Title	Supervised learning techniques and their ability to classify a change of direction task strategy using kinematic and kinetic features
Author(s)	Richter, Chris; King, Enda; Falvey, Éanna; Franklyn-Miller, Andrew
Publication date	2017-10-01
Original citation	Richter, C., King, E., Falvey, E. and Franklyn-Miller, A. (2017) 'Supervised learning techniques and their ability to classify a change of direction task strategy using kinematic and kinetic features', Journal of Biomechanics. In Press. doi:10.1016/j.jbiomech.2017.10.025
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://www.sciencedirect.com/science/article/pii/S0021929017305572 http://dx.doi.org/10.1016/j.jbiomech.2017.10.025 Access to the full text of the published version may require a subscription.
Rights	**© 2017 Elsevier B.V. This manuscript version is made available under the CC BY-NC-ND 4.0 license.** http://creativecommons.org/licenses/by-nc-nd/4.0/
Embargo information	Access to this article is restricted until 12 months after publication by request of the publisher.
Embargo lift date	2018-10-31
Item downloaded from	http://hdl.handle.net/10468/5011

Accepted Manuscript

Supervised learning techniques and their ability to classify a change of direction task strategy using kinematic and kinetic features

Chris Richter, Enda King, Eanna Falvey, Andrew Franklyn-Miller

PII: S0021-9290(17)30557-2
DOI: https://doi.org/10.1016/j.jbiomech.2017.10.025
Reference: BM 8426

To appear in: *Journal of Biomechanics*

Received Date: 24 August 2016
Accepted Date: 15 October 2017

Please cite this article as: C. Richter, E. King, E. Falvey, A. Franklyn-Miller, Supervised learning techniques and their ability to classify a change of direction task strategy using kinematic and kinetic features, *Journal of Biomechanics* (2017), doi: https://doi.org/10.1016/j.jbiomech.2017.10.025

SUPERVISED LEARNING TECHNIQUES AND THEIR ABILITY TO CLASSIFY A CHANGE OF DIRECTION TASK STRATEGY USING KINEMATIC AND KINETIC FEATURES

Chris Richter[a,*], Enda King[a,b], Eanna Falvey[a,c], Andrew Franklyn-Miller[a,d]

[a]*Sports Surgery Clinic, Santry Demense, Dublin 9, Ireland*
[b]*Department of Life Sciences, University of Roehampton, United Kingdom*
[c]*Department of Medicine, University College Cork, Ireland*
[d]*Centre for Health, Exercise and Sports Medicine, University of Melbourne, Australia*

Abstract

This study examines the ability of commonly used supervised learning techniques to classify the execution of a maximum effort change of direction task into predefined movement pattern as well as the influence of fuzzy executions and the impact of selected features (e.g. peak knee flexion) towards classification accuracy. The experiment utilized kinematic and kinetic data from 323 male subjects with chronic athletic groin pain. All subjects undertook a biomechanical assessment and had been divided previously into 3 different movement strategies in an earlier paper. Examined supervised learning techniques were: a decision tree, an ensemble of decision trees, a discriminant analysis model, a naive Bayes classifier, a k-nearest-neighbour model, a multi-class model for support vector machines, a stepwise forward regression model, a neural network and a correlation approach. Performance (measured by comparing the predefined and classified movement pattern) was highest for the correlation approach (82 % - CI 81 to 83 %) and support vector machine (80 % - CI 79 to 80 %). The percentage of fuzzy observations within the data was between 15 and 25 %. The most informative features for classification were: hip flexion

*Corresponding author
Email address: `mr.chris.richter@gmail.com` (Chris Richter)

angle, ankle rotation angle, a flexion moment [ankle and hip] and thorax flexion. Findings of this study support the assumption that multiple patterns are used to execute a movement task and demonstrate that classification models can predict movement patterns with a high accuracy (83 %).

Keywords: movement classification, subgroup analysis, change of direction manoeuvre

Word count: 4027

---

1. Introduction

It is believed that the underlying mechanism of musculoskeletal overuse pathologies, may be the result of a repetitive and poorly controlled loading of tissues rather than a single excessive movement or incident [6, 11]. Although a variety of intrinsic and extrinsic factors can be identified [29] it is hypothesised that the lack of variability in movement strategies (e.g. continuous load of the same structure) or the inappropriate execution of a movement are propagative of overuse injury.

Studies in the field of biomechanics often aim to identify a casual mechanism for musculoskeletal overuse pathologies and generally utilize a single group design [3], which proposes that a selection of individuals (with similar anthropometric measures, gender or age) is homogeneous in executing a movement task. However, studies highlight that differences exist within movement patterns across individuals [4, 6, 8, 13, 14, 20, 26, 28]. Consequently, utilizing a single group design has the potential to mask features that relate to a dependent variable (e.g. injury related factor; [3, 20, 24, 25]). An alternative to this model is the single case design, which examines one individual assuming a unique movement strategy with a unique injury related mechanism [3]. While providing an insight into potential injury mechanism, findings are dependent on the studied individual, their mental and physical condition during the data capture and the comparison of training interventions is difficult (as the ordering of interventions might affect results; [12, 2]), while the number of trials can introduce fatigue potentially affecting findings [2]. Consequently, the generalization of findings is problematic. An alternative design that combines both the generalizability of a single group and the flexibility of the single case design is the analysis of subgroups. The subgroup design presupposes that multiple strategies (movement pattern) can be used to execute a movement task. The number of strategies and the membership of a sample (subject or trial) to a movement strategy (movement pattern) can be identified using clustering techniques and the subsequent pattern specific analysis can identify injury related features for a specific movement strategy. In fact, a commonly employed version of a subgroup analysis is the differentiation between genders or age in studies.

A subgroup design has been advocated in a number of clinical studies investigating the aetiology of musculoskeletal pathologies [4, 8, 13, 14, 26, 28]. The subgroup design has also been reported to increase prediction accuracy by up to 11 % compared to a single group design when predicting jump height in counter-movement jumps using ground reaction forces [20], while reducing the amount of data (fewer features) used to predict jump height. As such a subgroup of analysis may prove particularly useful for clinicians, trainers and researchers as the identification of specific movement deficits may facilitate the development of individualized rehabilitation, injury prevention and performance programs. Furthermore the ability to examine movement pattern selection may also enhance the understanding of musculoskeletal pathologies.

Most recently, the subgroup design was used to evaluate movement patterns during a maximum effort change of direction task in patients with athletic groin pain [6]. However, the practical use of this study is limited because no classification guidelines or algorithms are provided and future studies have no ability to classify a sample's movement into the strategies defined by Franklyn-Miller et al. [6]. The most objective way to decide on the membership of a sample / trial to a predefined movement strategy is the use of a classification model that has been trained, using supervised learning, to recognize the relationship between input features (e.g. max. knee abduction) and the target (predefined movement strategy). The selection of the right learning technique can be challenging as a number of learning techniques / theories exist with different abilities to learn relationships between target and input features [7, 30]. In the past, linear models (e.g. regression models) have been proven to be very useful as they are easy to interpret (indication of importance of a feature) and implement. However, regression models assume that the input features and target have a linear relationship [7], the analysis of a mix of numerical and categorical features is problematic and the number of features is limited because a "input to observations" ratio is required to generate a robust prediction model (1:10 to 1:15; [1, 15]). If, for example, 12 features are fed into a regression model a minimum of 120 to 180 samples is required to develop a robust prediction model. This "input to observations" ratio can represent a problem in biomechanical studies because a relatively low number of samples but large number of input features can generate an overfitted classification model. Techniques that are frequently used outside the field of biomechanics are k-nearest-neighbour classifiers, naive bayesian classifiers, decision trees, neural networks and support vector machines (SVM). These techniques have different advantages and limitations, which are briefly described in this paper[1], and their performance is influenced by the characteristics of a dataset.

The k-nearest-neighbour classifier is similar to the regression model but is more flexible, as it utilizes the distance of neighbouring samples to classify input data, allowing non linear boundaries and continuous training. However, when using a k-nearest-neighbour classifier, input features should be treated (e.g. normalized) [7], which necessitates a normalization approach. Compared to the k-nearest-neighbour classifier, naive bayesian classifiers and decision trees use more complex statistics when classifying a sample. An advantage of bayesian classifiers is that they can be trained continuously and allow a easy interpretation[2] of what the classifier has learned [23]. A shortcoming in bayesian classifiers

[1] The interested reader is referred to [7, 10, 23] for a detailed description.

[2] Bayesian classifiers store probabilities of each feature - e.g. the ability of a feature to divide the target. These probabilities can be examined to determine the impact of a feature

is the ability to learn predicting targets based on the interaction of features [10]. This is not a problem for decision trees, which can easily cope with interaction between input features and also allow the interpretation of what the tree has learned. However, decision trees can become extremely complex if the number of branches is not limited. The most complex classifiers discussed here are neural networks and SVM's, which have been described to be extremely powerful as they have the ability to learn very complex linear and non-linear relationships [23]. However, it is difficult to understand what the model has learned for the user, which represents a huge limitation - e.g. when trying to understand injury mechanism.

In addition to the difficulty of selecting a learning technique, it is important to a) account for different membership types (fuzzy and logic) when examining movement pattern and b) determine high impact features (features with a significant impact to the prediction model) to give feedback to the user. The term fuzzy indicates that a sample holds characteristics of two strategies, while logical membership indicates that a sample holds solely characteristics of one strategy. Fuzzy memberships can influence the accuracy measures of prediction models (if they have not been considered in the cluster generation) as the membership of a fuzzy sample can change. Regarding the impact of a feature to the prediction process, studies have shown that a professional trainer and researcher are able judge risk of injury during a movement task due to experience and knowledge of a injury mechanism [16, 17]. Consequently, reporting features that impact the model most may allow the classification or screening outside the laboratory.

The primary aim of this study was to assess and compare the performance of commonly used supervised learning techniques to classify previously defined movement patterns using kinematic and kinetic input features during a maximal effort change of direction task. The secondary aim was to examine the influence of samples with fuzzy memberships and to explore how to "spot" a fuzzy membership. The third aim was to report which features have greatest

towards classification.

impact when classifying movement patterns of a change of direction in athletic groin pain patients.

2. Methods

Data Set. This study used a data set that has been published previously in Falvey et al. [5] and Franklyn-Miller et al. [6]. The data contained three hundred and twenty-three male subjects aged 27.6 (± 7.6) years old, 180 (± 6.0) cm tall and 81.9 (± 9.4) kg with athletic groin pain and a median time of 36 (IQR 16 to 75) weeks between onset of symptoms and presentation to the Sports Surgery Clinic. The Sports Surgery Clinic Hospital Ethics Committee approved the study (Ref 25EF011) and all subjects signed informed consent. The study was registered at Clinicaltrials.gov (NCT02437942).

Data Analysis. The methods used in this study can be described in 3 steps: (1) detection of movement strategies, (2) the selection of a classification model and (3) the identification of high impact features (see table 1).

Detection of Movement Strategies. The following paragraphs briefly describes the steps used to detect movement strategies within the data, while a detailed description is reported in Franklyn-Miller et al. [6]. A motion analysis system, including eight infra-red cameras (Bonita-B10, Vicon Motion Systems Ltd, UK) and two force platforms (BP400600, AMTI, USA), was used to capture trajectories of 24 markers that were attached to the subjects (according to Plug-In-Gait) and the forces applied to the ground. Both systems were synchronized and controlled simultaneously by the software package Nexus (Nexus 1.8.5, Vicon Motion Systems Ltd, Oxford, UK). Prior to data collection, every participant performed a standard warm-up routine, including two sub-maximal repetitions of a planned 110 degree high speed change of direction manoeuvre (left and right) before three change of direction manoeuvres were captured on each leg. Before applying inverse dynamics (to calculate tri-planar joint kinematics and kinetics [31]) the captured trajectories (200 Hz) and ground reaction forces (1000

Hz) were filtered using a fourth-order Butterworth filter (cut-off frequency of 15 Hz; [9]). Only the best trial of a subject's painful side (by clinical palpation) was used for analysis. The start and end of the change of direction manoeuvre was defined by the ground reaction force (>5 N). Curves were normalized to 101 frames and to bodyweight (only joint moments). To eliminate variations in the timing of the start of the acceleration phase, ensuring the comparability of physical conditions throughout the movement cycle, all curves were landmark registered to the start of the acceleration phase (first instance of forward ground reaction forces; [18]).

Before clustering the captured kinematic and kinetic measures, Franklyn-Miller et al. [6] reduced the data to its main components (capturing the behaviour of a samples trial using multiple features) using the idea of analysis of characterizing phases [22]. Analysis of characterizing phases detects phases of variation within a waveform, using VARIMAX rotated principal components[3] that together described 99 % of the variances in the waveform [21]. Subsequently, a score ($feature$) was generated to capture the samples behaviour for every identified phase (k) as the summed difference between a subject's waveform (p) and the average waveform (q) within every time point (i) between the start (n) and end (m) of a phase. This was completed for every kinematic and kinetic measure (j) creating a feature matrix (rows = number of phases across each measure; columns = number of subjects; see Equation 1).

$$feature_{j,k} = \sum_{i=n}^{m} p(i) - q(i) \tag{1}$$

To maximize the ability to identify movement strategies this feature matrix was transformed into its correlation matrix to change the proximity measure from a distance to a relationship measure [20]. Gap statistic[4] was then used to

[3]The VARIMAX rotation can be used to increase the interpretability of principal components by revealing more meaningful components of variation [19].

[4]Gap statistic is a method that can be used to determine how many classes or clusters are within a dataset by comparing the within-cluster dispersion of a data set for a number

identify number of movement strategies within the data set [27] - detecting 3 clusters in the data. Subsequently, hierarchical clustering was used to cluster each sample into three movement strategies (distribution: cluster 1 = 19 %; cluster 2 = 41 % and cluster 3 = 40 %). The assigned movements strategies of this methods were used as the target variable in this study for the classification techniques and treated as ground truth.

Classification Model Selection. To generate and validate a classification model, the dataset consisting of ankle, knee, hip, pelvis and thorax angles as well as ankle, knee and hip moments was divided into training (75 %) and test data (25 %). When dividing the data set, the differences in cluster distribution were considered. Input features were calculated based on the phases of variation within the training data as described in Equation 1. The reader should note that no transformation of the proximity measures was applied during the classifier selection as the aim was to classify not to cluster.

This study used learning techniques included in the statistical toolbox of MatLab (R2015a, MathWorks Inc., USA) as well as a correlation approach. Tested were the classification performance of a decision tree (fitctree), an ensemble of decision trees (n trees = 50; TreeBagger), a discriminant analysis model (fitcdiscr), a naive bayesian classifier (fitcnb), a k-nearest-neighbour model (fitcknn), a multiclass model for support vector machines (fitcecoc), a regression model (mnrfit; in stepwise forward) and a neural network (patternnet). In addition to the techniques included within the statistical toolbox, a correlation method (Corr2Mean) was also tested. This method classified a sample based on its correlation to the cluster average (determined from the training data). For example, a sample within the test data might be correlated to the mean of the training data samples of cluster 1 with $r = .04$, $r = .58$ for cluster 2

of requested cluster solution (e.g. 2 to 25) to the average within-cluster dispersion cluster solution computed from x reference data sets (uniform copy of the real data) that hold a null distribution (e.g. no underpinning pattern). The interested reader is referred to the text of Tibshirani et al. [27] or Martinez et al. [10] for further information.

and $r = .95$ for cluster 3, in which case the sample would be classified as cluster 3.

Each supervised classification technique was trained using the non-transformed feature matrix and previously defined movements strategy of the training data - as defined by Franklyn-Miller et al. [6]. The generated model was then used to predict the movement pattern of the samples within the test data, based on the fed feature vector of the sample. The predicted and pre-defined membership was compared to determine the accuracy of classification. The described process was repeated 100 times using different training and test datasets (subjects were randomly assigned into training and test data in every iteration) to achieve a robust measure of the expected accuracy. It should be noted that the features calculated from phases of variation may change during the simulation as they are dependent on the variations within the training data set (as illustrated in figure 4).

To allow the identification and examination of fuzzy samples, the following variables were stored throughout every repetition of the model training process: frequency of the selection of a sample as test data, predicted cluster and the probabilities / decision criteria to judge the cluster membership[5]. The frequency of selection of a sample as test data and predicted movement pattern were used to calculate the percentage of misclassification rate of each sample (n misclassified / n selected). Based on the misclassification rate was a zone of logic, fuzzy and potentially miss clustered observations defined (< 15; 15 - 85 %; > 85 % misclassified respectively). The reader should note that these ranges are subjective and where pick in a post hoc manner. The probabilities / decision criteria to assign membership were carried on to judge the strength of membership to a movement pattern of a sample and to allow the generation of an equation of how to "spot" a fuzzy samples.

[5]The probabilities or decision criteria refer to a probability estimate rather than hard classifications to a cluster, similar as the r values in the Corr2Mean approach.

Identification High Impact Features. To identify features with a high impact toward classification, two steps were performed: a) identification of robust phases of variation to calculate features and b) utilizing a stepwise forward approach to determine the importance of a feature towards classification.

To identify robust phases of variation, start and end points of phases of variation were recorded during the trainings process. A phase was considered "robust" if it occurred in at least 95 % of the repetitions of the trainings process and spanned over at least 5 % of the data.

The phases of variation satisfying these criteria were used to calculate features describing the behaviour of a sample (as described previously). Then every generated feature was individually used to train a classification model and to predict the movement pattern of the test data. This process was repeated 100 times for every feature separately using different training and test dataset (subjects were randomly assigned into training and test data) to achieve a robust measure of the expected accuracy. The feature with the highest mean accuracy was judged to be the first high impact feature and considered as "model-base". Before identifying other high impact features, all features that correlated greater than 0.7 with the model-base were excluded to increase interpretability. Subsequently, the same process was performed using every non-model-base in combination with the model-base feature. The "add-on-feature" with the highest mean accuracy was judged to be a high impact feature and added to the model base. Again all features that correlated greater than 0.7 with the identified feature were excluded. The process was repeated until every feature was added to the model base or excluded due to multicollinearity to a model-base feature. The combination of features that first resulted in a classification accuracy plateau was considered high impact. All processing and analysis as performed using MatLab (R2015a, MathWorks Inc., USA).

3. Results

Performance of Approaches. The classification models differed in their performance, when using all features within the features matrix, to classify the observations in the test sample (figure 1) and were ranked, based on their overall average accuracy as follow (decreasing accuracy): Corr2Mean (82 %; Range 76 to 89 %; CI 81 to 83 %), SVM (80 %; Range 69 to 87 %; CI 79 to 80 %), stepwise regression (72 %; Range 57 to 84 %; CI 71 to 73 %), naive bayesian (72 %; Range 55 to 79 %; CI 71 to 73 %), k-nearest neighbour (67 %; Range 55 to 79 %; CI 66 to 69 %), neural network (67 %; Range 49 to 78 %; CI 65 to 69 %), decision tree (63 %; Range 53 to 70 %; CI 62 to 64 %), random forest (61 %; Range 53 to 70 %; CI 61 to 62 %), discriminant analysis (53 %; Range 42 to 64 %; CI 51 to 54 %).

Fuzzy Observations. Only the models that performed best using all features within the features matrix are presented (Corr2Mean and SVM). For Corr2Mean, 11 % of the samples were included in the miss class zone, 15 % were within the fuzzy zone, while 74 % of the samples were located in the logic zone. Based on the stored membership decision criteria, throughout the simulations, the following equation was derived subjectively in a post hoc manner to best separate fussy and logic samples: $difference = distance_{2^{nd}choice} - distance_{chosen}$. Samples with a ratio between 0.876 and 0.954 can be considered fuzzy (6 % error: 18 logic samples were defined fuzzy while 3 fuzzy samples were defined logic; figure 2).

For the SVM technique, 5 % of the samples were located in the miss class zone, 25 % were included within the fuzzy zone, while that rest (70 %) was within the logic zone. Based on the stored membership decision criteria, the following equation was derived subjectively in a post hoc manner to be best to separate fussy and logic samples: $difference = distance_{2^{nd}choice} - distance_{chosen}$. The used equation and differences between -0.407 and -0.098 can be considered fuzzy (18 % error: 54 logic samples were defined fuzzy while 3 fuzzy samples were defined logic; figure 3).

High Impact Features. Sixty-seven key phases met the pre-set criteria to be a robust key phase (figure 4). Phases of variance were consistent across the randomly chosen training data for most joint angles, except of phases in knee and hip rotation angles as well as pelvis flexion and abduction angles. While moment data was stable in variability, only a few key phases spanned over 5 % of the movement (figure 4).

When using only a subset of features both the Corr2Mean and SVM approach reached classification accuracies within the confidence intervals observed before (when utilising all the features of the features matrix). The Corr2Mean generated even higher accuracies then using a subset of 17 features (+ 4%). High impact factors for the Corr2Mean, in order of selection, were (phase & mean accuracy): hip flexion angle (27 to 41 & 20 %), ankle rotation angle (77 to 86 & 53 %), ankle flexion moment (19 to 24 & 78 %), pelvic drop (94 to 100 & 79 %), and thorax flexion (92 to 100 & 81 %; figure 5). High impact factors for the SVM were (in order of selection; phase mean accuracy): hip flexion angle (27 to 41 & 69 %), ankle rotation angle (77 to 86 & 76 %), hip flexion moment (19 to 24 & 77 %) and thorax flexion (92 to 100 & 78 %; figure 6).

4. Discussion

The findings of this study indicate that the executed movement pattern during a maximal effort change of direction task can be classified using a simple correlation approach, which demonstrated the highest classification accuracy and the better ability to identify fuzzy memberships. The mean features of the cluster 1, 2, and 3 defined by Franklyn-Miller et al. [6] are reported in table 3 and can be used to estimate movement strategy in other studies that examine a 110 degree change of direction.

This allows the analysis of maximal effort change of direction data from new perspectives and the development of movement specific rehabilitation, injury prevention and performance programs. The generated information about fuzzy membership and their detection post classification can help build personalized training program based on the deficits described within a movement pattern. Further, variability (e.g. changes in used movement pattern) can be examined using the findings to understand if an inability to switch between different movement strategies is responsible for an overload of tissues or if there is a save zone for every strategy.

The identification of stable key phases highlights a consistency in variation of waveforms and may assist the establishment and interpretation of continuous analysis as well as feature generation in future studies. Furthermore, findings demonstrated that kinematic measures do follow an consistent variability pattern, while kinetic variables may present a challenge when trying to compare findings of studies using continuous analysis as phases of variation / neuromuscular capacities may be represented over a short period of time (5 %) or at multiple short phases (possibly due to multi modal shapes a the first 30 % of the movement cycle).

Performance of Approaches. The classification models examined differed in their accuracy when classifying the randomly chosen test samples. The Corr2Mean approach and the SVM demonstrated the highest abilities (82 and 80 %, respectively) to classify observations in the test data. In comparison naive bayes, stepwise regression, neural network, k-nearest neighbour demonstrated a slightly lower ability (74 to 68 %) to classify the test samples, while the decision tree, random forest and discriminant analysis were outperformed by up to 24 %. Further, it should be noted that the variability in prediction accuracy was lowest in Corr2Mean and SVM. While it might not be surprising that the SVM model outperformed other techniques (because of it ability to learn very complex linear and non-linear relationships), it was surprising that the Corr2Mean approach did - the Corr2Mean did not receive any training and is compared to other used techniques rather simple. Potential reasons for the differences in accuracy between the classifiers are: non-linear relationships, sample size, proximity measure or fuzzy memberships [6]. Non-linear relationships are likely to have not impacted the prediction accuracy hugely because the neural network, which has been reported to be able to cope with these issues, was ranked lower than the regression model (table 2). The sample size, which is important for a good training of the neural network, might be a reason for the decreased accuracy of the neural network. One possible explanation for the good performance of Corr2Mean might be the proximity measure (relationship rather than distance). When clustering movement strategies it has been reported that differences in magnitude between samples are not as effective as their interdependence at maximizing the ability to predict a dependent variable [20]. In this study, the Corr2Mean approach was the only technique that directly used interdependence of a sample to a cluster mean. Including such a measure to the feature matrix might have increased the ability to learn the relationship of input features and target for other techniques. Another factor that may have an impact to the accuracy is the influence of fuzzy memberships. An additional analysis of the distribution of membership types supports this assumption (see table 2). It is likely that the Corr2Mean and the SVM have performed better than other techniques as they handled "weak" fuzzy observations better.

Fuzzy Observations. Findings highlight the importance of considering fuzzy memberships. While samples with fuzzy memberships accounted for 15 to 25 % in the data, they are also likely to have impacted the classification accuracy of the examined learning techniques. The importance of accounting for fuzzy memberships becomes even more important when reporting / describing a movement strategy. For example, mean values the movement strategies can change by an average of 21 % when only considering logic samples. Further, the influence of fuzzy observations to the mean values seems also to have influenced

[6]While it could be argued that the landmark registration of the data has affected the performance of the classification approaches, it is not listed here because the landmark registration has a methodological reason and because the registered time domain was fed into each approach as features.

the selected high impact features. Only features were selected that were little influenced by fuzzy observations. As such differences in calculated mean values, using only logic or both fuzzy and logic observations, differed less than 8 % (mean = 3.4 %), except of the angle rotation angles (20 to 39 %). The selection of ankle rotation is very likely related to high ankle rotation angles in cluster 1, which was about 17 degree greater than in cluster 2 and 3.

When estimating the fuzziness of an observation, calculating the ratio between decision criteria for the chosen and second most likely movement pattern separated fuzzy from logic observation (error 6 %) for the Corr2Mean approach ($0.867 < \text{ratio} > 0.945 = \text{logic}$), while using the SVM decision criteria (difference) resulted in an 18 % error. This suggests that the Corr2Mean approach should be used when judging the kind of a membership for an observation.

High Impact Features. To identify features with high impact towards classification accuracy, this study first identified phases that consistently described variance in the data. Findings, suggest that most variances in joint angles are stable over 5 % of the movement cycle, while in moment data variances were stable but only over a short period of the cycle. Further generating a classification model utilizing only a few selected features, calculated from stable phases, can to repeat or even outperform a classification model using every computed feature. This suggests that a subset of the identified robust phases of variation can capture the movement pattern / strategy. Additionally, when using the stepwise forward approach kinematic data was chosen more frequently than kinetic data to classify. This indicates that movement pattern could be judged solely on kinematic data. However, this does not mean that kinetic data can be omitted when examining injury related factors. Further, Corr2Mat and SVM selected very similar variables (hip flexion angle, ankle rotation angle, flexion moment [ankle and hip] and thorax flexion), indicating their importance for classifying the data into the predefined clusters.

Practical Implications. In conclusion, correlating hip flexion angle, ankle rotation angle, ankle flexion moment and thorax flexion of a sample to predefined values of movement strategies can be used to estimate membership to a strategy, as a correlation approach appears to be the most suitable technique for classification of a maximum effort change of direction task in groin pain patients. Support vector machines also showed a high level of suitability. Further, when classifying unknown samples into movement strategies, it is extremely important to consider the possibility of fuzzy memberships as they can present 15 to 25 % of the captured data. This is an important finding as it could have a huge impact to future research. If for example a classification technique can differentiate between 75 % of a healthy and injured population, the correct classified samples could be used to represent a truly *healthy* and *unhealthy* behaviour. Further, findings of this study highlight the existence of multiple patterns to execute a movement task and demonstrate that classification models can predict movement strategy with a high accuracy (83 %).

Conflict of Interest Statement

The authors confirm that there are no financial or personal relationship with other people or organisations that could inappropriately influence (bias) this work.

References

[1] Babyak, M. A. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic medicine 66*, 3 (2004), 411–421.

[2] Backman, C., and Harris, S. Case studies, single-subject research, and N of 1 randomized trials: comparisons and contrasts. *Am J Phys Med Rehabil 78*, 2 (1999), 170–176.

[3] Bates, B. T. Single-subject methodology: an alternative approach. *Medicine & Science in Sports & Exercise 28*, 5 (1996), 631–638.

[4] Carriero, A., Zavatsky, A., Stebbins, J., Theologis, T., and Shefelbine, S. J. Determination of gait patterns in children with spastic diplegic cerebral palsy using principal components. *Gait & Posture 29*, 1 (2009), 71–75.

[5] Falvey, É. C., King, E., Kinsella, S., and Franklyn-Miller, A. Athletic groin pain (part 1): a prospective anatomical diagnosis of 382 patients?clinical findings, mri findings and patient-reported outcome measures at baseline. *British journal of sports medicine* (2015), bjsports–2015.

[6] Franklyn-Miller, A., Richter, C., King, E., Gore, S., Moran, K., Strike, S., and Falvey, E. Athletic groin pain (part 2): a prospective cohort study on the biomechanical evaluation of change of direction identifies three clusters of movement patterns. *British Journal of Sports Medicine* (2016), bjsports–2016.

[7] Hastie, T., and Tibshirani, R. Elements of statistical learning, 2002.

[8] Kienast, G., Bachmann, D., Steinwender, G., Zwick, E. B., and Saraph, V. Determination of gait patterns in children with cerebral palsy using cluster analysis. *Gait & Posture 10*, 1 (1999), 57.

[9] Kristianslund, E., Krosshaug, T., and van den Bogert, A. J. Effect of low pass filtering on joint moments from inverse dynamics: implications for injury prevention. *Journal of biomechanics 45*, 4 (2012), 666–671.

[10] Martinez, W., Martinez, A., and Solka, J. *Exploratory data analysis with MATLAB*. CRC Press, 2004.

[11] Meeuwisse, W. H., Tyreman, H., Hagel, B., and Emery, C. A dynamic model of etiology in sport injury: the recursive nature of risk and causation. *Clinical Journal of Sport Medicine 17*, 3 (2007), 215–219.

[12] Morgan, D., and Morgan, R. *Single-Case Research Methods for the Behavioral and Health Sciences*. SAGE Publications, 2008.

[13] O'Byrne, J. M., Jenkinson, A., and O'Brien, T. M. Quantitative analysis and classification of gait patterns in cerebral palsy using a three-dimensional motion analyzer. *Journal of Child Neurology 13*, 3 (1998), 101–108.

[14] O'Malley, M. J., Abel, M. F., Damiano, D. L., and Vaughan, C. L. Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters. *Rehabilitation Engineering, IEEE Transactions on, 5*, 4 (1997), 300–309.

[15] Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., and Feinstein, A. R. A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology 49*, 12 (1996), 1373–1379.

[16] Petushek, E. J., Cokely, E. T., Ward, P., Durocher, J. J., Wallace, S. J., and Myer, G. D. Injury risk estimation expertise assessing the acl injury risk estimation quiz. *The American journal of sports medicine 43*, 7 (2015), 1640–1647.

[17] Petushek, E. J., Cokely, E. T., Ward, P., and Myer, G. D. Injury risk estimation expertise: Cognitive-perceptual mechanisms of acl-iq. *Journal of sport and exercise psychology 37*, 3 (2015), 291–304.

[18] Ramsay, J. O. *Functional data analysis*. John Wiley & Sons, 2006.

[19] Ramsay, J. O., Hooker, G., and Graves, S. *Functional data analysis with R and MATLAB*. Springer Science & Business Media, 2009.

[20] Richter, C., Marshall, B., Moran, K., et al. Clustering vertical ground reaction force curves produced during countermovement jumps. *Journal of biomechanics 47*, 10 (2014), 2385–2390.

[21] Richter, C., McGuinness, K., O?Connor, N. E., and Moran, K. The variance needed to accurately describe jump height from vertical

ground reaction force data. *Journal of applied biomechanics 30*, 6 (2014), 732–736.

[22] Richter, C., O'Connor, N. E., Marshall, B., and Moran, K. Analysis of Characterizing Phases on Waveforms An Application to Vertical Jumps. *Journal of applied biomechanics (in press)* (2013).

[23] Segaran, T. *Programming collective intelligence: building smart web 2.0 applications*. O'Reilly Media, 2007.

[24] Stergiou, N. *Innovative Analyses of Human Movement* , 1 ed. Human Kinetics, Leeds, U.K., 2004.

[25] Stergiou, N., and Scott, M. Baseline measures are altered in biomechanical studies. *Journal of Biomechanics 38*, 1 (2005), 175–178.

[26] Stout, J. L., Bruce, B., Gage, J. R., and Schutte, L. Joint kinetic patterns in children with spastic hemiplegia cerebral palsy. *Gait & Posture 3*, 4 (1995), 274.

[27] Tibshirani, R., Walther, G., and Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society B*, 63 (2001), 411– 423.

[28] Toro, B., Nester, C. J., and Farren, P. C. Cluster analysis for the extraction of sagittal gait patterns in children with cerebral palsy. *Gait & Posture 25*, 2 (2007), 157–165.

[29] van Mechelen, W., Hlobil, H., and Kemper, H. C. Incidence, severity, aetiology and prevention of sports injuries. *Sports medicine 14*, 2 (1992), 82–99.

[30] Vapnik, V. N., and Vapnik, V. *Statistical learning theory*, vol. 1. Wiley New York, 1998.

[31] Winter, D. A. *Biomechanics and motor control of human movement*, 4th editio ed. Wiley & Sons, 2009.

Figures

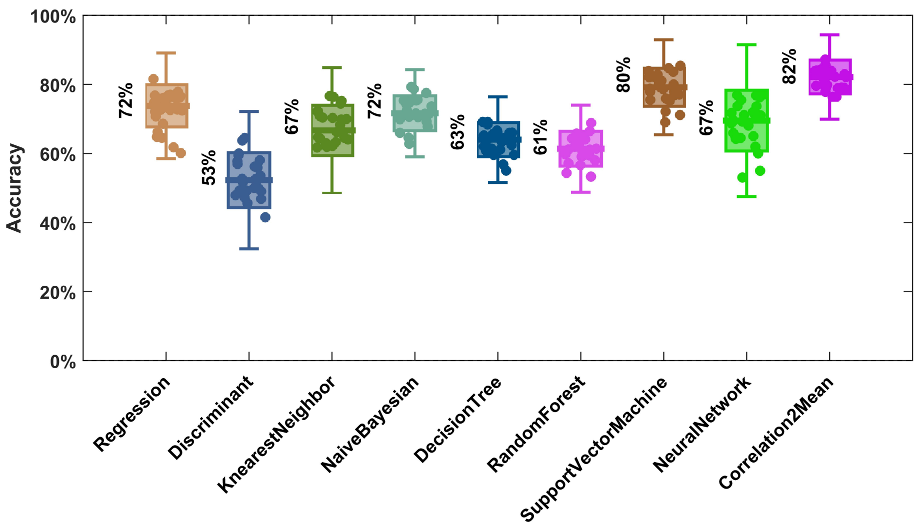

Figure 1: Box plots of the accuracy of each tested learning technique. It should be noted that for visualization purposes only 25 randomly selected instances are shown.

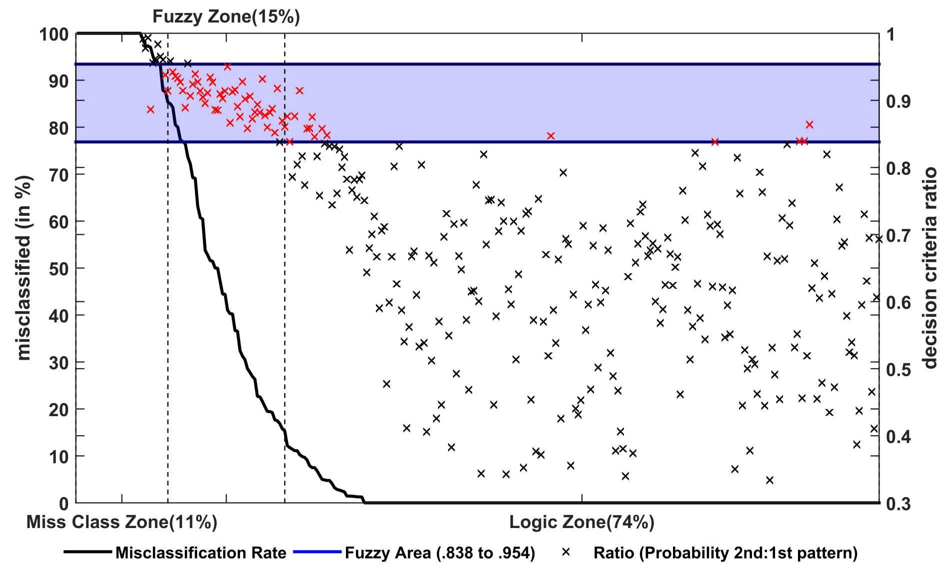

Figure 2: Illustration of Corr2Mean model, its misclassification rate (black line) and the ratio between the decision criteria for the chosen and second most likely movement pattern (x; $ratio = (r_{2^{nd}choice} + 1)/(r_{chosen} + 1)$). The red x is a sample that is likely to be fuzzy, based on the post hoc determined range of 0.876 to 0.954 (blue area). Each position on the x-axis represents a subjects, who have been sorted based in the magnitude of misclassification.

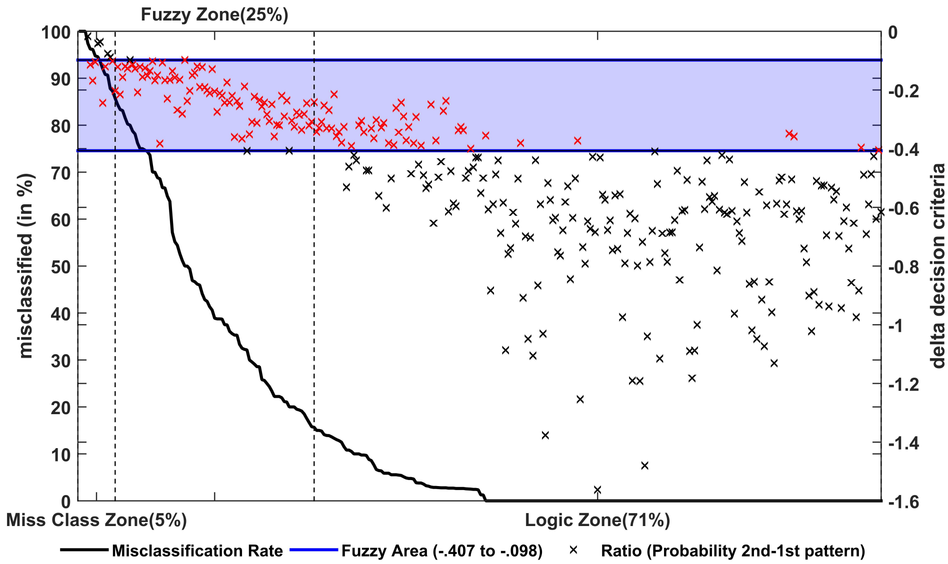

Figure 3: Illustration of SVM model, its misclassification rate (black line) and the difference between the decision criteria for the chosen and second most likely movement pattern (x; $difference = distance_{2^{nd}choice} - distance_{chosen}$). The red x is a sample that is likely to be fuzzy, based on the post hoc determined range of -0.407 and -0.098 (blue area). Each position on the x-axis represents a subjects, who have been sorted based in the magnitude of misclassification.

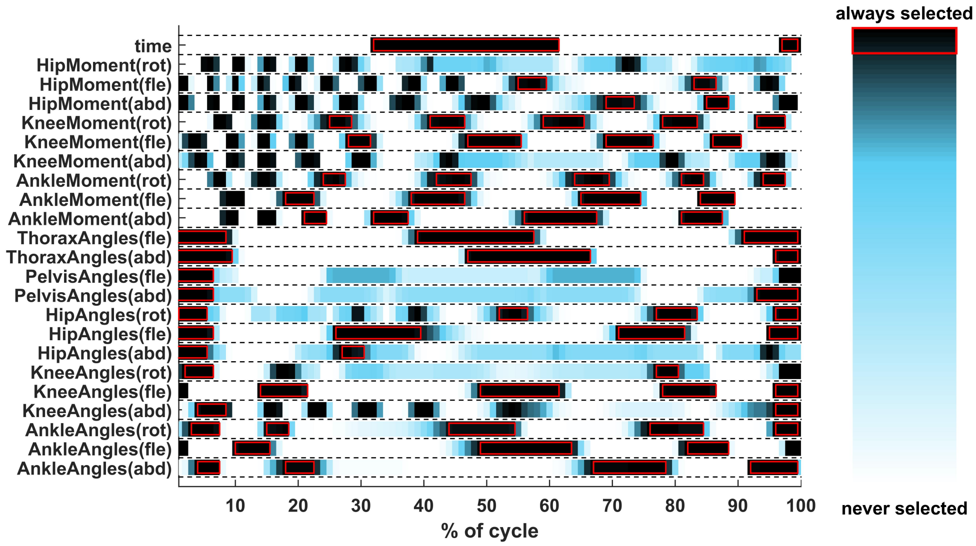

Figure 4: Illustrates the location and frequency of phases of variation that have been identified throughout the repeated model process. The color bar on the right represents the percentage of occurrence of a specific point in time within a phase of variation. The red boxes indicates that a phase has been identified as a stable phase of variation - e.g. a phase that represented variation in 95 % of simulations and spanned over 5 % of the movement cycle. The abbreviations: fle, abd and rot indicate the movement plane (sagittal, frontal and transversal) of the joint.

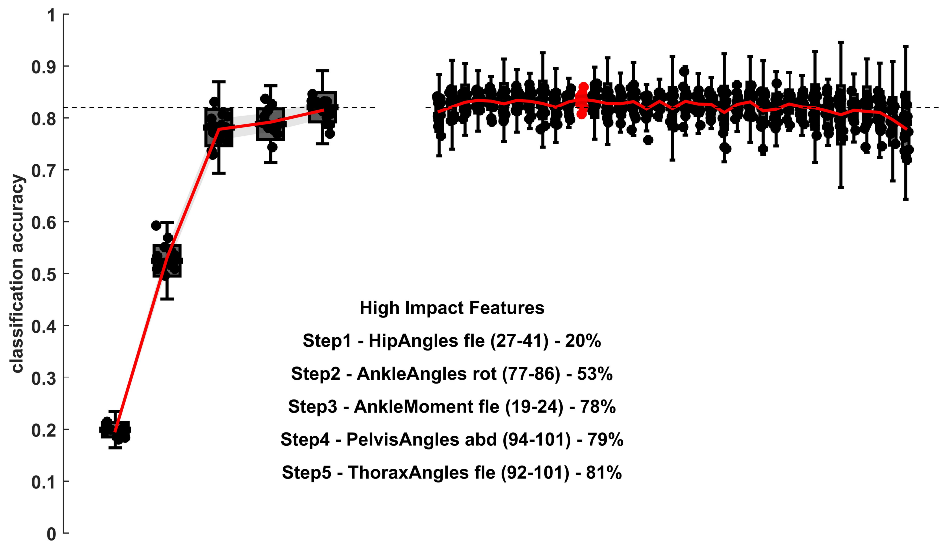

Figure 5: Illustration of the classification accuracy for the Corr2Mean model, when using a stepwise forward approach adding the most meaningful features to the model. The table within the figure can be decoded as: Step X - Joint - Joint plane (phase of variation used) - classification accuracy. The abbreviations: fle, abd and rot indicate the movement plane, sagittal, frontal and transversal of the joint.

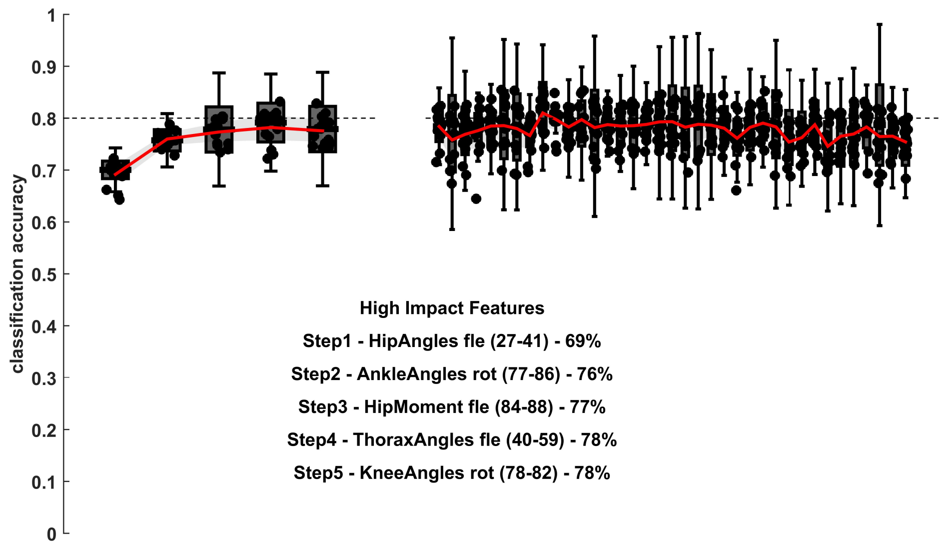

Figure 6: Illustration of the classification accuracy for the SVM model, when using a stepwise forward approach adding the most meaningful features to the model. The table within the figure can be decoded as: Step X - Joint - Joint plane (phase of variation used) - classification accuracy. The abbreviations: fle, abd and rot indicate the movement plane, sagittal, frontal and transversal of the joint.

Table 1: *Illustration of the workflow for the clustering, classifier selection and identification*

	Detect Movement Strategies	Select Best Classifier	Identify and Detect High Impact Features
Step 1. Generate Subject Scores	Data Reduction Calculation Subject scores Data Transformation	Data Reduction Calculation Subject scores	Assign Subjects Randomly Into Training And Testing Data Reduction Store Key Phases (loop) Select Stable Phases Calculation Subject scores
Outcome Step 1	*Subject scores that represent the relationship between subjects*	*Subject scores that represent the behaviour of a subject*	*"Stable" subject scores that represent the behaviour of a subject*
Step 2. Generate Data to be analyzed	Gap statistics Data Classification	Assign Subjects Randomly Into Training And Testing Training Classifier using all features Test Classifier (loop)	Assign Subjects Randomly Into Training And Testing Training Classifier using one feature and the model base Test Classifier (loop) Add best feature to model base (loop)
Outcome Step 2	*3 subgroups of the cohort*	*Accuracies of a tested classifier*	*An order of selection and accuracy of a combination of features*
Step 3. Analyze data	SPM to find difference between clusters	Identify highest accuracy to select best classifier	Identify sweet spot of accuracy to select the best combination of features

Table 2: Description of the distribution of defined zones.

Technique	Rank	Accuracy (in %)	% Miss Class	% Fuzzy
Corr2Mean	1	82	11	15
SVM	2	80	5	25
Naive Bayesian	3	72	5	36
Regression	4	72	11	24
Neural Network	5	67	8	32
K-Nearest-Neighbour	6	67	9	31
Decisions Tree	7	63	7	46
Random Forest	8	61	21	13
Discriminant	9	53	6	83

Table 3: Descriptive mean features (logic members only) of cluster 1, 2 and 3.

Feature	Mean	Cluster1	Cluster2	Cluster3
Sample	323 (208)	63 (34)	132 (90)	128 (84)
Hip Flexion Angle (27 - 41 %)	51.7 (50.7)	52.2 (51.1)	60.7 (61.9)	41.3 (38.5)
Ankle Rotation Angle (77 - 86 %)	-14.4 (-12.1)	-27.1 (-29.1)	-10.8 (-9.0)	-11.7 (-8.4)
Ankle Flexion Moment (19 - 24 %)	13.6 (13.3)	16.5 (16.0)	10.7 (9.9)	15.1 (15.9)
Pelvic Drop (94 - 100 %)	22.2 (22.3)	20.2 (20.2)	20.1 (19.6)	25.4 (26.0)
Thorax Flexion (92 - 100 %)	37.1 (37.1)	31.4 (31.8)	41.9 (43.4)	34.9 (32.6)