Detecting knee osteoarthritis with statistical and machine learning approaches.

1,2 Margarita Kotti, 1Lynsey D. Duffell, 2,3,4Aldo A. Faisal, and 1Alison H. McGregor
1Musculoskeletal (MSK) Laboratory, Division of Surgery, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, Charing Cross Hospital, London W6 8RF, UK
2Brain Behaviour Laboratory, Department of Bioengineering, Imperial College London, SW7 2AZ London, UK
3Department of Computing, Imperial College London, SW7 2AZ London, UK
4MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College London, Hammersmith Hospital Campus, London, UK

Knee osteoarthritis (OA) is the second leading cause of disability, with rates rising as a consequence of the expanding ageing population. Currently, we have a large volume of raw sensor data captured via a motion capture system, force plates, and electromyography. Additional data are derived from the subjects’ clinical examination. These high-dimensional, multimodal biomedical datasets are being processed using statistical and machine learning approaches to reveal the implicit information they contain.

Our aim is to (i) automatically detect the degree of knee OA (ii) extract parameters capable of differentiating between normal and osteoarthritic knees. Initial experiments exploited the joint angles at the pelvis, hip, knee, and ankle, as well as the direction of the foot during walking. The aforementioned parameters of 16 subjects were analysed by a support vector regressor. Clinical parameters were given as input to a regression tree. The combination yielded a perfect performance. Moreover, we studied a larger dataset comprising the ground reaction forces of 94 subjects. For this dataset, random forests were utilised in order to detect discriminating parameters with physical meaning and to induct rules. The bioinformatics tool evolved was able to identify approximately 20% of our “normal” population as having a high risk of knee osteoarthritis. The relevance of this in relation to a non-invasive early marker of disease requires further exploration. Currently, we are analysing a comprehensive database of the kinematics and kinetics of function in 180 subjects. In specific, two approaches are employed to analyse ground reaction forces. The first one built models exploiting probabilistic principal component analysis and computed the knee OA probability via a Bayes classifier. It achieved almost perfect accuracy. The second one utilised random forests and leads to a perfect accuracy.