Automatic recognition of gait patterns with machine learning

Zhou, Y. Hortobágyi, T. Lamoth, C.J.C.

University of Groningen, University Medical Centre Groningen, Centre of Human Movement Sciences, Groningen, The Netherlands

Email: y.zhou01@umcg.nl

PURPOSE AND BACKGROUND: Physical, emotional, cognitive, behavioral, and social factors and their interactions underlie gait and balance abilities in older adults. The International Classification of Functioning, Disability and Health (ICF) model can be used as a framework to examine the multifactorial nature of gait and balance decline and fall risk. Machine learning methods is becoming increasingly popular and significant to estimate fall risk and predict disorders. Here, we present a model using an aggregate analysis of gait accelerometer data from healthy adults and geriatric patients. The purpose of this study was to automatically classify gait patterns in groups of participants by comparing the classification accuracy between different machine learning methods, the final purpose for machine learning model is to early identify at-risk gait and evaluate treatment outcome.

METHODS: Healthy young and middle-aged adults (n=58, age: 42.7 ±16.60), healthy old adults (n=54, age: 74.6 ±5.71), and geriatric patient without cognitive impairment (n=126, age: 79.3 ±5.81) participated in the study. Trunk accelerations were measured with 3D accelerometers (DynaPort® MiniMod, McRoberts BV; iPod touch 4G, iOS 6, Apple Inc.; sample frequency ±100 Hz) during three minutes of walking. From the 3D accelerometer signals 23 dynamic gait variables were calculated quantifying gait speed, stability, regularity, variability, and regularity. A Kernel (Gaussian kernel functions) Principle Component Analysis (KPCA) was applied to extract underlying gait features and reduce the dimensionality of the data for Support Vector Machine (SVM) classification, compared with using the original gait features for other machine learning methods, namely Random Forest (RF) and Artificial Neural Network (ANN). Thereafter, a non-linear classification SVM approach was compared with the classification performances of RF and ANN. While previous studies used traditional unsupervised KPCA for feature selection with non-linear regular gait variables, SVM, RF, and AAN are suitable to capture the non-linear high dimensional structure of gait dynamics. While KPCA can reduce the dimensions of non-linear data and output the most significant information, SVM is powerful to classify non-linear dataset from KPCA. On the other hand, RF and ANN work well for non-linear and high dimensional data structures and for the determination of the relationships among gait variables. These methods also can be extended for other kinds of variables to predict different labels, such as fall risk and Parkinson disease.

RESULTS: KPCA reduced gait data dimensions efficiently from 23 dimensions to five dimensions, explaining 97 % of the variance, and representing gait features of pace, synchronization, regularity, and variability. Preliminary analyses showed that classification accuracy of SVM (accuracy=89%, AUC=0.91) and ANN (accuracy=90%, AUC=0.87) was similar and both could identify gait patterns for the three age-based groups. Compared to SVM and ANN, RF is sensitive to parameter selection and results in worse classification performance (accuracy=73% and AUC=0.86). With regard to SVM, different kernel functions resulted in similar classification performance with less parameter influence. All in all, SVM (89% accuracy) and ANN (90%) performed best with the type of data used in the

present study. The model was successful to predict the aging impact gait and geriatric disorder gait.

CONCLUSIONS: Aging affects specific gait features that can be identified by non-linear approaches like KPCA in combination with SVM or ANN. Both of these latter two methods determine temporal gait differences between age groups. Based on gait dynamics representing the quality of the gait, participants of different ages could be automatically labeled. We will add to the present model data streams and information about person's objective and subjective measures at the different levels of the ICF model. And different machine learning models will be added to investigate labelled disorders, to finally predict the risk of falling.