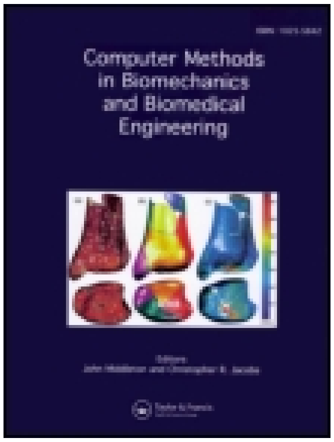

Computer Methods in Biomechanics and Biomedical Engineering
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gcmb20

Pattern classification of kinematic and kinetic running data to distinguish gender, shod/barefoot and injury groups with feature ranking

Bjoern M. Eskofier [a], Martin Kraus [b], Jay T. Worobets [a], Darren J. Stefanyshyn [a] & Benno M. Nigg [a]

[a] Human Performance Laboratory, Faculty of Kinesiology, University of Calgary, 2500 University Dr NW, Calgary, AB, Canada, T2N 1N4

[b] Pattern Recognition Laboratory, Department of Computer Science, University of Erlangen, Martensstrasse 3, 91058, Erlangen, Germany

Published online: 02 Feb 2011.

Computer Methods in Biomechanics and Biomedical Engineering
Vol. 15, No. 5, May 2012, 467–474

Taylor & Francis
Taylor & Francis Group

Pattern classification of kinematic and kinetic running data to distinguish gender, shod/barefoot and injury groups with feature ranking

Bjoern M. Eskofier[a]*, Martin Kraus[b1], Jay T. Worobets[a2], Darren J. Stefanyshyn[a3] and Benno M. Nigg[a4]

[a]*Human Performance Laboratory, Faculty of Kinesiology, University of Calgary, 2500 University Dr NW, Calgary, AB, Canada T2N 1N4;* [b]*Pattern Recognition Laboratory, Department of Computer Science, University of Erlangen, Martensstrasse 3, 91058 Erlangen, Germany*

(*Received 2 March 2010; final version received 18 November 2010*)

The identification of differences between groups is often important in biomechanics. This paper presents group classification tasks using kinetic and kinematic data from a prospective running injury study. Groups composed of gender, of shod/barefoot running and of runners who developed patellofemoral pain syndrome (PFPS) during the study, and asymptotic runners were classified.

The features computed from the biomechanical data were deliberately chosen to be generic. Therefore, they were suited for different biomechanical measurements and classification tasks without adaptation to the input signals. Feature ranking was applied to reveal the relevance of each feature to the classification task.

Data from 80 runners were analysed for gender and shod/barefoot classification, while 12 runners were investigated in the injury classification task. Gender groups could be differentiated with 84.7%, shod/barefoot running with 98.3%, and PFPS with 100% classification rate. For the latter group, one single variable could be identified that alone allowed discrimination.

Keywords: pattern classification; biomechanical group classification; generic features; feature ranking; AdaBoost; patellofemoral pain syndrome

1. Introduction

Biomechanical studies have been conducted to use kinematics, kinetics, soft tissue vibrations and/or EMG data to distinguish between groups such as gender (Von Tscharner and Goepfert 2003), age (Devita and Hortobagyi 2000), footwear (Nigg et al. 2003) and to identify injury mechanism characteristics (Stefanyshyn et al. 2006). The classification is often done by comparing means and standard deviations of discrete variables (e.g. peak impact force and maximal foot eversion). Recently, new methods have been proposed and applied for such classifications based on pattern recognition methods. For example, it has been shown (Janssen et al. 2008) that artificial neural networks can identify emotional state from human gait data. Furthermore, support vector machine classifiers have been used (Begg and Kamruzzaman 2005) to differentiate young and elderly gait patterns. These and other published classification methods (Schöllhorn et al. 2002; Wu et al. 2007) were successful in identifying groups using biomechanical data. However, the results depended on the analysed features, and these features were very specific to the input measurements and to the specific studies. Furthermore, the applied methods did not always provide information about what exactly characterised the differences between the groups identified.

Therefore, the purpose of this paper was to develop a pattern classification approach for typical tasks of biomechanical group classification that was not specific to the input measurements and additionally to provide information about the group differences. This general approach was applied to three different groups without adaptation. In particular, groups composed of gender, shod/barefoot running and injured/non-injured subjects were considered. While the first two groups were primarily tested for proof of concept, the last group had clinical relevance. The injury that was examined was the patellofemoral pain syndrome (PFPS), a condition affecting up to a quarter of all persons active in sporting activities (Malek and Mangine 1981; Devereaux and Lachmann 1984). For the purpose of classification, generic features were used which were not in any way specific to the collected measurements. These measurements consisted of kinetic and kinematic data collected during a longitudinal running injury study (Stefanyshyn et al. 2006).

2. Methods

The employed classification methods (Figure 1) followed a classical pattern recognition approach (Duda et al. 2001). Dynamic biomechanical measurements (Section 2.1) were

*Corresponding author. Email: bmeskofi@ucalgary.ca

ISSN 1025-5842 print/ISSN 1476-8259 online

http://dx.doi.org/10.1080/10255842.2010.542153
http://www.tandfonline.com

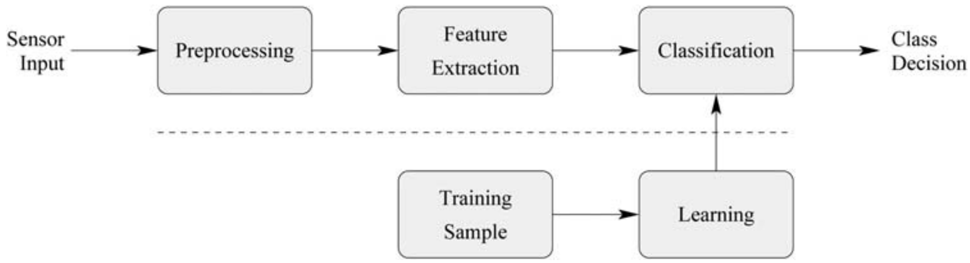

Figure 1. Overview of a pattern recognition system.

first subjected to preprocessing to enhance the signal properties. The output of the subsequent feature extraction step (Section 2.2), the feature vector $\mathbf{c} \in \Re^{N_c}$ described the input in N_c dimensional feature space.

All groups under investigation were differentiated using AdaBoost (Freund and Schapire 1997), a supervised classifier that has the inherent ability to select features according to their importance for the classification task (Section 2.3). Thus, it was possible to identify which features contributed most to the differentiation of the groups under investigation.

For AdaBoost training, a labelled sample S_{train} was used (Figure 1, below the dotted line). The label assigned a feature vector $\mathbf{c}$ to a class $\omega_k \in \Omega$, where Ω is set of N_Ω possible classes, $K = 1, \ldots, N_\Omega$ class number.

One example of class label was the runner gender, with $N_\Omega = 2$ in that particular case. The class-specific statistical properties were studied during training and decision criteria were established. In order to test the generalisation performance of the AdaBoost classifier, cross-validation was performed (Section 2.4).

2.1 Data for experimental evaluation

The process of collecting the experimental data for the prospective study was described earlier (Lun et al. 2004; Stefanyshyn et al. 2006). The study was reviewed and approved by the University of Calgary Conjoint Health Research Ethics Board. Shortly, at the beginning of the running season, 153 recreational runners (71 women, 82 men) were recruited; written consent was obtained and anthropometric and biomechanical measurements were conducted. During the following 6 months, any running-related injuries were recorded and a running journal was kept.

Of the initial 153 subjects, $N_r = 80$ (40 women, 40 men) were suitable for analysis (Table 1). A total of 73 subjects were dropped out of the study because they stopped running (4), work (2), incomplete running journal (16), injury at time of recruitment (2), loss to follow up (49).

2.1.1 Dynamic biomechanic measurements

All runners had data collected from both legs. The measurements included a shod condition, with the runners wearing their own running shoes, and a barefoot condition. For each subject, data were recorded during five trials in the shod condition and during five trials in the barefoot condition. The trials were only used for the subsequent experiments if the subjects hit the force plate (see below) properly. Therefore, between two to ten complete dynamic data sets were collected for each of the 80 subjects, with a mean number of 6.2 data sets per subject. In total, 496 data sets were thus used for experimental evaluation.

Three leg segments (upper leg, lower leg, foot) were prepared using a total of nine reflective markers. The 3D marker positions were collected using four electronically shuttered, high-speed video cameras (NAC MOS-TV, V-14B, Japan) equipped with 12.5–75 mm zoom lenses (Cosmicar, Japan) and a VP310 video processor (Motion Analysis Corp., Santa Rosa, CA, USA). 3D force data were collected using a force platform (Kistler AG, Winterthur, Switzerland) mounted flush with the floor

Table 1. Characteristics of the 80 runners who completed the prospective running injury study. Values are presented as mean (standard deviation).

	n	Age [y]	Mileage [km]	Experience [y]	Height [m]	Mass [kg]
Women	40	36.0 (8.8)	33.7 (16.9)	6.4 (6.5)	1.68 (0.08)	66.2 (10.0)
Men	40	41.1 (8.7)	36.3 (16.8)	11.4 (9.2)	1.79 (0.07)	85.7 (13.6)
Mean		38.5 (9.1)	35.0 (16.8)	8.9 (8.3)	1.74 (0.10)	75.9 (15.4)
Max		64.0	100.0	40.0	1.93	123.5
Min		22.0	6.0	0.5	1.55	47.6

in the centre of a 30-m runway. Running speed was controlled at $v = 4.0 \pm 0.2$ m/s using two photocells, 1.9 m apart, at shoulder height. Markers were tracked for 50 ms before and after force plate contact. Kinematic and kinetic data were imported into Kintrak 4.0 (Motion Analysis Corp.) for analysis. Joint attitude and angular motions were determined using a 3D joint coordinate system implemented in Kintrak 4.0.

A total of $N_{\text{dyn}} = 51$ variables were collected for each leg of the athlete (Table 2). Each of them consisted of $N_f = 101$ normalised time frames from touch-down on the force platform to toe-off. These variables are referred to as $d_i[k]$, with $i = 1, \ldots, N_{\text{dyn}}$ denoting the specific measurement (Table 2). These data of dimension 51∗101 were used as the input for feature extraction (see Section 2.2).

2.1.2 *Injury information*

After completing all measurements, the athletes were observed on a monthly basis for 6 months during their usual training routine from April to September (Lun et al. 2004). For all runners participating in the study the occurrence of any injury attributed to injury was documented. An injury was defined as any musculoskeletal symptom of the lower limb which required a reduction or stoppage of normal training. A weekly drop-in injury clinic was available to subjects for evaluation of injuries. The injuries were assessed by two experienced sports medicine doctors at the University of Calgary Sport Medicine Centre.

Table 2. Dynamic variables that were acquired during data collection.

Number i	Measured variable
1–3	Ground reaction force; vertical, medial–lateral, anterior–posterior
4,5	Centre of pressure location; medial–lateral, anterior–posterior
6	Free moment on the force plate
7–10	Ankle flexion–extension angle, velocity, moment, power
11–14	Ankle inversion–eversion angle, velocity, moment, power
15–18	Ankle abduction–adduction angle, velocity, moment, power
19–22	Knee flexion–extension angle, velocity, moment, power
23–26	Knee abduction–adduction angle, velocity, moment, power
27–30	Knee internal–external rotation angle, velocity, moment, power
31	Hip flexion–extension moment
32	Hip abduction–adduction moment
33	Hip internal–external rotation moment
34,35	Foot sagittal segment plane angle, angular velocity
36,37	Foot frontal segment plane angle, angular velocity
38,39	Foot transverse segment plane angle, angular velocity
40,41	Shank sagittal segment plane angle, angular velocity
42,43	Shank frontal segment plane angle, angular velocity
44,45	Shank transverse segment plane angle, angular velocity
46,47	Thigh sagittal segment plane angle, angular velocity
48,49	Thigh frontal segment plane angle, angular velocity
50,51	Thigh transverse segment plane angle, angular velocity

2.1.3 *Class labels*

The data were analysed with class labels based on gender, on the shod/barefoot condition and on whether the runners developed a specific injury type (PFPS). For each experiment, the dynamic data set of a single step cycle was labelled according to the membership of a certain class. For gender classification, 244 data sets were from females and 252 from males. For the shod and barefoot experiments, 217 and 279 data sets were assigned to each class, respectively.

The injury group consisted of runners that suffered from PFPS during the 6-month study period, compared with matched uninjured runners. This specific injury was selected for this project because (a) this was the most frequent injury in the 6 months prospective study and (b) it is a very common injury among runners, affecting up to a quarter of all persons active in sporting activities (Malek and Mangine 1981; Devereaux and Lachmann 1984). For this experiment, six patients who were diagnosed with PFPS by the clinicians were matched with respect to mass, gender, mileage and running experience (Table 3) with six subjects who remained injury free throughout the study (Stefanyshyn et al. 2006). These characteristics were proposed to be associated with injury and/or to have an influence on resultant joint moments (Van Mechelen 1992; Moisio et al. 2003). In this study, 28 data sets from the injury (PFPS) group were classified against 27 data sets from the matched asymptotic runners (asymptomatic matched (ASYM) group).

2.2 **Feature extraction**

Generic features for classification were calculated in order to be independent from specific characteristics of the original measurements. Generic in this context means that the features were not restricted to the calculation of key variables such as angles or forces at specific time points. Rather, their calculation considered the complete temporal information of the measurements. Consequently, the set of features was chosen so that it represented an arbitrary measurement as completely as possible. Therefore, the features can be straightforwardly applied in different group-classification tasks. The chosen features have already proven to perform well in other pattern

Table 3. Comparison of the characteristics of the six injured patients and the six asymptotic patients who were used as matched controls.

Patient	Injured side	Gender	Mileage [km]	Experience [y]	Mass [kg]
Injured #1	Left	Female	12.0	1.0	65.5
Match #1			15.0	0.7	59.0
Injured #2	Left	Female	40.0	16.0	59.3
Match #2			35.0	14.0	63.5
Injured #3	Right	Female	15.0	1.0	76.6
Match #3			15.0	2.0	76.2
Injured #4	Left	Male	55.0	20.0	82.5
Match #4			60.0	15.0	75.2
Injured #5	Right	Male	40.0	1.5	79.1
Match #5			40.0	4.0	84.0
Injured #6	Right	Male	30.0	1.5	76.0
Match #6			30.0	4.0	74.1
Mean					
Injured			32.0	6.8	73.2
Non-injured			32.5	6.6	72.0

recognition tasks (Furui 1986; Hafed and Levine 2001; Fan and Wang 2002).

2.2.1 *Basic features*

Basic features for time-dependent dynamic measurements $d_i[k]$ were mean

$$\mathbf{c}_{i,\mu} = \mu_i = \frac{1}{N_f}\sum_{k=1}^{N_f} d_i[k], \tag{1}$$

and variance

$$\mathbf{c}_{i,\sigma^2} = \sigma_i^2 = \frac{1}{N_f - 1}\sum_{k=1}^{N_f} (d_i[k] - \mu_i)^2, \tag{2}$$

for each measurement curve,

where N_f represents number of discrete time frames,
k represents discrete measurement time point, $k = 1, \dots, N_f$ and
i represents specific measurement.

Further basic features were derived from temporal positions and from the actual and absolute maxima and minima of the curves (Table 4).

2.2.2 *Transformation features*

Transformation features project the original function into a different space (e.g. frequency space). In this work, discrete cosine transform (DCT, (Ahmed et al. 1974)) features were used for the purpose of incorporating the information contained in the frequency components of the measurements. These features were successfully used for 2D image classification (Fan and Wang 2002), and face recognition (Hafed and Levine 2001). For the DCT, the original function values $d_i[k]$ were linearly transformed into real numbers $D_i[f]$, $f = 1, \dots, N_f$, which represented the measurement function in the frequency domain. The transformation formula was

$$\begin{aligned}\mathrm{DCT}(d_i[k]) &= D_i[f]\\ &= \sum_{k=1}^{N_f}\left(d_i[k]\cos\left[\frac{\pi}{N_f}\left((k-1)+\frac{1}{2}\right)(f-1)\right]\right),\end{aligned} \tag{3}$$

where f represents discrete frequency.

The values were used as features $\mathbf{c}_{i,\mathrm{dct}}$ by setting

$$\mathbf{c}_{i,\mathrm{dct},f} = \mathrm{DCT}(d_i[k]). \tag{4}$$

In this example, the high-frequency components contributed little information. Consequently, only the first $(f = 1, \dots, N_{\mathrm{dct\,cut}}, N_{\mathrm{dct\,cut}} < N_f)$ coefficients were considered. The parameter $N_{\mathrm{dct\,cut}}$ was set to 30 for this study, because no activity was observed in components with higher frequency.

2.2.3 *Regression features*

Cosine transform features described the measurement frequency characteristics, but did not sufficiently

Table 4. Basic extrema features that were generated.

Feature		Formula
Min	$\mathbf{c}_{i,\mathrm{min}}$	$\min D_i[k]$
Max	$\mathbf{c}_{i,\mathrm{max}}$	$\max D_i[k]$
Abs Min	$\mathbf{c}_{i,\mathrm{absmin}}$	$\min \lvert D_i[k]\rvert$
Abs Max	$\mathbf{c}_{i,\mathrm{absmax}}$	$\max \lvert D_i[k]\rvert$
Idx Min	$\mathbf{c}_{i,\mathrm{minidx}}$	$\arg\min_k D_i[k]$
Idx Max	$\mathbf{c}_{i,\mathrm{maxidx}}$	$\arg\max_k D_i[k]$
Idx Abs Min	$\mathbf{c}_{i,\mathrm{absminidx}}$	$\arg\min_k \lvert D_i[k]\rvert$
Idx Abs Max	$\mathbf{c}_{i,\mathrm{absmaxidx}}$	$\arg\max_k \lvert D_i[k]\rvert$

Notes: Min, Max, Abs and Idx are abbreviations of Minimum, Maximum, Absolute Value and Index, respectively. The variable *i* that is used to distinguish the different features corresponds to the number assigned to the measurements (Table 2).

characterise general trends. Therefore, polynomial regression (PR) features were calculated. Regression features have successfully been used in speech recognition (Furui 1986) as well as in gait classification tasks based on ground reaction force data (Mezghani et al. 2008). Depending on the degree N_p of the chosen polynomial, signal properties such as gradient or curvature are adequately described. For this purpose, each measurement $d_i[k]$ was approximated in a least squares sense by a polynomial function

$$p(x) = \sum_{m=0}^{N_p} r[m]x^m, \tag{5}$$

where

N_p represents number of polynomial coefficients,
$r[m]$ represents polynomial coefficients, $m = 0, \ldots, N_p$.

The coefficients were used as features by setting

$$\mathbf{c}_{i,pr,m} = r_i[m]. \tag{6}$$

The polynomial degree N_p was set to 4 for this study, which corresponds to a cubic fit of the original measurements. Experimental results showed no improvement when calculating fits of higher order.

2.3 Pattern classification using AdaBoost

AdaBoost ('adaptive boosting', (Freund and Schapire 1997)) is a meta algorithm for supervised learning that utilises other learning algorithms (i.e. classifiers). The idea of AdaBoost is to use a linear combination (ensemble) of multiple simple decision rules (weak classifiers), which when combined create a more accurate complex decision rule (a strong classifier). If each classifier predicts the correct class with an accuracy $\alpha_{\text{simple}} > 0.5$, then the ensemble classification accuracy approaches $\alpha_{\text{complex}} \rightarrow 1$ as the number of simple classifiers increases (Boland 1989).

The AdaBoost classifier was trained in several iterations, adding one simple classifier to the ensemble in each of these iterations. The iteration number for all experiments was set to 20 to prevent overadaptation of the classifier to the input data. During training, each feature vector $\mathbf{c}_t, \quad t = 1, \ldots, N_t$, where

N_t represents number of samples in the training set S_{train}, was assigned a weight w_t. These weights (Freund and Schapire (1997)) were initially equally distributed with

$$w_t = \frac{1}{N_t} \quad \forall \; t = 1, \ldots, N_t. \tag{7}$$

As weak classifier for AdaBoost, the so-called decision stump (Viola and Jones 2004) was used. This is a simple classifier that uses a threshold value in only one of the N_c feature space dimensions. Using the decision stump weak classifier, AdaBoost performed implicit feature selection. By counting how often each feature contributed to the final decision process, a measure of the ability of a certain feature to separate the trained classes was derived. As final hypothesis for classification, a weighted vote of all classifiers in the ensemble was cast.

2.4 Cross-validation

In order to test the classifier for generalisation performance and to prevent overadaptation to the training samples, cross-validation (Duda et al. 2001) was employed. The available data were partitioned into a fixed number N_{cv} of sets. Then the selected classifier was trained using $N_{cv} - 1$ sets. Following training, the remaining set S_{test} was used for classifier testing. By iterating this process until each of the N_{cv} sets was used as test set once, generalisation performance was tested because the information in the test set had always been unknown to the classifier. For each cross-validation iteration $v = 1, \ldots, N_{cv}$, the features selected for classification were recorded, and the test classification rates $\alpha_{\text{test},v}$ were also stored. The final classification rate α_{test} was calculated as the mean over all N_{cv} iterations.

The samples of one person were naturally more similar between each other, but the samples of different persons were not. Thus, the data were partitioned into $N_{cv} = 80$ disjunct sets for cross-validation for the gender and shod/barefoot experiments and into $N_{cv} = 12$ disjunct sets for cross-validation for the PFPS/ASYM group experiments. All data sets from one athlete were always in the same cross-validation set; therefore, we performed a leave-one-subject-out cross-validation. This means that it was ensured that data sets from the same subject were not both in the training and in the testing set. The cross-validation always left out all data sets from one subject, trained with the remaining data sets of all other subjects and tested the classification accuracy on the subject that was left out. The reported accuracies are the mean of all N_{cv} cross-validation runs.

2.5 Statistics

Classification accuracies were deemed significant if the null hypothesis which states that classification was random could be rejected using a binomial test with significance level $\alpha = 0.01$. A one-tailed t test was used for statements about the statistical significance of differences of single features between groups with significance level $\alpha = 0.01$.

3. Results

3.1 Gender classification

For gender, the mean classification accuracy was 84.7%. This is significantly different from random ($p < 0.001$).

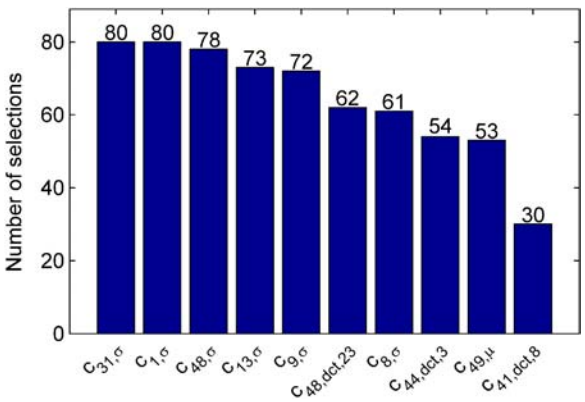

Figure 2. The 10 features that were most often selected for gender classification. Numbers above each bar represent how often each feature was selected. The respective feature identifiers are defined in Equations (1), (2), (4) and (6).

Two basic features were present in each cross-validation iteration (Figure 2); the variance c_σ of the hip flexion-extension moment and the variance c_σ of the vertical ground reaction force. The top 30 ranked features included 19 basic and 11 DCT features. The first PR feature appeared at rank 33.

3.2 *Shod/barefoot classification*

For shod/barefoot, the mean classification accuracy was 98.3%, with 8 of 496 sets misclassified. This is significantly different from random ($p < 0.001$). Two regression features were present in each cross-validation iteration (Figure 3); the quadratic polynomial component $c_{pr,2}$ of the foot sagittal plane angle and the linear polynomial component $c_{pr,1}$ of the shank sagittal plane angle. The top 30 ranked features included 16 cosine transform, 10 PR and 4 basic features.

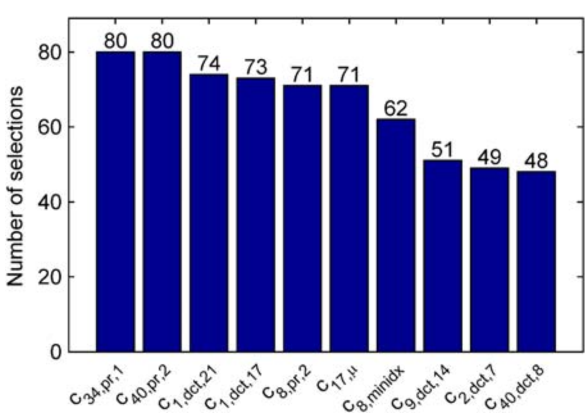

Figure 3. The 10 features that were most often selected for shod/barefoot classification. Numbers above each bar represent how often each feature was selected. The respective feature identifiers are defined in Equations (1), (2), (4) and (6).

3.3 *Injury/non-injury classification*

For the injury/non-injury classification task, the cross-validation consisted of only $N_{cv} = 12$ runs, because only 12 subjects were used for this experiment (see Section 2.1.3). As in the previous experiments, in each cross-validation run, all data from one single subject were left out.

For this classification task, the mean classification accuracy was 100%. Thus, every single data set and in effect every subject was assigned to the correct group. This is significantly different from random ($p < 0.001$). One single feature was selected in each cross-validation iteration, the mean c_μ of the hip abduction moment. This basic feature allowed classification without further combination with other features.

The mean hip abduction moments were significantly higher ($p < 0.001$) for all six PFPS group patients than those of the six matched ASYM group patients (Figure 4). This difference was also visible in the originally measured mean hip abduction moments (Figure 5).

4. Discussion

One purpose of this paper was to show that the applied methods are capable of pointing to the characteristics that discriminate groups. For the case of injury/non-injury classification, one single feature was found that distinguished runners from the injury group and the asymptomatic group with 100% accuracy. Because it is only a single feature, traditional methods should provide the same result. However, the method presented in this paper does not rely on assumptions that may be subjectively biased, but allows an objective, data-driven evaluation of the input variables. Stefanyshyn et al. (2006) had identified knee angular impulse as a predictor of PFPS in an earlier study on the same data using traditional methods.

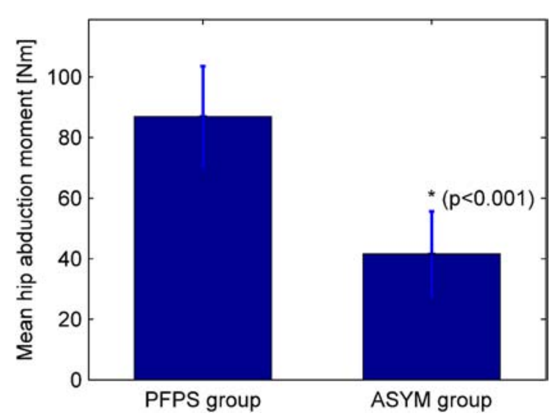

Figure 4. Mean (standard deviation) hip abduction moments for the six patients who developed patellofemoral pain syndrome (PFPS group) and the six asymptomatic matched controls (ASYM group).

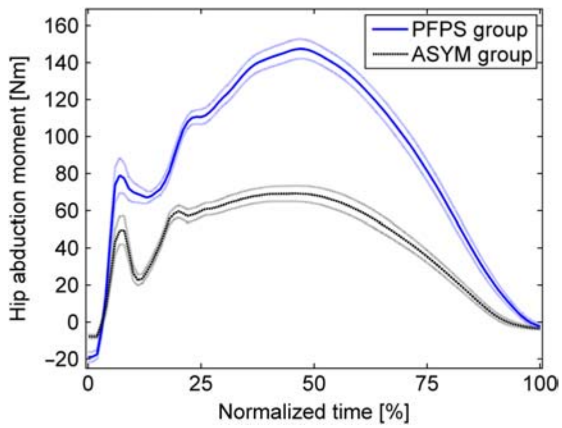

Figure 5. Mean (standard error) resultant hip abduction moments for the six patients who developed patellofemoral pain (PFPS group) and the six (ASYM group).

However, this variable only discriminated the runners on a one-to-one matched basis, and is individually true only for five of the six matched pairs. The result suggested by the presented algorithm shows that using the mean hip abduction moment as a discriminating predictor, all individuals can be classified simultaneously and correctly without the need to heed the matching.

Because prospective data were only available for six runners that developed PFPS, it remains to be evaluated whether this single indicator is also sufficient for a larger group of patients. However, even if a single variable is not sufficient for larger groups, the presented algorithms would be capable of finding those variables that are additionally needed for discrimination. Furthermore, the results are highly significant even for the relatively small group and indicate that increased hip abduction moments should be deemed risk factors that play a role in the development of patellofemoral pain in runners. Muscular deficits in the hip have already been presumed to be a major factor in the development of knee injuries in runners (James 1995; Fredericson et al. 2000; Astephen et al. 2008). Footwear and running style can influence hip abduction moments, and the appropriate manipulation of these variables may play a preventive role for patients who are predisposed to patellofemoral pain.

The ability to discriminate classes has also been shown for gender and shod/barefoot classification. Although these tasks lack the clinical relevance of the injury classification, a different strength of the methodology was shown. Although traditional methods often evaluate individual discrete variables, the presented algorithms are able to evaluate what combination of features is needed to discriminate classes. Those features that were most often selected contain the important indicators for the difference between the classes. For gender classification, two variables (hip flexion–extension moment and vertical ground reaction force) were always selected (Figure 2). The hip flexion–extension moment was already shown to be different for gender by Kerrigan et al. (1998) for walking. Another study (Vardaxis and Goulermas 2006) has also previously shown that hip features have high discriminatory accuracy when assessing gender-specific differences. The importance of the vertical ground reaction force for the gender differentiation can be explained by the higher weight of the male study participants (Table 1). For shod/barefoot classification, the variables that were most often selected for classification (Figure 3) were the foot and shank sagittal plane angles. These variables have already been reported to be significantly different for shod/barefoot running earlier (De Wit et al. 2000).

Highly significant classification accuracies could be achieved in all experiments. This shows that the generic features computed for the dynamic data contained the information content necessary for accurate group discrimination. Although the features did not directly represent key time points and associated variables such as angles and moments, they were computed for every measurement without modification. Thus, they can be used for a wide range of biomechanical measurements.

The ability of the presented methodology to rank features showed that typically a combination of the different feature types leads to correct classification. Although, for example, regression features are not mandatory for gender classification, they add important discrimination information for shod/barefoot classification. Consequently, it is a good idea to compute all the proposed features. This is particularly true because the inherent feature selection of AdaBoost will reject features for classification if they do not contribute indicators critical to the classification.

A further aspect of the presented methodology is that the applied algorithms are not restricted to a two-class problem. Freund and Schapire (1997) have already shown that AdaBoost is capable of discriminating multiple classes. Thus, the most important effects of, for instance, wearing different running shoes could be identified using biomechanical measurements.

One limitation of the methodology is the required parameter setting for the extracted features. However, these parameters can be set by simple experimental evaluation once the methodology is implemented. First, the number of necessary DCT features needs to be determined. This can be set by observing the amount of energy that is contained in higher frequency components. Second, the required polynomial fit order has to be identified. Experimental evaluation of the classification rates using different orders straightforwardly reveals the optimal setting.

A further limitation of the methodology is that the exact mechanisms responsible for the discrimination were not revealed by the different classification tasks. Nevertheless, the methodology pinpointed those variables that were most relative to a certain differentiation. From this starting point, more detailed experiments may be conducted in order to unveil the relationship between those variables and the specific classification task.

5. Summary

A classification approach using generic features and AdaBoost was shown which provides an effective tool for identifying variables that allow subject or patient group discrimination. Besides high classification accuracies for gender and shod/barefoot running, the results also suggested that the mean hip abduction moment may be a very important indicator connected to the development of PFPS in runners.

Acknowledgements

The authors thank Drs Willem Meeuwisse and Victor Lun of the Calgary Sport Medicine Centre for the injury assessment during the original data collection. The authors also thank Florian Hoenig (University of Erlangen), Heiko Schlarb and Dr Berthold Krabbe (adidas AG) for many helpful discussions.

Notes

1. Email: martin.kraus@informatik.uni-erlangen.de
2. Email: worobets@kin.ucalgary.ca
3. Email: darren@kin.ucalgary.ca
4. Email: nigg@kin.ucalgary.ca

References

Ahmed N, Natarajan T, Rao KR. 1974. Discrete cosine transform. IEEE Trans Comput. C-23:90–93.

Astephen JL, Deluzio KJ, Caldwell GE, Dunbar MJ. 2008. Biomechanical changes at the hip, knee, and ankle joints during gait are associated with knee osteoarthritis severity. J Orthop Res. 26:332–341.

Begg R, Kamruzzaman J. 2005. A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. J Biomech. 38:401–408.

Boland PJ. 1989. Majority systems and the condorcet jury theorem. J R Stat Soc Ser D (The Statistician). 38:181–189.

De Wit B, De Clercq D, Aerts P. 2000. Biomechanical analysis of the stance phase during barefoot and shod running. J Biomech. 33:269–278.

Devereaux MD, Lachmann SM. 1984. Patello-femoral arthralgia in athletes attending a sports injury clinic. Br J Sports Med. 18:18–21.

Devita P, Hortobagyi T. 2000. Age causes a redistribution of joint torques and powers during gait. J Appl Physiol. 88: 1804–1811.

Duda RO, Hart PE, Stork DG. 2001. Pattern classification. New York, NY: Wiley.

Fan Y, Wang R. 2002. An image retrieval method using DCT features. J Comput Sci Technol. 17:865–873.

Fredericson M, Cookingham CL, Chaudhari AM, Dowdell BC, Oestreicher N, Sahrmann SA. 2000. Hip abductor weakness in distance runners with iliotibial band syndrome. Clin J Sport Med. 10:169–175.

Freund Y, Schapire RE. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci. 55:119–139.

Furui S. 1986. Speaker-independent isolated word recognition using dynamic features of speech spectrum. IEEE Trans Acoust Speech Signal Process. 34:52–59.

Hafed ZM, Levine MD. 2001. Face recognition using the discrete cosine transform. Int J Comput Vis. 43:167–188.

James S. 1995. Running injuries to the knee. J Am Acad Orthop Surg. 3:309–318.

Janssen D, Schöllhorn WI, Lubienetzki J, Fölling K, Kokenge H, Davids K. 2008. Recognition of emotions in gait patterns by means of artificial neural nets. J Nonverbal Behav. 32: 79–92.

Kerrigan DC, Todd MK, Della Croce U. 1998. Gender differences in joint biomechanics during walking: normative study in young adults. Am J Phys Med Rehabil. 77:2–7.

Lun V, Meeuwisse WH, Stergiou P, Stefanyshyn D. 2004. Relation between running injury and static lower limb alignment in recreational runners. Br J Sports Med. 38: 576–580.

Malek MM, Mangine RE. 1981. Patellofemoral pain syndromes: a comprehensive and conservative approach. J Orthop Sports Phys Ther. 2:108–116.

Mezghani N, Husse S, Boivin K, Turcot K, Aissaoui R, Hagemeister N, De Guise JA. 2008. Automatic classification of asymptomatic and osteoarthritis knee gait patterns using kinematic data features and the nearest neighbor classifier. IEEE Trans Biomed Eng. 55:1230–1232.

Moisio KC, Sumner DR, Shott S, Hurwitz DE. 2003. Normalization of joint moments during gait: a comparison of two techniques. J Biomech. 36:599–603.

Nigg BM, Stefanyshyn D, Cole G, Stergiou P, Miller J. 2003. The effect of material characteristics of shoe soles on muscle activation and energy aspects during running. J Biomech. 36: 569–575.

Schöllhorn WI, Nigg BM, Stefanyshyn DJ, Liu W. 2002. Identification of individual walking patterns using time discrete and time continuous data sets. Gait Posture. 15: 180–186.

Stefanyshyn DJ, Stergiou P, Lun VM, Meeuwisse WH, Worobets JT. 2006. Knee angular impulse as a predictor of patellofemoral pain in runners. Am J Sports Med. 34: 1844–1851.

Van Mechelen W. 1992. Running injuries. A review of the epidemiological literature. Sports Med. 14:320–335.

Vardaxis VG, Goulermas JY. 2006. Gender specific gait patterns characterized by probabilistic neural networks. J Biomech. 39:S112.

Viola P, Jones MJ. 2004. Robust real-time face detection. Int J Comput Vis. 57:137–154.

Von Tscharner V, Goepfert B. 2003. Gender dependent EMGs of runners resolved by time/frequency and principal pattern analysis. J Electromyogr Kinesiol. 13:253–272.

Wu J, Wang J, Liu L. 2007. Feature extraction via KPCA for classification of gait patterns. Hum Mov Sci. 26:393–411.