Automatic Recognition of Low-Back Chronic Pain Level and Protective Movement Behaviour using Physical and Muscle Activity Information

Fasih Haider, Pierre Albert and Saturnino Luz
Usher Institute, Edinburgh Medical School, the University of Edinburgh, Scotland, United Kingdom

Abstract— **Automatic recognition of low-back chronic pain and movement behaviour in humans could be a useful technology in health monitoring and providing effective rehabilitation advice. Physical and muscle activity information can be used in automating this process in combination with machine learning and feature engineering methods. This paper presents a method for automatic recognition of chronic pain and movement behaviour using our recently proposed 'Active Data Representation' (ADR) method, and applies it to two tasks of the EmoPain 2020 Challenge using physical and muscle activity features. The ADR method is used for the transformation of the physical and muscle activity features for the classification tasks. Our results show that ADR outperforms the LSTM challenge baseline model in terms of Matthews correlation coefficient (0.43) and F score (61.21) for the recognition of chronic pain and movement behaviour respectively in hold-out validation settings. Although a decrease in performance is observed on the test dataset, ADR still outperforms the challenge baseline for the recognition of chronic pain and movement behaviour tasks.**

I. INTRODUCTION

Chronic Low-Back Pain (CLBP) is a continuous pain unrelated to a specific injury, with no existing lesion or persisting past healing. Physical activity is essential for rehabilitation [3] but subjective aspects (anxiety, perceived exacerbation, etc.) can hinder regular physical activity with negative consequences[15]. Psychological aspects are crucial to patient engagement in exercise plans, and should be taken into account in the provision of rehabilitation advice [14]. Detection of the patient's perception is essential in this effort to improve rehabilitation plans.

Avoidance of physical activity is expressed through protective behaviour, also referred as guarded movements – i.e. body movements aimed at avoiding strain [15]. Automatic detection of such movements can enable the development of support tools complementing the physiotherapist's intervention, in clinical settings or for self-management. Movement-based automatic recognition of protective behaviour is performed at different temporal scopes, namely: sequence labelling, which concerns overall classification, and frame by frame labelling, which aims to characterise specific movement. Posture and movements are recorded using sensors on the whole body. Movements are decomposed into frames, and joints angles and energy values are extracted for each time step.

In an overview of the literature, Aung et al. [1] found that although pain expression classification has been studied, body expression recognition has been neglected. Classification of pain related expression from body movement is based on the work in recognition of affective states from body movements and posture [6]. This work was expanded in an exploratory study after the collection of the EmoPain data set [1]. Random Forest based classification was performed on the main combinations of labels/exercises, suggesting contextual factors in the differences observed between the combinations themselves and between the type of exercise (instructed vs non instructed). In a follow-up study, a deep learning architecture (BodyAttentionNet) [16] was developed to capture spatial and temporal cues. Compared to the state of the art, the system drastically improved classification performance (F1-score: 0.572 vs 0.844). Compared to other LSTM-based architectures the BodyAttentionNet neural network required a much lower number of parameters (2.1k vs 40.9k). Separating the results of the spatial and temporal subsystems suggested a more important role for the former, although the combined performance hints to their complementarity.

Pain classification is based on similar body features. Initial classification using three pain levels (control, low, high) with Support Vector Machine on body motion and muscular activity combined with feature selection [11] showed good results. Feature sets of both modalities achieved similar F1 scores (movement: 0.63, muscles: 0.69), while their combination lead to a large improvement (F1 = 0.8). Investigation on the inclusion of a depression score as an additional input feature [12] did not improve the performances, however the results further stressed the prominence of the context (movement and type of exercise) in the classifier's performance. In their study on the relevance of features based on linear mixed model analysis, Olugabe et al. [13] further investigated the discriminative power of specific body features. The resulting optimised feature set allowed further improvement. Additionally, the study investigated the potential of ubiquitous monitoring through the use of a minimal set of features from low-cost sensors. In this last experiment, they achieved an F1-score of 0.78 on a reduced two-level pain classification.

The main contribution of the study presented in this paper is the demonstration of our recetly proposed Active Date Representation (ADR) [5], [4] method for pain and movement behaviour recognition tasks using physical and muscle activity information. The evaluation of the ADR method against the challenge baselines which are set using support vector machines for the pain recognition task, and stacked-Long Short-Term Memory (LSTM) for the movement behaviour recognition task.

II. Data Set Description

The movement challenge data set used for the 'Pain Recognition' and 'Movement Behaviour Classification' is based on the EmoPain dataset [1]. It contains movements from 30 participants carrying out physical activity, described by two types of features: full body motion capture and muscle activity. Body motion is tracked though bodily joints (13 angles and corresponding energies) and the muscle activity is tracked using Surface Electromyography (sEMG), a non-invasive method to record the electric activity of muscles. The dataset for pain recognition Task labels each movement for chronic pain into one of three levels: *none*, *low*, or *high*. The dataset for the movement behaviour recognition task is provided with continuous binary labels of protective behaviour (PB) for each of the 180 frames constituting a movement. PB labels were generated in the EmoPain data set from the fusion of six (5 + null) behaviour categories temporally segmented by experts in CLBP. The data set is slightly imbalanced (≈60/40) in terms of number of subjects, featuring 18 participants with CLBP and 12 healthy participants. The population was divided randomly into three sets (see table I).

TABLE I

Movement challenge dataset.

Set	Participants - total	CLBP	Healthy
Training	16	10	6
Validation	7	4	3
Test	7	4	3

III. Experimentation

This section describes the features sets of the of EmoPain Data for pain and movement behaviour recognition tasks, the training of the feature extraction model, the generation of a feature vector for the recognition tasks, the classification methods and the evaluation metrics.

A. Feature Sets

Each frame is a single data vector at each time step containing 30 features: 13 joint angles, 13 joint energies and 4 electromyography from lower and upper back. In total the data consists of 514,545 frames, with 356,107 and 158,438 in the training and validation sets respectively. The Classification is performed on time intervals (T), i.e. sequences of frames. The challenge organisers use a different time interval for each task, as follows:

Pain Recognition from Movement $(I \times T \times d)$

1) I = number of instances of exercise;
2) T = number of frames in each segment (variable, depending on the length of the exercise);
3) d = number of dimensions of each frame (30).

Movement Behavior Classification: $(W \times T \times d)$

1) W = total number of window segments over all exercise instances;
2) T = number of frames in each segment (180);
3) d = number of dimensions of each frame (30).

B. Active Data Representation

In this section, we describe our active data representation method briefly [4], [5]. This is the first study which evaluates the ADR using physical and muscle activity information. Previous studies [4], [5] used ADR with audio and visual information only. The ADR method involves the following steps:

1) *Clustering* of frames: Self-Organising Maps (SOM) [7] are employed for clustering of all the frames using 30 features. The number of clusters was determined through a grid search hold-out-validation procedure with a hyperparameter space of $m \in \{5, 10, \ldots, 100\}$. An example of clustering (i.e. feature extraction model) is shown in Figure 1.
2) *Generation* of the Active Data Representation (ADR_{Ai}) vector is done by first calculating the number of segments in each cluster for each I or W (Ai), that is, creating a histogram of the number of frames ($nADR_{Ai}$) present in each of the m clusters for each I/W. Then, to model temporal dynamics we calculate the mean and standard deviation of the rate of change with respect to the clusters associated with the frames for each I/W ($cADR_{Ai}$), where the rate of change is given by an approximation of the derivative
$$vADR_{Ai} = \frac{\partial cADR_{Ai}}{\partial t},$$
with respect to time (t).
$$nADR_{Ai_{norm}} = \frac{nADR_{Ai}}{\|nADR_{Ai}\|_1} \quad (1)$$
3) *Fusion*: the $ADR_{Ai_{norm}}$ feature set encompasses the features of $nADR_{Ai_{norm}}$, and $vADR_{Ai}$. Therefore a feature vector with dimensionality of $m+2$ is generated to represent each instances ($W or I$) for classification.

C. Classification Methods

The classification experiments were performed using four different methods, namely decision trees (DT, with leaf size of 20), nearest neighbour (KNN with K=1), linear discriminant analysis (LDA) and Random Forest (RF, with leaf size of 30 and 250 number of trees for pain recognition task, and with leaf size of 2 and 12 number of trees for movement behaviour recognition task). The classification methods are implemented in [9] using the statistics and machine learning toolbox. A hold-out validation procedure was adopted, where the training data do not contain any information of validation subjects.

The pain recognition task is a three class problem, where the classifier aims to assign an instance to one of the following categories:

1) 0: Healthy,
2) 1: Low-level pain,
3) 2: High-level pain.

The movement behaviour recognition task is a two class problem, with the following classes:

1) 0: Not protective,
2) 1: Protective.

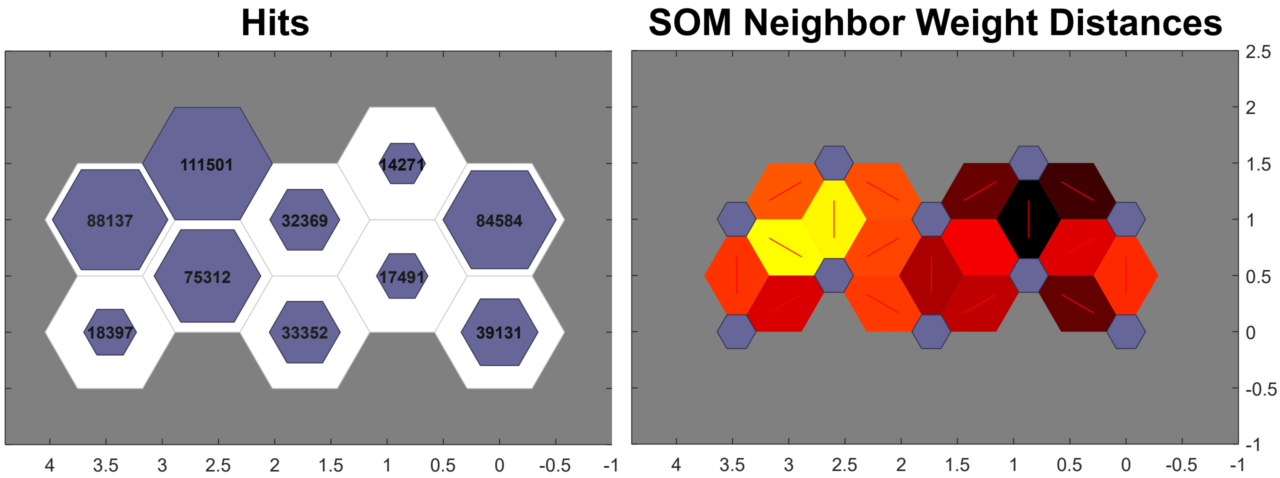

Fig. 1. ADR feature extraction model with $m = 10$ which provides the best result of 0.43 (MCC) for Pain Recognition task with RF classifier. Left figure indicates the number of frames present in each cluster (hexagon i.e. neuron) and right figure indicates the distance between clusters (blue dots i.e. neurons) and darker color indicates greater distance between clusters. The red lines connect neighboring neurons.

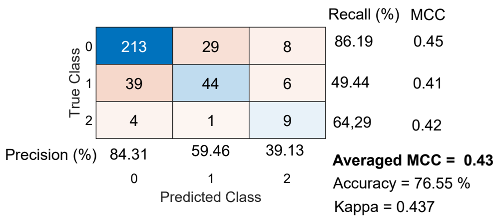

Fig. 2. Pain Recognition: Hold-out validation Results

TABLE II

RESULTS OF PAIN RECOGNITION TASK IN HOLD OUT VALIDATION: MCC

		Base-SVM	Base-KNN	LDA	KNN	DT	RF
	m	-	-	10	10	35	10
Valid.	class (0)	-	-	0.36	0.32	0.16	0.45
	class (1)	-	-	0.47	0.24	0.25	0.41
	class (2)	-	-	0.29	-0.06	0.11	0.42
	average	0.19	0.05	0.38	0.16	0.18	**0.43**
test	class (0)	-	-0.04	0.14	-	0.16	0.23
	class (1)	-	-0.06	0.03	-	-0.09	-0.01
	class (2)	-	0.16	0.03	-	0.16	0.14
	average	-	0.02	0.07	-	0.08	**0.12**

D. Evaluation Matrices

To assess the classification results, we used the average of Matthews Correlation coefficient (MCC) [10] for pain recognition task as shown in Equation 2 and Averaged F_{Score} for movement behaviour recognition task in hold-out-validation setting.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (2)$$

IV. RESULTS AND DISCUSSION

This section describes the results of our machine learning models for pain and movement behaviour recognition task of the EmoPain challenge.

A. Results: Pain Recognition

The MCC scores for the pain recognition task are shown in Table II. These results indicate that the RF (the feature extraction model is shown in Figure 1) provides the best averaged MCC (0.43) among the tested classifiers. The results also indicate that LDA provides the best MCC of 0.47 for class (1), while RF provides the best results for class (0) and class (2) with an MCC of 0.45 and 0.42 respectively. For further insight, the confusion matrices of the best results is shown in Figure 2 along with precision and recall scores of each class, overall accuracy, UAR and Kappa scores [8]. We have submitted the top three results (i.e. RF, LDA and DT) for independent testing to the challenge organisers. The results of these tests are also shown in Table II.

B. Results: Movement Behaviour Recognition

The F scores for the movement behaviour task are shown in Table III. The results indicate that the RF (0.61) provides better averaged F_{Score} (the feature extraction model is shown in Figure 3) than the other classifiers, exceeding also the challenge baseline of 0.48 [2]. RF provides the best scores for classes (0) and (1). The confusion matrix of the best results along with per-class and overall performance scores is shown in Figure 4. As in the provious task, the top three results (i.e. RF, LDA and DT) were submitted to the challenge organisers for testing, and the results are shown in Table III.

TABLE III

RESULTS OF MOVEMENT BEHAVIOUR RECOGNITION TASK ON VALIDATION AND TEST DATA : $F_{Score}(\%)$

		Base-LSTM	LDA	KNN	DT	RF
	m	-	75	25	15	25
Valid.	class (0)	96.22	95.98	94.55	95.06	96.77
	class (1)	-	25.13	21.40	23.01	25.64
	Averaged	48.11	60.56	57.97	59.03	**61.21**
test	class (0)	90.29	92.01	-	91.64	93.40
	class (1)	24.65	21.63	-	24.52	18.57
	Averaged	57.45	56.82	-	***58.08***	55.98

C. Discussion

In order to visualise the classification model in relation to the input features generated by the ADR method we drew

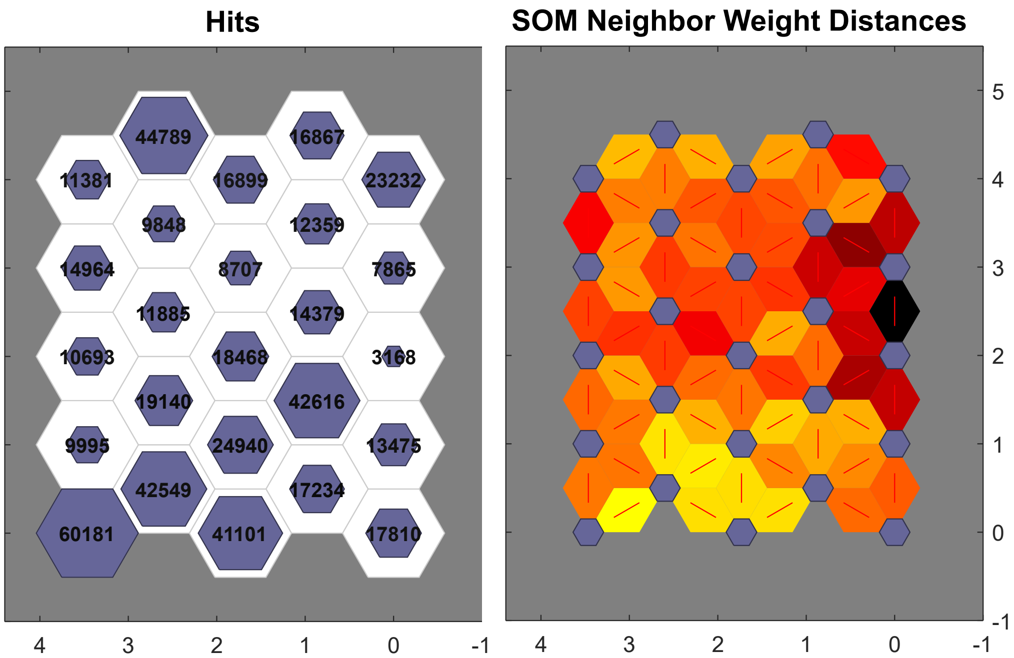

Fig. 3. ADR feature extraction model with $m = 25$ which provides the best result of 0.6121 (F_{Score}) for Movement Behaviour Recognition task with RF classifier. Left figure indicates the number of frames present in each cluster (hexagon i.e. neuron) and right figure indicates the distance between clusters (blue dots i.e. neurons) and darker color indicates greater distance between clusters. The red lines connect neighboring neurons.

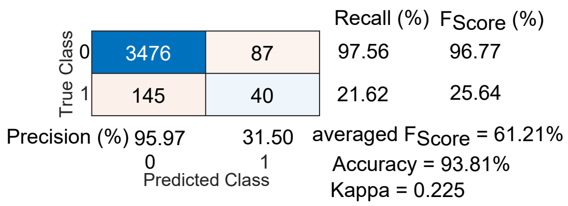

Fig. 4. Movement Behaviour Recognition: Hold-out Validation Results.

a decision tree from the RF classifier. This is shown in Figure 5. The ADR is generated with different numbers of clusters (m), and $m = 10$ provides the best results (MCC of 0.43) for pain recognition task. The dimensionlity of the ADR is $m + 2$ which is represented in Figure 5 as $x1, x2, x3, \ldots, x12$. Where $x11$ and $x12$ represent $vADR_{Ai}$ and $x1, x2, \ldots, x10$ represent $nADR_{Ai_{norm}}$.

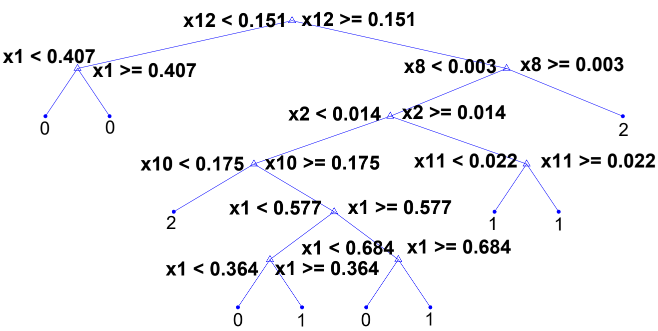

Fig. 5. Pain Recognition: An example of a tree from the RF which provides the best MCC of 0.43. In this case,the ADR has a dimensionality of 12 with m =10.

In the baseline study [2], the authors use KNN and SVM for the pain recognition tasks. They reported an MCC of 0.05 and 0.19 for KNN and SVM respectively in LOSOCV settings on training data. However a decrease in MCC is observed for KNN (0.02) and SVM on the test data. In this study, we performed hold-out validation. For movement behaviour recognition task, the authors also use hold-out validation and the stacked-LSTM algorithm for classification. The reported results are almost close to blind guess (i.e. $F_{Score} = 48.11\%$ on validation and $F_{Score} = 57.45\%$ on test). This is a very challenging machine learning problem as the classes of the data-set are highly imbalanced.

V. Conclusions

This paper presented the results of the 'Active Data Representation' (ADR) method for lower-back chronic pain and protective movement behaviour recognition tasks, as part of the EmoPain challenge. The results reported in this paper outperform the baseline results provided with an MCC of 0.43 and an averaged F_{Score} of 61.21% on the validation dataset, and an MCC of 0.12 and averaged F_{Score} of 58.08% on test data. In future we intend to evaluate the performance of the ADR method for multiple feature sets and compare the results with other feature extraction methods such as VGGNet and GoogleNet.

VI. Acknowledgments

This research is funded by the European Union's Horizon 2020 research programme, under grant agreement No 769661, towards the SAAM project. PA is supported by the Medical Research Council (MRC).

REFERENCES

[1] Min S. H. Aung, Sebastian Kaltwang, Bernardino Romera-Paredes, Brais Martinez, Aneesha Singh, Matteo Cella, Michel Valstar, Hongying Meng, Andrew Kemp, Moshen Shafizadeh, and et al. The automatic detection of chronic pain-related expression: Requirements, challenges and the multimodal emopain dataset. *IEEE Transactions on Affective Computing*, 7(4):435–451, Oct 2016.

[2] Joy Egede, Siyang Song, Temitayo Olugbade, Chongyang Wang, Hongyin Meng, Min Aung, Nicholas D. Lane, Amanda C. De C. Williams, Michel Valstar, and Nadia Bianchi-Berthouze. Emopain challenge 2020: Multimodal pain evaluation from facial and bodily expressions. *arXiv preprint*, 2020. arXiv:2001.07739.

[3] Robert J. Gatchel, Yuan Bo Peng, Madelon L. Peters, Perry N. Fuchs, and Dennis C. Turk. The biopsychosocial approach to chronic pain: scientific advances and future directions. *Psychological bulletin*, 133(4):581, 2007.

[4] Fasih Haider, Sofia De La Fuente, and Saturnino Luz. An assessment of paralinguistic acoustic features for detection of alzheimer's dementia in spontaneous speech. *IEEE Journal of Selected Topics in Signal Processing*, 2019.

[5] Fasih Haider, Maria Koutsombogera, Owen Conlan, Carl Vogel, Nick Campbell, and Saturnino Luz. An active data representation of videos for automatic scoring of oral presentation delivery skills and feedback generation. *Frontiers in Computer Science*, 2:1, 2020.

[6] Andrea Kleinsmith and Nadia Bianchi-Berthouze. Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing*, 4(1):15–33, 2012.

[7] Teuvo Kohonen. The self-organizing map. *Neurocomputing*, 21(1-3):1–6, 1998.

[8] J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.

[9] MATLAB. *version 9.6 (R2019a)*. The MathWorks Inc., Natick, Massachusetts, 2019.

[10] B.W. Matthews. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 405(2):442 – 451, 1975.

[11] Temitayo A. Olugbade, M.S. Hane Aung, Nadia Bianchi-Berthouze, Nicolai Marquardt, and Amanda C. Williams. Bi-modal detection of painful reaching for chronic pain rehabilitation systems. In *Proceedings of the 16th International Conference on Multimodal Interaction*, ICMI '14, page 455–458. Association for Computing Machinery, Nov 2014.

[12] Temitayo A. Olugbade, Nadia Bianchi-Berthouze, Nicolai Marquardt, and Amanda C. Williams. Pain level recognition using kinematics and muscle activity for physical rehabilitation in chronic pain. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, page 243–249, Sep 2015.

[13] Temitayo A. Olugbade, Aneesha Singh, Nadia Bianchi-Berthouze, Nicolai Marquardt, Min S. H. Aung, and Amanda C. De C. Williams. How can affect be detected and represented in technological support for physical rehabilitation? *ACM Transactions on Computer-Human Interaction (TOCHI)*, 26(1):1:1–1:29, Jan 2019.

[14] Dennis C. Turk and Akiko Okifuji. Psychological factors in chronic pain: Evolution and revolution. *Journal of consulting and clinical psychology*, 70(3):678, 2002.

[15] Johan WS Vlaeyen and Steven J. Linton. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain*, 85(3):317–332, 2000.

[16] Chongyang Wang, Min Peng, Temitayo A. Olugbade, Nicholas D. Lane, Amanda C. De C. Williams, and Nadia Bianchi-Berthouze. Learning bodily and temporal attention in protective movement behavior detection. *arXiv:1904.10824 [cs, stat]*, Jul 2019. arXiv: 1904.10824.