Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 176 (2020) 2822–2831

Procedia Computer Science

www.elsevier.com/locate/procedia

24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Automatic Low Back Pain Classification Using Inertial Measurement Units: A Preliminary Analysis

Zoë Bacon[a], Yulia Hicks[a, *], Mohammad Al-Amri[b, c], Liba Sheeran[b, c]

[a]Cardiff School of Engineering, Cardiff University, UK
[b]Cardiff School of Healthcare, Cardiff University, UK
[c] Biomechanics and Bioengineering Research Centre Versus Arthritis, Cardiff University, UK

Abstract

Low back pain (LBP) is a major health problem that has now become leading cause of disability worldwide. The majority of LBP has no specific pathological cause. Classification of non-specific LBP (NSLBP) into subgroups corresponding to the reported symptoms has been identified as an essential step towards the provision of personalised management and rehabilitation plans. Currently, clinicians classify low back pain patients into clinical subgroups based on clinical judgement and expertise, which is a time-consuming process open to human error. This paper introduces a novel approach for automatic classification of NSLBP patients into clinical subgroups on the basis of the MTw2 inertial measurement unit (MTw2 IMU tracker) motion data, which are portable units and thus desirable for clinical use. Four MTw2 IMU trackers tracking movement during a number of physical assessment tests were investigated in their ability to distinguish between clinically recognized NSLBP subgroups. Simple motion features such as the angular range of displacement were used in classification experiments to reflect how clinicians make decisions when classifying NSLBP. The achieved results were comparable to the state of art results in automatic NSLBP classification using optical motion capture data and demonstrated the feasibility of developing an automatic classification system on the basis of the MTw2 IMU tracker motion data obtained with an individual performing a battery of standard physical assessment tests. Further developments could address gaps in current medical and engineering literature and improve clinical outcomes.

* Corresponding author. Tel.: +44(0)2920875945.
E-mail address: hicksya@cardiff.ac.uk

1877-0509 © 2020 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the KES International.
10.1016/j.procs.2020.09.272

Keywords: Non-specific low back pain; inertial measurement units; automatic classification; personalised treatment.

1. Introduction and related work

According to the 2016 Global Burden of Disease Study, low back pain (NSLBP) is a major problem that has now become a leading cause of disability worldwide [1]. Between 1990 and 2015, the years lived with disability due to NSLBP increased by 54% [2]. NSLBP is one of the most prevalent musculoskeletal disorders, with 12% of the global population affected at any one time [3]. NSLBP can be defined as pain "localised below the costal margin and above the inferior gluteal folds" [4]. The majority of NSLBP (up to 85%) in non-specific (NSLBP) which means it has no specific pathological cause that can be cured or medicated. Instead, clinical guidelines recommend self-management and exercise tailored to individual needs [5]. Individually tailoring exercises to patients with NSLBP is a major clinical and diagnostic challenge [6]. This is because of the high complexity, biopsychosocial nature and extreme heterogeneity of the dominant factors driving the disorder. This highlights the importance of developing a solution to accurately and efficiently match treatments to a NSLBP person's profile.

Classification of NSLBP into subgroups corresponding to the reported symptoms has been identified as an essential step towards the provision of personalised management and rehabilitation plans. Currently, there are number of clinical classification systems in use. These are completed by healthcare practitioners based on clinical judgement and expertise. This can be a complex and lengthy process requiring high level of clinical expertise. It has been suggested that professionals need to undertake more than 100 hours of training for an inter-examiner agreement level of kappa coefficient 0.82 (i.e. almost perfect level of agreement kappa 0.82-1) [6]. For professionals who had completed less than 100 training hours, the inter-examiner agreement level was kappa coefficient 0.66 (i.e. substantial agreement). This demonstrates there can be considerable variability in the classification process between healthcare practitioners, which highlights the requirement of an objective classification system.

This paper aims to introduce a novel automatic objective classification method focussing on the physical characteristics across a battery of physical assessment tests previously established as clinically important to determine the spinal function of someone with NSLBP [7]. This will provide the next step towards individualising rehabilitation and assessment appropriate for clinical presentation of NSLBP on basis of the output from an objective classification system. This study also aims to determine which MTw2 IMU tracker and assessment tests are most useful to differentiate between the NSCLBP subgroups. By determining which data is the most useful, it will be possible to demonstrate the feasibility of developing an automatic classification tool.

There has been recent research efforts utilising objective classification methods to classify NSLBP patients into clinical subgroups on the basis of motion data captured using the Vicon optical motion capture system [8]. There, Cardiff Dempster-Shafter objective classification tool was applied to captured motion data to classify the patients into clinical subgroups. The study found a range of accuracies between 93.93% and 98.15% when differentiating NSLBP patients from the healthy controls. The classification accuracies differentiating between the NSCLBP subgroups were more variable with accuracies ranging between 70.27% and 96.8% [8]. The study concluded that automatic classification of NSLBP was possible using the Vicon motion capture system. However, as the Vicon motion capture system is usually available in research settings only, a portable device motion capture device would be a more convenient way to record a patients' motion in clinical settings. Inertial measurement units (MTw2 IMU tracker) are an example of portable motion analysis devices [9]

This paper introduces a novel approach for automatic classification of NSLBP patients into clinical subgroups on the basis of the MTw2 IMU trackers motion data. A number of MTw2 IMU trackers and a number of physical assessment tests were investigated to establish their ability to distinguish between clinical NSLBP subgroups. Simple motion features such as the angular range of displacement were used in classification experiments to reflect measures used by clinicians to classify NSLBP. The results were compared to a study completed by Sheeran L et al. [8].The achieved results were comparable to those presented in [8] and demonstrated the feasibility of developing an automatic classification system of NSLBP on the basis of the MTw2 IMU tracker motion data and simple physical assessment tests.

With further development, this research could reduce the challenge of physical assessment and tailoring rehabilitation methods for NSLBP patients. The ability to prescribe the most relevant rehabilitation methods early could lessen the likelihood of chronicity and the effects of associated psychosocial impact. Classifying NSLBP using IMU trackers could also be used to deliver personalised feedback during rehabilitation, which in turn would support self-management helping patients to do exercises without relying on direct supervision of a physio, thus reducing burden on the healthcare system.

2. Experimental setup

This section introduces the proposed approach in the collection, analysis and subclassification of the motion data for the participants in this study. It also outlines the basis for comparison to the study conducted in [8] and the assumptions that were made to complete the comparison.

2.1. Data Collection

The low back pain participants were recruited from the physiotherapy waiting list at Cardiff and the Vale University Health Board, Wales, U.K. Every participant gave informed consent prior to participating. There were 106 participants in the study (NSLBP n= 85, healthy n=21), who made one visit to the Research Centre of Clinical Kinesiology at Cardiff University, Wales, U.K. Table 1 displays clinical subgroups of NSLBP participants collected. The clinical features of the clinical subgroups are fully detailed elsewhere [8], and [10]. The anthropometric and demographic data of the healthy controls were matched to the participants with NSLBP. This helped reduce bias. An experienced physiotherapy expert in NSLBP management classified the patients into clinical subgroups using a validated and reliable multidimensional classification system [10].

Table 1- Sample size of each subgroup

Subgroup	Sample size
Healthy (H)	21
Flexion pattern MCI (FP)	33
Active extension pattern MCI (AEP)	26
Passive extension pattern MCI (PEP)	16
Multidirectional pattern (MDP)	7
Flexion lateral shift (FLS)	1
Flexion pattern movement impairment (FPMI)	2

The experienced physiotherapist placed four MTw2 trackers (Xsens Technologies) on the spinous process of C7 in the cervical spine (C7), the spinous process of L2 (upper lumbar/ UL), the spinous process of L4 (lower lumbar/ LL) and the sacral promontory (pelvis/ P). These locations are demonstrated in Fig. 2. The MTw2 trackers recorded three angles (roll, pitch and yaw) for each frame of motion at framerate 60 Hz. These angles are demonstrated in Fig. 1. The trackers recorded one angle for each frame of motion. The first frame of motion recorded was used as a reference point for all subsequent angles recorded by the MTw2 trackers for each frame of motion. Each participant was required to complete eight trials of six physical assessment tests (Table 2) with the trackers attached to their spine.

Table 2- Assessment tests completed by each study participant

Assessment test number	Assessment test
1	Flexion (forward bend)
2	Extension (backward bend)
3	Lateral flexion (side bend left and right)
4	Sit to stand
5	Mini squat
6	Sit to stand- with therapist feedback to instructing the participant to keep their

spine straight and lean forward before standing up

a b

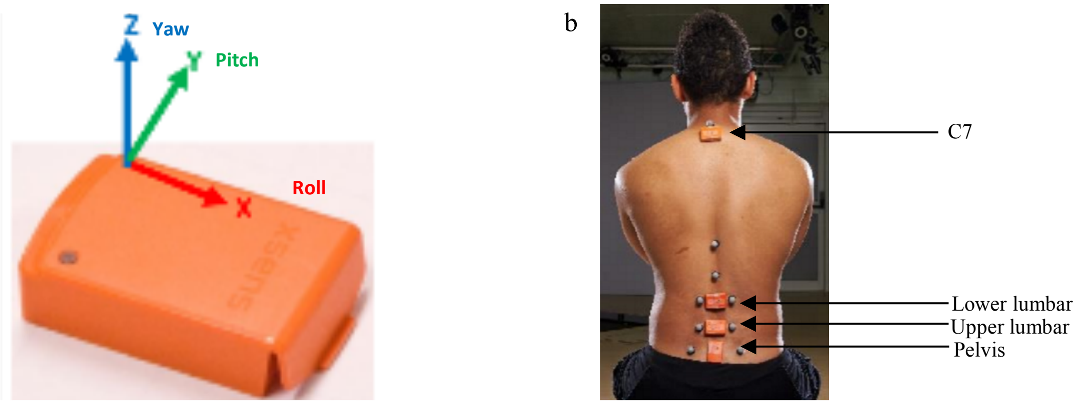

Fig. 1. (a) Yaw, pitch and roll of MTw2 trackers. Adapted from Xsens Technologies. (b) MTw2 trackers placement. The Cervical (C7), Upper lumbar (UL) and lower lumbar (LL) trackers were orientated horizontally. The Pelvis (P) tracker was orientated vertically.

2.2. Data processing

MATLAB (R2019b) was used to clean and process the data in this study. MATLAB Classification Learner was used to train the machine learning algorithms.

Before analysing the study participants motion data, the data was cleaned, and any outliers removed following the processes outlined below. Subgroups MDP, FLS and FPMI were removed from the analysis due to the limited number of samples for these subgroups (Table 1). The data was also removed from this study for those participants who did not have a min of eight repeats for each of the assessment tests (n=31). There were 18 participants with no MTw2 IMU tracker data recorded also excluded from the study. As a result, a total of 49 participants were identified as outliers and removed from the analysed data set. This process left healthy (H) (n=18), flexion pattern (FP) (n=21), active extension pattern (AEP) (n=17), passive extension pattern (PEP) (n=9) to be analysed.

In the remaining sample, the range of angular displacement was extracted for each of the four MTw2 IMU trackers, each of the angles (roll, pitch, yaw) and each of the assessment tests for each participant. The range of angular displacement was calculated by subtracting the smallest angle from the largest angle recorded by the MTw2 IMU trackers for each repetition of the assessment test separately. An average range was then found over the eight repeats of each assessment test completed by each participant.

Two types of classification sets (binary and multinomial) were produced to input into the MATLAB Classification Learner app [11]. In binary classification, only two subgroups were classified, whilst in multinomial classification more than two subgroups were classified. These classification sets contained the extracted average range of angular displacement and were created to determine which MTw2 IMU trackers and assessment tests were the most useful in classifying low back pain patients into clinical subgroups. Different subgroup combinations were created. These were NSLBP and healthy (H), FP and H, AEP and H, PEP and H, FP and AEP, FP and PEP, AEP and PEP. This method was repeated for each of the six assessment tests completed by the patients for each MTw2 IMU tracker position. As well as binary classification sets, a multinomial classification set was created to contain all subclassifications for each MTw2 IMU tracker position and assessment test to determine whether it would be possible to effectively distinguish between the subgroups using these classification sets.

An indirect comparison to the study conducted by Sheeran L et al. [8] was completed to determine whether the classification results achieved using the MTw2 IMU trackers were comparable to those achieved using the VICON motion capture system. The upper lumbar MTw2 IMU tracker classification results from this study were compared to the thoracic region classification results in [8], the lower lumbar MTw2 IMU tracker classification results were compared to the lumbar region classification results, and the pelvis MTw2 IMU tracker classification results were compared to the pelvis region classification results. There may have been discrepancies between the positions of the defined regions and positions of the MTw2 IMU trackers and as a result, different locations on the spine could have been assumed to be the same. This could have resulted in some inconsistencies within the analysis. Only the roll angle was used in the comparison as the predictor importance weights in Fig. 3 demonstrated that this was the most useful angle recorded by the MTw2 IMU trackers when classifying NSLBP patients. In [8], the VICON motion capture data obtained during the assessment test of sit and stand repositioning was analysed. As no MTw2 IMU tracker data had been collected for this assessment test, all assessment tests outlined in Table 2 were compared to the repositioning assessment test. The features used in [8] were repositioning sense absolute error (AE) representing error magnitude, constant error (CE), representing error direction and variable error (VE) representing consistency of error. The calculation of these variables is detailed in [8]. The results obtained using these features were compared to the results obtained in this study using the angular displacement features.

2.3. Data analysis

MATLAB classification learner [11] was used to determine the accuracy, sensitivity and specificity of automatic classification for the classification sets produced. This helped to identify which data was the most useful in determining the NSLBP subgroup for a previously unseen NSLBP patient. Before inputting the data into the classification learner, it was essential to ensure the classification sets were balanced [12]. This meant they were manipulated to have an equal number of each subclassification in the classification sets. If the classification sets were not balanced, this would result in the classification learner classifying into the dominant set when analysing future data. In Table1 it was evident the smallest sample of participants was in AEP (n=9), hence, some participant data had to be removed in order to balance the classification sets. The study participants included in each of the classification sets were randomly selected using the MATLAB *randperm* function [13]. The data for nine participants from each subgroup were randomly selected from the whole data set. As a result, in each binary classification set, there were 18 participants included. This process also ensured that if one researcher was collecting the data and consistently placed the MTw2 IMU tracker incorrectly, there would be less effect on the analysis of the data. This reduced the systematic bias. As a result, the subject-to-variable ratio (STV) was 14:1. This was due to the fact that 20% of the patients in the classification sets were used to test the model. Only one feature was used for binary classification to ensure that an appropriate STV ratio could be achieved as described below.

Only one feature (variable) was analysed during this research to ensure that with a limited sample size, an appropriate STV ratio could be achieved. Previous studies [14] have demonstrated that the STV ratio is a good measure of the machine learning performance. It is often evident that larger data sets have greater statistical power for pattern recognition [14]. It has also been suggested that the lower is the STV ratio, the higher the chance of the machine learning algorithm fitting to the noise within the model rather than the underlying pattern of the data [14]. Overfitting of data could affect the performance of the classifier when analysing unseen patient data.

To obtain the classification results within the MATLAB Classification Learner app, the response was chosen as the subgroup numbers and the extracted average angular range of displacements were selected as the predictors. Each classification set was trained using 5-fold cross-validation. As a result, one-fold was used to validate the model, while the remaining folds were used to train the model. The classification learner repeated the process five times, so each fold was used to validate the model once. The overall classifier accuracy was then calculated from averaging the accuracy of the 5 iterations. A decision tree model classifier was trained on the data [15]. The confusion matrix was analysed to calculate the sensitivity and specificity. These acted as a performance measure of the classifier models. The models were exported to the workspace so that they could be used in determining any future patients NSLBP subgroup automatically.

A classifier is an algorithm which allocates the input data to a specific category. Tree classifiers were used throughout the project, as they work well in complex classification problems due to their adaptivity [15]. The decision tree initially attempts to divide the scatter plot into sub-parts by creating nominal responses. The algorithm begins at a root node (representing the whole classification set) and applies a series of test conditions to an individual observation. This determines which category the data point belongs to. Tree classifiers are an example of a non-parametric classifiers, which is useful when the underlying data distribution is unknown as in this work. The disadvantage of this classifier is that it can be subject to overfitting the data. Tree classifiers are quick to train and therefore require less computational resources.

To identify which features had optimal predictor importance weights, the MATLAB *relieff* function was used. This separated the motion data from the subgroup numbers and identified which features of the motion data had the most optimal predictor effects in determining which subgroup a patient belonged to. Bar charts were produced to visually display the predictor weights (Fig. 3). Negative predictor importance weights were identified as negatively impacting the classification process. The features with negative predictor importance weights were not deemed as useful data as the inclusion of these features could potentially reduce classification accuracy. Predictor importance weights of 0 were identified as having a neutral effect on the subclassification process. In addition, binary classification sets with low trained model accuracies were interpreted as containing fewer useful features than classification sets with higher accuracies. It was investigated whether by combining the useful features, the classification set predictive abilities could be improved. This is known as feature engineering. Feature engineering removes features that do not contribute to the prediction or adversely affect the results by only supplying the model with the optimal form of input.

3. Results

The main findings are presented in Fig. 2 and Table 3 below. Fig. 2 displays the best achieved average accuracies across three trackers (Lower lumbar, Upper lumbar and Pelvis) in 5-fold experiments in binary and multinomial classification experiments along with the accuracies achieved in similar experiments in [8]. Table 3 presents the most useful tracker for each of the classification tasks across all assessment tests and all three MTw2 IMU trackers.

Fig. 3 displays the predictor effects for three angles: roll, pitch and yaw for different assessment tests. These were calculated separately for each tracker and then averaged over all trackers. The presented results are discussed in detail in the next Section.

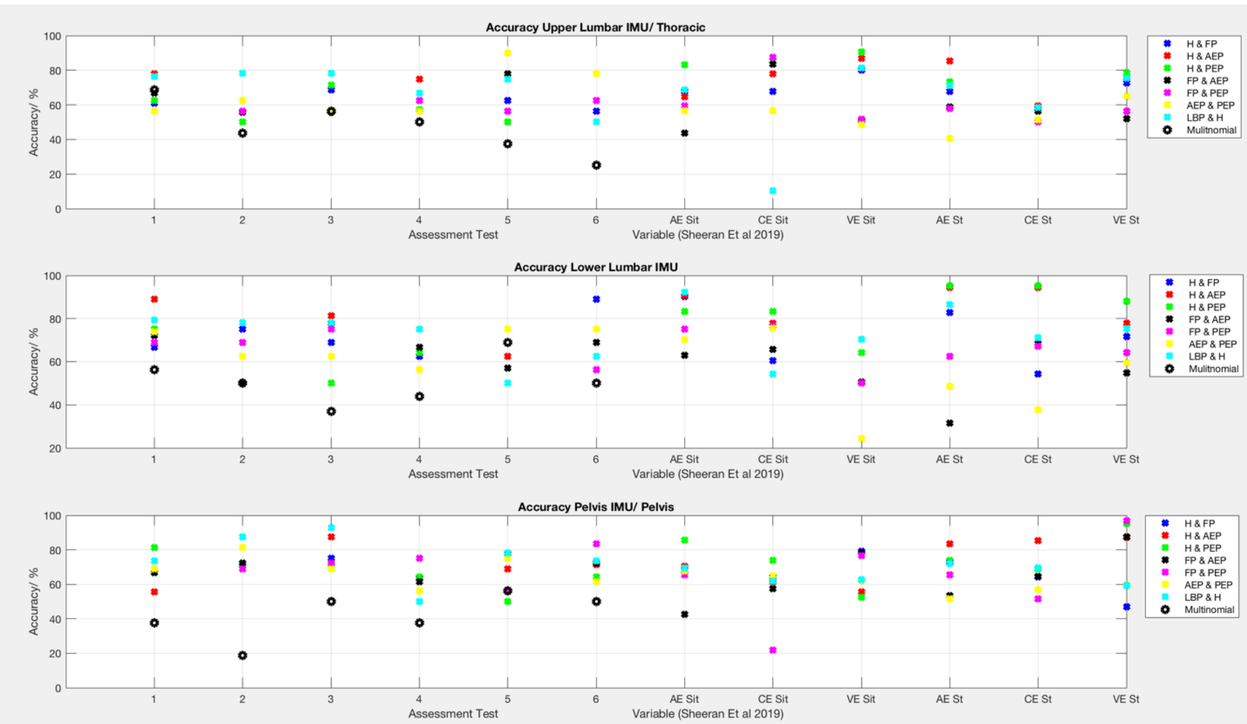

Fig. 2- Classification accuracy results and comparison to Sheeran et al. (2019) [8]

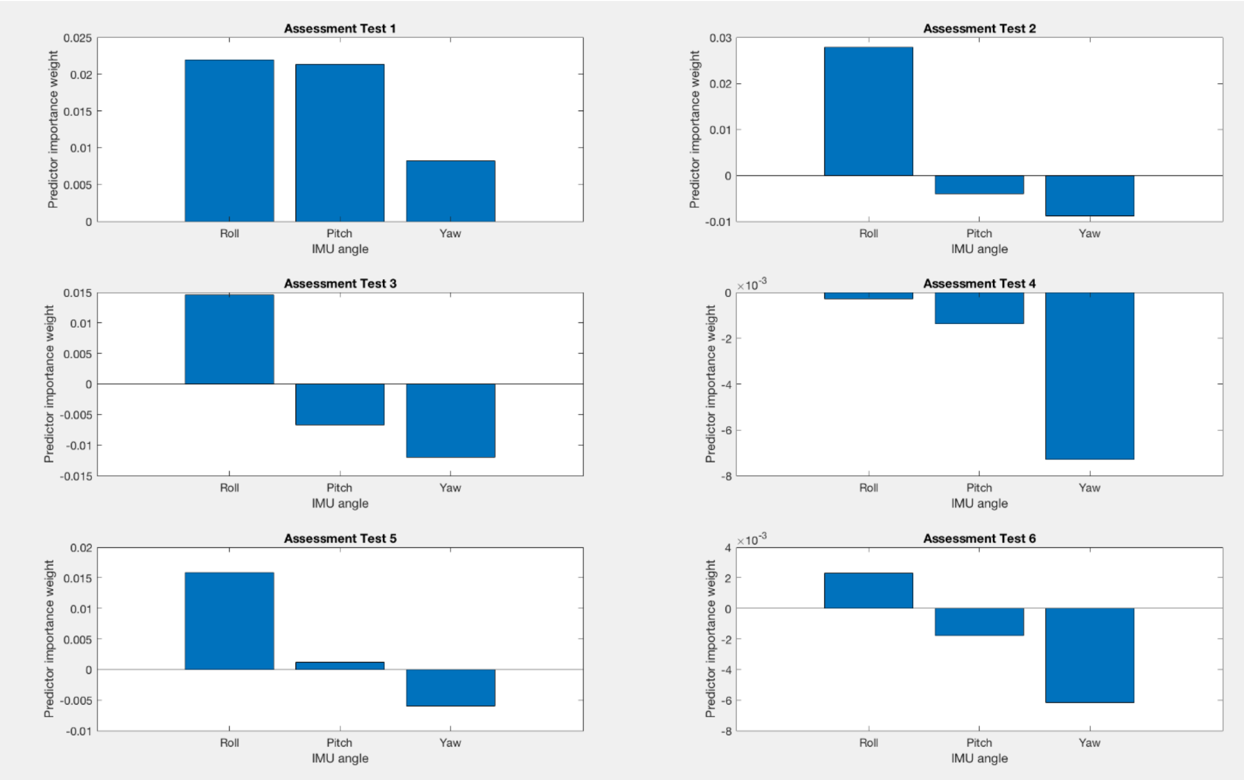

Fig. 3- Average (over four MTw2 trackers) predictor importance weights of the tracker angles (roll, pitch and yaw) for each assessment test

Table 3- the most useful trackers and assessment tests to distinguish between certain subclassification combinations

Subclassifications	MTw2 IMU trackers	Assessment test	Accuracy/ %
H & FP	Lower lumbar	6 (Sit to stand- with therapist feedback)	89
H & AEP	Lower lumbar	1 (Flexion- forward bend)	93.5
H & PEP	Pelvis	1 (Flexion- forward bend)	81.3
PEP & AEP	Upper lumbar	5 (Mini squat)	79
FP & PEP	Lower lumbar	5 (Mini squat)	81
FP & PEP	Pelvis	5 (Mini squat)	90
NSLBP & H	Pelvis	3 (Lateral flexion- side bend left and right)	92.3
Multinomial	Upper Lumbar	1 (Flexion- forward bend)	69

4. Discussion

This section discusses the application of the novel automatic objective classification method and the importance of different assessment tests and different MTw2 IMU trackers for distinguishing between the subgroups of NSLBP.

4.1. Determining the most useful data in subgrouping low back pain patients

The predictor importance weights in Fig. 2 demonstrated the roll angle recorded by the MTw2 trackers was on average (over four MTw2 trackers) the most useful angle in sub classifying patients for all assessment tests completed. In the majority of assessment tests, it was found the yaw had a negative impact on the classification process. From these findings, the roll was used for C7, UL and LL MTw2 trackers in the remaining analysis of results. The pelvis MTw2 tracker had different orientation to the rest of the trackers, with a different angle rather than roll likely to have been more useful in the analysis. This is to be addressed in future research.

Table 3 lists the most useful combinations of MTw2 IMU trackers and assessment tests for each of the subclassifications using the average displacement range of the roll angle as the predictor variable. The useful MTw2 IMU trackers in discerning between the binary classification subgroups were lower lumbar and pelvis. The fact that different assessment tests were more useful in detecting different subgroups fitted with the nature of the different movement and motor control impairment classification subgroups, each likely to present different challenges to patients belonging to different subgroups. For example, flexion- forward bend assessment test (test number 1) would be most challenging for patients with flexion related back pain (FP) whilst for patients with extension related back pain (EP), the most challenging test would be extension backward bend (test number 2). Nonetheless, as outlined in Table 3, it was evident that to discern between the low back pain subgroups sit to stand, flexion, lateral flexion and mini squat assessment tests were most useful. The implications of this evidence will be subject to future research. This, therefore, demonstrates the efficacy of feature engineering technique as different assessment tests may assist in identifying different subclassifications more successfully.

4.2. Binary and multinomial classification

From Fig. 3, it is evident that on average, the binary classification results achieved over the different subgroup combinations for each of the exercises and MTw2 IMU trackers produced higher accuracies compared to the multinomial classification set which was created to contain all subgroups of NSCLBP rather than only two of the subgroups. This suggests multinomial classification may not be suitable to distinguish between the NSCLBP subgroups analysed in this study. As demonstrated in Fig. 3, it was particularly evident that the classification learner found it challenging to differentiate between subgroups AEP and PEP. This may have been due to these two subgroups of low back pain having similar motion characteristics. Patients in these subgroups are clinically similar, both typically experiencing pain during activities involving spinal extension (e.g. walking, standing, bending backwards). Multinomial classification increases the complexity of the machine learning task and therefore, it was expected that a reduction in accuracy might have occurred compared to binary classification, especially with a limited sample size. When assessing future patient data, a patient could belong to any of the subgroups of low back pain. This means that multinomial classification would have been a promising way to identify a patient's unknown subgroup. However, when it comes to classifying NSCLBP into multiple known subsets, multinomial classification produced lower accuracies in comparison to the binary classification, rendering it currently not suitable way of developing an automatic classification tool in the preliminary stage of a more comprehensive research project.

4.3. Comparison to previous research

There were many variations between the study conducted by Sheeran et al. (2019) and this study, preventing us to make direct comparisons. The assumptions made could have also potentially impacted the effectiveness of the comparisons. Despite the differences between the two studies outlined in *Section 2.2., Data Processing*, the aim was to establish whether objective classification methods applied to MTw2 IMU tracker data can achieve the same accuracy as objective classification applied to Vicon data.

Fig. 3 demonstrated that across all three MTw2 IMU trackers/ regions of the spine, the MTw2 IMU trackers had lower standard deviations between the various subclassification combinations within the classification sets. This may have been explained by the variables measured. Clinically, the repositioning variable analysed in [8] was more likely to have a larger standard deviation than the range of angular displacement analysed in this study. In [8], the binary classification sets differentiating between one subgroup of low back pain and the healthy control resulted in higher accuracies than the classification sets differentiating between various subclassifications of low back pain. This means repositioning error is more likely to discern between H and NSLBP. The use of MTw2 IMU trackers resulted in higher accuracy results than achieved by [8] for binary classification when discerning between different subgroups of NSLBP. This is an important result as clinically, it is difficult to distinguish between the subgroups of NSLBP whereas it is easy to distinguish between with or without LBP through a patient’s report of their symptoms. When differentiating between healthy and NSLBP, it was particularly evident the pelvis MTw2 IMU tracker for the lateral flexion exercise was the most useful. This may have been explained by previous study observations that the lumbar pelvic rhythm is reversed in patients with low back pain (when compared to healthy control) during a lateral flexion motion [15]. Due to the close anatomical link between the spine and pelvis, it was unsurprising, the lower lumbar and pelvis MTw2 IMU tracker were useful in detecting the difference between NSLBP and healthy subjects.

Similarly, to [8], this study also found the lowest accuracies when differentiating between AEP and PEP. However, the accuracies discerning between AEP and PEP in this study (79%) exceeded that of the accuracy achieved in [8] (70.27%) when discerning between these subgroups. In Table 3, it was evident that the most useful MTw2 IMU tracker / assessment test combinations for each subgroup show clinically promising results as all binary classification subgroup combinations produced the same accuracy as inter-examiner agreement between clinicians undergoing more than 100 training hours (accuracy range 73-100%) [6] without the complexities of manual diagnosis.

For the MATLAB classification learner to reliably classify data, the classification sets were required to be balanced. It was evident the data sets were not balanced in the study conducted in [8]. This enabled more patient data to be analysed. Despite this, for this study, there was an STV ratio of 14:1 as the only feature analysed was the range of angular displacement, whilst an STV ratio of 10:1 was achieved in [8]. These STV ratios are comparable; however, the higher STV ratio in this study could have resulted in the MATLAB classification learner app having a slightly increased predictive ability compared to the Cardiff Dempster-Shafter classifier. The lower the STV ratio, the higher the chance the machine learning algorithm will fit to the noise within the supplied data, rather than the underlying patterns in the data. Models trained with low STV ratios are more likely to identify patterns that do not exist. Therefore, the achievement of a higher STV ratio with high accuracy results may be more clinically important.

To conclude this comparison, it is evident that the classification results achieved in this study were comparable to [8]. The results achieved were slightly below the classification results achieved in [8] when discerning healthy individuals from NSLBP. The classification results achieved in this study exceeded the classification results in [8] when discerning between the different subgroups of NSLBP. This indicates MTw2 IMU tracker motion data has the potential to be used to train objective classifiers to distinguish between different movement-controlled impairments of NSLBP. As MTw2 IMU trackers are more portable than the VICON motion capture system, there is a greater opportunity for them to be used in clinical settings as indicated above in order to tailor exercise rehabilitation therapies. This might suggest the results achieved in this study are useful when classifying low back pain patients into clinical subgroups given the scope and convenience of portable sensor technologies within in the healthcare system.

5. Conclusion

Currently, clinicians classify low back pain patients into clinical subgroups to offer personalised exercise rehabilitation programmes. Currently used clinical classification systems are based on clinical judgement and require substantial amount of expertise. This study explores the use of an automatic classification tool applied to MTw2 IMU tracker data collected whilst individuals performed a battery of physical assessment tests to distinguish between different NSLBP clinical subgroups in order to tailor exercise rehabilitation.

To the best of our best knowledge, this is the first study towards automatic classification of clinically recognised NSLBP patient subgroups on the basis of MTw2 IMU tracker data. It was evident that patients from different NSLBP subgroups have different motion characteristics, and by using machine learning techniques, it was possible to complete binary classification. In Table 3 it was identified which assessment tests and which MTw2 IMU trackers were the most useful to classify NSLBP patients into these subgroups. The useful assessment tests were sit to stand, lateral flexion, flexion and a mini squat. The roll angle was the most useful on average over four MTw2 IMU trackers, although this should be investigated for the four angles separately in future research as pelvis tracker was oriented differently to the three other trackers. The roll angle corresponded to the motion in saggital plane for the three spinal trackers and to the frontal plane for the pelvis tracker. The results achieved in this study are preliminary and require further analysis to increase the scope of the research. Future developments of this research may help reduce the complexities of diagnosing and treating this heterogeneous disorder.

Acknowledgements

The authors would like to acknowledge the support of the Centre for Artificial Intelligence, Robotics and Human-Machine Systems (IROHMS), operation C82092, part-funded by the European Regional Development Fund (ERDF) through the Welsh Government.

References

[1] Vos, T., Abajobir, A., Abate, K. et al. 2016 Disease and Injury Incidence and Prevalence Collaborators (2017) "Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016" *Lancet*.; **390**: 1211-1259.

[2] Hartvigsen, J., Hancock, M., Kongsted A. et al. (2018) "What low back pain is and why we need to pay attention" *The Lancet* **391(10137),** 2356–2367.

[3] Airaksinen, O., Brox, J., Cedraschi, C. et al. (2006) "Chapter 4 European guidelines for the management of chronic non-specific low back pain" *European Spine Journal* **15(S2)** s192–s300.

[4] Chou R. (2011) "Low Back Pain (Chronic)" *Am Fam Physician* **84 (4)**: 437-438.

[5] NICE (2016) "Low back pain and sciatica in over 16s: assessment management" *NICE Guideline NG59* Available at: https://www.nice.org.uk/guidance/ng59 [Accessed 24th April 2020]

[6] Fersum, K., O'Sullivan, P., Kvale, A. et al. (2009). "Inter-examiner reliability of a classification system for patients with non-specific low back pain" *Manual Therapy* **14(5):** 555-561.

[7] Sheeran and Robling, (2019) "Spinal function and exercise performance classification framework: Expert consensus study" Proceedings of the 10th Interdisciplinary World Congress on Low Back and Pelvic Girdle Pain, Antwerp, Belgium, 28 -31 October 2019.

[8] Sheeran, L., Sparkes, V., Whatling, G. et al. (2019). "Identifying non-specific low back pain clinical subgroups from sitting and standing repositioning tasks using a novel Cardiff Dempster-Shafter Theory Classifier" *Clinical Biomechanics* **70(2019)** 237-244.

[9] Al-Amri, M., Kevin, N., Button, K. et al. (2018) "Inertial Measurement Units for Clinical Movement Analysis: Reliability and Concurrent Validity." *Sensors*, ***18***, 719.

[10] O'Sullivan, P. "Diagnosis and classification of chronic low back pain disorders: maladaptive movement and motor control impairments as underlying mechanisms" Man Ther. 2005 Nov;10(4):242-55. Epub 2005 Sep 9.

[11] Mathworks. "Classification Learner" Available at: https://uk.mathworks.com/help/stats/classificationlearner-app.html [Accessed 24th April 2020]

[12] Picek, S., Heuser, A., Jovic, A. et al. (2018). "The Curse of Class Imbalance and Conflicting Metrics with Machine Learning for Side-channel Evaluations" *Ruhr Universitat Bochum.*

[13] Mathworks "randperm" Available: https://uk.mathworks.com/help/matlab/ref/randperm.html [Accessed 24th April 2020]

[14] Vabalas, A., Gowen, E., Poliakoff, E. et al. (2019) "Machine learning algorithm validation with a limited sample size" *PLOS ONE* **14(11)**: e0224365.

[15] Breiman, L., Friedman, J., Olshen, R. and Stone, C. *Classification and Regression Trees.* Boca Raton, FL: Chapman & Hall, 1984