Association between central sensitization and gait in chronic low back pain: Insights from a machine learning approach

Xiaoping Zheng a,*, Michiel F. Reneman b, Jone Ansategui Echeita b, Rita HR. Schiphorst Preuper b, Herbert Kruitbosch c, Egbert Otten a, Claudine JC. Lamoth a

a University of Groningen, University Medical Center Groningen, Department of Human Movement Sciences, Groningen, Netherlands
b University of Groningen, University Medical Center Groningen, Department of Rehabilitation, Groningen, the Netherlands
c University of Groningen, Center for Information Technology, Data Science Team, Groningen, the Netherlands

Keywords: Low back pain, Central sensitization, Supervised machine learning, Gait, Daily life

1. Introduction

Chronic low back pain (CLBP) is one of the most prevalent chronic musculoskeletal pains [1]. It is responsible for high treatment costs, sick leave and individual suffering and it represents a significant socioeconomic burden [2]. For 85%–90% of patients with CLBP, the relation between pathoanatomical and clinical presentations is absent [3] and, therefore, it is classified as nonspecific CLBP [4]. In CLBP, and other chronic musculoskeletal disorders, central sensitization (CS) might be present (reviewed in Ref. [5]). CS is defined as “increased responsiveness of nociceptive neurons in the central nervous system to their normal or subthreshold afferent input” [6] and manifests as mechanical hypersensitivity, allodynia and hyperalgesia [7]. A considerable number of people need treatment for CLBP. Although the overall efficacy of CLBP rehabilitation programs is positive, but the effect sizes are modest [8].

Correctly recognizing the physical and psychosocial factors perpetuating pain and physical disability of patients with CLBP remains a challenge [9]. Altered motor control of patients with CLBP could possibly contribute to the persistence of CLBP [10]. Altered motor control could affect daily-living activities, as patients with CLBP often exhibit altered movement patterns and motor control strategies; probably to avoid painful movement, such as walking [11]. Many clinicians may intuitively identify “abnormal” gait patterns in patients with CLBP, but identification and objectifying of specific “abnormal” gait outcomes is challenging. During walking, it is suggested that patients often adopt a “protective guarding” or “splinting” strategy [12] to avoid painful movements of the spine. These adaptations may lead to a slower and less flexible gait pattern [13]. Evidence for this, however, is ambiguous. Studies between patients with CLBP and healthy controls, observed inconsistent evidence regarding preferred walking velocity [13,14]...
stricte and non-linear data structures and take the interrelation of the symptoms (e.g. cancer, inflammatory diseases and/or spinal fractures); (b) had neuralgia and/or radicular pain in the legs; (c) were excluded if they: (a) had a specific diagnosis that would better account for the symptoms (e.g. cancer, inflammatory diseases and/or spinal fractures); (b) had neuralgia and/or radicular pain in the legs; (c) were pregnant; (d) in an acute phase of pain.

The study was approved by the Medical Research Ethics Committee of the University Medical Center Groningen (METc 2016/702) and conducted according to the principles expressed in the Declaration of Helsinki. The data used in this paper was derived from a larger study, of which protocol details were described elsewhere [19].

### 2.2. Data collection

Demographics were collected and standard clinical test were applied as part of the usual care of CLBP patients that are referred to the outpatient Pain Rehabilitation Department of the Center for Rehabilitation. Assessments included: Visual Analogue Scale for pain intensity (VAS Pain; 0–10), the Dictionary of Occupational Titles (DOT, the Pain Disability Index (PDI; 0–70), the physical functioning subscale of the Rand36 questionnaire (Rand36-PF; 0–100), the Pain Catastrophizing Scale (PCS, 0–52), the Injustice Experience Questionnaire (IEQ, 0–48), and the Brief Symptom Inventory (BSI global severity index t-score (GIST)); (see Table 3).

### Central sensitization (CS)

The presence of CS-related manifestations was assessed with section A of the Central Sensitization Inventory (CSI) [24]. Section A has 25-items to assess the presence of common CS-related symptoms. Scores can range from 0 to 100 where a higher scoring represents a higher level of CS. A score lower than 40 indicates lower CS levels (CLBP- group) and a score of 40–100 is interpreted as higher CS levels (CLBP + group) [25].

### Accelerometer data

The accelerometer data were collected between 2017 and 2019. Patients were instructed to wear a tri-axial accelerometer (Actigraph GT3X, Actigraph Corporation, Pensacola, FL) at all times for about one week, excluding sleeping or bathing times. The accelerometer was worn at the front right hip of the patient (at the anterior superior iliac spine). Assuming a standing and upright posture, the Y-axis pointed to the ground (vertical direction, V), Z-axis faced the walking direction (anteroposterior direction, AP), and the X-axis was perpendicular to the walking direction, pointing from a patient’s right to left (mediolateral direction, ML). These directions are approximate only. The sampling frequency of the accelerometer was set to 100 Hz and the dynamic range was ±6 gravity.

### 2.3. Data processing and analysis

#### 2.3.1. Raw data segmentation

Accelerometer data of each patient was segmented into 24 h span data segments (from 12:00 p.m. to next day 11:59 a.m.) to represent the activities during the days. Because the measurement started at 12:00 p.m., to make full use of the data, the 24 h span was between 12:00 p.m. until next day 12:00 p.m. Data that did not completely covered this 24 h span was discarded from the analysis. Because of technical errors or personal reasons, a full week of data could not be collected from all patients. To compare the data between different patients fairly, 4 segments (representing 4 days) of each patient were included in the analysis. Therefore, 7 patients who had less than 4 segments, were excluded. From patients with more than 4 segments, 4 segments were randomly sampled. Fig. 1a graphically shows the process of the raw data segmentation.

#### 2.3.2. Walking bouts extraction

The accelerometer data of the 4 segments were first smoothed by a low-pass filter with a 2nd order Butterworth and a 20 Hz cut-off frequency. Subsequently, potential walking events were detected by the Fast Fourier Transform (FFT) based method [26], which identified periods with 0.5–3.0 Hz power spectrum values. To remove false walking events from the potential walking periods, the zero-cross method [27] was employed. If the time interval between any two adjacent walking events was shorter than 2 s, these two walking events were merged into one walking bout. Finally, the walking bouts in each segment were extracted and their gait outcomes were calculated. Fig. 1b presents the walking bouts as the yellow vertical bars in the rectangle.

#### 2.3.3. Gait outcomes

All walking bouts in one 24 h segment were used to determine the total duration of walking, the total number of steps, the maximum duration of a walking bout and the maximum number of steps of a
walking bout. Subsequently, all walking bouts exceeding 10 s were selected and cut into non-overlapping 10 s windows [28]. From the segment, each 10 s window was used to calculate different gait outcomes, and these values were averaged over all 10 s windows in the segment representing the patient’s gait performance on that day.

Gait outcomes were divided into two categories, quantitative and qualitative gait outcomes. From one segment, we obtained one gait outcome vector, including 36 gait outcomes, based on the walking bouts (see Fig. 1c). The detailed descriptions of the quantitative and qualitative gait outcomes are presented in Table 1 and Table 2—for extended explanation of variables see Ref. [29].

Pearson-coefficient was calculated to examine relationship of gait outcomes between weekdays and weekend. The Pearson-coefficient ranges from –1 to 1, where 1 represents a perfect correlation.

The Mann-Whitney U test was used to statistically test the differences between CLBP- and CLBP + groups for demographics and CSI scores. To separate CLBP- and CLBP + groups by gait outcomes, RF was used.

2.3.4. Random Forest classifier

RF is considered as the optimal machine learning classification approach for the present data, because it performs well with (a) nonlinear and linear data; (b) high dimensional data; and (c) unbalanced and small datasets [30]. Apart from this, a comparison of different machine learning classifiers was performed to help to select RF as the best classifier for this study (details in Appendix A).

The input data of this approach was \(< S, L >\). S represents the gait outcome vectors of all patients and L is its corresponding label. The definition of \(S\) is: \(S = \{s_1, s_2, ..., s_i, ..., s_m\}\) and \(s_i = [d_1, ..., d_k]\), where \(s_i\) represents a gait outcome vector \(i\) and \(m\) is the number of all gait outcome vectors, \(d\) represents a gait outcome and \(k = 36\). \(L = [l_1, ..., l_m]\), where \(l \in \{CLBP, CLBP +\}\).

RF is constructed in four steps. Step one: Randomly sample \(n\) gait outcome vectors from \(S\) and \(n\) corresponding labels from \(L\), with replacement. These new set of gait outcome vectors and labels are called \(S_0\) and \(L_0\). In \(S_0\), \(s_i\) may appear more than one time or not appear. Step
### Table 2: Qualitative gait outcomes.

<table>
<thead>
<tr>
<th>Catalog</th>
<th>Gait characteristic</th>
<th>Description and method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regularity</td>
<td>Stride regularity (SR; V, ML, AP, All)</td>
<td>SR is computed by using the unbiased autocorrelation coefficient: ( Ad(m) = \frac{1}{N-m} \sum_{i=m}^{N} Acc(i)Acc(i+m) ), where ( m ) is the sample acceleration signal, ( Acc(i) ) is the number of samples, and ( N ) the number of time lag. The first peak of ( m ) is ( Ad(m) ) and it represents the stride regularity. Higher values (maximum 1.0) reflect repeatable patterns between strides.</td>
</tr>
<tr>
<td>Gait symmetry index (GSI)</td>
<td></td>
<td>GSI quantifies the ratio of the first and second peak of the ( Ad(c) ), ( Ad(d) ) of the gait outcome vectors of the remaining 20% of patients. It is a measure of the degree of symmetry of the left and right lower limbs during walking.</td>
</tr>
<tr>
<td>Smoothness</td>
<td>Index of harmonicit (IH; V, ML, AP, All)</td>
<td>IH = ( \sum_{i=1}^{P}AccV(i)) is the ratio of the power spectral density of the fundamental frequency ( P_0 ) and the sum of the power spectral density of the first six frequency ( P_i ). IH quantifies gait smoothness., with higher values representing a smoother (max 1.0) gait pattern.</td>
</tr>
<tr>
<td>Harmonic ratio (HR; V, ML, AP)</td>
<td>Max LyE of the input gait outcomes, ( \text{Max LyE} = \frac{\text{LyE}_1}{\text{LyE}_2} ), as calculated by the Max LyE, as measured by the accelerometer signal vector from time ( i ) to ( m + 1 ). Acc(i) is the Chebyshev distance, and ( r ) was set to 0.3. Sen quantifies the predictability of a time series. Small values (minimum 0) indicate better synchronization between acceleration signals.</td>
<td></td>
</tr>
<tr>
<td>Predictability</td>
<td>Sample entropy (Sen; V, ML, AP)</td>
<td>Sen = ( -\frac{N}{\sum_{i=1}^{N} r(i)} ), with ( A = \frac{1}{2}Ad(i) ), ( B = \frac{1}{2}Ad(d) ), ( Acc(i) ) and ( Acc(d) ) means the accelerometer signal vector from time ( i ) to ( m + 1 ), ( d ) is the average acceleration, ( Acc(i) ) is maximum, the correlation coefficient, and ( r ) was set to 0.3. Sen quantifies the predictability of a time series. Small values (minimum 0) indicate better synchronization between acceleration signals.</td>
</tr>
<tr>
<td>Stability</td>
<td>Maximal Lyapunov exponent (max LyE; V, ML, AP)</td>
<td>Max LyE, as calculated by the Rosenstein algorithm, quantifies the local stability of trunk acceleration patterns. The fitting window length was 60/100/1 by ( f ) is the sample frequency, and the embedding dimension was set to 7. The overall max LyE were calculated and normalized by per stride. Higher values represent greater sensitivity to local perturbations.</td>
</tr>
</tbody>
</table>

Two: In \( S_k \), randomly sample \( j \) \( (j \leq k) \) gait outcomes from \( s \). Therefore, \( S_k \) = \( S_1, S_2, S_3, \ldots, S_n \). Step three: Training a decision tree \( S_j \) on \( S_k \) \( \in S_k \). Step four: Repeat steps one to three 1000 times and combine the decision trees into an ensemble, called RF, that predicts by voting (see Fig. 2).

Before training RF, 80% of patients were randomly selected and their 4 corresponding gait outcome vectors were used as the training data. The gait outcome vectors of the remaining 20% of patients were used as the testing data. To avoid overfitting of the hyperparameters, a 5-fold cross-validation approach was used to estimate them, as shown in Fig. 1d. Four folds were used to train the model and the rest fold was used to estimate the performance of the current hyperparameters in RF. The performance reported by the 5-fold cross-validation was the average of the values computed in the 5 splits. After the best hyperparameters were determined, the testing dataset was used to evaluate the generalizability of the model.

#### 2.3.5. Accuracy evaluation

Accuracy, sensitivity, specificity, precision, F1-score, and maximum Youden index were calculated to evaluate the performance of the classification (Fig. 1f). In this study, CLBP+ was considered as the positive case and CLBP- was the negative case. Correct predictions of CLBP+ and CLBP- patients are considered as true positives (TP) and true negatives (TN), respectively. Incorrect classifications of CLBP- patients as CLBP+ or of CLBP+ patients as CLBP-, are considered false positives (FP) and false negatives (FN) respectively.

Accuracy was the proportion of all the correct classification results.

\[
\text{accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}
\]

Sensitivity represents the proportion of positive cases that are correctly assigned (true positive rate).

\[
\text{sensitivity} = \frac{TP}{TP + FN} \tag{2}
\]

Specificity refers to the rate of correctly predicted negative cases in all negative cases (true negative rate).

\[
\text{specificity} = \frac{TN}{TN + FP} \tag{3}
\]

Precision is the ratio of the correctly predicted positive cases in all predicted positive cases.

\[
\text{precision} = \frac{TP}{TP + FP} \tag{4}
\]

F1-score is the harmonic mean (average) of the precision and sensitivity.

\[
F_1 = \frac{2 \times \text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}} \tag{5}
\]

The receiver operating characteristic (ROC) curve was calculated to evaluate the performance of RF. The Y-axis of this curve represents the true positive rate (sensitivity) and the X-axis represents false positive rate (1-sensitivity). The overall classification performance of RF was evaluated by the area under the ROC curve (AUC). A classification model with a larger AUC value has a higher correct rate, and AUC = 1 represents perfect performance. The maximum Youden index was computed to measure the diagnostic test’s ability.

\[
J(c) = \text{Max}\left\{\text{sensitivity}(c) - (1 - \text{specificity}(c))\right\} \tag{6}
\]

where \( c \) is the cut-point. When the value \( J \) is maximum, the corresponding \( c \) is the optimal cut-point.

#### 2.3.6. Feature importance

SHapley Additive explaNations (SHAP) [31] was used to assess the gait outcomes’ importance to the classification model. SHAP connects optimal credit allocation with local explanations using the classic Shapley values from game theory. Shapley values, \( \varnothing_i \), explains the importance of gait outcome \( i \) for RF and is defined as:

\[
\varnothing_i = \frac{1}{N^2} \sum_{i \in \mathcal{S} \text{ and } s \in \mathcal{S}} (|i| - 1)(|N - |s||)F(R(s) - R(s - \{i\})) \tag{7}
\]

where \( N \) is the size of the full set of gait outcomes, \( s \) is the subset that includes \( i \) in \( N \), and \( R() \) is the accuracy of the input gait outcomes. Since computing the exact Shapley values is computationally expensive, SHAP uses a tree explainer to exploit the information stored in the tree structure to calculate the SHAP values which are highly approximate Shapley values. Therefore, higher SHAP values represent higher impact to classify CLBP- and CLBP + groups.
**3. Results**

Demographic characteristics are provided in Table 3. Out of a total of 60 patients, 11 were excluded because essential parts of their dataset were incomplete (CSI scores or/and accelerometry data), 7 were excluded because they had less than 4 segments data (3 had 1 segment, 2 had 2 segments, and 2 had 3 segments). Therefore, 42 patients were included in the data analysis. Differences between CLBP+ and CLBP- group characteristics (Table 3) were not statistically significant (p > 0.05), with exception of CSI score (p < 0.001) and BSI (p = 0.01).

Because 42 patients (23 CLBP- and 19 CLBP+) were included, and for every patient 4 segments were randomly selected, the total accelerometer data segments were 168. Therefore, the scales of training and testing dataset were 136 and 32. The mean Pearson-coefficient between workdays and weekend was 0.983, indicating almost perfect correlation.

Testing data were used to evaluate the generalizability of RF and the confusion matrix is shown in Fig. 3. From the confusion matrix, accuracy, sensitivity, specificity, precision, and the F1-score were calculated to evaluate the performance metrics of the model. RF achieved an accurate classification-result (84.4% accuracy), and the sensitivity and specificity were 75.0% and 93% respectively. The precision was 92% and the F1-score was 82.6%. The ROC curve is presented in Fig. 4 showing that RF achieved a 0.83 AUC and the maximum Youden index was 0.69.

The importance of the gait outcomes for RF is shown in Fig. 5. Based on the SHAP values, the 10 gait outcomes (above the red line in Fig. 5) were considered important to the classification model. For the gait outcomes below the red line, the SHAP values were too low. Important gait outcomes are IH-V, SF variability-ML/AP, SR-ML, Max LyE-V/ML, Sen-AP, Max LyE-per stride-V, HR-ML and SL variability.

Fig. 6 shows the violin-box plot of the 10 important gait outcomes. Violin-box plot is a hybrid of a kernel density plot and a box plot, and the dots show the individuals data. A box plot contains a set of whiskers, a box and a horizontal line in the middle of the box, representing the minimum, maximum, first quartile, third quartile and median of the data respectively. From this figure, it is easy to distinguish the differences of the median between groups. It shows that CLBP- group has more variations of the median between groups. The SHAP values, the 10 gait outcomes (above the red line in Fig. 5) were considered important to the classification model. For the gait outcomes below the red line, the SHAP values were too low. Important gait outcomes are IH-V, SF variability-ML/AP, SR-ML, Max LyE-V/ML, Sen-AP, Max LyE-per stride-V, HR-ML and SL variability.

**Table 3**

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Patient characteristics (n = 42).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CLBP- (n=23)</td>
</tr>
<tr>
<td>Gender</td>
<td>15 W/8 M</td>
</tr>
<tr>
<td>Age, years</td>
<td>40.8 ± 12.8</td>
</tr>
<tr>
<td>Height, cm</td>
<td>173.5 ± 10.6</td>
</tr>
<tr>
<td>Weight, kg</td>
<td>87 ± 17.7</td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
<td>28.9 ± 5.3</td>
</tr>
<tr>
<td>Central Sensitization Inventory (0–100)</td>
<td>31 ± 4.8</td>
</tr>
<tr>
<td>Time since pain onset (years)</td>
<td>4.5 ± 6.1</td>
</tr>
<tr>
<td>Educational Level</td>
<td>175 ± 4.9</td>
</tr>
<tr>
<td>Physical demands at work (DOT: Se/Li/Me/He)</td>
<td>3/11/8/1</td>
</tr>
<tr>
<td>Patient-reported Pain Intensity (VAS, 0–10)</td>
<td>5.5 ± 2</td>
</tr>
<tr>
<td>Disability (PDI, 0–70)</td>
<td>33.6 ± 11.2</td>
</tr>
<tr>
<td>Work Ability (WAS, 0–10)</td>
<td>4.5 ± 2.3</td>
</tr>
<tr>
<td>Physical Functioning (Rand36-PF, 0–100)</td>
<td>49.8 ± 22.3</td>
</tr>
<tr>
<td>Catastrophizing (PCS, 0–52)</td>
<td>16.3 ± 8.9</td>
</tr>
<tr>
<td>Injustice (IEQ, 0–48)</td>
<td>15.2 ± 8.9</td>
</tr>
<tr>
<td>Psychological traits Screening (BSI, t-score)</td>
<td>34.4 ± 5.9</td>
</tr>
</tbody>
</table>

Fig. 3. Classification results for Random Forest, and the mean accuracy was 84.4%. CLBP-, CLBP+: Patients with chronic low back pain with lower (−) and higher (+) central sensitization levels.

Fig. 4. The receiver operating characteristic (ROC) curve (in red) for Random Forest classifier. AUC: area under the curve.
wide range distribution. Similarly, in SR-ML, CLBP+ has a concentrating distribution while the peak of CLBP− is lower. In Max LyE-V and Max LyE per stride -V, CLBP− shows a log-normal distribution while CLBP+ shows a wider distribution. In HR-ML and SL Variability, the distributions are similar but CLBP+ has more outliers.

4. Discussion

The aim of this study was to analyze whether and how the presence of CS was related to differences in gait performance of patients with CLBP during daily life by using a machine learning approach. Based on quantitative and qualitative gait outcomes, using a RF, the two groups (CLBP− and CLBP+) could be classified with a high accuracy. The classification results indicated that CLBP− patients walk differently from CLBP+ patients. Furthermore, the SHAP values showed that the differences between CLBP− and CLBP+ groups were present in gait outcomes that represented smoothness, stability, predictability, regularity, and variability.

In the present study, we addressed the walking measurement of patients with CLBP in a daily-living environment. Walking in a controlled laboratory or during a clinical assessment is different from self-initiated gait, during activities of daily living. Walking in daily life, might be subject to environmental perturbations, quick changes while performing a task, and often involves the performance of several actions at the same time [32], e.g. walking when carrying a cup of coffee. These influences on gait are not present in controlled studies and are not captured by conventional gait outcomes that average outcomes over stride cycles, such as mean step length, mean step time, and number of steps.
In this study, SHAP was used to evaluate the importance of each gait outcome, instead of the conventionally used Gini impurity and information entropy. The value of Gini impurity is based on the tree structure in RF and information entropy reflects the level of “information” of a gait outcome. Gait outcomes are interrelated and interact in a complex nonlinear manner [33]. SHAP is based on the game theory and evaluates the contribution of each gait outcome to the classification accuracy by computing all possible combinations between gait outcomes. Therefore, SHAP provides a good method to explain the importance of gait outcomes to RF. The SHAP values suggest that the differences between CLBP- and CLBP+ groups are reflected in smoothness, stability, predictability, regularity, and variability of gait. Compared with CLBP-group, CLBP+ group exhibited lower smoothness and local stability of gait, while the CLBP+ group exhibited a more regular, less variable, and more predictable gait pattern.

Gait patterns of patients with CLBP, are usually compared with the gait pattern of healthy persons. To the best of our knowledge, this is the first study in patients with CLBP that addresses the difference in gait pattern between two CLBP groups based on low and high CS levels, which makes a direct comparison with other studies intricate. The results of different gait patterns between low and high CS levels support the notion that within the heterogeneous CLBP group, different motor control strategies are adopted. Two motor control strategies on a continuum have been suggested with “tight control” and “loose control” at each end, and normal trunk control in the middle [36].

The gait patterns of CLBP+ group might suggest that patients with CLBP+ adopt a more “tight control”. The “tight control” involves increased trunk muscle activation and enhanced muscle co-contraction, might enhance control over trunk posture and movement [36]. Increased muscle activation and enhanced co-contraction would help individuals to maintain the stability of lumbar spine [37] by restricting the movement amplitude of lumbar spine. However, in a complex daily-living environment, this strategy might impair patients’ ability to maintain balance during walking because of the unstable surfaces and environmental perturbations [38], and therefore has a lower gait stability (compared with CLBP− patients). Increased co-contraction would reduce the demand for the intricate control of the sequences of muscle activation. It might avoid the potential error raised by inaccurate sensory feedback of CLBP+ [36]. This might allow patients to control their trunks’ movement precisely [39] and, therefore, result in a lower variability and a higher regularity of gait of CLBP+ patients. Our results might infer thus that the CLBP+ group exhibited a more “tight control”. Therefore, the lower stability and variability, higher regularity and predictability in gait of the CLBP+ group could be the result of the adoption of “tight control”.

The gait patterns of CLBP− group, on the other hand, might be explained by “loose control” strategy. The “loose control” that involves reduced muscle excitability, might reduce the control over trunk movements [36]. The spine of which each spinal unit has $6^\circ$ of freedom, is controlled by its surrounding musculature. Reduced muscular excitability, leads to a reduced control over the spinal muscle, to larger amplitude movements, and to more movement variability during repeated tasks [36]. The increased variability in gait of CLBP− group might support this finding. Additionally, increased variability would lead to a lower regularity in gait which was also found in the CLBP− group. Apart from this, increased motor variability might probably prevent muscle fatigue [40] because it allows sharing the load between different structures or tissues. Motor variability makes it possible to explore new pain-free motor control solutions [41]. This is a possible explanation for the higher smoothness in gait of CLBP− patients, because it allowed them to flexibly adapt to the complex daily-living environment by using different movement solutions. So, the higher variability and smoothness, and lower regularity in gait patterns might hint that CLBP− group adopted a more “loose control” compared with CLBP+ group.

Although the “tight control” adapted strategy might have short-term
benefits, it may also contribute to a higher level of CS. The “tight control” present in CLBP + patients presumably increase muscle activation and co-contraction, and lead to larger forces acting on the spine and higher spinal loading. Moreover, it has been shown that even when patients are at rest, muscle co-contraction can be continuous [42]. These long-lasting peripheral noxious stimuli might explain the development and/or persistence of CS [43]. Additionally, it has been reported that a “tight control” strategy relates to negative pain cognitions [44], a psychological process that also might contribute to the higher CS scores of the CLBP + group.

Clinically, the gait outcomes identified as important to the classifier, may assist clinicians in providing them with a more accurate understanding of the gait performance of patients with CLBP, with low or high CS levels, and with an explicit operationalization of the observed “abnormal” gait pattern of patients with chronic pain. Whether “abnormal” should be interpreted as a functional or a dysfunctional motor control strategy in the short or long term, remains to be studied. RF and SHAP used in this study have presented a novel way to identify interacting features, and therefore, can be used for further studies. The presented accurate classification could become meaningful if this would lead to effective treatment approaches. The differences in gait patterns of CLBP- and CLBP + groups could be the results of the different motor control adapted strategies and the different motor control adapted strategies could be the causes, consequences, or both, of differences in CS levels on patients with CLBP. While this cross-sectional study has objectified a relation between CS and gait outcomes, the causality of this relation is unknown. Follow-up studies would benefit from a longitudinal design with multiple measurements to help further unraveling of this relation, as well as the relation to disability.

In line with most studies on walking and CLBP, we used cross-sectional data, thus we are not allowed to infer causality between motor control changes, CS and CLBP. Some patients had analgesic or anti-inflammatory treatment at the beginning of the study, and how these medicines interact with CS and gait outcomes is unknown. Moreover, we labeled the groups based on CSI score and the cut-off values from a previous study [25]. It should also be noted that a gold standard measure to diagnose CS is unavailable. The CSI is regarded as an indirect measure of CS, because higher scores are associated with the presence of CS syndromes [25]. In addition to gait assessment, it would be interesting to explore differences in physical activities between CLBP- and CLBP + groups, because several studies reported that relationship between CLBP and physical activity levels is heterogeneous [45].

5. Conclusion

The present study analyzed gait data during daily living of CLBP patients with low and high CS levels. RF and SHAP were applied for classification and for assessing the contribution of gait outcomes to the model. This analytic approach demonstrated that RF has the ability to accurately classify subgroups of patients with CLBP and low or high CS levels based on differences in gait outcomes. The results of SHAP showed that the differences of gait outcomes between low and high CS levels were in gait regularity, variability, predictability, smoothness, and stability. This may imply that patients with low and high CS levels adopted different motor control strategies. Patients with CLBP and low CS level (CLBP-) use a “loose control” and, therefore, exhibited more smoothness and stability in gait patterns. Patients with CLBP and high CS level (CLBP+) adopted a “tight control” and showed a more regular, less variable and more predictable gait pattern.

The results of this study may contribute to a better understanding of gait characteristics in patients with CLBP, its association with CS, and may in the future assist in better-personalized rehabilitation interventions [46].

Ethics approval and consent to participate

The study was approved by the Medical Research Ethics Committee of the University Medical Center Groningen (METc 2016/702) and conducted according to the principles expressed in the Declaration of Helsinki.

Authors’ contributions

XZ, MR, EO, and CL developed the conception and design of the study. JAE and RP acquired the data from participants. XZ analyzed the data and wrote the paper under supervision of MR, OB, and CL. HK reviewed the code. All authors reviewed and commented on the manuscript. All authors approved the final manuscript.

Research data

The raw data of all 42 participants and 168 acceleration segments are not publicly available. RF was available from the package scikit-learn 1.01 (https://scikit-learn.org/stable/) and the source code of SHAP is available from shap package (https://github.com/slundberg/shap) based on Python 3.8.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

XZ was supported by China Scholarship Council-University of Groningen Scholarship [Grant No.201906410084].

Appendix A

We did empirical work of comparing often applied approaches for classification: Random Frost (RF), Artificial Neural Network (ANN), Support Vector Machine (SVM), Naive Bayes (NB), and K-Nearest Neighbors (KNN). As can be seen in the Table below, RF and ANN had the best performance compared to SV, NB, and KNN. RF performed better in precision and specificity, while ANN performed better in sensitivity.

<table>
<thead>
<tr>
<th></th>
<th>RF</th>
<th>ANN</th>
<th>SVM</th>
<th>NB</th>
<th>KNN (n = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>84.4%</td>
<td>81.2%</td>
<td>68.8%</td>
<td>62.5%</td>
<td>62.5%</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>75%</td>
<td>93%</td>
<td>56%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Specificity</td>
<td>93%</td>
<td>68.8%</td>
<td>81.3%</td>
<td>75%</td>
<td>75%</td>
</tr>
<tr>
<td>Precision</td>
<td>92%</td>
<td>75%</td>
<td>75%</td>
<td>66.7%</td>
<td>66.7%</td>
</tr>
<tr>
<td>F1-score</td>
<td>82.6%</td>
<td>83%</td>
<td>64.1%</td>
<td>57%</td>
<td>57%</td>
</tr>
<tr>
<td>AUC</td>
<td>0.83</td>
<td>0.85</td>
<td>0.67</td>
<td>0.62</td>
<td>0.62</td>
</tr>
</tbody>
</table>
References


