performances of existing algorithms were evaluated by this research for time-series classification when adapted for pain prediction using vital signs. The results show that using machine learning algorithms with direct measurements gives the best result for correctly classifying instances. In terms of the AUROC score, extracted features give the best result as the time increases. In terms of the number of correctly classified instances, direct features give the best result as the time decreases.

By using computationally observable measures of variability of heart rate, interplays among the parasympathetic and sympathetic nervous systems can be learned. This effortless measure variable and can be used in both sedated and conscious patients by using ECG monitoring. However, age, sex, depth of anesthesia, surgical stimulation,

medications, and emotions can affect heart rate variability. Pupillary dilatation could be used to assess sympathetic stimulation because of pain. However, factors such as analgesics, antiemetic, anticholinergic and vasoactive agents, environmental luminance, age and uncommon conditions such as Horner's syndrome can alter pupillary response. Also, the process of pupillary response in sedated patients still remains to be unclear even though consistent results in response to transient painful stimuli, the pupillary light reflex amplitude has shown a variable correlation with numerical rating scores in the postoperative setting. Along with issues of practicality and according to these results, clinical application of pupillary response remains to be unknown particularly in unhelpful patients [13].

There are certain limitations in the field of computer-aided decision support in health domain [12]; these can be listed as; absence of enough data, missing/incorrect measurements and class imbalance in training data. Amount of labeled pain cases is not sufficient when time-series data are considered even though the MIMIC database offers plenty amount of anonymous data. Although resampling could have been another solution, the class imbalance problem was managed in this research at the assessment level by using AUROC as performance criteria instead of computing recall values and direct precision. In a nut shell, implementing beneficial techniques for managing imbalanced and small data and imputing missing values should be considered for a reasonable solution for pain prediction.

References

- [1] Watt-Watson J, Hunter J, Pennefather P, Librach L, Raman-Wilms L, et al. (2004) "An integrated undergraduate pain curriculum, based on IASP curricula, for six health science faculties." *Pain* **110**(1–2): 140–148.
- [2] Breivik H, Borchgrevink PC, Allen SM, Rosseland LA, Romundstad L, et al. (2008) "Assessment of pain" *British Journal of Anaesthesia* **101**: (17–24).
- [3] Ahlers SJ, van Gulik L, van der Veen AM, van Dongen HP, Bruins P, et al. (2008) "Comparison of different pain scoring systems in critically ill patients in a general ICU" *Crit Care* **12**: R15.
- [4] Mill Valley. (2017) "The Technical Writer's Handbook" AIMed: Artificial Intelligence in Medicine The Ritz-Carlton, Laguna Niguel, CAM. Young
- [5] A. E. W. Johnson, T. J. Pollard, L. Shen, L. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Celi, and R. G. Mark. (2016) "MIMIC-III, a freely accessible critical care database," *Scientific Data* **3**.
- [6] Ç.B Erdaş, I. Atasoy, K. Açııcı, and H. Oğul. "Integrating features for accelerometer-based activity recognition" (2016) *Procedia Comput. Sci.* **98**: (522–527).
- [7] T. Asuroglu, K. Acici, C. B. Erdaş, M. K. Toprak, H. Erdem, and H. Oğul. "Parkinson's disease monitoring from gait analysis via footworn sensors" (2018) *Biocybernetics and Biomedical Engineering* **38** (3): 760–772.
- [8] Jason Brownlee. "Machine Learning Algorithms" (2016) Retrieved from <https://machinelearningmastery.com/boosting-and-adaboost-for-machine-learning/>
- [9] Hastie, Trevor. Tibshirani, Robert. Friedman, Jerome. "The Elements of Statistical Learning: Data Mining, Inference, and Prediction". (2009) Springer.
- [10] Tibshirani, Robert. (2000) "Additive logistic regression: a statistical view of boosting". *Annals of Statistics* . **28** (2): 337–407.
- [11] L. Breiman, "Random Forests" (2001). "Machine Learning", **45**: 5–32
- [12] A. Rodríguez-González, J.E. Labra-Gayo, R. Colomo-Palacios et al. "SeDeLo: Using Semantics and Description Logics to Support Aided Clinical Diagnosis"(2012) *J Med Syst*, **36** (2471).
- [13] R. Cowen M. K. Stasiowska H. Laycock C. Bantel. "Assessing pain objectively: the use of physiological markers" (2015) *Anaesthesia* **70** (7)
- [14] MIMIC-III Critical Care Database (n.d) Retrieved from <https://mimic.physionet.org/about/mimic/>
- [15] H.Ogul, A.Baldominos, T.Asuroglu, R.Colomo-Palacios. (2019) "On computer-aided prognosis of septic shock from vital signs", 32rd IEEE International Symposium on Computer-Based Medical Systems (IEEE-CBMS'19), Cordoba.