Knee Function Classification

An overview of the Dempster-Shafer based classifier

Lianne Jones, Malcolm Beynon, Cathy Holt
Cardiff University,
Cardiff, Wales, UK

***Abstract*—Two major obstacles to using motion analysis as an aid to clinical decision making are the subjectivity of data interpretation and the difficulty in comprehending large amounts of both corroborating and conflicting information. In response to these findings a diagnostic tool was developed that is capable of producing an automated and objective diagnosis from large amounts of motion analysis data. This paper gives an overview of the potential of this objective diagnostic tool for classifying osteoarthritic knee function and quantifying recovery following total knee replacement surgery based on motion analysis techniques.**

Keywords-knee function; classification; osteoarthritis; total knee replacement

I. INTRODUCTION

With the introduction and advancement of computer technology the use of instrumented motion analysis in a research setting has become widespread. It is now possible to obtain and process highly accurate measurements of sophisticated movement in the space of a few hours. In contrast to the pioneers of motion analysis, one of the main challenges facing the biomechanical community today is no longer how to produce data that quantifies human movement; this is now commonplace. Rather, one of the greatest challenges is how to use this information so that it can be of clinical benefit [1].

The two major obstacles to using motion analysis as an aid to clinical decision making are the subjectivity of data interpretation and the difficulty in comprehending large amounts of both corroborating and conflicting information [2-7].

In response to these findings a diagnostic tool was developed that is capable of producing an automated and objective diagnosis from large amounts of motion analysis data [4-6]. This paper gives an overview of the potential of this objective diagnostic tool for classifying osteoarthritic (OA) knee function and quantifying recovery following total knee replacement (TKR) surgery based on motion analysis techniques.

II. METHODS

The diagnostic tool has two main elements: collection of data using motion analysis and the assessment of knee function using a classifier that is based around the Dempster Shafer Theory of evidence (DST).

A. Motion analysis technique

The protocol of [8] was used to analyze the knee function of 51 subjects (20 OA, 22 normal (NL) and 9 TKR). Knee rotations, ground reaction forces (GRF) and temporal-distance parameters during level walking were recorded and anthropometrical measurements were collected. At the end of their visit, patients completed The Activities of Daily Living Scale of the Knee Outcome Survey (KOS) [9].

Following [10] the temporal GRF and rotation waveforms were processed using Principal Component (PC) Analysis. Thus each waveform is represented by a pre-defined number of PC scores which relate to the waveform during specific portions of the gait cycle. This means that temporal information is retained. For interpretation of the PC Scores the reader is directed to [4].

B. Dempster-Shafer Method

The DST-based classifier enables decision-making in the presence of uncertain, inadequate and conflicting evidence, a common problem in the motion analysis laboratory. Using the data collected during the clinical knee trial, this novel approach enables automated, objective classification of subjects into an OA or NL group.

The DST-based classifier [4-6] transforms a subject's knee function data (Table I) into a set of exact belief values: a level of belief that a subject has OA knee function, denoted $m(\{OA\})$; a level of belief that a subject has NL knee function, denoted $m(\{NL\})$; and an associated level of uncertainty, denoted $m(\Theta)$. These belief values are then represented as a single point on a simplex plot, where the least distance from the point to each side of the equilateral triangle is in the same proportion to the three belief values (Figure 1a). This means that the nearer the point is to the NL vertex, the greater the level of NL knee function a subject has. Consequently, the simplex plot can be divided into four classification regions (Figure 1b), namely dominant NL ($m(\{NL\}) \geq 0.5$) (region 1), dominant OA ($m(\{OA\}) \geq 0.5$) (region 2), non-dominant NL ($m(\{OA\}) < m(\{NL\}) < 0.5$) (region 3) and non-dominant OA ($m(\{NL\}) < m(\{OA\}) < 0.5$) (region 4). It was shown in [4] that patients with knee OA tend to lie within the dominant OA region of the simplex plot and NL subjects within the dominant NL region. The control parameters of the DST-based classifier were designed using the knee function data of the 20 OA and 22 NL subjects [see 4].

Cardiff School of Engineering Scholarship; Depuy International; Medical Research Council (MRC); Engineering and Physical Sciences Research Council (EPSRC).

TABLE I. LIST OF THE VARIABLES, v_i (i = 1:18) USED IN THE CLASSIFICATION PROCESS

Variable, v_i	**Variable Description**
v_1	Body mass index (kgm^{-2})
v_2	Cadence (min^{-1})
v_3	Stance phase (% gait cycle)
v_4	PC score related to the peak anterior GRF and the peak posterior GRF
v_5	PC score related to the anterior-posterior GRF during the period from late mid-stance to mid-terminal stance
v_6	PC score related to the anterior-posterior GRF during late pre-swing
v_7	PC score related to the vertical GRF during a portion of mid-stance and the period from heel-rise to opposite initial contact
v_8	PC score related to the vertical GRF from loading response to mid-stance
v_9	PC score related to the vertical GRF during the phase from heel strike transient to the first peak vertical GRF
v_{10}	PC score related to knee flexion over the period from initial contact to opposite initial contact (°)
v_{11}	PC score related to knee flexion during the phase from 58% to 76% of the gait cycle (°)
v_{12}	PC score related to the knee abduction-adduction during the stance phase (°)
v_{13}	PC score related to the knee abduction-adduction during initial swing (°)
v_{14}	PC score related to the knee abduction-adduction during terminal swing (°)
v_{15}	PC score related to the internal-external rotation from loading response to mid swing (°)
v_{16}	Medio-lateral knee width (mm)
v_{17}	Anterior-posterior knee width (mm)
v_{18}	Thigh girth (mm)

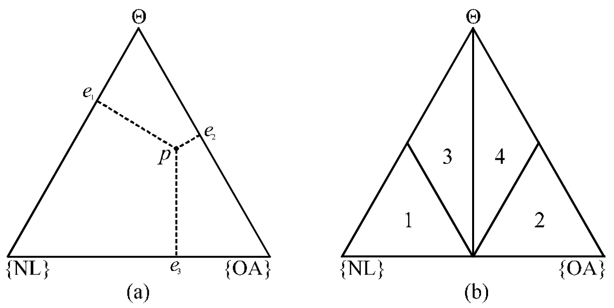

Figure 1. Simplex plot (a) showing the relationship between the position of a point in the simplex plot and the belief values, where $|pe_1|=h.m(\{OA\})$, $|pe_2|=h.m(\{NL\})$ and $|pe_3|=h.m(\Theta)$ (where h is the height of the triangle) and (b) showing the areas of dominant NL (region 1), dominant OA (region 2), non-dominant NL (region 3) and non-dominant OA (region 4) classification

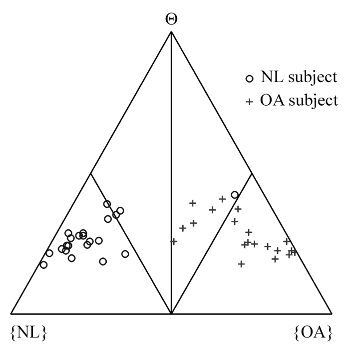

Figure 2. Classification of OA and NL subjects

III. RESULTS AND DISCUSSION

A. Classification of OA and NL knee function

The performance of the classification method in terms of the classification accuracy was determined using two different approaches, namely the resubstitution approach and the leave-one-out (LOO) cross-validation approach. Using resubstitution, the classifier is trained using a set of n subjects and the classification accuracy is calculated using the same n subjects that were used to design the classifier [11]. The classification accuracy calculated using resubstitution is subsequently referred to as the in-sample accuracy. The LOO approach is a special case of cross validation. Using a sample of n subjects, the classifier is trained using ($n-1$) training cases. These control variables are then used to transform the characteristic measurements of the remaining single subject into its associated BOEc, and thus used to classify the subject. This process is repeated n times and the LOO classification accuracy is calculated as the average of the classification accuracy of all of the individually left-out subjects [11]. The LOO classification is subsequently referred to as the out-of-sample accuracy.

The DST-based classifier method was able to classify both in-sample and out-of-sample subjects with an average accuracy of 97.6%. A visual representation of the classification is shown in Figure 2. Two separate clusters of simplex coordinates can be identified within the simplex plot. The NL subjects (circles) are clustered near to the vertex associated with {NL} whilst the OA subjects (crosses) are clustered nearer to the vertex associated with {OA}.

B. TKR Results

This study investigates the potential of the DST-based classifier as an objective tool for assessing the outcome of TKR surgery. Nine patients were followed before and at three stages after TKR surgery. The preoperative and postoperative $m(\{NL\})$ values for the 9 TKR patients are recorded in Table II and shown in terms of their simplex coordinates in Figure 3. The BOE_c values and simplex plots enable the level of benefit achieved by surgery to be established and make comparison of different subjects possible.

Although most of the patients showed some degree of recovery of NL knee function, none of the patients at any stage of recovery gained a dominant NL classification as seen in Figure 3. This suggests that none of the patients recovered complete NL knee function during level walking following TKR surgery. This is in agreement with [12-15] who reported that TKR patients do not achieve NL knee function over time.

Both preoperatively and postoperatively the patients have varying levels of NL knee function. The patients with the greatest levels of NL knee function before TKR surgery exhibit the greatest levels of NL knee function after surgery. Indeed, a Pearson Correlation test revealed a significant correlation between $m(\{NL\})$ at visit 1 and $m(\{NL\})$ at visit 4 (correlation coefficient, r = 0.738).

TABLE II. PRE AND POST-OPERATIVE $M(\{NL\})$ VALUES FOR THE 9 TKR PATIENTS

Patient	Visit 1	Visit 2	Visit 3	Visit 4
1	0.14	0.32	0.07	0.17
2	0.01	0.10	0.18	0.16
3	0.01	0.02	0.05	0.03
4	0.07	0.10	0.12	0.17
5	0.22	0.28	0.35	0.36
6	0.06	0.01	0.29	0.10
7	0.00	0.06	0.10	0.06
8	0.07	0.02	0.03	0.01
9	0.21	0.18	0.39	0.22

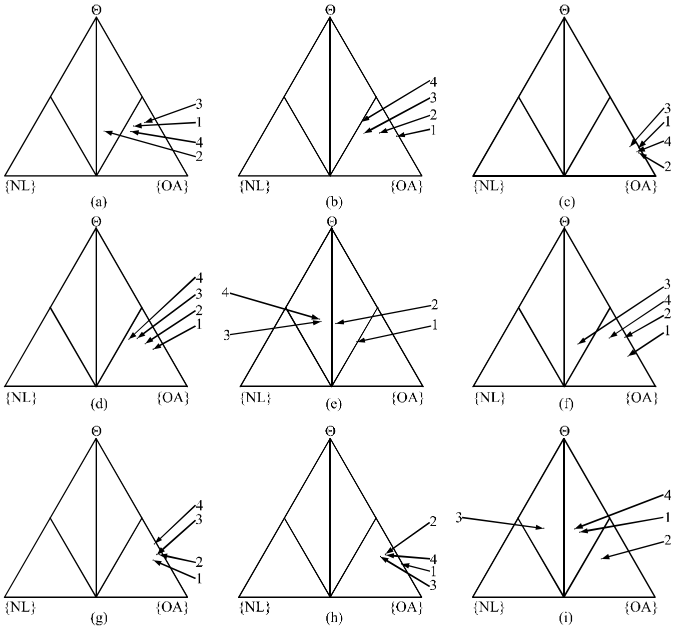

Figure 3. Simplex plots showing recovery following TKR surgery for patient (a) P1 (b) P2 (c) P3 (d) P4 (e) P5 (f) P6 (g) P7 (h) P8 and (i) P9. Numerals indicate visit: 1) pre-op, 2) 3months post-op, 3) 6 months post-op, 4) 12 months post-op.

The results of the DST-based classifier were compared to the results of the KOS [9]. Overall, there was no significant correlation between the BOE_c values and the KOS scores. In the only other study that attempted to relate gait analysis parameters to patient-related scoring systems, no correlation was found between the two measures of outcome [15]. The results raise the question of whether subjective outcome scores should be used to evaluate the outcome of TKR surgery.

C. Feature selection

A study was conducted to investigate the feasibility of using automatic feature selection in conjunction with the DST-based classifier. Three feature selection methods, Stepwise Linear Discriminant Analysis (SLDA), Sequential Selection Methods (SSM) and Genetic Algorithms (GA), are compared in order to address the question of whether it is best to use an automated approach or to rely on expert clinical opinion when choosing the input variables that should be used in the classification of OA and NL knee function. The in-sample and out-of-sample classification accuracy results of the three DST-based classifiers using feature selection are compared to the classifier that does not use feature selection in Table III.

TABLE III. COMPARISON OF THE PERFORMANCE OF THREE DST-BASED CLASSIFIERS USING FEATURE SELECTION METHODS WITH THE DST-BASED CLASSIFIER USING NO FEATURE SELECTION. SUBSETS OF VARIABLES SELECTED BY THE DIFFERENT FEATURE SELECTION METHODS.

Feature Selection Method	Accuracy, %		Subset of variables
	In-sample	*Out-of-sample*	
SLDA	99.01	100.00	$\{v_2, v_8, v_{15}, v_{17}\}$
SSM	100.00	100.00	$\{v_2, v_5, v_6, v_7, v_8, v_{10}, v_{11}, v_{13}, v_{14}, v_{16}, v_{17}\}$
GA	100.00	100.00	$\{v_2, v_5, v_6, v_7, v_8, v_{11}, v_{13}, v_{15}, v_{16}, v_{17}\}$
None	97.62	97.62	n/a

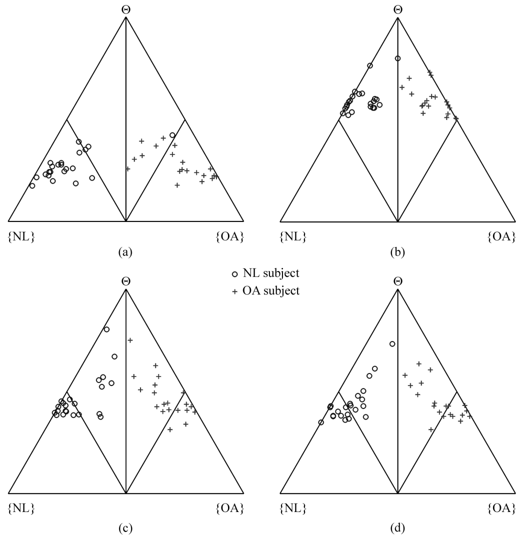

Figure 4. Simplex plots showing the simplex coordinates of the BOE_cs for the out-of-sample subjects for (a) DST-based classifier with no feature selection, (b) DST-based classifier with SLDA, (c) DST-based classifier with SSM and (d) DST-based classifier with GA.

Using feature selection produces a superior classifier in terms of the in-sample and out-of-sample classification accuracy. In each case, the use of feature selection reduced the number of input variables that were used as inputs to the DST-based classifier as shown in Table III. However, this decrease in the number of variables is also responsible for the increase in the level of uncertainty associated with the classification of subjects as shown in Figure 4. The use of the simplex plot highlights that although obtaining a high classification accuracy is of first importance, it is necessary to ensure that the method is still relevant and of practical use. In this study, the use of feature selection has meant that despite obtaining a highly accurate classifier, the lack of dominant OA and NL classifications signify that the simplex plot method is redundant and clinically irrelevant. The use of different feature selection methods can lead to different sets of “best” input variables. All of the feature selection methods identified common input variables as being important: v_2, v_8 and v_{17}, (cadence, PC score related to the vertical force during loading response and the anterior-posterior knee width). These variables have been reported to be clinically relevant [16-17].

D. Comparison with other methods

A study was conducted in order to compare the performance of the DST-based classifier with two other classifiers namely an artificial neural network (ANN) and linear discriminant analysis (LDA). The in-sample and out-of-sample accuracy results for each classifier are summarised in Table IV. Comparing the DST-based classifier and the ANN classifier, the results show that the DST-based classifier has a higher in-sample and out-of-sample accuracy than the ANN classifier. The DST-based classifier has a lower in-sample classification accuracy than the LDA classifier but a higher out-of-sample accuracy. These results suggest that the DST-based classifier is superior to both the linear LDA classifier and the non-linear ANN classifier in terms of out-of-sample classification accuracy (prediction). The five most important variables (in terms of influence on the classification) identified by the three classifiers are given in Table IV. Here a ranking of 1 is given to the most important variable. The DST and ANN classifiers identified similar variables as being important. These variables are often cited as being clinically relevant [16-17].

IV. CONCLUSIONS

The initial results of this application have demonstrated a logical, practical and visual approach that can be used to differentiate between the characteristics of NL and OA knee function and to diagnose the extent to which a patient has recovered after TKR surgery. Initial studies using this technique have provided encouraging results in terms of accuracy, validity and clinical relevance.

This objective tool may have significant clinical value as it could provide useful information on pre-operative disease progression; the effectiveness of surgical and therapeutic intervention; and on the functional analysis of joint prosthesis design. With an improvement to the clinical assessment process for common diseases, surgery to relieve the painful and functionally disabling symptoms could be more effectively tailored to suit patients. The tool could provide a powerful prediction of the extent to which a patient presenting a distinct set of pre-operative symptoms would respond to various treatment options.

The DST-based classifier is a generic method and, as such, it is applicable to a wide range of classification and predictive problems. The application of the DST-based classifier to other biomechanical and clinical problems should be investigated, e.g., identification of ankle, spine and hip pathology. Additionally, it is suggested that the method be used for the assessment of the relative merits of different treatment options, devices and surgery, e.g., identification of the differences between rotating platform and fixed bearing knee implants.

ACKNOWLEDGMENT

The authors would like to thank Mr. Richard Evans and Mr. Chris Wilson, Orthopaedic Consultants, UHW, Cardiff; Mr. Stuart Roy, Orthopaedic Consultant, Royal Glamorgan Hospital, Pontypridd; Prof. Colin Dent, UWCM, Cardiff; Medical Research Council; Depuy Johnson and Johnson.

TABLE IV. COMPARISON OF THE IN-SAMPLE AND OUT-OF-SAMPLE ACCURACY, %, AND FIVE MOST IMPORTANT VARIABLES (RANK 1 TO 5 WITH RANK 1 BEING MOST IMPORTANT) OF THE DST, ANN AND LDA CLASSIFIERS

Classifier	**Accuracy, %**		**Variable Rank**				
	In-sample	***Out-of-sample***	***1***	***2***	***3***	***4***	***5***
DST	97.62	97.62	v_8	v_2	v_4	v_{11}	v_{15}
ANN	77.82	63.89	v_8	v_2	v_4	v_{11}	v_{17}
LDA	100	95.24	v_{16}	v_1	v_2	v_3	v_9

REFERENCES

[1] M.W. Whittle, 1996. Gait analysis: an introduction. 2nd Edition. Oxford; Boston: Butterworth-Heinemann.

[2] G. Barton, A. Lees, P. Lisboa and S. Attfield, “Visualisation of gait data with Kohonen self-organising neural maps,” Gait Posture, in press.

[3] H.G. Watts, “Gait laboratory analysis for preoperative decision making in spastic cerebral palsy: is it all it’s cracked up to be?” J. Pediatr. Orthop., vol. 14, pp.703-704, 1994.

[4] L. Jones, “The development of a novel method for the classification of normal and osteoarthritic knee function,” PhD Thesis, Cardiff University, 2004.

[5] L. Jones, M.J. Beynon, C.A.Holt, S. Roy, “An application of the Dempster–Shafer theory of evidence to the classification of knee function and detection of improvement due to total knee replacement surgery,” J. Biomech, in press.

[6] M.J. Beynon, L. Jones, C.A. Holt, “Classification of osteoarthritic and normal knee function using three-dimensional motion analysis and the Dempster-Shafer theory of evidence,” IEEE Trans. Syst. Man Cybern. A., 36, 1, pp.173-186, 2006.

[7] S.Wolf et al., “Automated feature assessment in instrumented gait analysis,” Gait Posture, in press.

[8] C.A. Holt, N.J. Hayes, R.W.M. van Deursen and P.T. O’Callaghan, “Three-dimensional analysis of the tibiofemoral joint using external marker clusters and the JCS approach – Comparison of normal and osteoarthritic knee function,” *Computer Methods in Biomechanics and Biomedical Engineering 3*. Lisbon. Gordon and Breach Science Publishers SA. pp.289-294, 2000.

[9] J.J. Irrgang, L. Synder-Mackler, R.S. Wainner, F.H. Fu and C.D. Harner, “Development of a patient-reported measure of function of the knee,” JBJS, 80-A, 8, pp.1132-1145, 1998.

[10] K.J. Deluzio, U.P. Wyss, B. Zee, P.A. Costigan and C. Sorbie, “Principal component models of knee kinematics and kinetics: normal vs. pathological gait patterns,” Hum. Mov. Sci., 16, pp.201-217, 1997.

[11] S. M. Weiss and C. A. Kulikowski, “Computer Systems That Learn: Classification and Prediction Methods From Statistics, Neural Nets, Machine Learning, and Expert Systems,” San Mateo, CA: Morgan Kaufmann, 1991.

[12] M.G. Benedetti, F. Catani, T.W. Bilotta, M. Marcacci, E. Mariani and S. Giannini, “Muscle activation pattern and gait biomechanics after total knee replacement,” Clin. Biomech., 18, pp.871-876, 2003.

[13] T.P. Andriacchi, “Functional analysis of pre and post-knee surgery: total knee arthroplasty and ACL reconstruction,” J. Biomech. Eng.-Trans. ASME, 115, pp.575-581, 1993.

[14] C.M. Myles, P.J. Rowe, C.R.C. Walker and R.W. Nutton, “Knee joint functional range of movement prior to and following total knee arthroplasty measured using flexible electrogoniometry,” Gait Posture, 16, pp.46-54, 2002.

[15] S. Fuchs, M. Flören, A. Skwara and C.O. Tibesku, “Quantitative gait analysis in unconstrained total knee arthroplasty patients,” Int. J. Rehab Res., 25, pp.65-70, 2002.

[16] H. Gök, S. Ergin, and G. Yavuzer, “Kinetic and kinematic characteristics of gait in patients with medial knee arthrosis,” Acta Orthop Scand, 73, 6, pp.647-652, 2002.

[17] E. Schneider and E.Y. Chao, “Fourier analysis of ground reaction forces in normals and patients with knee joint disease,” J.Biomech, 16, 8, pp.591-601, 1983.