INTRODUCTION

Patellofemoral pain syndrome (PFPS) has been reported to be the most common running-related injury [1]. While the aetiology of PFPS is multifactorial, there remains limited understanding about the kinematic gait characteristics associated with PFPS [1]. Potentially injurious features related to PFPS, should be determined so that modifiable factors can be identified to treat and prevent this injury. The paucity of conclusive evidence in the current literature may be explained by the restricted ability of inferential statistics to discern between PFPS and healthy controls. Previous studies have typically presented high numbers of variables in conjunction with relatively small numbers of subjects and intergroup differences [1,2].

Support Vector Machines (SVM) have recently been shown to be accurate in classifying running kinematics patterns into two groups [2]. In contrast to inferential statistics, SVM techniques are able to classify group data using a combination of different kinematic features. Moreover, the SVM, when combined with a feature selection algorithm, could potentially provide a limited number of discriminative variables, thus allowing for a more robust understanding of kinematic factors associated with PFPS. The objective of this study is to observe the performance of the SVM in classifying PFPS and control subjects based on discrete kinematic gait variables.

METHODS

Sixteen symptomatic PFPS subjects (7 males, 9 females; age: 28±8 years; height: 167±9 cm; mass: 66±11 kg) and 14 pain-free subjects (7 males, 7 females; age: 28±10 years; height: 175±9 cm; mass: 65±10 kg) volunteered to participate in the study. All subjects were currently involved in running or jumping activities for 30 minutes at least three times/week. All subjects provided informed consent prior to participation. Six and ten subjects reported current right and left-sided PFPS symptoms, respectively over an average of 1.0±1.3 years. Inclusion and exclusion criteria for the PFPS subjects were based on a previous study [3].

To collect kinematic data, rigid clusters of retro-reflective markers were attached on the pelvis, thigh, shank and foot of the tested leg. Individual markers were also placed on bony anatomical landmarks to define the segmental coordinate systems during the standing calibration trial. After the standing calibration trial, the anatomical markers were removed and the subjects ran on a treadmill at 2.7 m/s. Eight Vicon cameras (Vicon, Oxford, UK) collected 3D marker co-ordinates at 200 Hz and the data were filtered at 10 Hz using a fourth order Butterworth filter. Data for ten complete footfalls were collected following a two-minute run to allow familiarization with the treadmill speed. The participants wore standard laboratory shoes (Nike Air Pegasus, Nike Inc, USA).

Two types of gait parameters were extracted for further analysis: basic and kinematic gait variables. Both types have been suggested as contributing factors in the development of PFPS [1]. For a better characterization of the PFPS groups, we selected the variables most commonly reported in previous studies [1]. Step width, stride rate, stride frequency and stance time (ST) were the basic gait parameters. The kinematic variables were extracted from the stance phase and included the joint angle at initial contact (IC) the peak value, the angular excursion (from IC to peak value) and percentage of stance to the peak angle. All kinematic variables were obtained for the hip, knee and ankle in all three anatomical planes (sagittal, frontal and transverse). The IC, peak and excursion of the foot angle with respect to the laboratory coordinate system were also quantified. The symptomatic leg for the PFPS group and the right or left leg for six and eight control subjects, respectively, were used for the comparison.

Altogether, 47 discrete basic and kinematic gait variables were chosen to train and test the SVM classification algorithm using linear and polynomial kernel function (d=3) implemented in Matlab 7.7 (Mathworks, MA, USA). In brief, the SVM classifier maps the features in a high-dimensional space using kernel functions and then constructs an optimal linear separating hyperplane providing a maximum distance between classes [2]. When the data is not linearly separable, a nonlinear function maps the data in the transformed high dimensional feature space. The only kernel independent parameter C that defines the tradeoff between margin width and misclassification rate was tested when the model was trained. Different values for C (0.1, 1, 10, 100, 1000) were used to assess the dependence of the classifier on this parameter. A 10-fold cross-validation was employed to assess the generalizability of the SVM algorithm. A minimum subset of features necessary to achieve the best performance of the classifier was determined by using a forward feature selection algorithm in combination with the SVM classifier [2]. In addition, unpaired t-tests were performed with the first three SVM-detected features to compare both results.

RESULTS AND DISCUSSION

The SVM classifier achieves at best 63% and 82% accuracy rate when the basic and kinematics features were input alone in the algorithm using linear kernel (C=0.1 and C=1000, respectively). In contrast, when those features were combined in an input matrix, the performance of the (C=1).
The basic variables alone provided limited discriminative information (63%). Likewise, the kinematic variables per se exhibited an improved accuracy rate (82%). However, when basic and kinematic variables were combined, the classifier achieved the best overall performance (85%). In fact, a previous study using the same classification approach has demonstrated improvement when different types of features were combined [4].

![Figure 1](image1.png)

**Figure 1**: Reliance of the accuracy rate (%) on the number of features selected by the forward selection algorithm using linear kernel (C=1).

Whilst the combination of features improves the ability of the classifier it appears that not all variables positively contributed to group discrimination, since with only three selected features it was possible to attain 100% accuracy for group allocation (Figure 1). Moreover, when more than 25 features were added into the algorithm the performance deteriorated (Figure 1). This behavior suggests that some features provided redundant information to the classifier. Among the optimal features detected by the SVM, two kinematics features: percentage to peak knee flexion and percentage to peak ankle rotation (Peak%KFLX and Peak%AROT); and one basic feature (ST) were firstly selected by the algorithm. In fact, the Peak%KFLX (p<0.001) and the Peak%AROT (p=0.003) occurred significantly earlier; and the ST (p=0.007) was greater for the PFPS group when the inferential statistics was applied in the SVM-detected variables, thus supporting the results of the SVM approach (Figure 2 and Table 1). Indeed, these results suggest that particular features were linearly separable since the recognition rate of the SVM classifier using linear kernel was better compared to the polynomial kernel (d=3). The scattergram in Figure 2 confirms the group allocation provided by the first two selected features. The earlier Peak%KFLX and Peak%AnkROT may negatively influence patellar tracking since an earlier peak angle would suddenly drive the patella to seat into the femoral trochlea, thus resulting in increased loading. It is possible that the combination of timings for peak knee flexion and peak ankle rotation may be important for detecting PFPS rather than the magnitude of their peaks. The prolonged ST has also been exhibited by PFPS subjects in a previous study [5]. This prolonged period with the foot on the ground may lead to an improper knee angular impulse which is considered a contributor factor to develop a PFPS in runners [6].

![Figure 2](image2.png)

**Figure 2**: Scatter plot graph showing the separability provided by the first two linear kernel SVM-detected features: Peak%KFLX and Peak%AROT.

**CONCLUSIONS**

The SVM was able to accurately and effectively group and identify PFPS running gait patterns based on a large set of combined gait features. Furthermore, the use of the SVM associated with a feature selection algorithm provided a smaller subset of optimal features (3 versus 47 features) necessary to achieve optimal performance. The SVM detected features that also exhibited group differences when inferential statistical were applied. These results suggest that the SVM was at least equivalent in its ability to detect running gait changes in PFPS subjects with the asset of using combined feature discriminative information. Future applications of the SVM in prospective studies, with a larger sample size, and the inclusion of kinetic data are encouraged to better understand the gait variables that precede PFPS.

**ACKNOWLEDGEMENTS**

We thank the Brazilian Ministry of Education (CAPES Foundation) and Alberta Innovates: Health Solutions for financial support.

**REFERENCES**


**Table 1**: Mean values (±1SD) and the P-values of the t-test comparisons between PFPS and Control participants.

<table>
<thead>
<tr>
<th>GROUP</th>
<th>VARIABLES</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PeakKFLX (%)</td>
<td>PeakAnkROT (%)</td>
<td>STANCE TIME (sec)</td>
<td></td>
</tr>
<tr>
<td>PFPS</td>
<td>37 ± 3</td>
<td>40 ± 10</td>
<td>0.28 ± 0.02</td>
<td></td>
</tr>
<tr>
<td>CONTROL</td>
<td>54 ± 11</td>
<td>55 ± 13</td>
<td>0.26 ± 0.01</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>&lt;0.001</td>
<td>0.003</td>
<td>0.007</td>
<td></td>
</tr>
</tbody>
</table>