Objective grading of the pivot shift phenomenon using a support vector machine approach

David R. Labbe a,b,*, Jacques A. de Guise a,b, Neila Mezghani a, Véronique Godbout c, Guy Grimard d, David Baillargeon e, Patrick Lavigne f, Julio Fernandes g, Pierre Ranger g, Nicola Hagemeister a,b

a Laboratoire de recherche en imagerie et orthopédie, Centre de recherche, Centre hospitalier de l’Université de Montréal (CHUM), Montréal, Canada
b École de technologie supérieure, Montréal, Canada
c Hôpital Notre-Dame, Montréal, Canada
d Hôpital Ste-Justine, Montréal, Canada
e Hôpital de la Cité-de-la-Santé, Laval, Canada
f Hôpital Maisonneuve-Rosemont, Montréal, Canada
g Hôpital du Sacré-Cœur, Montréal, Canada

A R T I C L E  I N F O

Article history:
Accepted 11 August 2010

Keywords:
Pivot shift
Objective grading
Classification
Knee joint instability
ACL rupture
Support vector machines

A B S T R A C T

The pivot shift test is the only clinical test that has been shown to correlate with subjective criteria of knee joint function following rupture of the anterior cruciate ligament. The grade of the pivot shift is important in predicting short- and long-term outcome. However, because this grade is established by a clinician in a subjective manner, the pivot shift’s value as a clinical tool is reduced. The purpose of this study was to develop a system that will objectively grade the pivot shift test based on recorded knee joint kinematics. Fifty-six subjects with different degrees of knee joint stability had the pivot shift test performed by one of eight different orthopaedic surgeons while their knee joint kinematics were recorded. A support vector machine based algorithm was used to objectively classify these recordings according to a clinical grade. The grades established by the surgeons were used as the gold standard for the development of the classifier. There was substantial agreement between our classifier and the surgeons in establishing the grade (weighted kappa = 0.68). Seventy-one of 107 recordings (66%) were given the same grade and 96% of the time our classifier was within one grade of that given by the surgeons. Moreover, grades 0 and 1 were distinguished from grade 2 to 3 with 86% sensitivity and 90% specificity.

Our results show the feasibility of automatically grading the pivot shift in a manner similar to that of an experienced clinician, based on knee joint kinematics.

1. Introduction

Rupture of the anterior cruciate ligament (ACL) typically leads to increased anteroposterior (AP) and rotational laxity, resulting in a functional instability of the knee joint. The AP laxity can be evaluated using the Lachman test (Torg et al., 1976) or the anterior drawer test (Turek, 1984). These tests, particularly the Lachman test, have been shown to be useful in establishing a diagnosis of ACL rupture but they are not related to subjective criteria of knee joint function (Lephart et al., 1992; Engstrom et al., 1993; Eastlack et al., 1999; Kocher et al., 2004; Pollet et al., 2005). The pivot shift test, which reproduces the functional instability, correlates with several subjective criteria such as patient satisfaction, giving way and activity limitation amongst others (Kocher et al., 2004; Leitze et al., 2005). It is generally accepted that ACL reconstruction should aim to eliminate the presence of a pivot shift in order to maximize patient outcome (Kaplan et al., 1990; Jonsson et al., 2004; Kocher et al., 2004; Leitze et al., 2005; Andriacchi et al., 2006).

The pivot shift test is performed with the patient supine and the examined leg lifted off the examining table. A gentle valgus force is applied to the knee and the knee is flexed in a controlled manner with slight internal rotation of the tibia. In the ACL-deficient knee, as flexion occurs, the tibia translates anteriorly and rotates internally. The joint is subluxed at this point. As the knee is flexed past 30°, soft tissues and joint geometry cause the joint to reduce (Galway and MacIntosh, 1980). This is the pivot shift,
graded subjectively according to the International Knee Document Committee (IKDC) criteria (0 = absent, 1 = glide, 2 = clunk and 3 = gross) (Hefni et al., 1993).

Such a grading system is poorly repeatable and varies greatly between clinicians, which may be due to their individual expertise (Bull and Amis, 1998). However, the grade is said to be critical in establishing which type of treatment to pursue (Bull and Amis, 1998) and has been directly correlated to the ability to return to normal sports participation (Kaplan et al., 1990). Different studies have established a link between the lingering post-operative pivot shift grade and poor long-term outcome following ACL reconstruction (Kaplan et al., 1990; Jonsson et al., 2004).

Because no objective and reliable pivot shift grading system is available, it is difficult to use pre- and post-operative pivot shift grades to evaluate results. This deficiency also makes it difficult to compare the outcomes of different treatments or studies by different authors. The Lachman test, which can be quantified using instruments such as the KT-1000, KT-2000 (MEDmetric Corp, San Diego, California) and the Rolimeter (Aircast Corp, San Diego, California) (Schuster et al., 2004), is thus still used for this purpose despite extensive literature showing the pivot shift to be the best predictor of short- and long-term outcome.

Previous studies have attempted to quantify the pivot shift test. These studies have focused on recording knee joint kinematics during the pivot shift test and establishing which parameters correlate with the pivot shift grade (Bull et al., 2002; Hoshino et al., 2007; Kubo et al., 2007; Amis et al., 2008; Lane et al., 2008a; Ishibashi et al., 2009; Lopomo et al., 2009). Some relatively strong correlations were found between the kinematic parameters and the pivot shift grade. However, the variability between recordings of subjects with the same pivot shift grade is such that none of these parameters can be used as a quantitative measure of the pivot shift.

Recently, statistical classification methods have been successfully used in the field of biomechanics to classify different types of gait patterns and diagnose different musculoskeletal pathologies (Chan et al., 2010; Lauer et al., 2005; Giansanti et al., 2008; Lau et al., 2009). Machine learning methods such as support vector machines (SVMs), artificial neural networks (ANNs) and the k-nearest neighbor algorithm use a set of training data for which the correct class is known in order to establish the class of a given patient’s data. SVMs are less prone to overfitting than ANNs and their simple geometric interpretation provides fertile ground for further investigation (Burges, 1998), making them particularly useful for classifying kinematic data.

In this study, we hypothesize that an SVM-based approach could be used to grade the pivot shift objectively in high agreement with the grading of experienced orthopaedic surgeons. This study was designed to develop such a method using knee joint kinematics recorded during an instrumented pivot shift test. If proven effective, this method would eliminate subjectivity as an important source of variability in the pivot shift grading.

2. Materials and methods

2.1. Kinematic data collection

The experimental protocol of this study was divided into two separate phases. The first involved 12 subjects (32.9 ± 9.2 years old), nine were male and three were female. Of these subjects, 4 were ACL-intact and 8 presented various degrees of knee joint instability caused by ACL rupture. Three orthopaedic surgeons each evaluated the ACL-deficient knee of the symptomatic subjects and a randomly chosen knee of ACL-intact subjects, resulting in 36 pivot shift evaluations. The purpose of this phase was to evaluate the inter-rater reliability of our grading method.

In the second phase, 44 additional subjects were evaluated by one of eight different orthopaedic surgeons. Of these subjects, 4 were ACL-intact and 40 were ACL-deficient. Subjects had a mean age of 26.6 years old (± 11.5); 26 were male, 18 were female. Subjects with an asymptomatic contralateral knee had both knees evaluated, which resulted in 71 additional pivot shift recordings for the second phase. These recordings were added to the 36 obtained in the first phase, resulting in 107 recordings used for the development and evaluation of the SVM classifier. Grade distribution is shown in Table 1.

To be included in the ACL-intact group, the subjects had to have been free of injury or pain in both knees. For ACL-deficient subjects, additional soft tissue injuries (i.e. meniscal and/or MCL tears) were not exclusion criteria as long as they did cause pain and guarding during the pivot shift evaluation. All subjects signed consent forms approved by the institutional ethics committees.

Each subject had electromagnetic motion sensors (Fastrak, Polhemus, Colchester, VT) attached to his Tibia and Femur using an attachment system developed by the authors with the objective of diminishing skin to bone movement artifacts (Fig. 1). This attachment system has been described in previous work (Labbe et al., in press). Another motion sensor was attached to a belt that was tightly apposed over the iliac crest and was used for anatomical calibration. With the patient in supine position, a passive functional calibration method was applied to identify hip, knee and ankle joint centers and align the anatomical axes through these joint centers. This calibration method was an adaptation of the functional postural (FP) method (Hagemeister et al., 2005), which is usually performed actively by the subject in a standing position.

An experienced orthopaedic surgeon then performed the pivot shift test while knee joint kinematics were recorded. The induced pivot shift was graded from 0 to 3 by the orthopaedist and this grade was considered the gold standard for further analysis. For subjects that were evaluated in the inter-rater phase, the clinicians were blinded to the grades attributed by their colleagues. The evaluations were conducted at 5 different hospitals.

2.2. Kinematic data extraction

The position data obtained from the pivot shift recordings were filtered using a second order Butterworth lowpass filter with a cutoff frequency of 6 Hz. The data from the calibration method were combined with the filtered kinematic data to express them in the anatomical axes, as described by Grood and Suntay (1983). This resulted in rotations (flexion/extension, internal/external tibial rotation and abduction/adduction) and translations in all three anatomical axes (anteroposterior, mediolateral and proximodistal). Velocities and accelerations were obtained by derivative and double-derivative of positional data.

2.3. SVM classifier

A second degree polynomial support vector machine (SVM) algorithm was implemented in Matlab (Mathworks, Natick, MA) and used for grade classification of the knee recordings. Because we aimed to replicate the grading of experienced clinicians, a supervised learning method such as an SVM was appropriate (Fig. 2). The kinematic features used as dimensions for the SVM classifier were identified using principal component analysis (PCA) (Labbe et al., in press). Features were added one at a time in order of their correlation to the principal components, as long as classification sensitivity was improving. The grades attributed by the clinicians were taken to be the gold standard for training the SVM.

Following a phase of supervised learning, a SVM constructs a hyperplane in n-dimensional space that separates a dataset into two subsets, representing different classes. As a first step, a SVM was trained to classify the recordings as being of grades 0, 1 or 2, 3 (SVM1) using all 107 recording as training data. This makes sense in the case of the pivot shift, as there is a clear distinction between these pairs of grades (absence vs presence of a clunk). Two additional SVMs were then trained to separate grades 0 from grades 1 (SVM2) and grades 2 from grades 3 (SVM3). Training data for these SVMs were the sets of recordings graded 0 or 1 and 2 or 3 by the clinicians, respectively. The combination of these three SVMs was then validated as a whole, using the leave-one-out cross-validation (LOOVC) method (Duda et al., 2001). In LOOVC, the subject dataset was separated into 107 training samples where one of the recordings was used as the test sample and the others were used as training data. The trained classifier was then tested with the test sample. This process was repeated 107 times using different test and training data so that each recording was included in both training and testing.

Table 1 Grade distribution of the pivot shift recordings, as attributed by the clinicians.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Number of knee recordings</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
</tr>
</tbody>
</table>

Grade 0 = absence, grade 1 = glide, grade 2 = clunk, grade 3 = gross.
Fleiss’ weighted Kappa coefficients ($k$) for multiple observers (Fleiss, 1971; Fleiss and Cohen, 1973) were calculated with quadratic weights to evaluate the inter-rater agreement between the clinicians in grading the pivot shift both subjectively and using the SVM classifier. Cohen’s weighted Kappa coefficients ($k$) (Cohen, 1968) with quadratic weights were calculated to obtain the agreement between the clinicians and our multiclass SVM-based classifier. The sensitivity and specificity of SVM 1, which separates recordings with a clunk from those without a clunk, were also calculated.

### 3. Results

Success rates for SVM 1 and SVM 3 were best when using the amplitude of anteroposterior (AP) translation and the five features that had the highest correlation to the two principal components. These features were, in order: the total linear acceleration, AP velocity, total linear velocity, mediolateral (ML) linear acceleration and AP linear acceleration. SVM 2, which separates grades 0 and 1 had the highest success rate using only the amplitudes of tibial translation and rotation as well as the velocity of tibial rotation.

In the inter-rater phase, one of the three clinicians established a pivot shift grade that was different than that established by his peers for 7 of the 12 subjects (Table 2). For all of these 7 subjects, the difference was of 1 grade. The Kappa coefficient for agreement between clinicians in subjectively grading the pivot shift was $k=0.83$, considered “almost perfect” (Landis and Koch, 1977). The agreement between the SVM-established grades of the different clinicians was $k=0.81$, also considered “almost perfect”. Agreement between the subjective grades and the SVM-established grades were $k=0.83$, 0.79 and 0.82 for clinicians 1, 2 and 3, respectively.

In the broad phase, the clinicians and the classifier were in agreement in 66% of cases (Table 3). Overwhelmingly, when the classifier was in disagreement with the clinicians, the difference was of only one grade. In fact, the classifier and the clinicians were in agreement to within one grade in 95% of cases. Agreement between the clinicians and the classifiers, as defined by a Cohen’s weighted Kappa, was $k=0.68$, considered to be substantial agreement (Landis and Koch, 1977).

SVM 1, trained to distinguish grades 0 and 1 from grades 2 and 3, also yielded noteworthy results. This SVM achieved a sensitivity of 86% and a specificity of 90%, using LOOCV (Table 4). Agreement with the clinicians was also found to be substantial as indicated by the Kappa coefficient of $k=0.74$. Of the 5 recordings wrongly classified as grades 2 or 3 by our classifier, in disagreement with the clinicians, 4 had been given a grade 1 by the evaluating clinician and 1 had been graded as a 0. All 9 of the recordings classified as being of grades 0 or 1, in disagreement with the clinicians, had been given a grade 2 by the evaluating clinician.

### 4. Discussion

Despite its demonstrated clinical relevance, the pivot shift grade remains highly subjective and unreliable (Bull and Amis, 1998; Lane et al., 2008b). In a study of the kinematics produced by 10 different examiners, Noyes et al. (1991) found that tibial rotations and translations induced on a single cadaver knee varied drastically. In the intra-rater phase of the current study, the grades established by 3 clinicians who all have experience grading

![Fig. 1. Electromagnetic motion capture device attached to a subject's lower limb using an attachment system developed with the objective of diminishing skin to bone movement artifacts.](image1)

![Fig. 2. Classification of the pivot shift recordings according to their clinical grade, using three SVMs.](image2)

### Table 2
Grades attributed by each clinician and by the SVM classification of the kinematic recording of their evaluations, in the inter-rater phase.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Clinician 1</th>
<th>SVMs</th>
<th>Clinician 2</th>
<th>SVMs</th>
<th>Clinician 3</th>
<th>SVMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Bold italics show clinical grades that are in disagreement with those established by peers; underlined grades show those where the SVM classifier graded in disagreement with the evaluating clinician.

### Table 3
Grading of the pivot shift recordings by clinicians (lines) and a SVM-based classifier (columns).

<table>
<thead>
<tr>
<th>Clinicians</th>
<th>Grade 0</th>
<th>Grade 1</th>
<th>Grade 2</th>
<th>Grade 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 0</td>
<td>22</td>
<td>5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Grade 1</td>
<td>8</td>
<td>12</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Grade 2</td>
<td>4</td>
<td>5</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>Grade 3</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>14</td>
</tr>
</tbody>
</table>
the pivot shift were shown to be in very high agreement statistically. Despite this fact, for 7 of the 12 subjects, one of the clinicians graded one grade higher or lower than did his peers.

Much effort has been put into quantifying the kinematics of the pivot shift (Bull et al., 2002; Hoshino et al., 2007; Kubo et al., 2007; Amis et al., 2008; Lane et al., 2008a; Ishibashi et al., 2009; Lopomo et al., 2009) but to date, there has not been a method proposed to establish the grade in an objective manner. In this study we were able to automatically grade the pivot shift in the same manner as the surgeon performing the pivot shift test for 71 of the 107 recordings. When the classifier was not in agreement with the clinicians, was there was a difference of more than one grade in only 5 cases. A weighted kappa was used to interpret the results because it accounts for partial agreement. It showed substantial agreement ($\kappa = 0.68$) between our SVM-based classifier and the interpretation of the orthopaedic surgeons, which were used as a gold standard. Moreover, agreement between the subjective grades and the SVM-established grades were similar for all three clinicians included in the phase ($\kappa = 0.83, 0.79, 0.82$) indicating that the classifier’s sensitivity did not differ significantly depending on the examining clinician.

The first step of our method (SVM 1) is able to classify the pivot shift recordings as being of grade 0, 1 or 2, 3 with high specificity (90%) and sensitivity (86%). These specificity and sensitivity values are calculated with regards to the agreement between our classification method and the grade established by experienced orthopaedic surgeons.

In the literature, distinction has been made between grades 0, 1 and 2, 3. Leitze et al. (2005) used the term pivot shift to describe grades 2 and 3 whereas they used the term pivot glide for grade 1. They showed that the correlation to subjective criteria of knee joint function was much lower for a pivot glide versus a pivot shift. Kaplan et al. (1990) found that a grade 1 pivot shift does not correlate to clinical instability and that the majority of knees that exhibit this type of pivot shift do not demonstrate giving way despite a high level of sports participation. In their study of 52 patients, none of the subjects with a grade 2 or 3 pivot shift were able to return to unlimited sports participation. On the other hand, only 29% of those with a grade 0 or 1 were unable to do so. Furthermore, it has been documented that patients that have increased joint laxity often display a grade 1 pivot shift in the absence of trauma (Lane et al., 2008b).

The kinematic parameters that were used by our classifier give insight into what a clinician is interpreting when he attributes a grade to the pivot shift. They show that the translational component is much more important than the rotational component in grading the pivot shift. Moreover, it is essential to consider the velocity and acceleration of the translation, not solely its amplitude. This indicates that the pivot shift grade is more closely related to the suddenness of the tibial reduction than to its amplitude. This makes sense as the current subjective scale describes grades 2 and 3 as a “clunk” and a “gross clunk”. These terms imply suddenness but contain no notion of amplitude of displacement.

The parameters used for SVM 2 differed. For separating grades 0 and 1, the tibial rotation and the amplitude of tibial translation are used. The velocity and acceleration of the translation are not.

This is also in accordance with the subjective scale as we are trying to distinguish between the presence or the absence of a “glide”. Here, there is no notion of suddenness, just of a small displacement.

Because there is no way to establish a true grade, our results are based on the grading of experienced clinicians. The subjective nature of this grading scale is what prompted efforts to quantify the pivot shift in the first place. Such an imperfect gold standard, coined a “fuzzy gold standard” (Phelps and Hutson, 1995), results in a paradox because we are attempting to reproduce the very scale which we aim to replace. This is highlighted by the fact that in the inter-rater phase, all three clinicians were in agreement for only 5 of 12 subjects. Taking this into consideration, it cannot be expected that the classifier establish the same grade as every clinician, every time. Moreover, the pivot shift grades are discrete but the actual kinematics that are produced are of a continuous nature. Therefore, in practice, there exist gray areas at the boundaries separating adjacent grades. For example, what one clinician interprets to be a grade 2, another may interpret to be a grade 3. While the lines our classifier draws between grades is not be identical to those of every clinician, it is objective and remains constant.

Consequently, it is to be expected that there will be some degree of disagreement with clinicians along these lines. In fact, using the first part of our classifier, 13 of the 14 recordings where there is disagreement were considered to be of grade 1 or 2 (the frontier grades) by the clinicians. In our full classifier, 31 of the 36 recordings where there is disagreement were assigned a grade adjacent to that given by the clinician.

An SVM-based approach for establishing the grade from recorded knee joint kinematics offers an objective alternative to grading the pivot shift test. Clinicians could use such a tool to establish the pivot shift grade in a replicable manner. It is worth noting however that our method classifies kinematic recordings and not patients. This distinction is important in that the method remains dependant on the clinician’s execution of the test. If a clinician only induces a low-grade pivot shift on an ACL-deficient knee, the recording will be classified as such. In that sense, our method could also be used as a tool for teaching the pivot shift maneuver and its grading. Before its transfer to clinical use, intra- and inter-observer reliability of the method will have to be further investigated. Future work will focus on the development of a quantitative measure based on the results of this study. Such a measure would be valuable in quantifying the outcomes of different ACL surgeries.

### Conflict of interest statement

There is a patent pending which is related to the attachment system which was used to attach the motion capture devices to the lower limb. This patent will be held by some of the authors of this study (Labbéd H, Hagemeister N, de Guise JA. Harness System for Kinematic Analysis of the Knee. Provisional Patent: USA (60/990,074): November 2007). Furthermore, some of the authors of this work hold stock in Emovi Inc., which funded part of the study and owns commercial rights to the attachment system that was used.

None of the authors have other potential financial conflict of interest.

### Acknowledgements

The authors would like to thank NSERC, Canada Research Chair, Emovi Inc. and the MENTOR program of the CIHR for
References


