Learning Bodily and Temporal Attention in Protective Movement Behavior Detection

Chongyang Wang, Min Peng, Temitayo A. Olugbade, Nicholas D. Lane, Amanda C. De C. Williams, Nadia Bianchi-Berthouze

1UCL interaction centre, University College London, London, United Kingdom
2Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Science, Chongqing, China
3Department of Clinical Health, University College London, London, United Kingdom
4Department of Computer Science, University of Oxford, Oxford, United Kingdom

chongyang.wang.17@ucl.ac.uk, pengmin@eigcit.ac.cn, temitayo.olugbade.13@ucl.ac.uk, nicholas.lane@cs.ox.ac.uk, amanda.williams@ucl.ac.uk, nadia.berthouze@ucl.ac.uk

Abstract—For people with chronic pain (CP), the assessment of protective behavior during physical functioning is essential to understand their subjective pain-related experiences (e.g., fear and anxiety toward pain and injury) and how they deal with such experiences (avoidance or reliance on specific body joints), with the ultimate goal of guiding intervention. Advances in deep learning (DL) can enable the development of such intervention. Using the EmoPain MoCap dataset, we investigate how attention-based DL architectures can be used to improve the detection of protective behavior by capturing the most informative biomechanical cues characterizing specific movements and the strategies used to execute them to cope with pain-related experience. We propose an end-to-end neural network architecture based on attention mechanism, named BodyAttentionNet (BANet). BANet is designed to learn temporal and body-joint regions that are informative to the detection of protective behavior. The approach can consider the variety of ways people execute one movement (including healthy people) and it is independent of the type of movement analyzed. We also explore variants of this architecture to understand the contribution of both temporal and bodily attention mechanisms. Through extensive experiments with other state-of-the-art machine learning techniques used with motion capture data, we show a statistically significant improvement achieved by combining the two attention mechanisms. In addition, the BANet architecture requires a much lower number of parameters than the state-of-the-art ones for comparable if not higher performances.

Keywords—chronic pain, protective behavior, body movement, neural network, attention mechanism

I. INTRODUCTION

Physical rehabilitation is important for the management of chronic pain (CP) [4] [5] [8]. Fear of pain and/or injury results in people with CP reducing physical activity or using movement strategies (e.g., guarding, stiffness, hesitation, bracing) [1] [2], collectively called protective behaviors [3]. Such behaviors cause further debilitation and reduced participation in valued activities, e.g. employment or social life [2] [4], so are important targets for intervention. For example, in clinical settings, physiotherapists modify psychological support, feedback, and exercise or movement to change the CP patient’s specific behavior [7] [8].

As physical rehabilitation for CP moves from clinical settings to self-management, ubiquitous technology is targeted as providing support in lieu of a physiotherapist’s affect-based personalized support [9]. In order to do so, it is essential that technology automatically assesses movement and detects protective behavior. Studies of automatic analysis of emotion-influenced movement behavior are rare, as most work in the pain scenario focuses on facial expression of pain or physiological responses to acute pain [10] [11]. This is partly due to scarce data on this aspect and possibly also to limited appreciation, in the computing field, of the importance of bodily behavior over facial expression in providing information about a person’s psychological capability to manage his/her condition [5] [12].

To drive the research on protective behavior, the EmoPain dataset [6] was created, which includes full-body motion capture data for 26 anatomical joints and surface electromyography (sEMG) data collected from 4 locations on the back. Participants were 24 people with chronic lower back pain (CLBP) and 30 people without chronic pain. Recorded movements included: One-leg-stand, Stand-to-sit, Sit-to-stand, Reach-forward and Bend - typical everyday
movements that are generally challenging for those with CLBP [7] [9]. Previous studies based on the EmoPain dataset mainly employed feature engineering approaches [6] [13] [14] and vanilla neural networks [15] on the motion capture and sEMG data, where the dynamic biomechanics of movements are only used to guide feature design. Unlike acute rehabilitation where a gold-standard movement trajectory (and its deviation) informs intervention, in chronic pain, fear of injury, fear of pain, and anxiety lead the person to engage body parts in ways that are not biomechanically necessary or efficient but may increase sense of control and reduce fear. Thus, in this paper we present an end-to-end neural network architecture called BANet that can, across different types of movement, self-learn when (temporal attention) and what (bodily attention) subsets of the anatomical joints contribute most to detection of protective behavior. With the analysis of the learned attention scores, we explore how the various behaviors observed lead to the network shifting attention to body parts as necessary.

To avoid ambiguity, we partially adopt the definition of EmoPain data terms provided in [15]. Specifically, ‘sample’ refers to a single data vector at each single timestep (1/60 of a second as the sensor captured at 60Hz), ‘segment’ refers to a small data window containing several samples within a data instance, ‘instance’ refers to the full data sequence containing all the movements (sit-to-stand, bent, etc.) performed by a subject during one trial (each participants underwent two trials under two difficulty levels, for details see [6]). Given the five types of protective behaviors (hesitation, guarding, stiffness, bracing and support) [6] and limited volume of the full dataset, we use the experience learned from [15] about treating them as one special category called **protective behavior**. Furthermore, each attention mechanism is built on local joint angles and its angular velocity (referred to as energy) calculated using three relevant anatomical points (e.g., the angle formed by joining the neck, knee and ankle joints). The selection of joint angles is based on previous studies [6][27]. The contribution of this work can be summarized as:

- We propose a novel deep learning architecture performing end-to-end temporal and bodily attention learning given data streams collected from different joint angles. Here, we focus only on MoCap data, but the network can be easily adapted to sensors of different types (e.g., EMG) or based on various spatial locations across the body.
- Through a range of experiments using the EmoPain dataset, we demonstrate that our method can achieve state-of-the-art results, if not slightly higher, with fewer trainable parameters for the detection of protective behavior.
- With attention score visualization and analysis, we discuss how such mechanisms could help better understand protective behavior from real-life measurements rather than just lab-based observations.

II. RELATED WORKS

In this section, we first present the state-of-the-art on the automatic detection of protective behavior in chronic pain. Then we review studies on attention mechanisms, especially those dealing with Human Activity Recognition (HAR) using wearable sensors.

A. Automatic Detection of Protective Behavior

An earlier study on the automatic detection of protective behavior was conducted by Aung et al. [6], using Random Forest (RF), a traditional machine learning algorithm [16], on features hand-crafted from body movement data (the EmoPain dataset). The features extracted from these were 13 joint angles (based on location of incidental points and two immediate neighbors) and the corresponding energies (defined as the sum square of angular velocity, assuming unit mass). These, together with the means of sEMG data, were used to predict protective (movement) behavior in each instance, with mean square error between 0.019 to 0.034 (mean = 0.44, standard deviation = 0.16) for different movement types.

A more recent study by Wang et al. [15] explored the use of recurrent neural networks considering the time dimension of movement and still using angles, energies and sEMG data from [6]. Their protect-LSTM architecture produced higher performance than state-of-the-art deep learning architectures in detecting protective behavior over sliding windows of different movement types, rather than having to build movement-dependent models as in [6]. To address the higher data demand typically required by neural network algorithms, data augmentation was carried out on training segments by adding Gaussian noise and randomly discarding data points, and obtained an F1-score of 0.815, a level appreciably better than using the original dataset. Experimentation also showed that 3 seconds with 0-padding was the optimal sliding-window setting parameter for protective behavior detection in the EmoPain dataset. Their study suggests the applicability of LSTM networks to the detection of protective behavior, based on motion capture and sEMG data. The limitation of both [6] and [15] is that they consider all body-parts with equal importance across time and movement whilst protective behavior may occur in a specific section of the movement and involve only a specific set of body parts, possibly different across movement types and partially different across people suffering from CP. Hence, by processing the full-body motion capture (and sEMG) data in a traversal manner, redundant information and un-informative data are retained, possibly constituting noise thus reducing the performance of the model. In our work, we aim to let the architecture focus on the relevant biomechanics of protective behavior strategies by integrating attention mechanisms with LSTM. Further, we aim to detect protective behavior without prior knowledge of the type of movement performed as in [15]. We also aim to exploit motion capture data only for two reasons. We want to better understand how the model shifts between different body areas that are easily visually understood and compare them with physiotherapists’ observations who do not rely on sEMG during typical consultations. Second, we are keen to understand what such modality alone (without sEMG) can achieve using attention mechanisms with respect to the state of the art. Previous studies [15] [33] have shown that sEMG data are critical for high performance. We leave exploration of the combination of the two types of sensor data for future work.

B. Attention Mechanism for HAR with wearable sensors

Attention mechanisms have recently been explored to improve HAR performance from movement data. A common scenario in HAR is the acquisition of data with motion capture sensors (e.g., gyroscope and accelerometer)
attached to different body locations. To encapsulate understanding of the relevance of each sensor in this scenario, Zeng et al. [19] proposed an attention-based LSTM framework, where a sensor attention module was used at the input level for each timestep, with an additional temporal attention module at a further layer. Their sensor attention module was implemented with a softmax and bilinear function with input from different sensors at single timestep, while tanh and softmax activation functions were together applied to compute the temporal attention based on the output of the LSTM layer. This improved performance on three HAR benchmarks (PAMAP2 [22], DG [23] and Skoda [24]). Visualization of the attention scores also matched the expectation of subset learning of sensors at important moments. Along the same lines, Murahari et al. [20] focused on temporal attention which they embedded at the end of a convolutional LSTM network (Conv-LSTM) [25]. Similar to [19], they used tanh and softmax functions with the difference being that they used the weighted sum of all previous LSTM hidden states to the last one (the output of Conv-LSTM) for classification [19]. A related study by Yao et al. [21] was motivated by the problem of sensor reliability in mobile sensing, where multiple sensors are deployed at same time but only a hidden subset provides reliable information. The deep sense framework [26] used employed a convolutional neural network at a lower level to extract information from each sensor at each timestep and a Gated Recurrent Unit (GRU) network to learn the temporal dynamics through all timesteps at a higher level. They further used two softmax functions to compute the attention score specific to sensor attention at lower level and temporal attention at higher level. This led to better performance on the HAR dataset [32] compared with results achieved with the original DeepSense framework.

This suggests that explicitly designing an attention mechanism can help a neural network better learn patterns in data from multiple sources (e.g., multiple anatomical points). However, we noticed two key limitations: i) sensor attention and temporal attention are computed based on different scales of information, i.e., the computation of sensor attention with the low-level input data at each single timestep and temporal attention with the output of LSTM/GRU layer across a prior of time spanning multiple timesteps, which created a gap between the two attention results making it inappropriate to combine them together directly; ii) computing sensor attention straight from the input data at a single timestep is questionable as the time attended to is then too local. We argue that the relevance of a sensor (joint angles in our case) can only be understood over a time period (i.e. over a movement segment), rather than at single timesteps. Further, where in HAR activities can perhaps be recognized based on temporally-local relationships, protective behavior detection is a higher-level analysis that may thus require a longer period of perception before a clear judgment can be made, even more because of the variability in the way people express protection from harm or pain. As the size of the dataset for protective behavior is much smaller than typical HAR datasets, more care needs to be taken in designing the learning network.

III. METHODOLOGY

As discussed earlier, we propose BANet an end-to-end neural network integrating attention mechanisms, for the protective behavior detection. In this section, the BANet architecture is presented. We also describe the data preparation used including its segmentation and augmentation as well as important experimental settings.

A. The BodyAttentionNet (BANet) Architecture

An overview of BANet is given in Fig. 1. The input to the BANet is a $2 \times 13$ low-level movement data (angles and energies from 13 body parts as described later), for each sample in a movement segment. A shared vanilla LSTM subnetwork is used to extract the temporal information independently for each of the 13 body parts, i.e. given a data segment, the output of such temporal decoder is a matrix of hidden states $H^c_t = [H_1^c, ..., H_{13}^c]$ with $H_1^c_t = [H_1^{c,1}, ..., H_{13}^{c,1}]$, where $C \in \{1, 2, ..., 13\}$ for 13 body parts, $T = 1, 2, ..., t$ is the temporal length of input data segment, $K = 1, 2, ..., k$ is the number of hidden units in the temporal decoder.

The attention mechanism of BANet is implemented in two stages: a temporal attention module for each body part following at a higher-level by global body attention module. As we can see in Fig. 1, the computation of temporal and body attention is originated from the same information source, the temporal information extracted with LSTM network from each body part. The two attention modules are described below:

1) Temporal Attention. To learn the attention score $a_t^c$ for $H_t^c$, we use a $1 \times 1$ convolutional layer followed by a softmax layer to compute:

$$a_t^{c,k} = \text{SoftMax}(W_a \ast H_t^{c,k})$$

$$\text{SoftMax}(x_t) = \frac{\exp(x_t)}{\sum_{k=1}^{K} \exp(x_t)}$$

where $W_a$ is a learnable weight matrix, and $\ast$ is the convolution operation. Fig. 2 (above) illustrate such computation. Unlike a fully-connected layer, the $1 \times 1$ convolution layer acts as a linear embedding which limits irrelevant connections among the input segment (in our case is the connection of samples at different timesteps within a segment). Thus, the $1 \times 1$ convolution layer can also help to minimize the size of parameters to be trained. Nevertheless, we also experiment with the fully connected layer to compare the performances of the two alternatives. The temporal attention further includes a merge of this attention score with the original output of the temporal decoder through a multiplication followed by a sum over samples:

$$H_t^c = \sum_{t=1}^{T} a_t^{c,k} H_t^{c,k}$$

The output of the temporal attention module is a matrix of the weighted-sum of the temporal information from the 13 body parts, which can be written as $H_t^c = [(H_1^1, ..., H_1^1), ..., (H_{13}^1, ..., H_{13}^1)]$.

Fig. 2. Above: temporal attention block; Below: the body attention block.
2) **Body Attention.** So far, the network has processed the data segment separately for each body part. A global attention score for each body part is now computed, based on the output from temporal attention layer. This is in order to learn the subset of the body parts that play a key role in the detection of protective behavior during a given movement segment. For this stage, we use two fully-connected layers with tanh and softmax activations respectively:

\[
e^K = \tanh(W_p H^K) \tag{4}
\]

\[
b^K = \frac{\exp(e^K)}{\sum^K_{i=1} \exp(e^K)} \tag{5}
\]

where \(W_p\) is a learnable weight matrix. The body attention module is completed with a merge of the body attention scores with the original output of the temporal attention module using a multiplication operation:

\[
\text{atten} H^K = b^K \circ H^K \tag{6}
\]

The **attention-over-attention** structure of BANet finally produces a \(K \times 13\) matrix \(\text{atten} H^K\) which encodes the importance of each body parts at important moments (samples) for each input segment. With such output, the detection is completed by a final fully-connected layer using softmax activation.

**B. Data Preparation and Experimental Settings**

1) **Data Segmentation and Augmentation.** As introduced earlier, we use only the motion capture data of the EmoPain dataset collected from 18 CLBP patients and 12 healthy people performing functional movements [6]. In total, there are 46 movement instances where each instance is about 15 minutes long and contains sequences of sit-to-stand, stand-to-sit, bending, reaching forward and one-leg-stand exercise). Such limited data size is typical when real data (as opposed to acted) is collected from patient cohorts.

Following the approach in [6], each sample of a sequence is characterized by 13 full-body joint-angles as well as the angular energies of these. Each joint angle is the angle formed by connecting three body-joints in the 3D Cartesian space, and they are invariant to transition and change of reference frame [28]. Due to the limited space, the description of the 13 joint-angles is shown in the bottom-right of Fig. 3.

To create the training and test data for our experiments, we run a sliding window, length = 3 seconds and overlapping ratio = 75% based on findings in [15], within each movement type of data instance. 0-padding is used when the window slides across different movement types. This amounts to a total of 2,569 segments. The groundtruth for each segment is set based on majority-voting, where a segment is labelled with protective behavior if at least 50% of the samples within it had been rated as protective by at least 2 raters, and non-protective behavior otherwise.

For the training of BANet and of other architectures evaluated for comparison, we apply two augmentation techniques, both previously used in [15]. The first technique is based on creating new instances by adding normalized Gaussian noise to the original data using 3 different standard deviations: 0.05, 0.1 and 0.15. The second approach creates new instances by randomly altering (set to 0) the movement data of the 13 body parts using selection probability of 0.05, 0.1 and 0.15. This latter method aims to simulate the presence of incomplete data. The use of the two approaches lead 18,653 segments.

2) **Experimental Settings.** The BANet was implemented with Keras and TensorFlow. For the LSTM network used for temporal information encoding, we used a 3-layer LSTM network with 8 hidden units in each layer. Dropout with probability of 0.5 was used after each LSTM layer. In the full network, weights were updated with the Adam optimization algorithm [31], a learning rate of 0.003 and batch size of 40 was applied.

The validation method used in this study is the standard Leave-One-Subject-Out cross-validation (LOSO) to test the generalization ability of a model to unseen subjects. In our work, the detection of protective behavior is a binary task as we merged the 5 categories of protective behaviors [6] [15] in the EmoPain dataset into one. Statistical tests with repeated-measures ANOVA and post-hoc paired t-tests are used to compare the performances of the architectures.

**IV. EXPERIMENT RESULTS AND DISCUSSION**

In this section, we first report the results achieved in the comparison experiments. Then, we visually analyze and discuss the movement cues of protective behavior that emerge from the attention mechanisms.

**A. Comparison Algorithms**

As the only study on using neural networks for the protective behavior detection was proposed in [15], we compare BANet with the vanilla neural networks used in this earlier study: i) Convolutional LSTM (Conv-LSTM) [29], with the convolution kernel size 1×10, max pooling size 1×2, filter number 10, 28 LSTM hidden units and batch size of 50; ii) Bi-directional LSTM (bi-LSTM), with 14 LSTM hidden units followed by a Dropout with probability of 0.5, and batch size set to 40; iii) Protect-LSTM, a vanilla 3-layer LSTM network [15] with each LSTM layer set with 28 hidden units followed by Dropout with probability of 0.5, the batch size was set to 20. For all the neural networks, the hyperparameters were tuned through grid-search, and the Adam algorithm [31] was used for weight updating with learning rate of 0.003.

We also compare with a variant of the BANet with a fully-connected layer used in the temporal attention computation (BANet-dense) instead of a 1×1 convolution layer. In addition, we compare our work with the approach used in related HAR studies [19] [20] [21], where the sensor attention was computed before the extraction of temporal information. We create a BANet variant (BANet-compat, for BANet compatibility version) where the computation of body attention was done at input level instead of at feature fusion level with the same attention algorithm presented in last section. For BANet-compat, at each timestep, the body attention scores were computed for the 13 body parts and after multiplication, all the timesteps are concatenated for the temporal information extraction and temporal attention computation that follow. Finally, the output to be classified has the same size of \(k \times 13 \) as the BANet (\(k\) is the number of hidden units of the LSTM network within). Additionally, to show the impact of the two attention modules introduced in our BANet, we provide the results of BANet-body where only the body attention is implemented and BANet-time where only the temporal attention is computed.
hyperparameters and training strategy stay the same with BANet.

Results for the comparison experiments are shown in Table I. As shown, the proposed BANet achieved the best results (accuracy of 0.8688, mean F1-score of 0.844), using a smaller parameter-size of 2,131 in comparison to other tested LSTM-based architectures (parameter sizes ranging from 14,000 to 40,000). The parameter reduction is obtained in BANet through the use of: i) the temporal information extraction strategy independent of body parts, providing data with a smaller dimension and allowing the LSTM layer within BANet to have a smaller number of hidden units; ii) a 1 × 1 convolution instead of fully-connected operation for computing temporal attention, the former being a critical advantage due to the many timesteps (180 timesteps) of the input to this layer. The second best is achieved with BANet-body which show the importance of focusing on a subset of joints-angle (rather than all) for the detection of protective behavior. Instead, the BANet-time that learns separately the temporal attention within each joint does not achieve high accuracy results. This could be expected as this network, lack a global processing between all body joints.

The next best result is achieved by the Protect-LSTM [15] (accuracy of 0.8534, mean F1-score of 0.812). Although the results are very similar with BANet (see also their confusion matrices), Protect-LSTM requires the much larger number of parameters (18,986). Table II shows the confusion matrices of BANet and Protect-LSTM respectively. Although the BANet-compa is only the representative of the architectures used in [19][20][21], this result implies that encoding the importance of body parts at a single timestep is not valuable to the detection of protective behavior, but it should be delayed as in BANet for a higher-level processing given a period of data input.

| Architecture (Bold: best performance) | Accuracy | Mean F1 | p-value | Parameters Size |
|---------------------------------------|----------|---------|---------|-----------------|
| Conv-LSTM [29]                        | 0.8059   | 0.737   | 0.049   | 40,940          |
| bi-LSTM [30]                          | 0.8460   | 0.804   | 0.05    | 14,282          |
| Protect-LSTM [15]                     | 0.8534   | 0.812   | 0.026   | 18,986          |

| Proposed architecture                | Accuracy | Mean F1 | p-value | Parameters Size |
|--------------------------------------|----------|---------|---------|-----------------|
| BANet-compa                          | 0.6630   | 0.572   | 0.0001  | 6,204           |
| BANet-dense                          | 0.8167   | 0.789   | 0.019   | 65,430          |
| BANet-time                           | 0.806    | 0.758   | 0.09    | 1,767           |
| BANet-body                           | 0.867    | 0.831   | 0.167   | 2,023           |
| BANet                                | 0.8688   | 0.844   | -       | 2,131           |

A repeated-measure ANOVA shows statistically significant differences in comparing the performances across folds between proposed BANet and the others: F(3.072, 89.099)=15.612, p<0.0001, $\mu^2 =0.519$. Post-hoc paired t-test with Bonferroni corrections show the result is not significant only between BANet and BANet-body (p=0.167). However, when BANet-body is compared with the other architectures, significant differences are found only with BANet-time (p=0.022). This suggests that the impact introduced by the body attention module is more significant than the temporal attention one. In addition, the combination of the two attention mechanisms leads more consistently to better results.

### B. Attention Scores Analysis

While a full analysis of attention scores is out of the scope of this paper due to space limit, we highlight some of the trends to understand to what extent the two attention mechanisms capture aspects of protective movement strategies. Beside improving classification performance, such scores may help further understanding of pain behavior, still predominantly based on observations rather than objective measurements. Fig. 3-Top shows boxplots of the joint attention scores for all test segments over all folds organized by movement type. It is interesting that boxplots for healthy participants (green) are quite narrow compared to those of participants with CLBP (blue and orange). This can be related back to the literature in two ways.

First, even when people were not asked to perform a movement according to ideal trajectories [34], healthy subjects tend to perform simple everyday movements in a quite similar way. Only a few healthy participants’ boxplots are slightly wider, especially in the bend and reach-forward movements that may be more demanding. Instead, pain literatures (e.g., [38]) show that people with CLBP do tend to show a wide variety of strategies when performing simple movements, these may be depend on what part of the body they perceive as vulnerable (e.g. in stand-to-sit, avoiding bending the trunk by bending the legs more (e.g. Fig. 3-bottom-P17), or reducing weight on the legs by twisting the trunk (e.g. Fig. 3-bottom-P14)).

Second, the limited width of the healthy participants’ boxplots supports the fact that each joint maintains fairly constant relevance throughout the movement, highlighting synergetic use of engaged body parts in performing the movement [6] [35]. See, for example, the temporal attention scores of control participant C16 in the stand-to-sit (Fig. 3-bottom). This constancy is not the case for CLBP participants, as the wider boxplots suggest. As discussed in [33] [35] [36], people suffering from CLBP tend to engage different body parts at different stages of the movement rather than in a synergetic way despite making the movement more difficult to execute. This could be due to more conscious control over the movement, and/or the fear of pain or injury, or reduced proprioception/coordination capabilities as discussed in [33] [38]. Again, let’s analyze more in details P14’s darker temporal attention scores during stand-to-sit (Fig. 3-bottom). P14’s initial engagement of various body parts without really starting the sit-down suggests hesitation (as indicated by physiotherapists in [6] [37]). Initial hesitation is followed by a horizontal twist of the shoulder (captured by the right shoulder score) followed by the bending of the neck to check for the chair position, then still the twisting of the shoulder (captured by the left shoulder score) to use the arm (left elbow bent beside the
of trunk) for support on the chair to minimize the load on trunk and on legs. In C16, the participant also uses the arms, but behind the body (rather than on the side) and these reach for the chair together with the trunk, offering no much support for legs or back.

V. CONCLUSION
This paper investigates the use of both body attention and temporal attention mechanisms combining LSTM to improve the quality of detection of protective behavior in people with CLBP. The study is based on the EmoPain motion capture data. In comparison to the state-of-the-art, our architecture delays the attention processes to the second and third layer of the architecture to enable a first learning of low-level features as the movement is processed. In doing so, both attention mechanisms work on a higher-level representation of the movement. The results show that such an approach leads to a substantial improvement (F1 from 0.572 to 0.844). In addition, it shows results slightly higher than LSTM-based architectures with a critical decrease in number of hyper-parameters (from 40,940 to 2,131). The results also show that the joint-body attention mechanism plays a more important role than the temporal attention mechanisms (F1 of 0.831 vs 0.758), still the combination of the two mechanisms leads to much higher performance (F1 of 0.844). This may suggest that the temporal attention mechanism may capture more detailed local temporal dynamics missed by the joint-body attention. The paper concludes discussing some examples of how the two types of attention mechanism scores capture aspects of protective movement, suggesting that such systems, deployed in everyday rehabilitation activity, could be used not only to provide a more effective therapy but also contribute to the pain literature by enabling a better understanding of protective behavior in everyday life.

Fig. 3. Above: Boxplots for the distribution of body attention computed by BANet for each testing data organized by movement type. Bottom-Left: Temporal attention map from BANet for testing instance of patient number 14, 17 and healthy subject number 16 with their respective MoCap data (stick figures). Bottom-Right: Angle description for the 13 joint angles used in this paper.
REFERENCES

[1] Vlaeyen, J. W. S., & Linton, S. J. Fear-avoidance and its consequences in chronic musculoskeletal pain: A state of the art. Pain, 85(3), 317–332. (2000).

[2] Vlaeyen, J. W. S., Morley, S., & Crombez, G. The experimental analysis of the interruptive, interfering, and identity-distorting effects of chronic pain. Behaviour Research and Therapy, 86, 23–34. In (2016).

[3] Keele, F. J., & Block, A. R. Development of an observation method for assessing pain behavior in chronic low back pain patients. Pain Management Nursing, 13(4), 363–375. (1992).

[4] Breevink, H., Collett, B., Venatafridda, V., Cohen, R., & Gallacher, D. Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment. European Journal of Pain, 10(4), 287–333. (2006).

[5] Sullivan, M. J. L., Thibault, P., Savard, A., Catchlove, R., Kozej, J., & Stanish, W. D. (2006). The influence of communication goals and physical demands on different dimensions of pain behavior. Pain, 125(3), 270–277.

[6] Aung, M. S. H., Kaltwang, S., Romero-paredes, B., Martinez, B., Cella, M., Valstar, M., ... Bianchi-Berthouze, N. The automatic detection of chronic pain-related expression: requirements, challenges and a multimodal dataset. IEEE Transactions on Affective Computing, 7(4), 1–18. (2016).

[7] Singh, A., Klapper, A., Jia, J., Fidalgo, A., Jimenez, A. T., Kanakam, N., ... Williams, A. Motivating People with Chronic Pain to do Physical Activity: Opportunities for Technology Design. In International Conference on Human Factors in Computing Systems (CHI) (pp. 2803–2812). (2014).

[8] Singh, A., Piana, S., Pollarolo, D., Volpe, G., Varni, G., Tajadura-Jimenez, A., ... Bianchi-Berthouze, N. Go-with-the-Flow: Tracking, Analysis and Sonification of Movement and Breathing to Build Confidence in Activity Despite Chronic Pain. Human-Computer Interaction, 31(3–4), 1–49. (2016).

[9] Singh, A., Bianchi-Berthouze, N., & Williams, A. C. Supporting Everyday Function in Chronic Pain Using Wearable Technology. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 3903–3915. (2017).

[10] Fasel, Beat, and Juergen Luettin. "Automatic facial expression analysis: a survey." Pattern recognition 36.1 (2003): 259-275.

[11] Zeng, Zhong, et al. "A survey of affect recognition methods: Audio, visual, and spontaneous expressions." IEEE transactions on pattern analysis and machine intelligence 31.1 (2009): 39-58.

[12] de Gelder, B. Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1535), 3475–3484. (2009).

[13] Olugbade, T. A., Aung, M. S. H., Bianchi-Berthouze, N., Marquardt, N., ... & Williams, A. C. Bi-Modal Detection of Painful Reaching for Chronic Pain Rehabilitation Systems. Proceedings of the 16th International Conference on Multimodal Interaction - ICM '14, 455–458. (2014).

[14] Olugbade, T. A., Bianchi-Berthouze, N., Marquardt, N., & Williams, A. Pain Level Recognition using Kinematics and Muscle Activity for Physical Rehabilitation in Chronic Pain. Proc. 6th Conf. 243, 243–249. (2015).

[15] Chongyang Wang, Temitayo A. Olugbade, Akhil Mathur, Amanda C. De C. Williams, Nicholas D. Lane, and Nadia Bianchi-Berthouze. “Automatic Detection of Protective Behavior in Chronic Pain Physical Rehabilitation: A Recurrent Neural Network Approach.” arXiv preprint arXiv:1902.08990 (2019).

[16] Breiman, Leo. "Random forests." Machine learning 45, no. 1 (2001): 5-32.

[17] Zeng, Ming, et al. "Automatic recognition of fear-avoidance behavior in chronic pain physical rehabilitation." Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), (2014).

[18] Andreas Bulling, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. Comput. Surveys 46, 3 (Jan. 2014), 1–33. 2014.

[19] Zeng, Ming, Haoxiang Gao, Tong Yu, Ole J. Mengshoel, Helge Langseth, Ian Lane, and Xiaobing Liu. "Understanding and improving recurrent networks for human activity recognition by continuous attention." In Proceedings of the 2018 ACM International Symposium on Wearable Computers, pp. 56-63. ACM, 2018.

[20] Murahari, Vishvak S., and Thomas Plotz. “On attention models for human activity recognition.” In Proceedings of the 2018 ACM International Symposium on Wearable Computers, pp. 100-103. ACM, 2018.

[21] Yao, Shuochoa, Yiran Zhao, Shaohan Hu, and Tarek Abdelzaher. "QualityDeepSense: Quality-Aware Deep Learning Framework for Internet of Things Applications with Sensor-Temporal Attention." In Proceedings of the 2nd International Workshop on Embedded and Mobile Deep Learning, pp. 42-47. ACM, 2018.

[22] Attila Reiss and Didier Stricker. Introducing a new benchmarked dataset for activity monitoring. In ISWC. IEEE, 108–109. 2012.

[23] Marc Bachlin, Daniel Roggen, Gerhard Troster, Meir Plotnik, Noit Inbar, Inbal Meidan, Taliya Herman, Marina Brozgol, Eliya Shavir, Nir Giladi, and others. Potential of enhanced context awareness in wearable assistants for Parkinson’s disease patients with the freezing of gait syndrome. In ISWC. IEEE, 123–130. 2009.

[24] Piero Zappi, Clemens Lombriser, Thomas Stiefmeier, Elisabetta Farella, Daniel Roggen, Luca Benini, and Gerhard Troster. Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection. In Wireless sensor networks. Springer, 17–33. 2008.

[25] FJ Oródez and D Roggen. 2016. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. Sensors 16, 1, 115. (2016).

[26] Yao, Shuochoa, Shaohan Hu, Yiran Zhao, Aston Zhang, and Tarek Abdelzaher. "DeepSense: A unified deep learning framework for time-series mobile sensing data processing." In Proceedings of the 26th International Conference on world wide web. (2017).

[27] A. Kleinsmith and N. Bianchi-Berthouze, “Affective body expression perception and recognition: a survey,” IEEE Trans. Affective Comput., vol. 4, no. 1, pp. 15–33, 2013.

[28] Falco, Pietro, Matteo Saveriani, Eka Gibran Hasany, Nicholas H. Kirk, and Dongheu Lee. “A human action descriptor based on motion coordination.” IEEE Robotics and Automation Letters 2, no. 2, 811–818. (2017).

[29] Morales, Francisco Javier Oródez, and Daniel Roggen. “Deep convolutional feature transfer across mobile activity recognition domains, sensor modalities and locations.” Proceedings of the 2016 ACM International Symposium on Wearable Computers. ACM, 2016.

[30] Hammerla, Nils Y., Shane Halloran, and Thomas Ploetz. “Deep, convolutional, and recurrent models for human activity recognition using wearables.” arXiv preprint arXiv:1604.08880 (2016).

[31] Kingma, Diederik P., and Jimmy Ba. “Adam: A method for stochastic optimization.” arXiv preprint arXiv:1412.6980 (2014).

[32] A. Sisem, H. Blundic, S. Bhattacharya, T. S. Pentow, M. B. Kjergaard, A. Dey, T. Sonne, and M. M. Jensen. Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition. In Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems, pages 127–140. ACM, 2015.

[33] Olugbade, Temitayo A., Aneessa Singh, Nadia Bianchi-Berthouze, Nicolai Marquardt, Min SH 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, no. 1 (2019).

[34] Sigkö, T., Gliobowski, E., Barczyk-Pawelec, K., & Kuczyński, M. The Effect of Chronic Pain Intensity on Sit-to-Stand Strategy in Patients with Herniated Lumbar Disks. Journal of Manipulative and Physiological Therapeutics, 39(3), 169–175. (2016).

[35] Shafizadeh, Mohsen. 'Movement coordination during sit-to-stand in low back pain people.' Human Movement 17, no. 2 (2016): 107-111.

[36] Watson, Paul J., C. J. Booker, C. J. Main, and A. C. N. Chen. "Surface electromyography in the identification of chronic low back pain patients: the development of the flexion relaxation ratio." Clinical Biomechanics 12, no. 3 (1997): 165-171.

[37] Olugbade, T. A., Bianchi-Berthouze, N., Marquardt, N., & Williams, A. C. Human Observer and Automatic Assessment of Movement Related Self-Efficacy in Chronic Pain: from Movement to Functional Activity. IEEE Transaction on Affective Computing. (2018).

[38] Ippersiel, Patrick, Shawn Robbins, and Richard Preuss. "Movement variability in adults with low back pain during sit-to-stand-to-sit." Clinical Biomechanics 58 (2018): 90-95.