ISB 2013 BRAZIL
XXIV CONGRESS OF THE INTERNATIONAL SOCIETY OF BIOMECHANICS
XV BRAZILIAN CONGRESS OF BIOMECHANICS

TOWARDS AUTOMATICALLY ASSESSING OSTEOARTHRITIS SEVERITY BY REGRESSION TREES & SVMS

[1,2] Margarita Kotti, [1]Lynsey D. Duffell, [2,3,4]Aldo A. Faisal, and [1]Alison H. McGregor
[1]Musculoskeletal (MSK) Laboratory, Division of Surgery, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, Charing Cross Hospital, London W6 8RF, UK
[2] Brain Behaviour Laboratory, Department of Bioengineering, Imperial College London, SW7 2AZ London, UK
[3]Department of Computing, Imperial College London, SW7 2AZ London, UK
[4]MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College London, Hammersmith Hospital Campus, London,UK
email: mkotti@imperial.ac.uk, webadderss: http://www1.imperial.ac.uk/medicine/people/a.mcgregor/

SUMMARY
Osteoarthritis (OA) is a degenerative knee joint disease which causes chronic pain and affects approximately 8.5 million people in the UK. In this paper, a novel fully automated framework is proposed which computes the likelihood and degree to which a subject may have OA of the knee. This study aims to provide an automated tool for the clinical environment that can support decision making particularly diagnosis and subsequent orthopaedic management of OA. Specifically this tool focuses on; a. generic subject attributes (like age, sex, assessment of the Knee Injury, Osteoarthritic Outcome Score (KOOS)) and b. kinematic data derived during a gait cycle to automatically classify and diagnose knee OA. For the generic subject data, a hierarchical regression tree was built, whilst the kinematic data was inputted into a support vector machine (SVM) regressor (a robust state-of-the-art machine learning technique) which produces a likelihood value of knee OA. On 8 knee OA patients and 8 controls, 100% correct prediction is achieved.

INTRODUCTION
A challenge in analyzing behaviour is its variability [1] and key to studying behaviour is the ability to identify an underlying simplicity in the data that is reflective of the data mechanisms [2] (here the effects of knee osteoarthritis (OA) on the musculoskeletal system). We believe that a purely data-driven approach yields objective measures and patterns [3]. We call this view bioinformatics of behavior.
This study forms part of a larger, ongoing study examining movement patterns of normal and OA subjects with respect to gait, stair ascent/decent, sit-stand and stand-sit, and squat. So far generic attributes from 100 patients have been recorded. Kinetic data is collected for a subset of this group that includes 16 subjects. Of these 8 presented with knee OA. The remaining 8 were considered normal, presenting with no current pain or diagnosis of OA. However, this does not preclude that they may have early OA symptoms.
Kinematic parameters of gait data were collected. This analysis generated high volumes of data making direct exploitation unfeasible. Consequently we opted to utilize machine learning techniques more specifically regression trees and support vector machine (SVM) regressors. Regression trees exhibit the advantage that they demonstrate apparent closeness to the clinical reasoning processes where at each step a sub-decision is made based on a set of observations. For the kinematic data (SVMs) were employed. SVMs are modern and robust machine learning techniques [4].
The final purpose of this study is to offer clinicians an automated tool that calculates a regression value that ranges from 0 to 2, in order to support them with their clinical decisions. We focus on regression instead of classification since clinician's value a continuous value rather than a single yes/no answer. This approach is also more pertinent to OA which is a slow degenerative disease process that progresses with time, where a value closer to 0 corresponds to a healthy subject, whereas a value of 2 is indicative of severe OA at both knees. In general, a patient may be considered to exhibit no OA if the system calculates a value less than 0.5

METHODS
With respect to generic subject attributes, each subject filled in questionnaires about their age, sex, and dominant leg, whether they have experienced an injury or a surgery at knee and their Tegner activity score. They all completed the Knee injury and Osteoarthritis Outcome Score (KOOS) questionnaire. Additionally, the body mass index (BMI) as well as the leg length was measured. Subjects were asked to walk at their normal speed along a walkway with their motion captured using a Vicon (Oxford, UK) motion capture system. Each subject wore 34 reflective markers and was asked to walk along the walkway 3 times while data was captured at 100 Hz using 10 cameras to track the motion. Trial data where subjects did not cleanly strike the force plate was excluded from the analysis. Joint angles at the pelvis, hip, knee, ankle, and foot were determined using a custom model written in Body Builder software [5]. For all the aforementioned angles, we consider three axes: sagittal, frontal, and transverse. The data for one complete gait cycle was time normalized by linear interpolation for each trial. Subsequently the generic subject attributes were inputted to a regression tree with the training and the test sets kept

disjointed. Specifically, we tested the regression tree on the 16 subjects for which kinematic data are available, whereas we trained it on the remaining 84 subjects. That is there are no common subjects among the training and the test dataset. The resulting regression tree can be seen in Fig. 1. A subset of the original set of 58 generic subject attributes is retained, namely symptoms, ADLs (as they are defined in KOOS), age, Tegner activity score, and existence of previous injury.

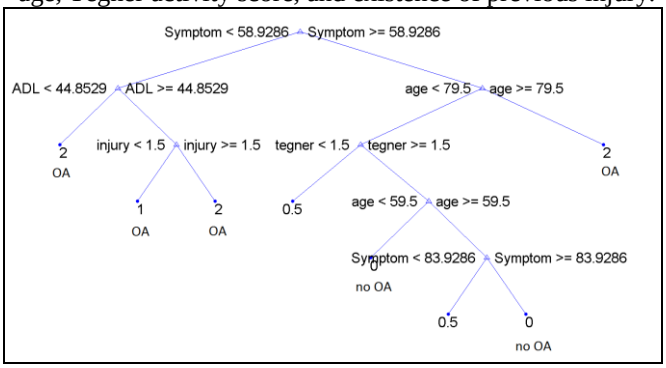

Figure 1: The regression tree trained on 84 subjects and tested on 16 for which kinematic data was available.

With respect to kinematic data, we utilized the data from 16 subjects to provide an input to the SVM regressor [6]. Feature vectors are composed as: 2 knees x 5 joint angles x 3 axes (X,Y,Z) x 101 samples per gait cycle, resulting to 3030 dimensions. Both knees are considered as it is assumed that the gait patterns for both legs will change even if one of the knees suffers from OA. Since each subject provides 2-3 gait cycles, the output is averaged over the gait cycles. The experimental protocol is subject-independent. If a subject's trial is included in the training set, then all the trials of this subject are part of the training set and are consequently not used in the test set. This way, the system is able to handle efficiently a new subject, not encountered during training. The protocol adopted here is a 50%-50% training/testing spilt with two-fold cross validation. Initially the system is trained on the first half of the data (8 subjects) and tested on the second half and then the system is trained on the second half of the data and tested on the first half. This way, every subject is used exactly one time for training and one time for testing. When partitioning the data into two halves extra care was taken in order to have 4 subjects with no OA, 2 with OA at one knee, and 2 with OA at both knees on every half. At a final processing step the outputs of the two techniques (i.e. regression tree and SVM regressor) are linearly combined with equal weights.

RESULTS AND DISCUSSION

An overview of the system's performance is shown in Table 1. Output values below our 0.5 threshold correspond to subjects with no OA. The larger the value is, the more distinctive the sighs of early OA. For all 8 of the patients that have no OA the proposed system produces values with a range of 0.0613 to 0.3626. All those values are below the OA threshold of 0.5. For the remaining subjects that suffer from OA to one or both knees, the values produced belong to range 0.5866 to 1.6122. The larger the value is above 0.5, the more severe the OA. Note that 2 subjects have OA both knees, but the values calculated by the system are 0.5866 and 0.5895, respectively. For one of them the medical records indicate that he/she has mild OA at both knees.

CONCLUSIONS

Our bioinformatics of behavior approach yields an effective assessment of whether the subject suffers from OA and to what degree. The advantage of our method is that it computes a regression value that belongs to a predetermined range, i.e. 0-2, so as to support clinicians when making clinical decisions.

ACKNOWLEDGEMENTS

The authors acknowledge support from the Medical Engineering Solutions in Osteoarthritis Centre of Excellence funded by the Wellcome Trust and the EPSRC.

REFERENCES

1. Faisal A. A. et al, *Nature Reviews Neuroscience* **9**:292-303, 2008.
2. Faisal A. A. et al, *COSYNE* 2009.
3. Faisal, A.A. et al, *PLoS ONE* **5** (11:): e13718, 2010.
4. Cortes C. et al, *Machine Learning* **20**:273–297, 1995.
5. Hope N. et al, *EBS* 2011.
6. Chang C. C. et al, *EBS* 2011.

Table 1: Subject's actual description, along with values produced by the regression tree, the SVM regressor, and the final output of the proposed system.

Subject ID	Subject OA (verbal description)	Subject OA (numerical value)	Regression tree output value	SVM regressor output value	Final output value	Correctly predicted
1	No OA	0	0	0.1226	0.0613	Yes
2	No OA	0	0	0.3812	0.1906	Yes
3	OA at the left knee	1	0	1.4821	0.7410	Yes
4	No OA	0	0	0.7107	0.3554	Yes
5	OA at both knees	2	2	1.2244	1.6122	Yes
6	OA at the left knee	1	1	0.7041	0.8520	Yes
7	OA at the right knee	1	0.5	0.8810	0.6905	Yes
8	OA at both knees	2	2	0.8367	1.4184	Yes
9	No OA	0	0	0.7252	0.3626	Yes
10	No OA	0	0	0.4580	0.2290	Yes
11	OA at the left knee	1	0.5	0.8490	0.6745	Yes
12	OA at both knees	2	0.5	0.6732	0.5866	Yes, but underestimated
13	No OA	0	0	0.4893	0.2447	Yes
14	No OA	0	0	0.4892	0.2446	Yes
15	OA at both knees	2	0	1.1790	0.5895	Yes, but underestimated
16	No OA	0	0	0.1656	0.0828	Yes