EmoPain(at)Home: Dataset and Automatic Assessment within Functional Activity for Chronic Pain Rehabilitation

Temitayo Olugbade*, Raffaele Andrea Buono*, Amanda C de C Williams*, Santiago de Ossorno Garcia*
Nicolas Gold*, Catherine Holloway*§, Nadia Bianchi-Berthouze*†
*University College London, UK, §Global Disability Hub, UK, †n.berthouze@ucl.ac.uk

***Abstract*—While there is growing interest in developing technology to support pain assessment, pain-related self-management, and healthcare personalisation, there are currently no datasets on nonverbal pain behaviour in the context of functional activities. To address this gap, we introduce the EmoPain(at)Home dataset which consists of motion capture data and self-reported pain, worry, and confidence intensities captured from people with chronic pain. The data were recorded during self-selected functional activities in the home, e.g. vacuuming. We include analysis of the dataset as well as baseline classification of pain levels with average F1 score of 0.61 for two classes. We additionally discuss inclusivity considerations for capture of datasets in naturalistic settings, based on lessons learnt within our study.**

***Index Terms*—affect, body movement, confidence, dataset, home, pain, worry**

I. Introduction

Automatic assessment of pain experience and behavior is an established area of research that ultimately aims to support personalization of care, empowerment of patients, and self-management of chronic conditions [1]–[3]. As with the general field of affective computing, datasets are key to advance in the area. They are a critical resource for developing and benchmarking machine learning algorithms, and can additionally be valuable for better understanding of pain experience. In this paper, we introduce *EmoPain@Home*, a new dataset that captures body movement data together with self-reported levels of pain and related worry and confidence from people with chronic pain during functional activities at home.

Several datasets for automatic recognition of pain and related affect exist (see Table I for an overview) but only a much smaller number cover chronic pain and these have been limited to the context of stillness and instructed movement in lab settings [4], [6], [8], [10], [16], [21]. Chronic pain is of particular interest due to its significant effect on the sense of self, engagement in valued activities, and interaction with others [20], [26]–[28]. It is important for technology that aims to support personalization of care and self-management for people with chronic pain to be built on data that is more representative of pain experience and behaviour in the context of everyday activities where their challenges lie [20]. This requirement has motivated our new dataset, which shifts data capture to naturalistic functional activity settings and covers activities that people with chronic pain find challenging.

Overall, our paper contributes the following to the area of automatic recognition of pain and related affective states:

1) **The first functional activity pain dataset**, *EmoPain@Home*, that consists of motion capture data and self-reports of pain, pain-related worry, and movement-related confidence captured from 9 people with chronic pain during their normal home activities over multiple days (mean=2.78 days per participant). This dataset builds on our previous dataset (EmoPain [16]) captured from people with chronic pain in exercise movements in lab settings. While the EmoPain dataset further includes data from healthy participants, with the EmoPain@Home dataset we focus on capturing a wide range of pain experiences (e.g. from no pain to extreme pain) in everyday functioning with chronic pain.
2) **A discussion of lessons learnt around inclusivity during our data collection process**. The importance of capturing and understanding issues around the acquisition of sensor data and related self-report is increasingly being appreciated by different stakeholder groups. Insight into issues that affect inclusivity for specific populations particularly have the potential to shape the design of technology so that it is feasible for use by these people groups. Such insights can further be useful in helping dataset creators maximise participation and data quality via the use of informed data collection methods and protocols (our discussion in this paper centres on the sensor selection and participant training).
3) **Exploration of automatic detection based on the EmoPain@Home dataset**, with analysis of features of pain, worry, and confidence captured from different movement timescales, as well as baseline classification of pain levels. This extends previous pain level classification studies based the EmoPain dataset with a more challenging dataset captured in functional movement settings at home and with a more difficult task of detecting pain levels in continuous time during each activity rather than only at the end of the activity. Further, the findings of our analysis provide new insights into relationships between movement behavior and pain experience.

This work was supported by the EU Future and Emerging Technologies Proactive Programme H2020 (Grant No. 824160: EnTimeMent - https://entimement.dibris.unige.it.

Our studies were approved by our local research ethics committee (reference: 5095/001).

II. Background: Nonverbal Measures of Pain and The Significance of The Body Movement Modality

Table I shows the variety of pain expression modalities represented across the multiple pain datasets that exist. The facial modality of expression has been the most widely captured in this area as can be seen in the table. Findings in [29] indeed highlight the relevance of this modality, showing statistically significant difference in the activation of facial actions between the use of the pain-affected arm in people with shoulder pain and the use of their non-affected arm. Significant although low correlation was also found between the aggregate scoring of these facial actions and self-reported pain intensity.

Vocal/paraverbal expressions have also been shown to be valuable for capturing pain experience. For example, [34] found significant higher base frequencies, insistence and periodicity, as well as duration of maximum pressure in infant cries where there were other expressions of pain, compared to cries associated with mild or no non-vocal expression. In [35] with adults with severe intellectual and developmental disabilities, vocalizations (such as moaning, crying, yelling) were observed in both painful and non-painful activities. However, there were significant differences in the number of pulses and the shimmer level of the vocalizations for the two sets of activities.

There is also evidence of physiological response to painful stimulus. For instance, [30] found significant effect of heat stimuli intensity (higher than subjective pain threshold) on skin conductance level and number of skin conductance responses. In [20] (based on the EmoPain dataset), significant delayed and lower back muscle relaxation was found in people with chronic low back pain and high level pain during trunk flexion movements compared with those experiencing lower level pain or healthy participants. Relaxation was also significantly lower in their upper back muscles. These findings are similar to the finding of low but significant correlation between self-reported pain levels and flexion relaxation ratio in lower back muscle activity in people with chronic low back pain [31].

Closely related to muscle activity is overt bodily behaviour which has also been associated with pain experience. [32] found significant differences in range and speed of head movement between people experiencing pain (heat or cold pressor) and those not experiencing pain. Their finding was consistent across two datasets (BioVid Heat Pain [11] and BP4D-Spontaneous [12]). They also explored a third dataset UNBC-McMaster Shoulder Pain Expression [10] but their findings were less evident with this. This may be due to the constraint on head movement in the UNBC-McMaster dataset as there were several instances in which the participant was lying down. Unlike experimentally induced pain (i.e. for BioVid Heat Pain and BP4D-Spontaneous datasets) where speed was higher in people experiencing pain, speed was lower in the movement of the pain-affected region (arm) for people with clinical pain in the UNBC-McMaster dataset. [20] similarly found both lower speed and lower range of motion (in the head and trunk) for participants with chronic low back pain experiencing high level pain, in the EmoPain dataset.

While the use of multiple nonverbal measures of pain is ideal [1], logistical constraints such as the practicality of data capture is a critical consideration. For example, facial expressions are more practical for sedentary settings (e.g. [9], [11]–[13], [18], [23], [25]) or when mobility is limited to a single space (e.g. [10], [16]). Context and/or purpose of assessment are further important in selecting the modalities to employ. Findings in [33], for instance, suggest that body movement may be particularly valuable when the goal of assessment includes judgement of task demands and coping strategy during physical activity for people with chronic pain.

For our new EmoPain@Home dataset, we captured body movement data because this is especially relevant in physical activity context, and it provides insight into strategies used for executing challenging activities. Additional modalities such as muscle activity could be useful for a more comprehensive capture of pain experience. However, we focused on body movement in this dataset to minimize the burden (learning to use unfamiliar sensors, self-attaching sensor units with pain, and dealing with technical sensor issues) on participants.

III. The EmoPain@Home Dataset

Participants were recruited using ads on social media as well as by directly contacting community pain support groups across the UK. 10 people signified interest in taking part in the study, but one person withdrew from taking part (due to other unforeseen commitments) resulting in 9 participants in total. The participants gave informed consent for collection and processing of their data as well as for sharing pseudonymised data with the research community. Participants were reimbursed for their time at the rate of £10 per hour.

All (5 female, 4 male) self-identified as living with chronic musculoskeletal pain involving the lower back area: 4 reported sciatica, 2 reported chronic pain resulting from an old spinal injury, and 3 reported other forms of chronic pain. The participants were between 27 and 59 years old (mean=45.11, standard deviation=11.50). The participants were in the UK at the time of the study (March-August 2021).

A. Data Capture Settings

Data was captured in the context of everyday physical functioning at home. The sensor system used was sent to participants by courier and the participants were trained (remotely) to attach the sensors and record the data themselves.

Rather than recreating activities for the purpose of the study, the participants performed tasks that they needed to do in their own homes (see Table II). In order to have a good representation of pain experiences across the range of these activities for each participant, they were asked to include activities that they usually found particularly challenging as well as those that they did not find challenging. Participants were asked to fill in a diary over the days before the first data capture session to identify these activities beforehand. Data capture sessions were then arranged for the days when the

TABLE I
EXISTING PAIN RECOGNITION DATASETS

Dataset	Year	Type of pain	Context	Modalities
Gioftsos and Grieve [4]	1996	chronic musculoskeletal	instructed exercise movements in lab settings	body movement, feet force
Bishop et al. [5]	1997	acute	constrained exercise movements in lab settings	spine movement
Dickey et al. [6]	2002	chronic musculoskeletal	instructed exercise movements in lab settings	spine movement
Brahnam et al. [7]	2006	puncture	hospital neonatal unit	facial expression
Levinger and Gilleard [8]	2007	chronic musculoskeletal	instructed walking in lab settings	lower limb movement, ground reaction force
Hi4D-ADSIP [9]	2011	acted	seated	facial expression
UNBC-McMaster Shoulder Pain Expression [10]	2011	acute, chronic musculoskeletal	instructed exercise movements standing or laid down in the lab	facial expression
BioVid Heat Pain [11]	2013	heat	seated in lab settings	physiological signals, facial expression, upper body movement
BP4D-Spontaneous [12]	2014	cold	seated in lab settings	facial expression, head movement
Rivas et al. [13]	2015	acute	seated exergaming	hand movement, hand pressure, facial expression
Infant Cry Sounds [14]	2015	acute	hospital visit	vocal expression
Zhang et al. [15]	2016	cold	seated in lab settings	physiological signals, facial expression, head movement
Triage Pain-Level Multimodal Database	2016	acute	triaging in hospital emergency unit	physiological signals, vocal expression, facial expression
EmoPain [16]	2016	chronic musculoskeletal	instructed exercise movements in lab settings	physiological signals, facial expression, body movement
SenseEmotion [17]	2017	heat	seated looking at affective images augmented with sound	physiological signals, vocal expression, facial expression, body movement
Multimodal Intensity Pain (MIntPAIN) [18]	2018	electrical stimulation	seated in lab settings	physiological signals, facial expression
Ubi-EmoPain [19], [20]	2018	chronic musculoskeletal, widespread	instructed and naturalistic movement in lab settings	physiological signal, body movement
Hu et al. [21]	2018	chronic	instructed standing in lab settings	physiological signals, ground reaction force, spine movement
Clinical Valid Pain [22]	2018	acute	hospital emergency unit visit	blood test data, facial expression
X-ITE Pain [23]	2019	electrical stimulation, heat	laid down in lab settings	physiological signals, vocal expression, facial expression
iCOPEvid [24]	2019	puncture	neonatal hospital unit	facial expression
Intelligent Sight and Sound Chronic Cancer Pain I [25]	2021	cancer	reading aloud text, describing current feelings	vocal expression, facial expression
EmoPain@Home (current paper)	***2022***	***chronic***	***functional activities at home***	***body movement***

participant planned to engage in the noted activities. Within each session, the participant engaged in a number of the self-selected activities (shown in Table II).

Each participant took part in multiple data capture sessions across multiple days. The number of session days per participant ranged between 1 and 4 days (median=3). For each session day for a participant, the session was limited to an hour to minimise fatigue. We additionally limited capture of each activity in a session to 15 minutes roughly and participants were encouraged to take breaks earlier if needed. These considerations were decided together with a clinician within the research team and further discussed with the participant.

In order to enable recording of end-of-session interviews as well as to facilitate continuous capture of self-report based on experience sampling, the researcher was present during the sessions. However, the researcher only attended them remotely, i.e. via video-conferencing. Three participants further took part in sessions without the researcher present.

B. Data Description

1) Body movement data: In each data capture session, body movement data was captured using wearable inertia-based sensor units (Notch sensors) that record 3D joint angle and position data. To further limit the burden on the participants, only 6 sensor units were used; lab tests suggested that sensor attachment time and possibility of technical issues increased with the number of sensor units. We captured movement data for the right elbow and wrist, mid spine, hip, and right knee and ankle. Findings in previous work [19], [20], [36] suggest

TABLE II
ACTIVITIES CAPTURED IN THE EMOPAIN@HOME DATASET
(SELF-SELECTED BY PARTICIPANTS)

PID	Challenging activities	Non-challenging activities
2	Changing bedsheets**, Washing up**, Loading washing machine, Unloading washing machine, Window cleaning	Walking exercise, Sweeping, Dusting**, Vacuuming*
3	Hoovering, Washing up**, Bathroom cleaning, Unloading shopping, Cleaning windows, Tidying up*	Washing up**, Unloading washing machine**, Loading washing machine**
4	Hoovering**, Changing bedsheets**, Vacuuming (car), Watering garden	Bathroom cleaning, Dusting (car), Preparing food*, Cleaning parrot cage*
5	Painting shelves, Painting a wall, Walking exercise, Bathroom cleaning	-
6	Changing bedsheets	Unload dishwasher, Sorting out boxes of stuff
7	Unloading washing machine, Unloading dishwasher, Changing bedsheets	Loading washing machine, Loading dishwasher, Tidying up room
8	Washing up, Hanging clothes to dry, Vacuuming, Changing bedsheets, Cleaning windows	Yoga, Unloading washing machine
9	Ironing, Preparing lunch	Filing documents
10	Tidying up room	-

Note that the participant with PID=3 originally specified 'Washing up' as non-challenging, but referred to it as challenging for a different session when they experienced (chronic) hand pain.
PID - Participant ID
* - performed without the researcher (remotely) present
** - performed both with and without the researcher (remotely) present

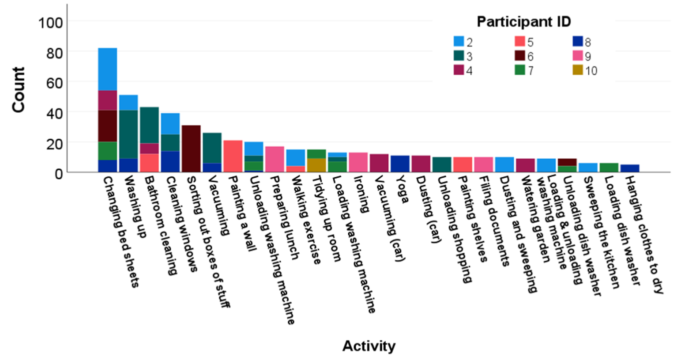

Fig. 1. Distribution of the activity instance segments by activity type.

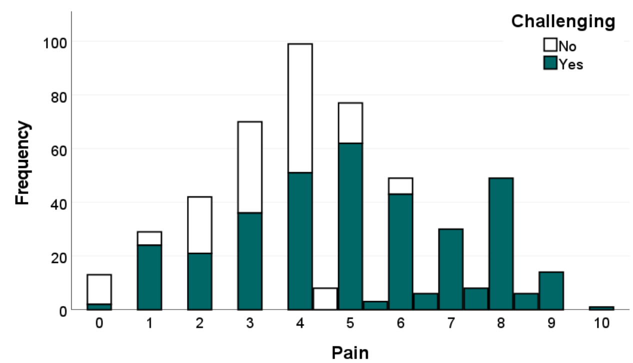

Fig. 2. Distribution of the activity instance segments by pain intensity.

that capture of data from only one side of the body can be informative for automatic assessment.

2) Self-reported pain and related affect labels: In the data capture sessions where the researcher was present (remotely), audiovisual data was recorded using the participant's webcam. This recording was used to capture from the participant at every minute verbal self-report of pain, worry about pain, and confidence about being able to perform the rest of the activity. Pain and worry were assessed on a numerical scale from 0 to 10, with 0 for 'none' and 10 for 'very severe' as was done in the EmoPain dataset. Confidence was assessed using an ordinal scale of: no confidence, less than average confidence, average confidence, more than average confidence, and max confidence. While there is a lot of precedent for the 0-10 scale (especially for pain) that is the standard [37], there is little evidence that people make up to 11 (or more) distinctions between levels of confidence, so we kept the scale for confidence simpler. The fewer distinctions for the confidence scale made it feasible to use a verbal (rather than numeric) scale, which people prefer [38]. An additional rationale for using the ordinal scale for confidence was to help differentiate it from the pain and worry scales for which, unlike with confidence, a higher value represents a more negative experience, and so make it as minimally challenging as possible, cognitively, to self-report all three constructs continuously through the activities.

Participants responded to prompts for current pain, worry, and confidence from the researcher. In order to limit disruption of the activities, the verbal prompt was shortened to "time" once the self-report constructs and procedure were extensively described to the participant. The method and frequency of self-report was based on discussion in [39] of the value of the method for both research and the participants themselves.

For capture sessions where the researcher was not present, the participant recorded in written form the pain, worry, and confidence experienced at the start and end of each activity.

C. Descriptive Analysis

For analysis, we focused on the activity instances where self-report was provided during the activity (at every minute), rather than only at the start and end. The mean duration of these activity instances (i.e. the time between the first and last self-report in each activity instance) is 9.62 minutes (standard deviation = 5.89). Each self-report point within these activity instances was then used to define a unique activity instance segment. An activity instance segment thus has as labels the pain, worry, and confidence reported at the corresponding self-report point as well as the activity type in which the self-report was made. We obtained 504 activity instance segments. We recoded the confidence labels to integer values with 1 for 'no confidence' and 5 for 'max confidence' for analysis.

Fig. 1 shows the distribution of the activity instance segments across activity types highlighting the participant that

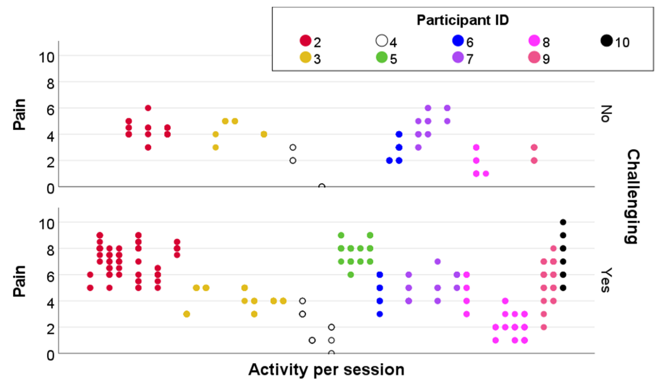

Fig. 3. Distribution of pain intensity within each activity instance.

performed the respective segment. There are 26 activity types. More than 30% of these have 20 or more segments. Further, while most of the activity types are exclusive to individual participants (e.g. only one participant had a *yoga* activity instance), a number of activity types are common to 3 or more participants. For instance, 5 participants had *bedsheet changing* activity instances, and 3 participants had *loading washing machine* activity instances.

Plot of the pain intensities across the activity instance segments (Fig. 2) shows that although most of the segments have mid-level pain labels, higher level pain labels are well-represented. Further inspection of the range of pain intensities within each activity instance (Fig. 3) shows that for most of the activity instances, there are changes in pain intensities across segments from the same activity instance. The findings were similar for worry and confidence (plots not shown due to space constraints) although there is a right skew for worry with a peak at intensity of 2 and a skew to the positive side for confidence with its peak at 4 ('more than average confidence').

IV. Automatic Detection of Pain Levels: Baseline

We report here baseline classification of pain levels based on the EmoPain@Home dataset using standard machine learning algorithms and features based on existing work. We further present findings of correlation analysis of the features with respect to pain, worry, and confidence.

A. Methods

1) Feature Extraction: We extracted body movement features corresponding to each of the activity instance segments. We extracted features only from activity instance segments for which all 6 anatomical joints of interest were recorded without missing data ($n = 226$ out of 504 segments).

We computed 6 sets of features: average speed, average jerk, average energy, normalised amount of movement, minimum distance from the forearm (to capture self-adaptor behaviour), and angular range of motion. These features were informed by previous work in [19], [20] based on the EmoPain dataset, but here they were computed them across two different timescales. For the *activity instance timescale*, the feature covered the time period from the first frame of the corresponding activity instance to the frame of the corresponding self-report point of the given activity instance segment. With the *activity instance segment timescale*, the same feature was computed over the one-minute window preceding this self-report point.

2) Learning Algorithm: We used a standard machine learning algorithm, *Bagging* [42] which enables (random) selection of a subset of features to build each learner in the ensemble. This is particularly relevant given the dimensionality of the extracted features ($m = 60$) in the context of the available data ($n = 226$). We set the maximum number c_{max} of features used to build each learner to 18 with replacement. We did not obtain better performance with other values ($c_{max} = 12, 30, 60$). We used decision tree [43] as the *Bagging* learner as it performed better than others such as support vector machines, k-nearest neighbours, logistic regression, and multilayer perceptron in experiments. The *Bagging* model was evaluated using leave-one-activity-instance-out (LOAIO) cross-validation where all segments from the same activity instance are held out in each fold. The number of learners to use for the model was selected in each fold using nested LOAIO cross-validation. We did not test for generalisation to unseen participants (leave-one-subject-out cross-validation) because of the difference in the types of activities performed by the participants (Fig. 1).

3) Training Data Boost: To boost the size of the training data, in each fold we included data from participants with chronic pain in the EmoPain dataset. The context of the EmoPain dataset was not functional activities at home but rather exercise movements captured in lab settings. However, the exercise movements (sit-to-stand, stand-to-sit, sitting still, standing still, standing on one leg, forward reach, bend, walk) are representative of movements involved in functional activities. For example, changing bedsheets may involve forward reaching; vacuuming may involve walking and bending.

The same pain scale used in the EmoPain@Home dataset was used in the EmoPain dataset. While the motion capture sensor types for the two datasets are different, we expected that the features computed would be comparable across the two datasets. We further scale the feature set for each dataset separately to account for any differences.

In the EmoPain dataset, exercise instances were of short durations (under a minute [40]) and pain intensity was reported for each instance only at its end. Thus, each exercise instance specified an activity instance segment for the training data. This resulted in 330 additional segments in the training data for each fold. While we extracted two timescales of features per segment as was done with the EmoPain@Home dataset, for the shorter timescale features were computed over the second half of the corresponding exercise instance, and computed over the full exercise instance for the longer timescale.

B. Results

1) Classification Results: Given the limited size of the data, we focused on two levels of pain: lower level pain, defined as pain intensity less than 5 on the pain scale, and higher level pain otherwise. Table III shows the confusion matrix for the automatic classification of pain levels across all folds, with

TABLE III
BASELINE PAIN LEVEL CLASSIFICATION RESULTS: CONFUSION MATRICES

WITH TRAINING DATA BOOST

		Prediction	
		Lower level pain	Higher level pain
Ground truth	Lower level pain	**81**	28
	Higher level pain	61	**56**

WITHOUT TRAINING DATA BOOST

		Prediction	
		Lower level pain	Higher level pain
Ground truth	Lower level pain	**81**	28
	Higher level pain	64	**53**

TABLE IV
FEATURE CORRELATIONS

Variable	**Joint**	**Timescale**	**Pain**	**Worry**	**Confidence**
Speed	trunk	*Activity instance segment*	NS	NS	NS
	thigh		NS	-0.14*	0.14*
	upper arm		NS	NS	0.14*
	lower leg		NS	-0.17*	0.18**
	forearm		NS	NS	NS
	hip		NS	NS	NS
	trunk	*Activity instance*	-0.15*	-0.14*	-0.15*
	thigh		-0.24**	-0.21**	0.23**
	upper arm		-0.14*	-0.13*	0.16*
	lower leg		-0.25**	-0.20**	0.23**
	forearm		NS	NS	NS
	hip		NS	NS	NS
Jerk	trunk	*Activity instance segment*	-0.30**	-0.24**	0.31**
	thigh		NS	NS	NS
	upper arm		-0.28**	-0.22**	0.32**
	lower leg		-0.24**	-0.21**	0.23**
	forearm		-0.18**	-0.16*	0.23**
	hip		NS	NS	NS
	trunk	*Activity instance*	-0.36**	-0.29**	0.37**
	thigh		-0.14*	NS	NS
	upper arm		-0.33**	-0.31**	0.38**
	lower leg		-0.32**	-0.25**	0.28**
	forearm		-0.20**	-0.22**	0.28**
	hip		NS	NS	NS
Energy	trunk	*Activity instance segment*	-0.15*	-0.14**	0.14*
	thigh		NS	NS	NS
	upper arm		NS	NS	0.18**
	lower leg		NS	NS	0.15*
	forearm		NS	NS	0.15**
	hip		NS	NS	NS
	trunk	*Activity instance*	-0.26**	-0.21**	0.21**
	thigh		-0.26**	-0.18**	0.22**
	upper arm		-0.22**	-0.23**	0.24**
	lower leg		-0.27**	-0.19**	0.23**
	forearm		NS	NS	NS
	hip		NS	NS	NS
Amount of movement	trunk	*Activity instance segment*	NS	NS	NS
	thigh		NS	-0.14*	0.14*
	upper arm		NS	NS	0.14*
	lower leg		NS	-0.17*	0.18*
	forearm		NS	NS	NS
	hip		NS	NS	NS
	trunk	*Activity instance*	-0.15*	-0.14*	0.15*
	thigh		-0.24**	-0.21**	0.23**
	upper arm		-0.14*	-0.13*	0.16*
	lower leg		-0.25**	-0.20**	0.23**
	forearm		NS	NS	NS
	hip		NS	NS	NS
Minimum distance to the forearm	trunk	*Activity instance segment*	-0.27**	-0.24**	0.28**
	thigh		NS	NS	NS
	hip		NS	NS	NS
	trunk	*Activity instance*	-0.34**	-0.18**	0.25**
	thigh		-0.16*	-0.14*	0.15*
	hip		-0.21**	-0.17*	0.18**

Range of hip and knee motion was not significant for the two timescales.
* - significant at $p < 0.05$; ** - significant at $p < 0.01$;
NS - not significant at $p = 0.05$

and without the training data boost (with test data from the EmoPain@Home dataset only in both cases). We obtained F1 scores of 0.65 and 0.56 for the lower and higher pain levels respectively with the multi-dataset training data. As can be seen, the boost improves performance albeit marginally. This improvement is despite the differences in the timescales of self-reporting and extracted features for the EmoPain@Home and EmoPain datasets.

This finding suggests that more data could improve performance on pain level classification for the EmoPain@Home dataset. It additionally highlights the possibility of combining multiple pain datasets that include motion capture data from people with chronic pain to create larger datasets. Beyond additional body movement data collected in the context of pain, novel data augmentation/generation techniques and careful use of existing body movement datasets outside the experience of pain could perhaps also be valuable.

2) Correlation Results: In further exploration of the EmoPain@Home dataset, we performed Spearman's correlation analysis. This was done using the EmoPain@Home data alone. For the correlations between constructs ($n = 504$), we found strong association with worry: confidence having the strongest correlation of -0.76 ($p < 0.01$) and the correlation for pain lower at 0.65 ($p < 0.01$). The correlation between pain and confidence was only moderate, -0.50 ($p < 0.01$). The correlations of each of these with the extracted features ($n = 226$) are shown in Table IV. We found statistically significant, although small, correlations with several features.

We found that correlation was consistently stronger for the *activity instance timescale* than the *activity instance segment timescale*, particularly for the trunk, thigh, and lower leg. This suggests that non-verbal behaviour associated with a given experience may manifest over a longer timescale than the sampling period that the verbalisation of the experience covers. Another finding that was consistent, especially for speed and amount of movement, is that while there was less significant correlation for pain intensity than for confidence levels in the shorter timescale, correlation was significant and stronger for pain than confidence in the longer timescale. This further suggests that non-verbal behaviour may be more related to pain-related confidence in the short term, but with stronger association with pain intensity in retrospect.

Further, there was generally stronger correlation for jerk than for the other sets of features. The reason for this is not clear, but we additionally found an inverse relationship between pain (or worry) and jerk, with lower jerk (and so more smoothness) associated with high pain intensity. This departs from intuition. However, it is consistent with similar findings, e.g., with acute shoulder pain in [41] which highlight relationship between jerk and speed, with higher jerk (less smoothness) associated with greater speed. People with higher levels of pain were slower and so less jerky. We found the same relationship in the EmoPain@Home dataset.

No significant correlation was found with the hip and knee range of motion features. This could be due to the wide variety in activity types which inherently involve different ranges of motion. For example, a small range of motion would be expected in the *washing up* activity, whereas *vacuuming* is more likely to be executed using a much larger range of motion. Indeed, a boxplot of range of motion across activity types showed a limited range of motion in the *washing up* activity particularly for the knee, while activities such as *yoga*, *vacuuming*, and *sweeping* had much larger ranges of motion. Visual inspection of range of motion in lower and higher pain levels by activity type showed no conclusive patterns.

V. Discussion: Recommendations Addressing Inclusivity in Data Capture outside Lab Settings

Beyond the pain dataset captured in our study, our data collection approach led to additional insights that highlight important considerations for datasets creation, especially for data capture in the wild. We focus our discussion here on things to consider for inclusivity, particularly with respect to sensor selection/design and participant training.

A. Sensor Selection

We found that certain kinds of motor manipulations (e.g. turning on the sensor by pinching in a specific way) that the sensors in our study required were difficult for the participants with (additional) upper limb pain. Such experiences are significant as ethically, the burden on the individual participant must be considered and balanced with the benefits for both the participant and the larger society. Further, sensor use difficulties could contribute to *data cascades* [45] that negatively affect the performance and use of technology, e.g. by deterring participants from (effectively) engaging with it.

Dataset creators thus need to intentionally incorporate inclusivity in their selection of sensors, together with lobbying for sensor designs that are more robustly inclusive. This will include uncovering sensor-related barriers for the specific participant populations and in the scenarios of interest. Direct observation in situ allowed us to capture some of such issues in our study. As a contrast, a similar study [44] with participants with movement impairment but not using direct observation found no population-specific sensor issues. While large scale studies cannot afford to have the researcher present in all data capture sessions as done in our study, a few of such 'supervised' sessions can be included in the data collection protocol to enable direct observation of limitations with respect to sensor inclusivity in context.

B. Participant Training

At the start of our study, we evaluated the sensors as easy to use by participants with limited technology experience, given training. Our training approach with each participant consisted of a one-on-one half-hour session in which they had hands-on exploration of how to use the sensor system. We chose this approach to minimise the burden on participants, leveraging the direct availability of the researcher to troubleshoot issues during actual data capture. However, we found that participants often forgot specific procedures at the time of data capture.

The approach in [44] using formal post-training tests to check (and require) participant understanding of how to use the sensor(s) can be valuable in addressing this, especially for large scale studies where researcher availability will be limited. However, with the approach comes ethical questions that the dataset creator needs to carefully address. For example, to what extent should participants be unable to grasp how to use the sensors to be excluded? Also, how many potential participants might drop out from being overwhelmed by a long and cumbersome briefing procedure?

These further underline the significance of sensor design that sufficiently meets a broad range of needs in the given population. Thus, a critical outcome of research should be an explicit agenda challenging designers to make sensors easily usable by a broader range of potential users, and at the same time also challenging researchers to select sensors that support this in their studies. A good sensor design should at least ensure that the right amount of information is conveyed at the right time for data collection. It could perhaps also be adaptable to available cognitive resources, e.g. through interface options (such as standard versus enhanced). Failing to drive engagement with usability issues will limit the scope and utility of affect-aware systems in the long term.

VI. Conclusion

We present the EmoPain@Home dataset of motion capture data with pain and related worry and confidence labels captured in everyday activities at home with people with chronic pain. The dataset can be accessed on request to the last author.

Our analysis shows that the dataset has a good distribution of pain, worry, and confidence levels as well as an adequate variety of activity types. Further, baseline performance of 0.61 average F1 score for continuous pain classification across different activities demonstrates that the dataset can be valuable for developing automatic assessment technology for the context of pain. This dataset thus represents a starting point in building a corpus of ecologically valid data for research and development in the area of pain. Extension of the dataset is already ongoing. Additionally, we have highlighted important considerations that can affect inclusivity as a guide for other creators aiming to expand such corpus.

The outcome of our additional analysis of the relationship between movement behaviour and pain and related experience at different timescales highlights a need to explore in more detail the timescales of influence of pain increase on movement. Insights gained from such investigations can be useful for more appropriate interpretation of movement behavior that further facilitates helpful tailoring of intervention.

REFERENCES

[1] P. Werner, D. Lopez-Martinez, S. Walter, A. Al-Hamadi, S. Gruss, and R. Picard. "Automatic recognition methods supporting pain assessment: A survey." IEEE Trans. Affect. Comput. (2019).

[2] S. Walter, S. Gruss, S. Frisch, J. Liter, L. Jerg-Bretzke, B. Zujalovic et al. "“What About Automated Pain Recognition for Routine Clinical Use?” A Survey of Physicians and Nursing Staff on Expectations, Requirements, and Acceptance." Frontiers in Medicine 7 (2020).

[3] S. Felipe, A. Singh, C. Bradley, A. Williams, and N. Bianchi-Berthouze. "Roles for personal informatics in chronic pain." In Int Conf Pervasive Comput Technol Healthc (2015), 161-168.

[4] G. Gioftsos and D. Grieve. "The use of artificial neural networks to identify patients with chronic low-back pain conditions from patterns of sit-to-stand manoeuvres." Clin Biomech 11, no. 5 (1996): 275-280.

[5] J. Bishop, M. Szpalski, S. Ananthraman, D. McIntyre, and M. Pope. "Classification of low back pain from dynamic motion characteristics using an artificial neural network." Spine 22, no. 24 (1997): 2991-2998.

[6] James P. Dickey, Michael R. Pierrynowski, Drew A. Bednar, and Simon X. Yang. "Relationship between pain and vertebral motion in chronic low-back pain subjects." Clin Biomech 17, no. 5 (2002): 345-352.

[7] Sheryl Brahnam, Chao-Fa Chuang, Frank Y. Shih, and Melinda R. Slack. "Machine recognition and representation of neonatal facial displays of acute pain." Artificial intelligence in medicine 36, no. 3 (2006): 211-222.

[8] Pazit Levinger and Wendy Gilleard. "Tibia and rearfoot motion and ground reaction forces in subjects with patellofemoral pain syndrome during walking." Gait posture 25, no. 1 (2007): 2-8.

[9] Bogdan Matuszewski, Wei Quan, and Lik-Kwan Shark. "High-resolution comprehensive 3-D dynamic database for facial articulation analysis." In ICCV Workshops (2011), 2128-2135.

[10] P. Lucey, J. Cohn, K. Prkachin, P. Solomon, and I. Matthews. "Painful data: The UNBC-McMaster shoulder pain expression archive database." In Int Conf Autom Face Gesture Recognit (2011), 57-64.

[11] S. Walter, S. Gruss, H. Ehleiter, J. Tan, H. Traue, P. Werner, et al. "The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system." In Int Conf Cybernetics (2013), 128-131.

[12] X. Zhang, L. Yin, J. Cohn, S. Canavan, M. Reale, A. Horowitz, et al. "Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database." Image Vis Comput 32, 10 (2014): 692-706.

[13] Jesús Rivas, Felipe Orihuela-Espina, Enrique Sucar, Lorena Palafox, Jorge Hernández-Franco, and Nadia Bianchi-Berthouze. "Detecting affective states in virtual rehabilitation." In International Conference on Pervasive Computing Technologies for Healthcare (2015), 287-292.

[14] Shivam Sharma, Shubham Asthana, and Vinay Mittal. "A database of infant cry sounds to study the likely cause of cry." In International Conference on Natural Language Processing (2015), 112-117.

[15] Z. Zhang, J. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, et al. "Multimodal spontaneous emotion corpus for human behavior analysis." In CVPR (2016), 3438-3446.

[16] M. S. H. Aung, S. Kaltwang, B. Romera-Paredes, B. Martinez, A. Singh, M. Cella, et al. "The automatic detection of chronic pain-related expression: requirements, challenges and the multimodal EmoPain dataset." IEEE Trans. Affect. Comput. 7, no. 4 (2015): 435-451.

[17] M. Velana, S. Gruss, G. Layher, P. Thiam, Y. Zhang, D. Schork, et al. "The senseemotion database: A multimodal database for the development and systematic validation of an automatic pain-and emotion-recognition system." In IAPR Workshop (2016), 127-139.

[18] M. Haque, R. Bautista, F. Noroozi, K. Kulkarni, C. Laursen, R. Irani, et al. "Deep multimodal pain recognition: a database and comparison of spatio-temporal visual modalities." In International Conference on Automatic Face Gesture Recognition (2018), 250-257.

[19] Temitayo Olugbade, Nadia Bianchi-Berthouze, Nicolai Marquardt, and Amanda Williams. "Human observer and automatic assessment of movement related self-efficacy in chronic pain: from exercise to functional activity." Trans. Affect. Comput. 11, no. 2 (2018): 214-229.

[20] Temitayo Olugbade, Aneesha Singh, Nadia Bianchi-Berthouze, Nicolai Marquardt, Min Aung, and Amanda Williams. "How can affect be detected and represented in technological support for physical rehabilitation?" ACM Trans Comput Hum Interact 26, no. 1 (2019): 1-29.

[21] B. Hu, C. Kim, X. Ning, and X. Xu. "Using a deep learning network to recognise low back pain in static standing." Ergonomics 61, no. 10 (2018): 1374-1381.

[22] P. Liu, I. Yazgan, S. Olsen, A. Moser, U. Ciftci, S. Bajwa, et al. "Clinical valid pain database with biomarker and visual information for pain level analysis." In Int Conf Autom Face Gesture Recognit (2018), 525-529.

[23] S. Gruss, M. Geiger, P. Werner, O. Wilhelm, H. Traue, A. Al-Hamadi, et al. "Multi-modal signals for analyzing pain responses to thermal and electrical stimuli." J Vis Exp 146 (2019): e59057.

[24] S. Brahnam, L. Nanni, S. McMurtrey, A. Lumini, R. Brattin, M. Slack, and T. Barrier. "Neonatal pain detection in videos using the iCOPEvid dataset and an ensemble of descriptors extracted from Gaussian of Local Descriptors." Appl. Comput. Inform. (2020).

[25] C. Ordun, A. Cha, E. Raff, B. Gaskin, A.r Hanson, M. Rule, et al. "Intelligent Sight and Sound: A Chronic Cancer Facial Pain Dataset." In NeurIPS Datasets and Benchmarks Track (2021).

[26] J. Smith and M. Osborn. "Pain as an assault on the self: An interpretative phenomenological analysis of the psychological impact of chronic benign low back pain." Psychol Health 22, no. 5 (2007): 517-534.

[27] C. Murray, C. Groenewald, R. de la Vega, and T. Palermo. "Long-term impact of adolescent chronic pain on young adult educational, vocational, and social outcomes." Pain 161, no. 2 (2020): 439.

[28] K. Karos, A. Williams, A. Meulders et al. "Pain as a threat to the social self: a motivational account." Pain 159, no. 9 (2018): 1690-1695.

[29] K. Prkachin and P. Solomon. "The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain." Pain 139, no. 2 (2008): 267-274.

[30] M. Nickel, E. May, L. Tiemann, M. Postorino, S. Dinh, and M. Ploner. "Autonomic responses to tonic pain are more closely related to stimulus intensity than to pain intensity." Pain 158, no. 11 (2017): 2129-2136.

[31] Paul Watson, Kerry Booker, and Chris Main. "Evidence for the role of psychological factors in abnormal paraspinal activity in patients with chronic low back pain." J. Musculoskelet. Pain 5, no. 4 (1997): 41-56.

[32] Philipp Werner, Ayoub Al-Hamadi, Kerstin Limbrecht-Ecklundt, Steffen Walter, and Harald C. Traue. "Head movements and postures as pain behavior." PloS one 13, no. 2 (2018): e0192767.

[33] M. Sullivan, P. Thibault, A. Savard, R. Catchlove, J. Kozey, and W. Stanish. "The influence of communication goals and physical demands on different dimensions of pain behavior." Pain 125, no. 3 (2006): 270-277.

[34] Carlo V. Bellieni, Renata Sisto, Duccio M. Cordelli, and Giuseppe Buonocore. "Cry features reflect pain intensity in term newborns: an alarm threshold." Pediatric research 55, no. 1 (2004): 142-146.

[35] Michal Icht, Hadar Wiznitser Ressis-tal, and Meir Lotan. "Can the Vocal Expression of Intellectually Disabled Individuals Be Used as a Pain Indicator? Initial Findings Supporting a Possible Novice Assessment Method." Frontiers in Psychology (2021): 2926.

[36] C. Wang, Y. Gao, A. Mathur, A. Williams, N. Lane, and N. Bianchi-Berthouze. "Leveraging activity recognition to enable protective behavior detection in continuous data." In IMWUT 5, no. 2 (2021): 1-27.

[37] Mark P Jensen, and Paul Karoly. "Self-report scales and procedures for assessing pain in adults." In Dennis C. Turk Ronald Melzack (Eds.), Handbook of pain assessment, 3rd ed., Guilford Press (2011): 19-44.

[38] Norman Cliff, and John A Keats. Ordinal measurement in the behavioral sciences, New York: Psychology Press (2003).

[39] T. Bi, R. Buono, T. Olugbade, A. Singh, C. Holloway, E. Costanza, et al. "Towards Chatbot-Supported Self-Reporting for Increased Reliability and Richness of Ground Truth for Automatic Pain Recognition: Reflections on Long-Distance Runners and People with Chronic Pain." In Int Conf Multimodal Interact Workshop (2021), 43-53.

[40] Chongyang Wang, Temitayo Olugbade, Akhil Mathur, Amanda Williams, Nicholas Lane, and Nadia Bianchi-Berthouze. "Recurrent network based automatic detection of chronic pain protective behavior using mocap and semg data." In ISWC (2019), 225-230.

[41] J. C. Slaboda, J. R. Boston, T. E. Rudy, S. J. Lieber, D. M. Rasetshwane, and D. K. Weiner. "Effects of pain on shoulder jerk in elderly subjects during repetitive lifting." In EMBC (2003), 1728-1731.

[42] L. Breiman. "Bagging predictors." Mach. Learn. 24, 2 (1996): 123-140.

[43] Leo Breiman, Jerome Friedman, Richard Olshen, and Charles Stone. "Classification and regression trees". New York: Routledge (2017).

[44] S. Rutkove, K. Qi, K. Shelton, J. Liss, V. Berisha, and J. Shefner. "ALS longitudinal studies with frequent data collection at home: study design and baseline data." Amyotroph Lateral Scler Frontotemporal Degener 20, no. 1-2 (2019): 61-67.

[45] N. Sambasivan, S. Kapania, H. Highfill, D. Akrong, P. Paritosh, and L. Aroyo. "“Everyone wants to do the model work, not the data work”: Data Cascades in High-Stakes AI." In CHI (2021): 1-15.