Disparity in Frontal Lobe Connectivity on a Complex Bimanual Motor Task Aids Classification of Operator Skill Level

Javier Andreu-Perez, PhD
The Hamlyn Centre,
Imperial College London,
United Kingdom
javier.andreu@imperial.ac.uk

Daniel Richard Leff, PhD, FRCS
The Hamlyn Centre,
Imperial College London,
United Kingdom
d.leff@imperial.ac.uk

Kunal Shetty, MRCS
The Hamlyn Centre,
Imperial College London,
United Kingdom
k.shetty@imperial.ac.uk

Ara Darzi, FRCS, FACS, FMedSci
Department of Surgery and Cancer,
Imperial College London,
United Kingdom
a.darzi@imperial.ac.uk

Guang-Zhong Yang, PhD, FIEEE, FIET, FAIMBE, FMICCAI, FREng
The Hamlyn Centre,
Imperial College London,
United Kingdom
g.z.yang@imperial.ac.uk

Address for correspondence:
The Hamlyn Centre
Imperial College London,
Level 4, Bessemer Building, South Kensington Campus,
London, SW7 2AZ, United Kingdom
Phone: +447771425969

Keywords

Functional Near-infrared Spectroscopy (fNIRS), Optical Topography, Functional Connectivity, Motor Learning, Operator skill level, Technical skill levels assessment

Abstract

Objective metrics of technical performance (e.g. dexterity, time and path length) are insufficient to fully characterize operator skill level, which may be encoded deep within neural function. Unlike reports that capture plasticity across days or weeks, this paper studies long-term plasticity in functional connectivity that occurs over years of professional task practice. Optical neuroimaging data are acquired from professional surgeons of varying experience on a complex bimanual co-ordination task with the aim of investigating learning-related disparity in frontal lobe functional connectivity that arises as a consequence of motor skill level. The results suggest that prefrontal and premotor seed connectivity is more critical during naïve versus expert performance. Given learning-related differences in connectivity, a least-squares support vector machine with a radial basis function kernel is employed to evaluate skill level using connectivity data. The results demonstrate discrimination of operator skill level with accuracy ≥ 0.82 and Multiclass Mathew's Correlation Coefficient ≥ 0.70. Furthermore, these indices are improved when local (i.e. within-region) rather than inter-regional (i.e. between-region) frontal connectivity is considered ($p=0.002$). The results suggest that it is possible to classify operator skill level with good accuracy from functional connectivity data, upon which objective assessment and neuro-feedback may be used to improve operator performance during technical skill training.

.

Acronyms List

PFC	Prefrontal cortex	**HR**	Heart rate
SMA	Supplementary motor area	**ML**	Machine learning
PMC	Premotor cortex	**LS-SVM**	Least-squares support vector machine
M1	Primary motor	**OVA**	One-vs-all
ROI	Region of interest	**OVO**	One-vs-one
OT	Optical topography	**MMCC**	Multiclass Mathew's correlation coefficient
$HB0_2$	Oxygenated hemoglobin	**TP**	True positive
HHb	Deoxygenated hemoglobin	**TN**	True negative
MIS	Minimally invasive surgery	**FP**	False positive
LS	Laparoscopic suturing	**FN**	False negative
FLS	Fundaments of laparoscopic suturing	**KW**	Kruskal wallis
NHS	National Health Services	**DT**	Dunn's Test
MHR	Mean heart rate	**Tuckey's HSD**	Tukey's honest significance differences
MHRV	Mean heart rate variability	**SW**	Shapiro-Wilk's test

1 Introduction

Neuroergonomics captures the study of brain behaviour at work with the goal of improving performance, safety and efficiency (Parasuraman et al., 2012; Parasuraman and Rizzo, 2003). This is particularly relevant to fields that warrant operator vigilance, technical skill levels and decision-making such as command control in aviation (Izzetoglu et al., 2004) and surgery (James et al., 2010). Of these, surgery is unique in that the operator's technical and cognitive skill levels have a direct impact on patient safety. In theory, enhanced neuronal efficiency in surgeons acquired through practice and training or through improved ergonomics may manifest as improved patient safety. Deepening our understanding of operator brain behaviour may have value when applied to improve performance, particularly in high-risk procedures. Advances in functional neuroimaging technology have made it possible to monitor operators in more realistic settings and track evolution in brain behaviour that accompanies motor skill levels learning. To this end, there has been an increasing number of research studies focused on studying evoked cortical response to complex motor behaviour in the context of open and minimally invasive surgery (MIS) (Bahrami et al., 2011; Duty et al., 2012; Leff et al., 2007a; Leff et al., 2008a; Leff et al., 2008b; Ohuchida et al., 2009; Zhu et al., 2011). In this study we used functional near-infrared spectroscopy (fNIRS), which has raised increasing interest in recent years for performing less constricted, hence more naturalistic neuroscience experiments (Rodrigo et al., 2014).

Previous studies have exposed skill level-related differences in brain behaviour and longitudinal changes in cortical excitation in line with technical skill levels acquisition (Leff et al., 2007a; Leff et al., 2007b, Leff et al., 2011). For example, data suggest that cortical responses may be skill level-dependent (Leff et al., 2008b). Greater activation within executive control centres such as the prefrontal cortex (PFC) has been observed in novices and PFC excitation appears to attenuate following practice (Leff et al., 2008a). In a study with 18 subjects, decreasing ratios of oxygenated haemoglobin were observed in supplementary motor area (SMA) and preSMA when learning a motor skill (Hatenaka et al., 2007). Similarly, a decrease in cortical activation of the sensorimotor cortex was reported during the learning of a multi-joint discrete motor task (Ikegami and Taga, 2008). This is commensurate with evidence that re-organisation of brain function (i.e. neuroplasticity) accompanies motor skill levels learning, such that operator skill level may be best reflected in the magnitude of regional brain excitation or shifts in activation foci (Draganski and May, 2008; Halsband and Lange, 2006; Kelly and Garavan, 2005). This notwithstanding, investigations of highly complex motor skill levels such as MIS, have failed to demonstrate differences between naïve subjects and expert operators (Ohuchida et al., 2009). One theory is that for highly complex motor tasks such as MIS that require 2D to 3D perceptual transformation and precise inter-manual co-ordination, skill *level*-related disparity may manifest as differences in frontal lobe connectivity rather than changes in activation per se. This can be deduced from the cortical network differences reported in Sun et al. (2007) across early and late motor learning states, as well as the variations in cortical connectivity presented in James et al. (2013) during skill acquisition for a surgical task.

Whilst longitudinal changes in network topologies have previously been studied in surgical learners, cortical interactions have only been tracked over days (James et al., 2013) and not months or years. In this study from our group (James et al., 2013), changes in the network cost across days of the practice were found amongst subjects of similar skill level. One advantage of evaluations that incorporate master operators is that they facilitate interrogation of motor plasticity as a result of repeated practice of a technical skill level over many years. Indeed, few studies have investigated learning-related changes in connectivity across such timescales, and current reports are restricted to minutes or weeks (Coynel et al., 2010; Dayan and Cohen, 2011; Heitger et al., 2013; Heitger et al., 2012; Sun et al., 2007). Nevertheless, existing literature seems to suggest greater connectivity between frontal and cortical motor regions in ‘early’ versus ‘late’ learning (Sun et al., 2007) and longitudinal attenuation in functional integration in motor-related networks following extended practice (Coynel et al., 2010). Therefore, conceivably functional connectivity between associative and premotor regions may be expected to decrease in line with continued practice and increasing operator skill level.

Differences in functional connectivity may help discriminate operators based on their skill level. Classification of operator proficiency based on brain behaviour may prove invaluable for objective assessment of technical skill levels, evaluation of trainee progress, and if “interfaced” to the operator or team, may improve performance or aid patient safety through cognitive biofeedback (Nan et al., 2012; Wang and Hsieh, 2013; Zoefel et al., 2011). Whilst amplitude of the evoked response has been used for classification of operator states previously (Coyle et al., 2007; Naito et al., 2007; Power et al., 2012), there have been no such reports in surgeons. Indeed, whilst differences in amplitude signal change in executive control and motor cortical regions related to surgical skill level have been previously observed (Leff et al., 2008b; Ohuchida et al., 2009), it has not been possible to discriminate operators' performance based solely on these signal characteristics. Classification of operator proficiency based on functional connectivity data represents a non-trivial high dimensionality problem for which conventional statistics are ill-posed to solve. Machine learning (ML) techniques such as least-squares support vector machine (LS-SVM) generate predictive models that can discriminate data based on the complexity of multivariate relationships such as those arising from functional connectivity.

In this study, we aim to investigate and discriminate operator skill level during laparoscopic (key-hole) surgical manoeuvres using brain connectivity derived from evoked optical imaging responses. Specifically, we employ Optical Topography (OT) to monitor operator brain function during a highly complex surgically relevant motor task (i.e. key-hole surgical suturing). We hypothesize that frontal lobe connectivity in associative (prefrontal) and premotor seed regions will decrease in line with increasing operator experience. Functional connectivity computed from one-second epochs of filtered oxygenated and deoxygenated haemoglobin signals (HbO_2 and HHb, respectively) was used to classify operator skill level. Additionally, an analysis of the discriminatory performance of local (within-region) and inter-regional connectivity was performed. To facilitate this comparison, a single metric of performance named Multiclass Mathew's Correlation Coefficient (MMCC) (Jurman et al., 2012) was employed. The results demonstrate that ML-based discrimination of operator skill level is feasible and accurate, upon which educators may capitalise in the form of neural feedback training.

2 Materials and methods

2.1 Experimental set-up and neuroimaging data acquisition

Local Research Ethical Council approval was obtained (05/Q0403/142). Participants were screened for handedness (Oldfield, 1971), gender and neuropsychiatric illness. Participants abstained from consumption of alcohol and caffeine for 24 hours before the study date (Orihuela-Espina et al., 2010). Thirty-two right-handed male surgeons were recruited from the National Health Service (NHS) and Imperial College London to perform a complex visual-spatial task, namely simulated laparoscopic suturing (LS) (i.e. key-hole surgical stitching), in a box trainer (i-SIM, iSurgicals, UK). The cohort included 12 novices (with no prior experience of MIS, mean age 22.4 ± 1.6 years), 11 trainees (limited MIS experience of less than 50 cases involving LS, mean age 33.8 ± 2.9 years) and 9 expert consultants (minimum of 50 independent cases requiring LS, mean age 42.7 ± 3.6 years). Participants were required to perform the task three times as per the fundamentals of laparoscopic surgery curriculum (FLS) (FLS, 2015). Statistical analysis on the FLS was performed using Kruskal Wallis (KW), followed by Dunn's Test (DT) for multi-comparisons. During each of the three sessions, the surgical procedure was segregated sequentially into three subtasks, namely 'needle insertion', 'double-throw knot tying' and 'single-throw knot tying' as highlighted in Figure 1.

[INSERT FIGURE 1 HERE]

Prior to recording, the participants were allowed a brief familiarisation session (15 minutes). The experimental design consisted of a sequence of continuous episodes of baseline motor rest (30 seconds), subtask phases during which given LS manoeuvres were executed (variable time i.e. self-paced) and inter-trial recovery periods (40 seconds). During the rest episodes the participants were asked to remain still and regard the centre of the monitor. Observable technical performance in LS was assessed by adoption of the FLS score [Score = 600 – (time in seconds) – (penalties $\times$10)]. Penalties were measured as per needle entry and exit distance points from pre-marked points (mms) as well as gap between the edges of wound (mms), i.e. an objective quantitative assessment. The derived score based on time and accuracy of performance was compared across experience groups using statistical tests of significance. Cortical hemodynamic data were recorded at 10 Hz using a 44-channel OT system (ETG-4000, Hitachi Medical Corp, Japan). To capture functional behaviour relating to motor skill levels acquisition, channels were positioned over the PFC, premotor cortex (PMC), SMA and motor cortex (M1) regions of the brain according to landmarks of the International 10-10 system (Jurcak et al., 2007) as illustrated in Figure 2. Optode positions were measured using a 3D digitizer. A stand-alone registration method was used to project NIRS probe positions into a MNI (Montreal Neurological Institute) coordinate space (Tsuzuki et al., 2012). As OT data may be influenced by changes in the systemic circulation (Elwell et al., 1994; Obrig et al., 2000), a portable electrocardiogram sensor was attached to the chest of each subject to continuously monitor the heart rate (HR) during the course of the experiment. Mean heart rate (MHR) and heart rate variability (MHRV) was extracted from the electrocardiogram sensor readings using R-R wave intervals. Statistical significance was computed using analysis of variance ANOVA over these measurements.

[INSERT FIGURE 2 HERE]

2.2 De-noising and filtering

For each channel of data, an average baseline was computed considering 10s of data prior to each task onset to allow for haemodynamic normalisation after task offset. During this period, subjects were asked to refrain from any motion prior to task onset and not to engage in deliberate cognitive work. For each subtask, the first 6s of data were discarded, thereby allowing for the temporal delay between task onset and cortical haemodynamic change. At each channel, changes in the concentration of HbO_2 and HHb were reconstructed from variations in light attenuation using the modified Beer-Lambert law (Delpy et al., 1988). To reduce systemic interference, the haemodynamic data were low-pass filtered, detrended for eliminating system drift and integrity checked for eliminating noisy channels using Imperial College Neuroimage Analysis (ICNA) software (Orihuela-Espina et al., 2010).

2.3 Between-group differences in functional connectivity

The methodology for computing the between-group differences consists of the following sequential sets:

1) For every sample of haemoglobin data, the difference between the current value and the average baseline is computed. The resultant metrics are denoted by ΔHbO_2 or ΔHHb.
2) ΔHbO_2 and ΔHHb from channels belonging to the same region of interest (ROI) are grouped and averaged, resulting in $\mu_{\mathrm{HbO_2}}$ and μ_{HHb} for each ROI.
3) To obtain a single metric of inter-connectivity, the Rv coefficients (Josse et al., 2008; Abdi, 2007) are calculated between the pairs of $\mu_{HbO_2}$and μ_{HHb} from two ROIs. Rv coefficients form a metric of multivariate statistical relationship to measure the dissimilarity between two sets of data. It is formulated as follows:

$$r = \frac{trace(\alpha\,\alpha^T\,\beta\,\beta^T)}{\sqrt{trace((\alpha\,\alpha^T)^2)\times trace((\beta\,\beta^T)^2)}}; \qquad \begin{array}{l} \alpha = \left[\mu_{HbO_2}^{\mathrm{ROI}^1}, \mu_{HHb}^{\mathrm{ROI}^1}\right] \\ \beta = \left[\mu_{HbO_2}^{\mathrm{ROI}^2}, \mu_{HHb}^{\mathrm{ROI}^2}\right] \end{array} \tag{1}$$

where we denote Rv coefficients with the function *r* taking as inputs the mean-centered α and β matrices; and the indexes ROI^1 and ROI^2 represent two distinct regions of interest. Once computed, Rv coefficients return values between 0 and 1 that can be interpreted as an approximation of the squared Pearson correlation coefficient (Josse et al., 2008; Abdi, 2007).

4) Fisher transformation is applied to the Rv coefficients yielding the Gaussian distributed z scores. In order to obtain a single informational value, the z scores are averaged and the inverse Fisher transformation is applied, transforming the average back to its original Rv coefficient. The resultant average value is then considered as the inter-regional connectivity between two brain areas of a subject performing a given LS trial (session).

Statistical analysis of normalized z-scores for inter-regional connectivity was performed using the ANOVA test, followed by post-hoc analysis using Tukey's honest significance difference (HSD).

2.4 Machine learning from functional connectivity for operator skill level discrimination

2.4.1 Functional connectivity datasets for automated discrimination

Two sub-datasets with different granularity are studied to classify operator skill level from fNIRS data: the *session based networks* and *time-course based networks*. The former aims to classify operators based on data from an entire session while the later from every one-second time epoch within a session. Both are derived from the ΔHbO_2 and ΔHHb readings of the original dataset captured at 10 Hz.

a) *Session based networks:* Rv coefficient is a scalar that determines the relationship of the joined signals, ΔHbO_2 and ΔHHb, between two channels during a session (i.e. the duration of one trial). Each correlation matrix represents a session. Each element within the matrix corresponds to a scalar that represents the 2D relationship between two channels when jointly considering both signals of interest, namely ΔHbO_2 and ΔHHb. The resulting dimensionality of the correlation matrix is 44×44, i.e. one element for each channel, out of which 22 are on the PFC, 10 are on the SMA, 8 are on the PMC, and 4 on the M1.

b) *Time-course based networks*: Due to the averaging used to construct an exemplar network across a whole session, some latent discriminatory information might be omitted. A short period connectivity network can be constructed using Spearman's correlations between channels, this time considering ΔHbO_2 and ΔHHb separately. To increase the granularity, correlations are computed within one-second epochs across the whole session. Hence, each correlation matrix corresponds to a single epoch and each element within the matrix is a single correlation of either ΔHbO_2 or ΔHHb between two channels. As a result, two 44x44 correlation matrices are generated, one for each haemoglobin species.

2.4.2 Classifier and parameter settings

For the analysis of the functional connectivity datasets, a least-squares support vector machine (LS-SVM) with a non-linear radial basis function kernel is used. In LS-SVMs the parameters of the separating hyperplane are formulated as a closed-form linear system of equations (Gaonkar and Davatzikos, 2013; Suykens et al., 2002). The entire data are first divided into 5 cross-validation subsets and one is held out for testing. Before each test, optimization of the L2 internal regularization parameter is performed over a 5-fold cross-validation process using only the training set. This prevents over-fitting over the presence of irrelevant features. The optimization algorithm used is simulated annealing (Press et al., 2007).

The present study is related to a multiclass problem (with $K > 2$ as the number of classes) since we aim to sub-classify operator skill level into one of three groups (novices, trainees and experts). Several binary LS-SVM models are trained and their outputs are later combined using a voting scheme. These individual binary LS-SVM models can be designed in two ways: (1) one class against the rest or one-versus-all (OVA), which requires training K large models; (2) one-versus-one (OVO), where one binary LS-SVM is trained for each pairwise class and consists of $K(K-1)/2$ simpler models (via a directed acyclic graph). Both methods can lead to similar results although they can be used for different purposes (Galar et al., 2011). In Section 3.3 the OVO method is used, while in Section 3.4 both methods are employed. Training and testing sets were separated via randomized 5-fold cross-validation to ensure generalization to independent sets.

2.4.3 Performance measures for classification

To evaluate the performance of the classifier for each of the two sub-datasets defined in 2.4.1, standard metrics are computed from the true/false positive/negative (i.e. TP, TN, FP, FN) frequencies. These metrics of performance are:

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN}; \;\; precision = \frac{TP}{TP + FP}$$
$$sensitivity = \frac{TP}{TP + FN}; \;\; specificity = \frac{TN}{TN + FP} \tag{2}$$

Considering the metrics presented in (2), every class would lead to an independent metric of performance, which is not desirable for assessing the overall performance. Instead, a practical metric of classification performance is the Mathews' correlation coefficient (MCC), which considers all TP, TN, FP and FN simultaneously. For binary class problems the MCC is formulated as:

$$\begin{matrix} MCC \\ binary \end{matrix} = \frac{(TP * TN - FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}; \tag{3}$$

MCC can be re-formulated to be multiclass and define a global performance metric representative of all classes (Gorodkin, 2004; Jurman et al., 2012). In order to make it multiclass, the whole multiclass confusion matrix can be used as the reference to compute the MCC statistics. The multiclass MCC (MMCC) is obtained from marginalizing the different dimensions of the confusion matrix C, which elements include all true/false positive/negative frequencies for each class. Each row of C represents a class prediction while each column accounts for the true class. Hence, formula (3) becomes:

$$\begin{matrix} MMCC \\ multiclass \end{matrix} = \frac{\sum_{q,e,p} C_{q,q} * C_{e,p} - C_{q,e} * C_{p,q}}{\sqrt{\sum_q \left(\sum_e C_{e,q}\right)\left(\sum_{e',q' \neq q} C_{e',q'}\right)} \sqrt{\sum_q \left(\sum_e C_{q,e}\right)\left(\sum_{e',q' \neq q} C_{q',e'}\right)}} \tag{4}$$

Where $C_{(.)(.)}$ is a single element of the confusion matrix and all matrix indexes (q, e, p) start with 1 and finish with K, unless when specified. Formula (4) generalises MCC and can be used either for multiclass or binary problems. The value obtained with MMCC is always between -1 and 1. If the coefficient is -1, the MMCC portrays a total disagreement between predictions and the ground truth, whereas 0 indicates a random prediction and 1 means a ground truth match.

2.4.4 Predicting operator skill level

After training the classifier using *a priori* data, operator skill level can then be predicted for any new, unseen (i.e. unlabelled) set of fNIRS data. Figure 3 summarises the processing steps necessary to obtain these predictions:

1) A signal processing step is composed of the following tasks:
    a. Signal extraction: 1Hz of haemoglobin data captured by each OT channel is stored in an independent buffer.
    b. Signal filtering (section 2.2): Each signal in a buffer is denoised, smoothed and baseline-subtracted.
2) Functional connectivity: coefficients are computed as a proxy for functional connectivity for all between-channel signal pairs, thus generating a functional connectivity matrix. The elements of the resulting matrix are translated into a sample vector, in which each vector dimension corresponds to an element of the matrix. In the case where two functional connectivity matrices are generated (one for each haemoglobin species), the elements of both matrices are linked together within a same sample vector.
3) Skill level estimation: In advance of testing, a LS-SVM algorithm is trained and parameter optimization is performed as described in 2.4.2. The sample vector obtained in step 3 is tested against a support vector machine model by using one of the multiclass settings described in 2.4.2. New data samples are analysed and a prediction is made regarding the likely skill level.

[INSERT FIGURE 3 HERE]

2.5 Evaluating discrimination capabilities by groups of connectivity

It is valuable to explore the relative importance of each brain area, by considering the connectivity within and between brain areas in the frontal lobe, as opposed to solely considering classifier performance based on the fully connected adjacency matrix. The extended dataset is used in this analysis. In this regard, the classifier is adapted to consider only the within-region (local) or between-region (inter-regional) connectivity for a given brain region of interest (ROI). As illustrated in Figure 4, for a given ROI the local, inter-regional or combined functional connectivity data may be considered and defined as:

a) *Local connectivity (within-region)*: Considering observed connections within a given brain region. In other words, it refers to the internal functional connectivity of a particular brain region without considering long-range connections to other brain regions.
b) *Inter-regional connectivity:* Involves observed connections between channels of different brain areas, capturing distant functional relationships between areas of the brain, pruning local connections within the region.
c) *Combined (within-region and inter-regional)*: Includes all possible functional connections, both within a given ROI and between that ROI and other brain regions, thereby emphasizing all possible functional relations exhibited by a specific area.

Two different classifier strategies are used in this study to analyse the performance of each classification for each sub-region: (1) an overall multiclass classification following a one-vs-one strategy, (2) an independent binary skill level classification based on a one-vs-all strategy.

[INSERT FIGURE 4 HERE]

3 Results

3.1 Behavioural performance and cognitive load

Table 1 summaries the results of task-related change in MHR, MHRV and assessment of motor laparoscopic skill levels (FLS). As anticipated, technical performance was significantly different between groups ($p<0.001$, Chi-square=58, df=2, KW). Experts displayed superior performance (lower FLS scores) compared to trainees ($p<0.001$, $z=-32.2$, DT) who in turn outperformed novices ($p=0.002$, $z=-22.8$, DT). No statistically significant differences in task-related MHR ($p=0.87$, F-value=0.145, df=2; ANOVA) and MHRV ($p=0.83$, F-value=0.182, df=2; ANOVA) were found between groups. For all groups, the null hypothesis is retained under Shapiro-Wilk test of normality for HRV (Novices: $p=0.42$, Trainees: $p=0.42$, Experts: $p=0.50$) and MHRV (Novices: $p=0.61$, Trainees: $p=0.21$, Experts: $p=0.33$).

[INSERT TABLE 1 HERE]

3.2 Learning-related changes in local and inter-regional connectivity

For all groups, the null hypothesis is retained under Shapiro-Wilk's (SW) test of normality (Novices: p=0.31, Trainees: p=0.11, Experts: 0.09).

3.2.1 Needle insertion

Frontal lobe connectivity during needle insertion varied according to surgical skill level (Figure 5). Correlations between PFC and PMC seed regions and other frontal motor cortical regions were observed to be lower in trainees and experts than novices. Overall, a significant main effect of skill level was observed for the interactions PFC-SMA (p<0.01, F-value=20.06, df=2; ANOVA), PFC-PMC (p<0.01, F-value=16.05, df=2; ANOVA), SMA-PMC (p<0.01, F-value=36.67, df=2; ANOVA). Upon post-hoc analysis no significant differences between trainees and experts were observed (PFC-SMA: p=0.56, PFC-PMC: p=0.97, SMA-PMC: p=0.09; Tukey's HSD). However, comparisons between novices and trainees (PFC-SMA: p<0.01, PFC-PMC: p<0.01, SMA-PMC: p<0.01; Tukey's HSD), and novices and experts were statistically significant (PFC-SMA: p<0.01, PFC-PMC: p<0.01; SMA-PMC: p<0.01; Tukey's HSD). No main effect of skill level was observed for M1-PFC or M1-SMA seed interactions. However, a main effect of skill level was observed for M1-PMC interactions (p=0.01; ANOVA). Post-hoc analysis revealed no significant difference between trainees and experts while, differences between novices and trainees (p=0.01; Tukey's HSD) and novices and experts (p=0.03; Tukey's HSD) reached statistical threshold.

3.2.2 Double-throw knot tying

The strength of PFC-PMC interactions was not observed to depend on skill level (Figure 5) for double-throw knot tying. No significant differences were found for PFC seed interactions. However, interactions between the PMC and other frontal lobe seed regions varied with skill level. For example, a main effect of skill level was observed for interactions between the SMA and PMC ($p<0.01$, F-value=8.574, df=2; ANOVA). The interaction strength of this connection was significantly lower in experts than novices (SMA-PMC: $p<0.01$; Tukey's HSD). Of all frontal lobe interactions, differences between trainees and experts were only observed in SMA-PMC, the interaction strength being significantly lower in experts (SMA-PMC: $p=0.04$; Tukey's HSD).

3.2.3 Single-throw knot tying

A significant main effect of skill level was observed for the interactions PFC-PMC ($p<0.01$, F-value= 11, df= 2; ANOVA), PFC-SMA ($p=0.03$, F-value= 4.14, df= 2; ANOVA) and SMA-M1 ($p=0.01$, F-value=6.49, df=2; ANOVA). For PFC seed interactions, the strength of association was observed to be significantly lower in experts than novices (PFC-SMA: $p=0.03$; PFC-PMC, $p<0.01$; Tukey's HSD) and also lower for trainees between PFC-PMC ($p=0.03$, Tukey's HSD) (Figure 5). Interestingly, differences between experts and trainees were only evident in SMA-M1 interactions, the strength of association being significantly greater in experts (SMA-M1: $p=0.01$; Tukey's HSD).

[INSERT FIGURE 5 HERE]

3.3 Classification of operator skill level

Tables 2 and 3 show the results of the classifier using the "session based networks'' (Table 2) and the 'time-course based networks' (Table 3).

[INSERT TABLE 2 HERE]

[INSERT TABLE 3 HERE]

Classification using *session based networks* yields a greater number of true positives than false positives in all subtasks, with low MMCC scores (<0.5) for certain trials. In contrast, the *time-course based networks* consistently achieves high MMCC scores (≥ 0.70) across subtasks. For this latter case, the classifier appears to perform equally well in each LS subtask. Overall, novice identification was highly precise (≥ 0.8 for all three subtasks) and sensitive (0.92 for needle insertion, 0.85 for double-throw knot and 0.91 for single-throw knot) whilst trainee and expert identification was precise and specific (e.g. precision for trainees ≥ 0.83 and specificity ≥ 0.86 across all subtasks). The discrimination of experts also exhibits high specificity (≥ 0.95) and precision (≥ 0.76), but with lower sensitivity versus the other operator groups (0.63-0.67). These results suggest that the variability in cortical haemodynamic responses between groups based on experience is potentially greater than the variability in responses across subtasks within a given experience group.

3.4 Discrimination capability: local versus inter-regional connectivity

Performance results for each brain area and connectivity type are presented in Table 4 for the OVO multiclass setting which summarises the overall classification over the entire set of categories, i.e. a combination of three one-vs-one classifiers are used to determine the class of each sample (multiclass approach). Tables 5, 6, and 7 display results from the OVA setting which summarises the classification of operator skill level, i.e. three one-vs-all classifiers are used as three separate classifiers (independent binary classification of a determined skill level against the rest). MMCC data were analysed using the Friedman test followed by post-hoc analysis using the Wilcoxon signed-rank test with a Bonferroni correction to establish statistical significance at $p < 0.05$.

Analysis from OVO strategy, overall classification. The subset of connectivity (i.e. local vs. inter-regional vs. combined) was statistically significant (p<0.001, chi-square=18.167, df=2; Friedman). Post-hoc analysis revealed statistically significant differences in MMCC between local and inter-regional connectivity (p=0.002, z-score=-3.059; Wilcoxon), combined and inter-regional connectivity (p=0.002, z-score=-3.059; Wilcoxon) but not local versus combined (p=0.53, z-score: -0.628; Wilcoxon). Statistically significant differences were observed in the MMCC between brain ROIs (p=0.001, chi-square=16.6, df=2; Friedman), regardless of the type of connectivity considered. Upon post-hoc analysis, MMCC values for M1 were significantly lower versus other frontal ROIs for M1 vs. SMA (p=0.008, z-score=-

2.666; Wilcoxon), M1 vs. PMC (p=0.008, z-score=-2.666; Wilcoxon). However, no other significant differences were observed in MMCC data between fontal ROIs.

[INSERT TABLE 4 HERE]

Analysis from OVA strategy, one-skill level-against-the-rest classification. The type of network connectivity was compared using a within-skill level approach. Regardless of the skill level of the operator, the type of connectivity significantly affected performance of the classifier (p<0.001, chi-square=54.056, df=2; Friedman). Indeed, performance was significantly poorer when inter-regional rather than local connectivity was considered (p<0.001, z-score=-5.232; Wilcoxon). Similar to the results obtain for the OVO model, regardless of the connectivity data used, statistically significant differences were observed between regions in the performance of the classifier. Specifically, for all skill level groups, MMCC values from M1 were significantly lower than several other frontal brain regions for novices: M1 vs. PFC (p=0.008, z-score=-2.666; Wilcoxon), M1 vs. SMA (p=0.008, z-score=-2.666; Wilcoxon), M1 vs. PMC (p=0.008, z-score=-2.666; Wilcoxon); trainees: M1 vs. SMA (p=0.008, z-score=-2.666; Wilcoxon); experts: M1 vs. PFC (p=0.008, z-score=-2.666; Wilcoxon); M1 vs. SMA, (p=0.008, z-score=-2.666; Wilcoxon).

[INSERT TABLE 5 HERE]

[INSERT TABLE 6 HERE]

[INSERT TABLE 7 HERE]

4 Discussion and conclusion

Notwithstanding established models of motor skill levels learning (Hikosaka et al., 2002) and evidence supporting learning-related changes in brain function (Coynel et al., 2010; Heitger et al., 2012; Sun et al., 2007), disparate technical performance does not necessarily translate into differences in functional activations on highly complex bimanual co-ordination tasks

(Ohuchida et al., 2009). Our hypothesis is that differences in technical skill level on tasks of such complexity may be better reflected in differences in functional connectivity within associative and/or sensorimotor brain networks (Bullmore and Bassett, 2011; Coynel et al., 2010; Heitger et al., 2012; Hikosaka et al., 2002; Rissman et al., 2004). Here, skill level-related differences in frontal lobe connectivity were revealed, summarised as a reduction in connectivity strength between cortical regions involved in the associative network. This suggests that frontal lobe cortico-cortical connectivity on a bimanual co-ordination task varies according to operator skill level and technical skill level, thereby confirming the dynamic nature of coupling in cognitive-motor circuitry. In particular, interaction strength between prefrontal and premotor seeds and other motor-related cortical regions were found to attenuate with skill level. In other words, novices appear to depend on the interactions between associative and motor cortical networks more than experts. Of these frontal lobe interactions, PFC-SMA, PFC-PMC and SMA-PMC connections were consistently stronger in novice operators whereas SMA-M1 interactions remained stable or even increased in expert compared with trainee operators. These data align with models of motor learning implying a high level of integration which decreases with practice in the associative/premotor network and between this and the sensorimotor network (Hikosaka et al., 2002). Moreover, the results are consistent with empirical data that suggest large-scale functional re-organisation accompanies motor skill levels learning (Bassett and Bullmore, 2006; Coynel et al., 2010; Heitger et al., 2012; Rissman et al., 2004; Sun et al., 2007).

Among studies investigating spatial and temporal changes in recruitment of brain regions involved in motor skill levels learning (Hikosaka et al., 2002; Kelly and Garavan, 2005), few have interrogated functional connectivity (Bassett and Bullmore, 2006; Bernardi et al 2013, Coynel et al., 2010; Heitger et al., 2012; Rissman et al., 2004; Sun et al., 2007). Whilst investigations have explored changes in modularity and allegiance of network nodes (Bassett and Bullmore, 2006), variation in functional network econometrics (Heitger et al., 2012) and hierarchical integration within associative and sensorimotor networks (Coynel et al., 2010), only one study specifically interrogated changes in cortico-cortical connectivity during explicit bimanual motor skill level learning (Sun et al., 2007). Commensurate with our findings, this study by Sun and colleagues (Sun et al., 2007) demonstrated enhanced cortico-cortical network connectivity during early phases of explicit bimanual motor sequence learning. Interestingly, connectivity between higher cognitive centres (PFC) and the motor network (PMC) increased only during early learning and subsequently decreased during late within-session learning when subjects improved their performance (Sun et al., 2007). Fast motor learning is characterised by increased functional connectivity between dorsolateral PFC and PMC, possibly related to the heightened attentional demands required at this stage of skill levels acquisition (Dayan and Cohen, 2011; Hikosaka et al., 2002).

Consistent with our findings, studies of slow motor learning suggest longitudinal changes in functional connectivity in premotor associative networks over extended practice schedules (Coynel et al., 2010). Slow learning (weeks) is accompanied by decreased integration, a metric reflecting functional interactions amongst several brain regions, in a premotor-associative-striatum cerebellar network (Coynel et al., 2010). The current analysis both supports and extends these findings, suggesting that attenuation in associative cortico-cortical network connectivity is durable across many years of practice, i.e. from novice to trainee. More importantly, it appears that progression from intermediate (trainee) to advanced (expert) phases of skill level execution are potentially independent of further changes in connectivity strength within this associative cortico-cortical network. Contrary to variation observed in the strength of connectivity between prefrontal and premotor regions associated with skill level, SMA-M1 connectivity strength was more stable and skill level-related differences were less apparent. Despite models of learning suggesting that integration in the sensorimotor network is anticipated to increase (Hikosaka et al., 2002), stability in the strength of SMA-M1 integration across motor skill levels learning has been observed previously, further supporting the results of the current analysis (Coynel et al., 2010). It is plausible that gradual on-going refinements in complex motor skill levels do not require sustained or progressive increase in SMA-M1 connectivity even if the representation of the trained task continues to gradually expand in these brain regions in association with slow skill levels learning (Karni et al., 1998).

Moreover, correlation data obtained at one-second intervals of the time course were subsequently used as an input to a LS-SVM algorithm towards automated discrimination of operator skill level. Overall, the approach was found to precisely classify skill level, although it was significantly influenced by the subset of connectivity data under consideration. Specifically, local (within-region) connectivity significantly improved classifier performance. Passive Brain Computer Interfaces (pBCI) offers the potential to feedback implicitly derived information regarding user states. In this regard, through immediate automated categorisation of short epochs of brain inputs, optical neuroimaging data if appropriately decoded may be interfaced with a learner toward improvements in motor skill levels training (Ros et al., 2009). Previous research on fNIRS-BCI has focused on modelling patterns of signal amplitude (Coyle et al., 2007; Naito et al., 2007; Power et al., 2012), failing to incorporate functional connectivity data. Here, discriminatory patterns of operator skill level based on functional connectivity have been exposed through the multi-class machine learning approach, and were not revealed with a conventional analytical framework (Bahrami et al., 2011; Leff et al., 2008c; Leff et al., 2008d; Ohuchida et al., 2009). The algorithm for classification of operator skill level states has proven to be accurate, sensitive and specific for detection of operator skill level during LS, regardless of the subtask. This suggests that the algorithm is capable of capturing experience-related differences in frontal lobe connectivity equally well in each task phase. Whilst each LS subtask has a unique and specific goal, the rudimentary bimanual co-ordination required to achieve these goals may be similar and hence generic learning-related differences in frontal lobe connectivity may exist, thereby aiding discrimination. Moreover, the results highlight the importance of within-region correlations in discrimination of operator skill level. Indeed, in comparison to the results obtained with inter-regional connectivity data, classification was significantly superior when local connectivity was considered. Critically, these results were consistent across models (i.e. OVA and OVO). Finally, classification results using M1 connectivity appeared inferior to the results obtained from other frontal regions possibly owing to a relative paucity of channel information and hence a smaller number of nodes. More importantly, the fact that classifier results were not consistently superior for any specific ROI suggests that the entire frontal lobe is equally informative.

When results of the classifier are provided online, these may be used as a pBCI to feedback data regarding skill level to the operator in the hope of improving the cognitive abilities required to execute the task. Automated classification of skill level based on cortical connectivity rather than technical performance or abstracted end-points may serve to improve skill levels learning by interfacing decoded brain data to the learner, trainer or team. Incorporation of instrument motion tracking in conjunction with brain function may further improve workflow segmentation and proficiency detection. Future work will focus on measuring the impact of neural feedback training to determine if learning and performance can be improved. In this way, online neural feedback regarding operator skill level may minimise the dependency on human assessors allowing learners to rapidly access objective assessments, thereby facilitating self-directed training. In addition, we acknowledge potential confounds due to individual characteristics that were not controlled for, such as the potential influence of intellectual ability (IQ), which could be of interest to consider in future studies. Likewise, incorporating electrophysiological signals that overcome some of the limitations of optical imaging technology such as inferior temporal resolution may improve performance and be more practical for pBCI in surgery. Finally, future research could address the limitations associated with a cross-sectional approach by implementing longitudinal studies to extend on the present findings.

In summary, the current study highlights differences in frontal brain connectivity underpinning operator skill level on a complex bimanual co-ordination task. Operators well versed in bimanual co-ordination appear to depend less on associative and premotor connectivity than novices. Moreover, the disparity in learning-related changes in connectivity facilitates an algorithmic solution toward online classification of operator skill level which is both highly accurate and precise. Automated classification of operator states may facilitate interfaces designed to enhance learning, performance and patient safety.

Acknowledgements

The authors wish to thank the anonymous reviewers as well as the editorial team of Brain Connectivity for their detailed feedback and very helpful comments in enhancing the quality of this manuscript.

References

Abdi H.. 2007. Rv coefficient and congruence coefficient. In: Salkind N.J (eds.), *Encyclopedia of measurement and statistics*, 1st ed. New York, NY: Sage Publications; p. 849.

Bahrami P, Schweizer TA, Tam F, Grantcharov TP, Cusimano MD, Graham SJ. 2011. Functional MRI-compatible laparoscopic surgery training simulator. Magn Reson Med 65: 873-881.

Bassett DS, Bullmore E. 2006. Small-world brain networks. Neuroscientist 12: 512-523.

Bernardi, G., Ricciardi, E., Sani, L., Gaglianese, A., Papasogli, A., Ceccarelli, R., Franzoni, F., Galetta, F., Santoro, G., Goebel, R., 2013. How skill expertise shapes the brain functional architecture: an fMRI study of visuo-spatial and motor processing in professional racing-car and naïve drivers. PloS one 8: 1-11.

Bullmore ET, Bassett DS. 2011. Brain graphs: graphical models of the human brain connectome. Annu Rev Clin Psychol 7: 113-140.

Coyle SM, Ward TE, Markham CM. 2007. Brain–computer interface using a simplified functional near-infrared spectroscopy system. J Neural Eng 4: 219-226.

Coynel D, Marrelec G, Perlbarg V, Pelegrini-Issac M., Van de Moortele PF, Ugurbil K, Doyon J, Benali H, Lehericy S. 2010. Dynamics of motor-related functional integration during motor sequence learning. Neuroimage 49: 759-766.

Dayan E., Cohen LG. 2011. Neuroplasticity subserving motor skill learning. Neuron 72, 443-454.

Draganski B, May A. 2008. Training-induced structural changes in the adult human brain. Behavioural brain research 192: 137-142.

Duty B, Andonian S, Ma Y, Peng S, Shapiro E, Dhawan V, Richstone L, Eidelberg D, Kavoussi LR. 2012. Correlation of laparoscopic experience with differential functional brain activation: a positron emission tomography study with oxygen 15-labeled water. Arch Surg 147: 627-632.

Elwell CE, Cope M, Edwards AD, Wyatt JS, Delpy DT, Reynolds EO. 1994. Quantification of adult cerebral hemodynamics by near-infrared spectroscopy. J Appl Physiol: 77, 2753-2760.

FLS, Fundamentals of Laparoscopic Surgery. www.flsprogram.org Last accessed Jan. 15, 2015

Galar M, Fernández A, Barrenechea E, Bustince H, Herrera F. 2011. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. Pattern Recognition 44: 1761-1776.

Gaonkar B, Davatzikos C. 2013. Analytic estimation of statistical significance maps for support vector machine based multi-variate image analysis and classification. Neuroimage 78: 270-283.

Gorodkin J. 2004. Comparing two K-category assignments by a K-category correlation coefficient. Computational biology and chemistry 28: 367-374.

Halsband U, Lange RK. 2006. Motor learning in man: a review of functional and clinical studies. J Physiol Paris 99: 414-424.

Hatakenaka M, Miyai I, Mihara M, Sakoda S, Kubota K. 2007. Frontal regions involved in learning of motor skill—a functional NIRS study. Neuroimage 34: 109-116.

Heitger MH, Goble DJ, Dhollander T, Dupont P, Caeyenberghs K, Leemans A, Sunaert S, Swinnen SP, 2013. Bimanual Motor Coordination in Older Adults Is Associated with Increased Functional Brain Connectivity–A Graph-Theoretical Analysis. PLoS One 8(4): 1-16.

Heitger MH, Ronsse R, Dhollander T, Dupont P, Caeyenberghs K, Swinnen SP. 2012. Motor learning-induced changes in functional brain connectivity as revealed by means of graph-theoretical network analysis. Neuroimage 61: 633-650.

Hikosaka O, Nakamura K, Sakai K, Nakahara H. 2002. Central mechanisms of motor skill level learning. Current opinion in neurobiology 12: 217-222.

Ikegami, T., & Taga, G. 2008. Decrease in cortical activation during learning of a multi-joint discrete motor task. Experimental brain research 191: 221-236.

Izzetoglu K, Bunce S, Onaral B, Pourrezaei K, Chance B. 2004. Functional optical brain imaging using near-infrared during cognitive tasks. International Journal of Human-Computer Interaction 17: 211-227.

James DR, Leff DR, Orihuela-Espina F, Kwok KW, Mylonas GP, Athanasiou T, Darzi AW, Yang GZ. 2013. Enhanced frontoparietal network architectures following "gaze-contingent" versus "free-hand" motor learning. Neuroimage 64: 267-276.

James DR, Orihuela-Espina F, Leff DR, Mylonas GP, Kwok KW, Darzi AW, Yang GZ. Cognitive burden estimation for visuomotor learning with fNIRS. In Proceedings of Medical Image Computing and Computer-Assisted Intervention, Beiging, China, 2010, p. 319.

Josse J, Pagès J, Husson F. 2008. Testing the significance of the Rv coefficient. Comput Stat Data Anal, 53: 82-91.

Jurcak V, Tsuzuki D, Dan I. 2007. 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. Neuroimage 34: 1600-1611.

Jurman G, Riccadonna S, Furlanello C, 2012. A comparison of MCC and CEN error measures in multi-class prediction. PLoS One 7(8): 1-8.

Karni A, Meyer G, Rey-Hipolito C, Jezzard P, Adams MM, Turner R, Ungerleider LG, 1998. The acquisition of skill leveled motor performance: fast and slow experience-driven changes in primary motor cortex. Proceedings of the National Academy of Sciences 95: 861-868.

Kelly AM, Garavan H, 2005. Human functional neuroimaging of brain changes associated with practice. Cereb Cortex 15: 1089-1102.

Leff D, Leong J, Atallah L, Athanasiou T, Hui Koh P, Elwell C, Delpy D, Yang GZ, Darzi A. Open Surgical Knot-Tying Induces a Lateralised Prefrontal Brain Response in Surgical

Novices. In Proceedings of Organization for Human Brain Mapping, Chicago, Illinois, USA. 2007a.

Leff DR, Elwell CE, Orihuela-Espina, F., Atallah, L., Delpy, D.T., Darzi, A.W., Yang, G.Z., 2008a. Changes in prefrontal cortical behaviour depend upon familiarity on a bimanual co-ordination task: an fNIRS study. Neuroimage 39: 805-813.

Leff DR, Leong JJ, Aggarwal R, Yang GZ, Darzi A, 2008b. Could variations in technical skill levels acquisition in surgery be explained by differences in cortical plasticity? Annals of surgery 247: 540-543.

Leff DR, Orihuela-Espina F, Atallah L, Athanasiou T, Leong JJ, Darzi AW, Yang GZ, 2008c. Functional prefrontal reorganization accompanies learning-associated refinements in surgery: a manifold embedding approach. Computer Aided Surgery 13: 325-339.

Leff DR, Orihuela-Espina F, Atallah L, Darzi A, Yang GZ. Functional near infrared spectroscopy in novice and expert surgeons - a manifold embedding approach. In Proceedings of Medical Image Computing and Computer-Assisted Intervention, Brisbane, Australia, 2007b, p. 270.

Leff DR, Orihuela-Espina F, Leong J, Darzi A, Yang GZ. Modelling dynamic fronto-parietal behaviour during minimally invasive surgery--a Markovian trip distribution approach. In Proceedings of Medical Image Computing and Computer-Assisted Intervention, New York, USA, 2008d, p. 595.

Leff, D. R., Orihuela-Espina, F., Elwell, C. E., Athanasiou, T., Delpy, D. T., Darzi, A. W., & Yang, G. Z. 2011. Assessment of the cerebral cortex during motor task behaviours in adults: a systematic review of functional near infrared spectroscopy (fNIRS) studies. NeuroImage 54: 2922-2936.

Naito M, Michioka Y, Ozawa K, Kiguchi M, Kanazawa T. 2007. A communication means for totally locked-in ALS patients based on changes in cerebral blood volume measured with near-infrared light. IEICE transactions on information and systems 90: 1028-1037.

Nan W, Rodrigues JP, Ma J, Qu X, Wan F, Mak PI, Mak PU, Vai MI, Rosa A. 2012. Individual alpha neurofeedback training effect on short term memory. International Journal of Psychophysiology 86(1): 83-87.

Obrig H, Wenzel R, Kohl M, Horst S, Wobst P, Steinbrink J, Thomas F, Villringer A. 2000. Near-infrared spectroscopy: does it function in functional activation studies of the adult brain? International Journal of Psychophysiology 35: 125-142.

Ohuchida K, Kenmotsu H, Yamamoto A, Sawada K, Hayami T, Morooka K, Takasugi S, Konishi K, Ieiri S, Tanoue K. 2009. The frontal cortex is activated during learning of endoscopic procedures. Surg Endosc 23: 2296-2301.

Oldfield RC. 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9: 97-113.

Orihuela-Espina F, Leff D, James D, Darzi A, Yang GZ, 2010. Quality control and assurance in functional near infrared spectroscopy (fNIRS) experimentation. Physics in medicine and biology 55: 3701-3724.

Parasuraman, R., Christensen, J., Grafton, S., 2012. Neuroergonomics: The brain in action and at work. Neuroimage 59: 1-3.

Parasuraman R, Rizzo M, 2003. *Neuroergonomics*. Oxford, UK: Oxford University Press

Power SD, Kushki A, Chau T, 2012. Automatic single-trial discrimination of mental arithmetic, mental singing and the no-control state from prefrontal activity: toward a three-state NIRS-BCI. BMC research notes 5: 141-151

Press W, Teukolsky S, Vetterling W, Flannery B, 2007. Section 10.12. Simulated Annealing Methods. In: *Numerical Recipes: The Art of Scientific Computting,* 3rd ed. New york, NY: Cambridge University Press. p. 549.

Rissman J, Gazzaley A, D'Esposito M. 2004. Measuring functional connectivity during distinct stages of a cognitive task. Neuroimage 23: 752-763.

Rodrigo AH, Di Domenico SI, Ayaz H, Gulrajani S, Lam J, Ruocco AC. 2014. Differentiating functions of the lateral and medial prefrontal cortex in motor response inhibition. Neuroimage 85: 423-431.

Ros T, Moseley MJ, Bloom PA, Benjamin L, Parkinson LA, Gruzelier JH. 2009. Optimizing microsurgical skill levels with EEG neurofeedback. BMC neuroscience 10: 87-97.

Sun FT, Miller LM, Rao AA, D'Esposito M, 2007. Functional connectivity of cortical networks involved in bimanual motor sequence learning. Cerebral Cortex 17: 1227-1234.

Suykens JA, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J, 2002. *Least squares support vector machines*. Singapore, World Scientific Pub.

Tsuzuki D, Cai DS, Dan H, Kyutoku Y, Fujita A, Watanabe E, Dan I, 2012. Stable and convenient spatial registration of stand-alone NIRS data through anchor-based probabilistic registration. Neurosci Res 72: 163-171.

Wang JR, Hsieh S, 2013. Neurofeedback training improves attention and working memory performance. Clin Neurophysiol 124: 2406-2420.

Zhu FF, Poolton JM, Wilson MR, Hu Y, Maxwell JP, Masters RS. 2011. Implicit motor learning promotes neural efficiency during laparoscopy. Surg Endosc 25: 2950-2955.

Zoefel B, Huster RJ, Herrmann CS, 2011. Neurofeedback training of the upper alpha frequency band in EEG improves cognitive performance. Neuroimage 54: 1427-1431.

Figure captions

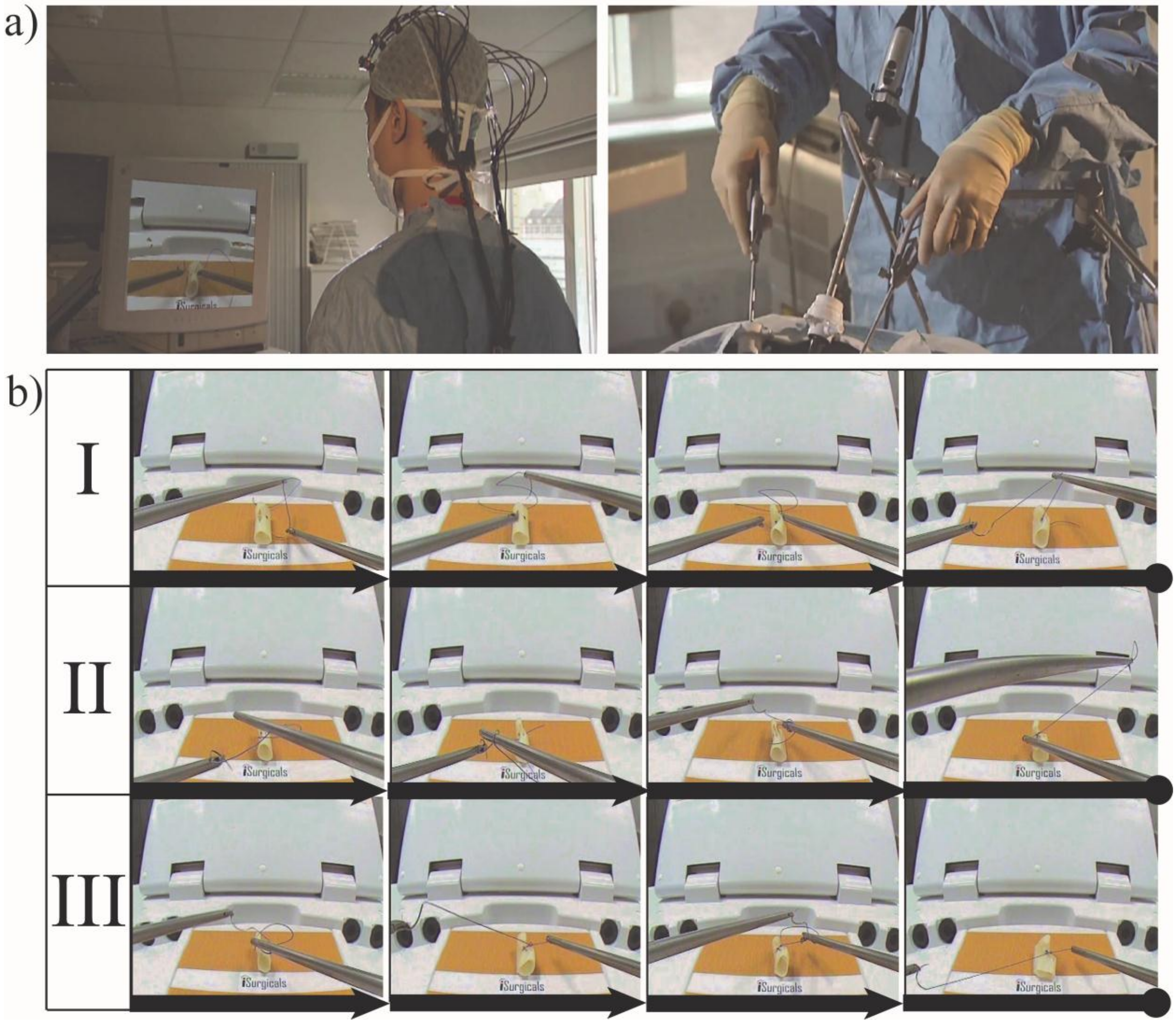

Fig. 1 (a-b). (a) A participant performing LS in a box trainer with simultaneous recordings of cortical brain activity using OT; (b) The three subtasks of the surgical procedure, namely 'needle insertion' (I), 'double-throw knot' (II) and 'single-throw knot' (III).

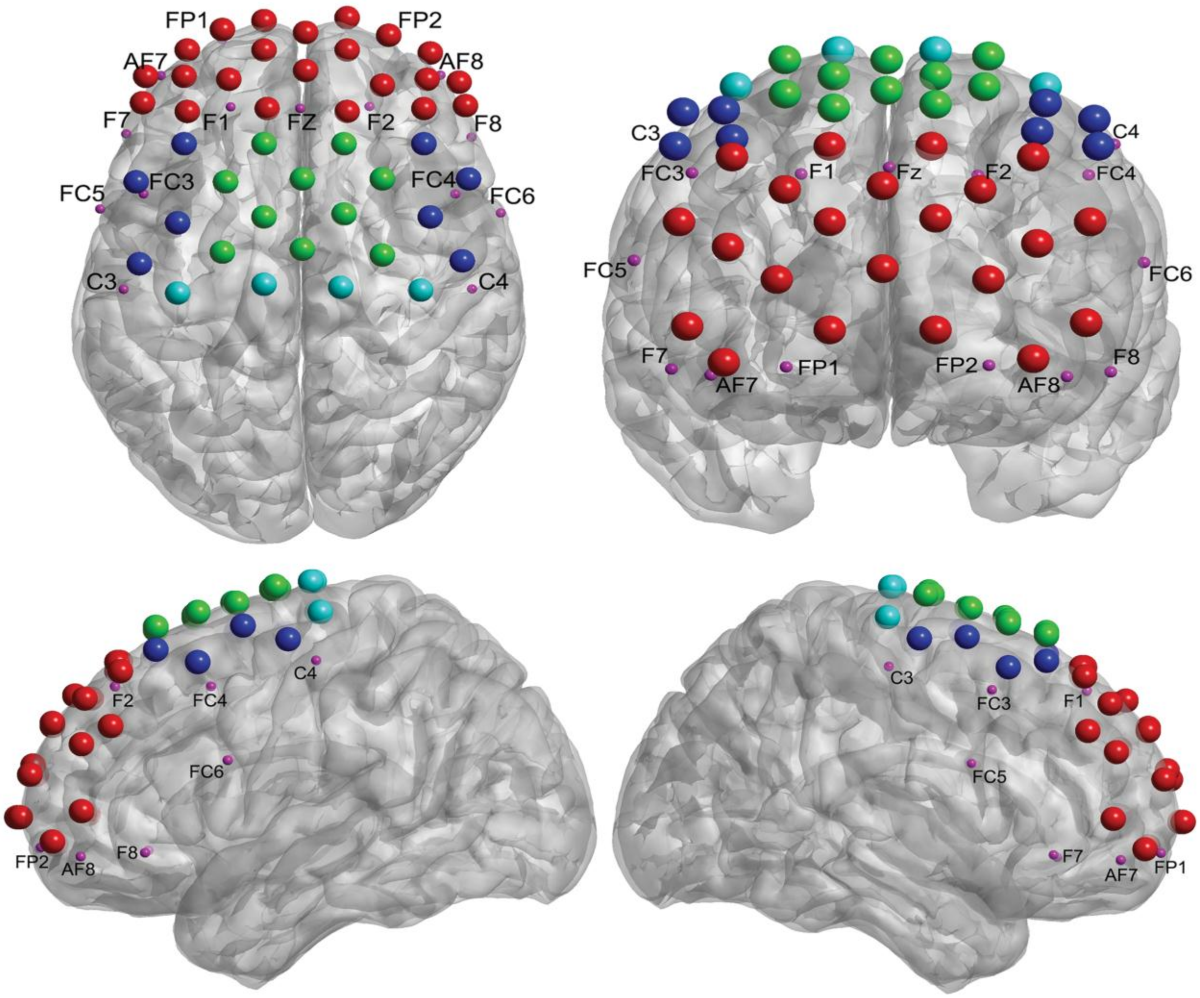

Fig. 2. Registration of channel positions in MNI space, illustrating the approximate locations over the prefrontal cortex (red), supplementary motor area (green), premotor cortex (dark blue) and primary motor cortex (soft blue) relative to International 10-10 markers (magenta).

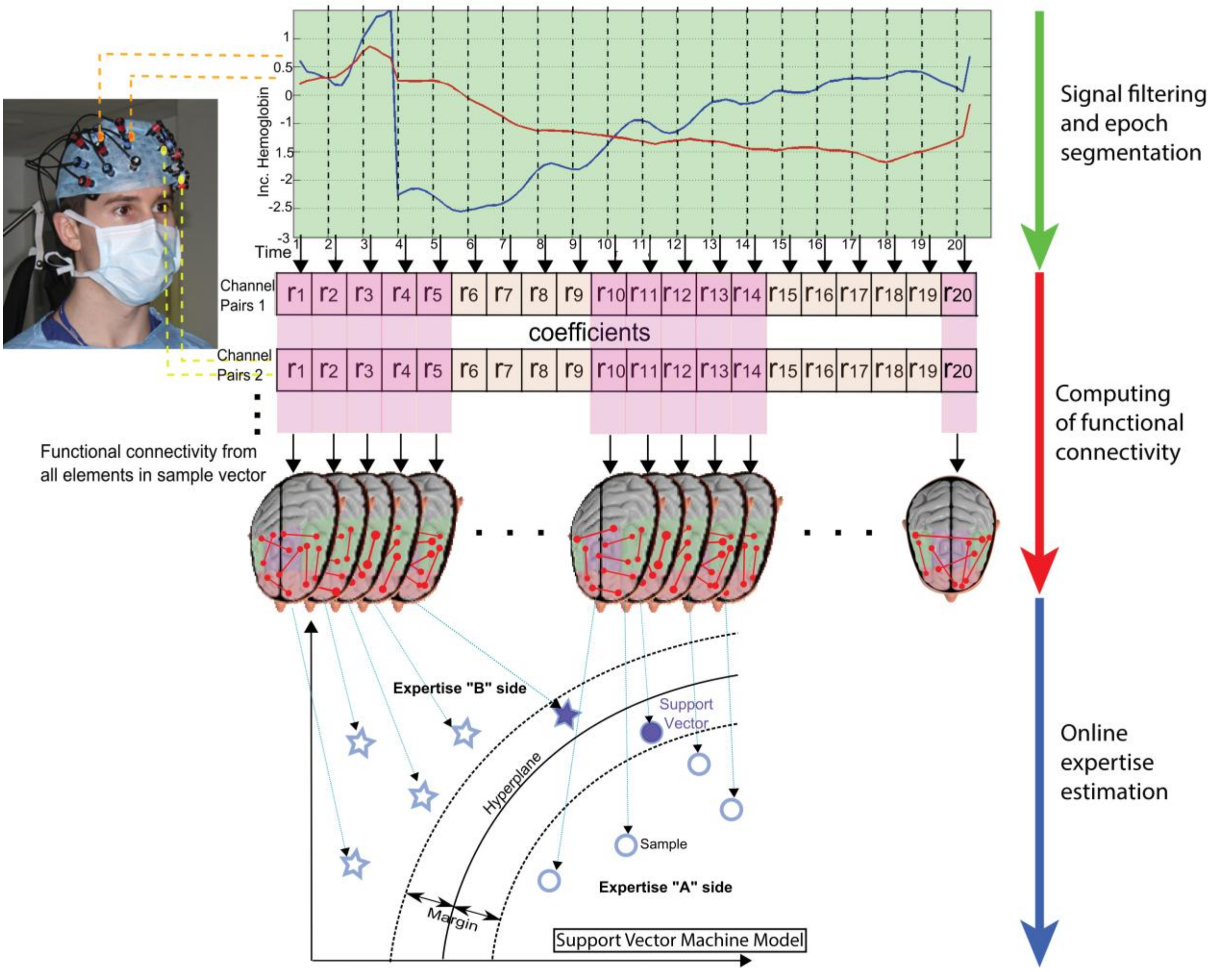

Fig. 3. Schematic illustration of the processing steps and classification of the extended dataset using a large-margin classifier, i.e. LS-SVM.

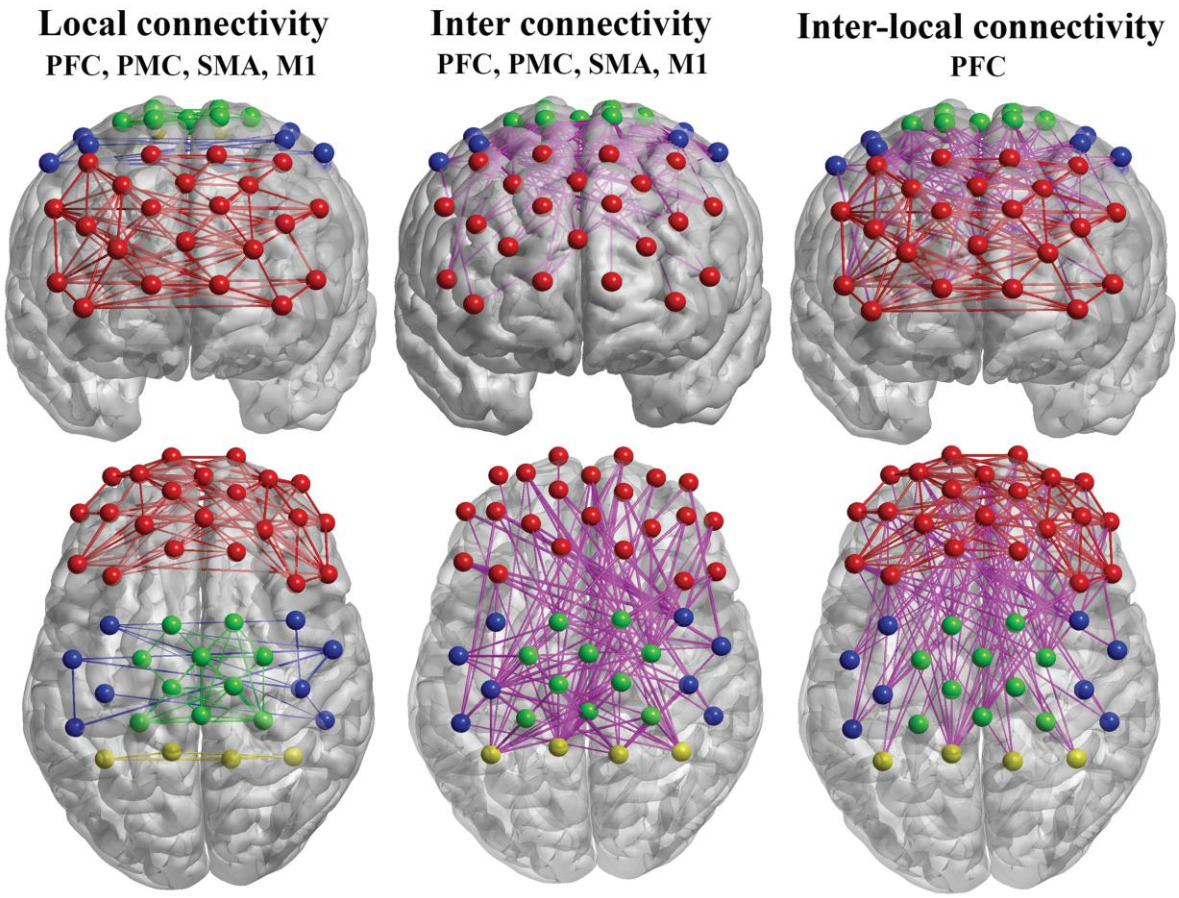

Fig. 4. Graphical representation of local, inter-regional and combined (inter-local regional) connectivity showing channels over the PFC (red), PMC (blue), SMA (green) and M1 (yellow). Each edge represents a functional connection between channels. Within-region connections (colour coded by sub-region) and inter-regional connections (magenta) are highlighted.

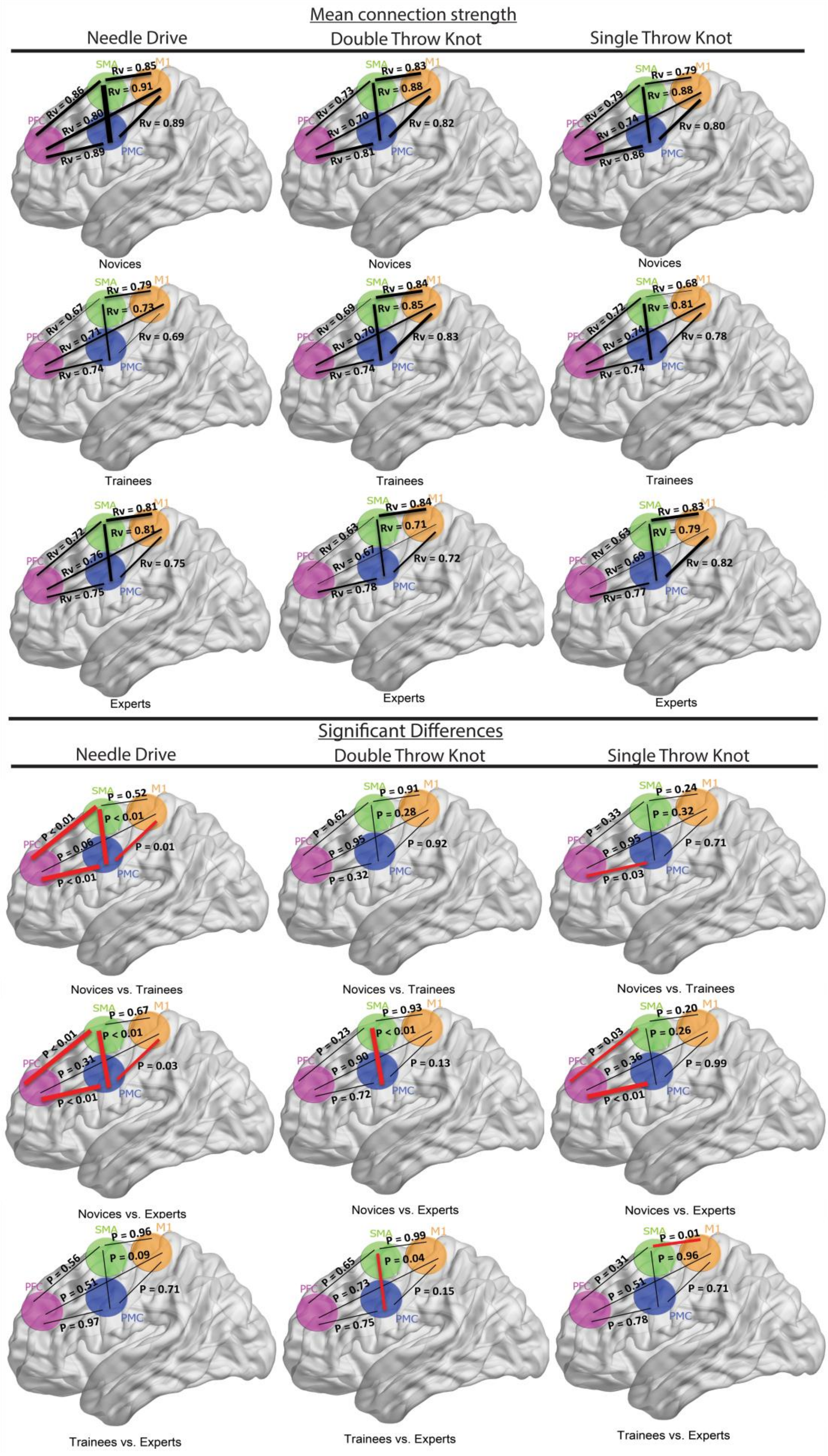
Mean connection strength
Needle Drive
Double Throw Knot
Single Throw Knot
Novices
Trainees
Experts
Significant Differences
Novices vs. Trainees
Novices vs. Experts
Trainees vs. Experts
SMA
M1
PFC
PMC

Fig. 5. Sagittal brain views of learning-related changes in connectivity for the needle insertion, double-throw knot and single-throw know subtasks. The three sagittal brains at the top display the mean Rv connectivity strength, while the following three at the bottom show the significance strength of the statistical test. Areas are depicted as PFC (magenta), SMA (green), M1 (orange) and PMC (blue).

Table 1. Between-group differences in task-related heart rate and technical skill (FLS).

Values		Experts [E]	Trainees [T]	Novices [N]	Significance test
Heart Rate	Mean beats per min.	80.71 ± 12.03 (mean ± standard deviation)	84.31 ± 11.2 (mean ± standard deviation)	82.98 ± 15.66 (mean ± standard deviation)	p=0.87 *F-value*=0.145 df=2
Heart Rate Variability	Mean R to R wave interval	0.76 ± 0.12 (mean ± standard deviation)	0.72 ± 0.1 (mean ± standard deviation)	0.75 ± 0.13 (mean ± standard deviation)	p=0.83 *F-value*=0.182 df=2
Motor Performance &Technical Skill (FLS score)		487 ± 53 (median ± inter-quartile range)	400 ± 90 (median ± inter-quartile range)	334 ± 76 (median ± inter-quartile range)	p<0.001 *Chi-square*=58 df=2

Table 2. Classification results for the session based networks.

a) Needle insertion	**Precision**	**Specificity**	**Sensitivity**
Novices	0.66	0.83	0.66
Trainees	0.65	0.76	0.76
Experts	0.77	0.92	0.60
Model accuracy: 0.68	F-measure: 0.68	MMCC: 0.52	
a) Double-throw knot	**Precision**	**Specificity**	**Sensitivity**
Novices	0.71	0.84	0.74
Trainees	0.65	0.77	0.70
Experts	0.68	0.89	0.59
Model accuracy: 0.68	F-measure: 0.68	MMCC: 0.50	
a) Single-throw knot	**Precision**	**Specificity**	**Sensitivity**
Novices	0.66	0.85	0.61
Trainees	0.59	0.70	0.73
Experts	0.57	0.85	0.45
Model accuracy: 0.61	F-measure: 0.61	MMCC: 0.41	

Table 3. Classification results for the time-course based networks.

a) Needle insertion	**Precision**	**Specificity**	**Sensitivity**
Novices	0.83	0.80	0.92
Trainees	0.85	0.94	0.78
Experts	0.8	0.8	0.67
Model accuracy: 0.83	F-measure: 0.81	MMCC: 0.71	
a) Double-throw knot	**Precision**	**Specificity**	**Sensitivity**
Novices	0.83	0.90	0.85
Trainees	0.83	0.86	0.86
Experts	0.76	0.95	0.67
Model accuracy: 0.82	F-measure: 0.81	MMCC: 0.70	
a) Single-throw knot	**Precision**	**Specificity**	**Sensitivity**
Novices	0.82	0.82	0.91
Trainees	0.85	0.92	0.81
Experts	0.8	0.97	0.63
Model accuracy: 0.82	F-measure: 0.81	MMCC: 0.70	

Table 4. Multiclass Mathew's correlation coefficient and accuracy for the OVO scheme*.

	PFC		**SMA**		**PMC**		**M1**	
	MMCC	Accuracy	MMCC	Accuracy	MMCC	Accuracy	MMCC	Accuracy
Local	0.63 ± 0.01	0.78 ± 0.01	0.63 ± 0.01	0.78 ± 0.01	0.62 ± 0.01	0.77 ± 0.01	0.58 ± 0.02	0.75 ± 0.01
Inter-regional	0.45 ± 0.02	0.67 ± 0.01	0.33 ± 0.02	0.60 ± 0.02	0.21 ± 0.01	0.54 ± 0.02	0.11 ± 0.01	0.48 ± 0.02
Combined	0.60 ± 0.01	0.77 ± 0.01	0.63 ± 0.02	0.78 ± 0.01	0.63 ± 0.01	0.78 ± 0.01	0.58 ± 0.02	0.75 ± 0.01

*Results averaged from all subtasks, means and standard deviations (±) are presented in this table.

Table 5. Multiclass Mathew's correlation coefficient and accuracy for Novices in OVA scheme*.

	PFC		**SMA**		**PMC**		**M1**	
	MMCC	Accuracy	MMCC	Accuracy	MMCC	Accuracy	MMCC	Accuracy
Local	0.69 ± 0.01	0.85 ± 0.01	0.66 ± 0.01	0.83 ± 0.01	0.66 ± 0.01	0.83 ± 0.01	0.61 ± 0.01	0.81 ± 0.01
Inter-regional	0.50 ± 0.01	0.76 ± 0.01	0.32 ± 0.01	0.67 ± 0.02	0.21 ± 0.02	0.62 ± 0.03	0.14 ± 0.04	0.60 ± 0.03
Combined	0.66 ± 0.01	0.83 ± 0.01	0.66 ± 0.01	0.84 ± 0.01	0.65 ± 0.02	0.83 ± 0.01	0.62 ± 0.01	0.81 ± 0.01

*Results averaged from all subtasks, means and standard deviations (±) are presented in this table.

Table 6. Multiclass Mathew's correlation coefficient and accuracy for Trainees in OVA scheme*.

	PFC		**SMA**		**PMC**		**M1**	
	MMCC	Accuracy	MMCC	Accuracy	MMCC	Accuracy	MMCC	Accuracy
Local	0.67 ± 0.03	0.78 ± 0.09	0.68 ± 0.01	0.86 ± 0.01	0.60 ± 0.09	0.82 ± 0.06	0.62 ± 0.02	0.83 ± 0.02
Inter-regional	0.47 ± 0.02	0.77 ± 0.01	0.42 ± 0.03	0.75 ± 0.03	0.26 ± 0.01	0.70 ± 0.04	0.10 ± 0.04	0.65 ± 0.05
Combined	0.62 ± 0.01	0.83 ± 0.01	0.67 ± 0.01	0.85 ± 0.02	0.66 ± 0.01	0.85 ± 0.01	0.63 ± 0.02	0.84 ± 0.02

*Results averaged from all subtasks, means and standard deviations (±) are presented in this table.

Table 7. Multiclass Mathew's correlation coefficient and accuracy for Experts in OVA scheme*.

	PFC		**SMA**		**PMC**		**M1**	
	MMCC	Accuracy	MMCC	Accuracy	MMCC	Accuracy	MMCC	Accuracy
Local	0.61 ± 0.02	0.89 ± 0.01	0.57 ± 0.01	0.88 ± 0.01	0.54 ± 0.01	0.87 ± 0.01	0.50 ± 0.02	0.87 ± 0.01
Inter-regional	0.37 ± 0.06	0.84 ± 0.01	0.15 ± 0.03	0.82 ± 0.01	0.07 ± 0.03	0.81 ± 0.01	0.08 ± 0.02	0.81 ± 0.01
Combined	0.57 ± 0.02	0.88 ± 0.01	0.57 ± 0.01	0.88 ± 0.01	0.55 ± 0.01	0.88 ± 0.01	0.52 ± 0.02	0.87 ± 0.01

*Results averaged from all subtasks, means and standard deviations (±) are presented in this table.