Journal of Biomechanics 44 (2011) 540–542

Contents lists available at ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech
www.JBiomech.com

Short communication

Support vector machines for detecting age-related changes in running kinematics ☆

Reginaldo K. Fukuchi [a,*], Bjoern M. Eskofier [b], Marcos Duarte [c], Reed Ferber [a]

[a] Running Injury Clinic, Faculty of Kinesiology, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4
[b] Human Performance Laboratory, Faculty of Kinesiology, University of Calgary, Canada
[c] School of Physical Education and Sport, University of São Paulo, São Paulo, Brazil

ARTICLE INFO

Article history:
Accepted 27 September 2010

Keywords:
SVM
Ageing
Kinematics
Running

ABSTRACT

Age-related changes in running kinematics have been reported in the literature using classical inferential statistics. However, this approach has been hampered by the increased number of biomechanical gait variables reported and subsequently the lack of differences presented in these studies. Data mining techniques have been applied in recent biomedical studies to solve this problem using a more general approach. In the present work, we re-analyzed lower extremity running kinematic data of 17 young and 17 elderly male runners using the Support Vector Machine (SVM) classification approach. In total, 31 kinematic variables were extracted to train the classification algorithm and test the generalized performance. The results revealed different accuracy rates across three different kernel methods adopted in the classifier, with the linear kernel performing the best. A subsequent forward feature selection algorithm demonstrated that with only six features, the linear kernel SVM achieved 100% classification performance rate, showing that these features provided powerful combined information to distinguish age groups. The results of the present work demonstrate potential in applying this approach to improve knowledge about the age-related differences in running gait biomechanics and encourages the use of the SVM in other clinical contexts.

1. Introduction

There has been an increase in the number of elderly people engaged in recreational and competitive running (Jokl et al., 2004). However, some studies have considered biological ageing as a significant risk factor for running-related injuries (Satterthwaite et al., 1999; Wen et al., 1998).

The greater incidence of injuries among older runners might be due to age-related changes in musculoskeletal properties such as muscle weakness, joint stiffness, and/or changes in running movement patterns (Bus, 2003; Fukuchi and Duarte, 2008). Therefore, identification of changes in movement patterns in elderly people would be helpful for injury prevention.

In a previous study, we employed classical statistics to describe differences in the running kinematic patterns between young and elderly adults (Fukuchi and Duarte, 2008). However, these statistical techniques are limited in their ability to discriminate between age groups based on the high number of variables present in biomechanical studies contrasted with the small number of between-group differences. Therefore, Support Vector Machines (SVM) have recently arisen as an innovative approach to solve classification problems in the biomedical area (Begg and Kamruzzaman, 2005; Chan et al., 2010).

SVM aim to find a hyperplane that maximizes the distance between groups, thereby creating discriminatory parameters to distinguish groups from one another (Vapnik, 1995). In fact, previous studies have used SVM successfully to discriminate walking biomechanical patterns between age groups (Begg and Kamruzzaman, 2005; Wu and Wang, 2008). However, to date we are unaware of any study that has used this approach for running movement patterns in elderly adults.

In the present work, we hypothesized that SVM will be able to discriminate running patterns across age groups. We re-analyzed our previous data (Fukuchi and Duarte, 2008) to investigate the possibility of detecting differences in gait patterns, using an SVM approach, based on kinematic data for young and elderly runners.

2. Methods

2.1. Subjects

Seventeen elderly male adults (age 69 ± 2 yr) and 17 young male adults (age 31 ± 6 yr), regular runners and all injury-free, volunteered to participate in this study. All subjects met the inclusion criteria, outlined in our previous study, and

☆ http://www.runninginjuryclinic.com/

* Corresponding author. Tel.: +403 210 7091; fax: +403 220 0546.
E-mail addresses: r.fukuchi@ucalgary.ca, regifukuchi@gmail.com (R.K. Fukuchi).

0021-9290/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jbiomech.2010.09.031

were able to complete a recent 10 km run in less than one hour. This study was approved by the ethics committee of the University of Sao Paulo, Brazil.

2.2. Data collection

Three-dimensional kinematics of the right leg were obtained using four digital cameras (GRDVL9800U, JVC Inc., Wayne, NJ). The marker positions were filmed and digitized at 120 Hz using APAS software (Ariel, Inc., Trabuco-Canyon, CA) while the subjects were running at 3.1 m s^{-1}. The mean values across five trials for the right knee and rearfoot angles were used for feature calculation. Further procedures adopted are described in more detail in our previous study (Fukuchi and Duarte, 2008).

2.3. Feature extraction

Thirty-one running kinematic features were extracted from the recorded data, resulting in a feature matrix consisting of 34 subject rows (17 young and 17 elderly) and 31 feature (F) columns. Initial contact (*IC*), peak and excursion of rearfoot and knee joint angles, in all anatomical planes, were recorded resulting in F1–F18 (3 planes $\times$ 3 variables $\times$ 2 joints). In addition the tibial rotation angles at those stance events (F19-F21) were quantified according to our previous study (Fukuchi and Duarte, 2008). The time-to-peak of rearfoot eversion, tibial internal rotation, knee internal rotation, and knee flexion were also recorded (F22–F25). The ratio between the rearfoot eversion and tibial rotation excursions (F26), the absolute segmental angles of rotation of the femur, tibia, and foot at *IC* (F27–F29), as well as the stride length and frequency (F30–F31) were additionally quantified.

2.4. Classification approach

In brief, the SVM algorithm (Vapnik, 1998) found the optimal separating hyperplane, which generated the maximum margin of separation between dataset (young vs. elderly). First, the SVM transforms the input features data into a higher dimensional space using kernel functions and then constructs the linear hyperplane in this transformed space. If the dataset was non-linearly separable a modification of the minimization problem had to be employed in order to allow classification error by considering some non-negative variables often called *slack variables*. The only kernel independent parameter of the SVM was the *C*-parameter that defined the trade-off between margin width and misclassification rate. Different values for *C* (0.1, 1, 10, 100, 1000) have been used in the evaluation to test the dependence of the approach on the *C*-parameter (Appendix A).

Linear, polynomial (d=3) and Gaussian radial basis function (RBF) kernels (Schölkopf and Smola, 2002) were adopted in the present study, since the performance of the SVM may vary accordingly to the chosen kernel. Moreover, to our knowledge there is no theoretical result that supports the performance of one kernel over another.

A 10-fold cross-validation (Duda et al., 2001) was performed and the dataset was divided into ten equal subsets with nine subsets used to train the classifier and one subset used to test. The generalization performance of the classifier in labeling unknown data was assessed during this testing procedure. Overall accuracy was computed using all input features to assess the general ability of the algorithm in discriminating the young–elderly groups successfully.

Furthermore a forward feature selection approach was adopted in the present study. This method works by creating a subset of features and then subsequently adding one new feature at a time, choosing the subset that most increased the classification accuracy (Kohavi and John, 1997). All the computations were made in custom software implemented in Matlab 7.7 (Mathworks, MA, USA).

3. Results

The linear kernel (C=1) exhibited the best performance reaching an overall accuracy rate, when all 31 features were used, of 91% compared to polynomial (d=3; C=100) and RBF (σ=1; C=100) kernels with 85% and 50%, respectively. The chosen *C*-parameters influenced the accuracy rate and the best performance was achieved by the linear kernel when C=1 opposed to C=100 for the polynomial and RBF kernels, respectively (Appendix).

The behavior of the SVM classifier, using the linear kernel (C=1), was assessed using the forward feature selection approach (Fig. 1). This approach demonstrated that with only six selected features, the classifier achieved 100% performance in distinguishing young and elderly runners. The features containing the most discriminative information were the knee flexion excursion angle (KFLXRoM), knee abduction angle at *IC* (ICKABD), ankle peak dorsiflexion angle (PEAKAnkDF), peak knee abduction angle (KABDPEAK), tibial rotation excursion (TibRotRoM), and toe-out angle at *IC* (ICToeOut). Moreover, it can be observed (Fig. 1) that adding more than 18 features decreases the performance of the classifier.

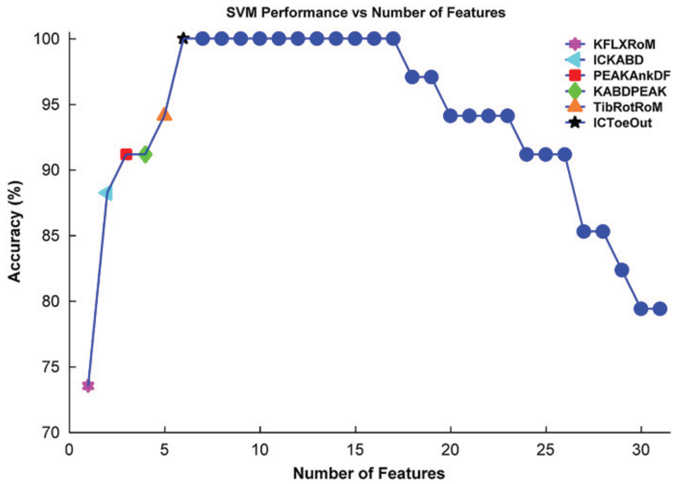

Fig. 1. Performance of the SVM classifier using linear kernel on the number of features. The best performance was achieved with 6 features: KFLXRoM, ICKABD, PEAKAnkDF, KABDPEAK, TibRotRoM, and ICToeOut.

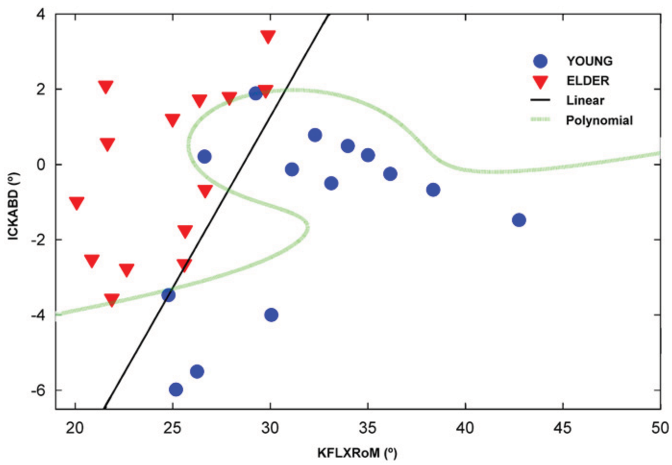

Fig. 2. Scatter plot graph showing the distribution of the best two discriminating features (KFLXRoM (horizontal axis) and ICKABD (vertical axis)) and the separating line (hyperplane) with linear and polynomial (d=3) kernel SVM.

The first two features (ICKABD and KFLXRoM) that were selected as well as the decision boundary (linear and polynomial) are shown in the 2D scatter plot (Fig. 2).

4. Discussion

This classification approach has demonstrated that the SVM algorithm can distinguish young and elderly runners using running kinematic data. Indeed, we have previously reported differences between age groups, when each feature was compared using inferential statistics approach (Fukuchi and Duarte, 2008). However, the present results suggest that not all features have good discriminatory information, since the SVM approach required only 6 features for maximum accuracy. Moreover, when more than 18 features were added, the classification performance deteriorated. This characteristic of overfitting has also been reported in previous studies and is attributed to the redundancy of information (Begg and Kamruzzaman, 2005; Wu and Wang, 2008).

The overall accuracy was best using a linear kernel (91%) and least using RBF (50%) when the *C*-parameters were 1 and 100, respectively (Appendix). Therefore the *C*-parameter should be carefully selected to achieve the best performance. Lai et al. (2009) reported similar behavior when gait kinematic variables from subjects with and without patellofemoral pain syndrome were input in their model. Hence, the linear kernel might be a suitable option since it is the simplest and the computationally fastest to solve the optimization problem.

The forward feature selection algorithm consistently selected KFLXRoM and ICKABD using linear kernel method. In fact, a maximum separation of 88.2% was achieved when these features were combined alone. The distribution of the subjects' data in the 2D scatter plot (Fig. 2) also demonstrates the ability to accurately discriminate young and elderly runners using this feature combination. Indeed, the KFLXRoM exhibited the most significant difference ($p < 0.001$), when age groups were compared in our previous study (Fukuchi and Duarte, 2008). Furthermore, a decrease in KFLXRoM among older runners has been reported in other studies (Bus, 2003; Karamanidis and Arampatzis, 2005). Therefore, feature selection can detect potential candidate features as well as function to avoid redundant information.

The SVM presented ability in detecting ageing effects, using 31 kinematics features with accuracy rate comparable with previous study (Begg and Kamruzzaman, 2005). Nevertheless, this method has demonstrated 100% accuracy with suitably fewer selected kinematic features. Given that there is a higher incidence of injuries among older runners, future clinical applications using SVM approach are envisioned to investigate the relationship between movement pattern and injury development. However, future prospective studies are required to answer this question.

Conflict of interest statement

We wish to confirm that there are no known conflicts of interest associated with this publication and there was no significant financial support for this work that could have influenced its outcome.

Acknowledgements

We thank the Ministry of Education of Brazil (CAPES Foundation) and Fundação de Amparo à Pesquisa do Estado de Sao Paulo (FAPESP/Brazil) for financial support (Grant nos. 0614/092 and 08/10461-7).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.biomech.2010.09.031.

References

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.
Bus, S.A., 2003. Ground reaction forces and kinematics in distance running in older-aged men. Med. Sci. Sports Exercise 35, 1167–1175.
Chan, Y.Y., Fong, D.T., Chung, M.M., Li, W.J., Liao, W.H., Yung, P.S., et al., 2010. Identification of ankle sprain motion from common sporting activities by dorsal foot kinematics data. J. Biomech. 43, 1965–1969.
Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification2nd ed. Wiley & Sons, New York, NY.
Fukuchi, R.K., Duarte, M., 2008. Comparison of three-dimensional lower extremity running kinematics of young adult and elderly runners. J. Sports Sci. 26, 1447–1454.
Jokl, P., Sethi, P.M., Cooper, A.J., 2004. Master's performance in the New York City Marathon 1983–1999. Br. J. Sports Med. 38, 408–412.
Karamanidis, K., Arampatzis, A., 2005. Mechanical and morphological properties of different muscle–tendon units in the lower extremity and running mechanics: effect of aging and physical activity. J. Exp. Biol. 208, 3907–3923.
Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. Artif. Intell. 97, 273–324.
Lai, D.T., Levinger, P., Begg, R.K., Gilleard, W.L., Palaniswami, M., 2009. Automatic recognition of gait patterns exhibiting patellofemoral pain syndrome using a support vector machine approach. IEEE Trans. Inf. Technol. Biomed. 13, 810–817.
Satterthwaite, P., Norton, R., Larmer, P., Robinson, E., 1999. Risk factors for injuries and other health problems sustained in a marathon. Br. J. Sports Med. 33, 22–26.
Schölkopf, B., Smola, A.J., 2002. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, 1st ed. MIT Press, Cambridge, MA.
Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer, New York.
Vapnik, V.N., 1998. Statistical Learning Theory, 1st ed. Wiley & Sons, New York, NY.
Wen, D.Y., Puffer, J.C., Schmalzried, T.P., 1998. Injuries in runners: a prospective study of alignment. Clin. J. Sport Med. 8, 187–194.
Wu, J., Wang, J., 2008. PCA-based SVM for automatic recognition of gait patterns. J. Appl. Biomech. 24, 83–87.