

Technical Correspondence

Sensory-Glove-Based Open Surgery Skill Evaluation

Laura Sbernini, Lucia Rita Quitadamo, Francesco Riillo, Nicola Di Lorenzo, Achille Lucio Gaspari, and Giovanni Saggio

Abstract—Manual dexterity is one of the most important surgical skills, and yet there are limited instruments to evaluate this ability objectively. In this paper, we propose a system designed to track surgeons' hand movements during simulated open surgery tasks and to evaluate their manual expertise. Eighteen participants, grouped according to their surgical experience, performed repetitions of two basic surgical tasks, namely single interrupted suture and simple running suture. Subjects' hand movements were measured with a sensory glove equipped with flex and inertial sensors, tracking flexion/extension of hand joints, and wrist movement. The participants' level of experience was evaluated discriminating manual performances using linear discriminant analysis, support vector machines, and artificial neural network classifiers. Artificial neural networks showed the best performance, with a median error rate of 0.61% on the classification of single interrupted sutures and of 0.57% on simple running sutures. Strategies to reduce sensory glove complexity and increase its comfort did not affect system performances substantially.

Index Terms—Gesture recognition, manual dexterity, motion capture, training evaluation, wearable systems.

I. INTRODUCTION

The most effective educational approach for trainee surgeons is to experience directly on patients in the operating room (OR) under the supervision of expert surgeons. However, this approach can represent a potential issue for patients' safety. Laboratory-based surgical education has received increasing interest, allowing trainees to practice without risks to patients. Outside the OR, simulators can successfully enhance surgeon's skill [1], [2]; however, there are still significant limitations due to the variability in the requirements associated with different surgical techniques. For instance, minimally invasive surgery (MIS) has a limited workspace and the fulcrum-mediated motion restricts surgeon's hand movements to four degrees of freedom (DoFs) only [3]. In open surgery (OS), instead, the workspace is the entire surrounding environment, and surgeon's hand operates surgical instruments in a full DoF scenario. As a result of this, simulators have been more widely accepted for MIS procedures, such as laparoscopic [4] and robotic [5] surgery, rather than the OS setting [6].

Manuscript received July 28, 2016; revised January 14, 2017, June 13, 2017, and August 24, 2017; accepted October 25, 2017. This paper was recommended by Associate Editor X. Xu. (Corresponding author: Giovanni Saggio.)

L. Sbernini, F. Riillo, and G. Saggio are with the Department of Electronics Engineering, University of Rome "Tor Vergata," 00173, Rome, Italy (e-mail: laura.sbernini@uniroma2.it; fr.riillo@gmail.com; saggio@uniroma2.it).

L. R. Quitadamo is with the Department of Electronics Engineering, University of Rome "Tor Vergata," 00173, Rome, Italy, and also with the School of Life and Health Sciences, Aston Brain Centre, Aston University, Birmingham B4 7ET, U.K (e-mail: lucia.quitadamo@gmail.com).

N. Di Lorenzo and A. L. Gaspari are with the Department of Experimental Medicine and Surgery, University of Rome "Tor Vergata," 00173, Rome, Italy (e-mail: nicola.di.lorenzo@uniroma2.it; achille.gaspari@gmail.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/THMS.2017.2776603

Independently of the educational approach, the main goal is to increase surgeon's skills, since clinical outcome is strongly related to skill levels [7]. Manual dexterity is one of the most valuable of those skills [3], and its *objective* evaluation should be a critical part of a surgeon's assessment.

The aim of this study was to evaluate manual performance while participants manipulated real surgical instruments in a laboratory setting, in the context of OS tasks. Using classification algorithms, we evaluated the relationship between *quantitative metrics* obtained by a hand tracking system and *qualitative metrics* expressing the participants' surgical "level of competence."

To overcome some of the limitations and costs of the most frequently used tracking technologies based on camera capture and video analysis [8], [9], our hand tracking system was a 20-DoF sensory glove that measured flexion/extension of hand joints and wrist movements. In fact, camera-based technologies need an expensive dedicated environment, a free line of sight, and model the hand as a single point in the 3-D space, with consequent loss of detail on movements of individual fingers. Alternative hand tracking technologies are based on accelerometers [10] and on measurement of electromagnetic fields [11]. Despite their low invasiveness and cost, these systems are only able to measure a reduced number of hand's DoF with reduced measurement accuracy.

We evaluated surgeons' maneuvering expertise (*qualitative metrics*) attributing hand movements to a surgical "class of competence," e.g., novice, intermediate, and expert, detected by means of classification techniques [12], [13]. Classifiers based on the Markov model and the hidden Markov model have been successfully used in the literature; however, they present a high computational complexity [3], [12], [14]. Alternative less demanding approaches based on linear discriminant analysis (LDA) [13], [15], support vector machines (SVMs) [16], artificial neural networks (ANNs), or artificial neuro-fuzzy inference systems [17] have been explored. In this study, we compared the performance of three of these classifiers, LDA, SVM, and ANN, being the most suitable for future real-time implementations on sensory gloves. Moreover, the classification framework was designed to address the limitations of our previous studies [18]–[20], in which we investigated the possibility to objectively assess manual performances during OS training, although they were evaluated on a simple task and with a less accurate classification scheme. A further aim of our study was to explore strategies to reduce the overall number of sensors and to make the system as comfortable as possible.

II. MATERIALS AND METHODS

A. Participants

Eighteen subjects signed an informed consent, in accordance with our local ethical committee, and took part to the study. Participants were grouped into two classes according to their surgical experience. In

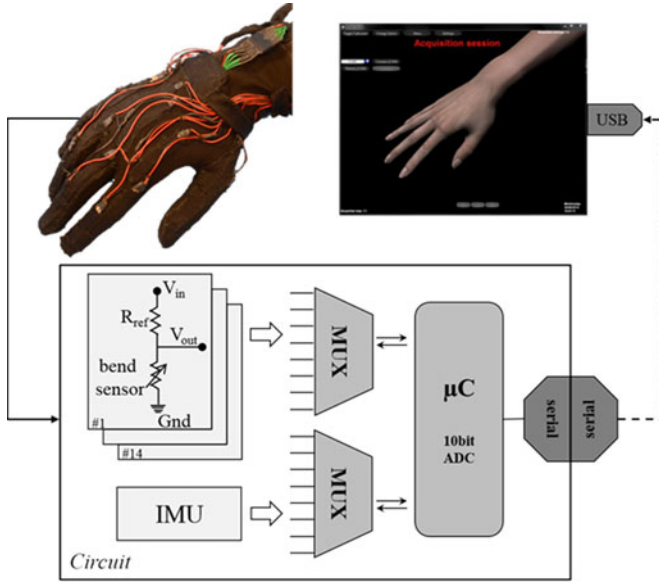


Fig. 1. System. A sensory glove equipped with flex and inertial sensors fed an electronic circuit, which elaborates and sends data to a computer, which elaborates information and virtually represents the hand movements by an avatar. Solid path line (left): glove/circuit interface; dotted path line (right): circuit/computer interface. R_{ref} : reference resistance; V_{in} and V_{out} : input and output voltages, respectively; Gnd: ground; IMU: inertial measurement unit; MUX: multiplexer; μC : microcontroller; ADC: analog-to-digital converter.

particular, three men and six women (aged 34–37 years) were allocated to the reference group “Expert,” and seven men and two women (aged 26–30 years) to the testing group “Novice.” Subjects of the first group were senior surgical residents in general surgery in their final three-year period of training, whereas subjects of the second group were students without any surgical experience.

All the subjects declared to be right handed and performed the experiments with the dominant hand.

B. Sensory Glove and Electronics

We designed and developed two sensory gloves (a small size and a medium size to best fit users’ hands) and the electronics to acquire and elaborate data from sensors.

Each glove was implemented with a polyester/elastane mixed fabric and was equipped with 15 sensors [21]. Of these, 14 resistive flex sensors (RFSs) (by Flexpoint Sensor Systems, Inc., Draper, UT, USA) were placed on the finger joints, and one six-DoF inertial measurement unit (IMU) (Analog Combo Board Razor by SparkFun Electronics, Niwot, CO, USA) was placed on the back of the glove (see Fig. 1).

The increase of RFS electrical resistance due to fingers’ bending [22] allowed measuring the flexion/extension of the metacarpophalangeal (MCP), proximal interphalangeal (PIP), and distal interphalangeal (DIP) joints of the index, middle, ring, and little fingers and the MCP and interphalangeal joints of the thumb. The IMU, equipped with a three-axis accelerometer and a three-axis gyroscope, allowed measuring the flexion/extension, pronation/supination, and radial/ulnar deviation of the wrist. All the sensors were individually hosted into suitable fabric pockets sewn on the dorsal aspect of the glove. Thanks to the adopted sensors, we obtained measurements with an angular resolution of the order of one degree for each DoF [22].

The resistance values coming from the 14 flex sensors were individually converted to voltage values by means of 14 series resistors, which formed 14 voltage dividers. Resistance values of the series resistors,

$R_{ref,i}$ ($i = 1, 2, \dots, 14$), were obtained according to the following formula [23]:

$$R_{ref,i} = \sqrt{R_{max,i} R_{min,i}} \quad (1)$$

where $R_{max,i}$ and $R_{min,i}$ are the maximum and minimum resistance values of the i th RFS, within its working range, respectively.

The 20 analog voltages (14 from the flex sensors through the voltage dividers and six directly from the IMU) were multiplexed (two 16-channel analog multiplexers MC14067 by Motorola, Inc.) and sent to a microcontroller (PIC18F4550 by Microchip Technology Inc., Chandler, AZ, USA). The microcontroller managed the multiplexers and performed the A/D conversion and the data transmission to a personal computer at a frequency of 57.5 Hz.

A custom-made software received and saved data to allow, both in real-time and offline, the reproduction of hand movements by means of an avatar [24] (see Fig. 1). Differently from a video-based reproduction of the movements of a real hand, the avatar-based reproduction allows viewing fingers’ details and hand’s movements from any possible point of view, thanks to the availability of image rotation options.

In order to take into account differences among users’ hands and to map the analog sensor values to the corresponding joint angles, a calibration procedure of the sensory glove was required [25]. We decided to omit this time-consuming phase, adopting an uncalibrated approach, which was successfully used in previous glove-based classification studies [26]. The 20 raw analog sensor readings, which indirectly describe hand movements, were used as direct input to the data analysis, avoiding the conversion of their voltage data into angle data.

C. Experimental Setup and Protocol

Basic technical-training OS suturing tasks were performed on a simulated cutaneous wound, which was developed using a $7.5 \text{ cm} \times 10.0 \text{ cm}$ soft foam pad with a vertical incision. Participants used standard surgical tools, such as a needle holder and a needle with suture thread.

Participants were asked to perform two tasks: a single interrupted suture (SIS), which consists in joining the two foam-pad vertical edges together by sewing, and a simple running suture (SRS), which consists in five single sutures without interruptions and without knot tying. We chose these two tasks since they are key to correct surgical manual dexterity.

Each participant sat in front of a desk with the foam pad on top and wore the sensory glove of an appropriate size. A contour of the hand was drawn on the desk to indicate the start/end position of the hand in each trial. “Experts” performed the tasks based on their own experience, while “Novices” were trained to perform the tasks by the “Experts.” No restrictions were set about the suture placements and the execution time. Fig. 2 shows the experimental setup.

Each participant performed 24 trials consisting of 12 SISs and 12 SRSs. Tasks were performed in two different days, but only data recorded on the second day were analyzed, after participants had gained familiarity with the assignment. The first and the last repetitions of each task were discarded, as they might have been affected by subjects’ low practice (the first repetition) and high fatigue (the last repetition).

D. Data Analysis

Manual dexterity was evaluated using a trial-based classification and a subject-based classification. For the former, each task repetition was classified based on whether it was performed by a trained or by an untrained user; for the latter, each subject was classified as an expert or a novice. Classification procedures were performed considering both



Fig. 2. Experimental setup. A soft foam pad with a precut vertical incision was used to simulate a cutaneous wound. Participants were asked to wear the sensory glove and to perform sutures with standard surgical instruments.

the full-set of sensors and a subset of them to investigate the possibility of reducing the number of sensors without significantly affecting the classification performance.

All the analyses were performed using ad-hoc routines written in MATLAB (R2013a, The MathWorks, Inc., Natick, MA, USA), and each step of the process is described in the following sections.

1) *Preprocessing*: To increase the signal-to-noise ratio, data from flex and inertial sensors were preprocessed with a five-sample-sized moving average filter. A data-cutting algorithm based on a discrete moving average filter was then applied to discard data that preceded and followed in time the real task execution. Each recording was time-normalized by a resampling procedure to obtain N uniformly spaced time points, with N set to 1000 for the SIS and to 4000 for the SRS task, respectively.

2) *Feature Extraction*: Each data recording was empirically segmented into a series of 300 sample-length windows, overlapping by 75 samples. Each time-normalized trial registration was then segmented into ten for the first task and 50 windows for the second.

Mean, variance, root mean square, waveform length, slope sign change, and zero crossing were computed to represent typical time-domain features of the signal [27], [28]. For each subject, features were extracted from each window of each sensor and arranged into a feature vector. Forty-one feature vectors composed either of single features or of multiple combinations of up to three different features were then tested.

3) *Expertise Classification*: Three different classifiers, LDA, SVM, and ANN, were adopted and compared.

LDA allows rearranging input data into one linear dimension maximizing the distance between classes by means of a hyperplane that linearly separates input data. To reduce misclassification errors that can occur due to bad estimation of covariance matrices (bias), a shrinkage-based regularization was implemented [29].

The SVM tries to find the best hyperplane that separates all data points of one class from those of the other class, that is, the hyperplane with the maximum margin between the classes. In the case of data characterized by intrinsic nonlinearity, the SVM makes use of kernel functions, which project data into a higher dimensionality space, in which data can be linearly separated again [28]. We adopted an SVM with a linear kernel. We set the maximum number of iterations to solve the optimization problem to 10^7 and then used the default setting

available in the *svmtrain* MATLAB function in which training data points are shifted and scaled to have zero mean and unit variance.

An ANN is based on a network with nodes (artificial neurons) interconnected by weighted and directed connections. It is suitable to solve complex problems and capable to update itself according to the training data [30]. In this study, we implemented a multilayer perceptron ANN, with one hidden layer composed of ten neurons, one neuron in the output layer, and the Levenberg–Marquardt error backpropagation algorithm [31], [32] as the training algorithm. The activation functions were sigmoid for the hidden layer and linear for the output layer. Weight and bias values for the i th layer were initialized using the Nguyen–Widrow initialization method [33]. In order to find the best ANN architecture, 70% of the whole training test samples were used for the network training, 15% to validate that the network is generalizing and to stop training before overfitting, and the last 15% was used as a completely independent test of network generalization. Potential overfitting was avoided, thanks to the early stopping procedure provided in MATLAB.

We evaluated the performances of each classifier using a leave-one-out cross-validation approach [30]: each trial of the whole set was used as a testing set and the remaining trials as a training set. The procedure was iterated until each trial was used at least once for the testing.

Classification results were evaluated for each combination of feature vector/classifier computing the misclassification rate, i.e., the error rate associated with the classification of all the windows of all the trials. As a result, the classification stage determined a $[41 \times 3]$ matrix, rows being the feature vectors and columns being the three classifiers. The normal distribution of the error rates relative to each classifier (columns) was assessed with the Kolmogorov–Smirnov test. The Friedman test (i.e., the nonparametric equivalent of the repeated-measures analysis of variance) was used to detect significant differences in performance of the three classifiers. Alpha value was set to 0.05. When the null hypothesis was rejected post-hoc analysis with Wilcoxon signed-rank tests, a Bonferroni multiple-comparison correction ($p < 0.05/3 = 0.017$) was further performed to identify which of the dependent variables was influenced by classifiers.

The “best” classifier was determined by comparing the medians computed on the 41 error rates obtained for each classifier and selecting the classifier with the lowest value. Then, we identified the “best” feature vector of the best classifier as the one with the lowest error rate among the 41 available ones.

The best classifier and the best feature vector were used to determine whether a single task repetition was performed by an expert or by a novice (trial-based classification). To this aim, we evaluated each trial according to the number of correctly classified windows. Let w be the number of windows for each trial, with $w = 10$ for the SIS and $w = 50$ for the SRS; we considered a trial as performed by an expert/novice subject if at least the 90% of w was assigned to the correct class. Results filled a confusion matrix with experienced/inexperienced classes [28], furnishing the error distribution across the 180 trials (i.e., 18 subjects and ten repetitions). Similarly, we evaluated each subject according to the number of his/her correctly classified trials (subject-based classification). Thus, we considered a participant as an expert or a novice if at least nine out of the ten trials were correctly assigned to their respective class; the entire procedure was performed for each of the two tasks.

4) *Sensor Selection*: For each task and each subject, we used the principal component analysis to reduce datasets’ dimensionality, while maintaining the 90% of the original data variability [34]. We obtained k significant principal components (PCs) for each dataset, with k_{MAX} being the highest value of k . Linear correlation coefficients were computed between each of the k PCs and the original dataset,

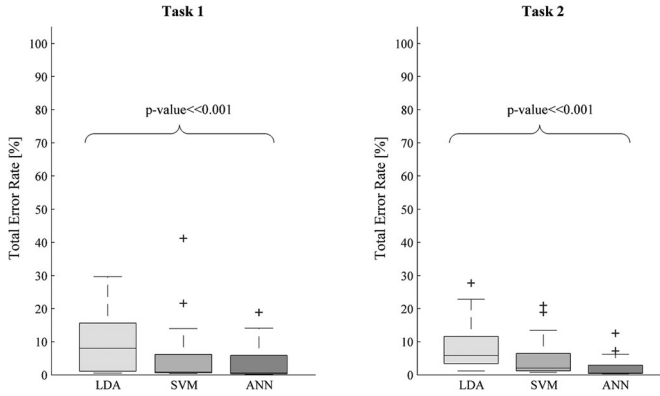


Fig. 3. Total error rates for each classifier and for both tasks (task 1: SIS, task 2: SRS). Friedman test results are reported as p -values (Alpha value was set to $p = 0.05$).

TABLE I
BEST ANN CLASSIFICATION RESULTS*

Sensors	Task	Error [%]	BFV
All	1 [#]	0.17	Var + M + ZC
	2 [§]	0.30	M
Subset	1 [#]	0.56	RMS + SSC
	2 [§]	0.48	M + WL

Abbreviations: BFV, best feature vector; Var, variance; M, mean; ZC, zero crossing; RMS, root mean square; SSC, slope sign change; WL, waveform length.

*The best feature vector and its error rate are shown for each case.

[#]Task 1: SIS.

[§]Task 2: SRS.

and one sensor was associated with each of the k PCs according to the highest correlation value with the PC itself [34]. We then counted the number of times each sensor was identified as the one with the highest correlation among all the subjects. Variables were, therefore, ranked in descending order, and the first k_{MAX} sensors were selected as the new subset. To generalize the results to both tasks, we kept into account the sensors included in at least one out of the two new subsets. Following this procedure, a new dataset with a lower number of variables was obtained and constituted the new input data for the classification.

III. RESULTS AND DISCUSSION

Test for normality showed that at least two out of three sets of error rates were not coming from a continuous probability distribution ($p < 0.05$). The Friedman test showed that, for each task, the three classifiers performed differently ($p < 0.05$), and post-hoc analysis demonstrated significant differences among each pair of classifiers revealing their independence ($p < 0.017$).

Fig. 3 summarizes classification results by means of box plots referred to median (M) and interquartile range (IQR) values of the total error rates. In particular, for SIS, we had M [IQR] equal to 8.06% [1.11–15.64], 0.89% [0.66–6.21], and 0.61% [0.33–5.90], respectively, for LDA, SVM, and ANN; for SRS, we had 5.86% [3.43–11.66], 2.08% [1.25–6.48], and 0.57% [0.43–2.95], respectively, for LDA, SVM, and ANN.

The ANN showed the lowest median values and outperformed LDA and SVM in all the cases. Table I reports the resulting best classifier/feature vector couple for each task.

TABLE II
TRIAL-BASED CLASSIFICATION RESULTS: CONFUSION MATRICES*

Sensors	Task	Exp	Nov
All	1 [#]	Exp	90
		Nov	0
	2 [§]	Exp	90
		Nov	0
Subset	1 [#]	Exp	89
		Nov	2
	2 [§]	Exp	89
		Nov	1
		Exp	89
		Nov	1

Abbreviations: Exp, “Expert” group; Nov, “Novice” group.

*Trials are 180 for each case: 18 subjects and ten repetitions.

[#]Task 1: SIS.

[§]Task 2: SRS.

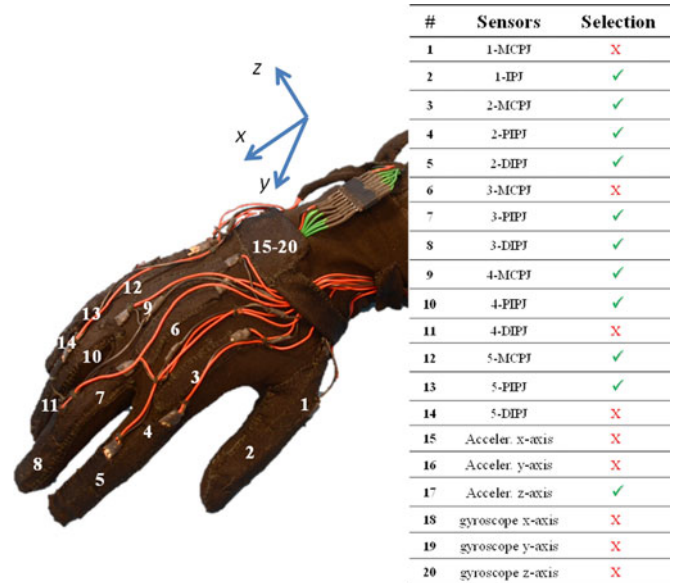


Fig. 4. Sensor selection results. Sensors are coded following the table (right) and placed on the glove as shown in the picture (left). In the table, ticks indicate the selected sensor outputs and crosses indicate the discarded ones. Joints nomenclature: DIPJ, distal interphalangeal joint; PIPJ, proximal interphalangeal joint; MCPJ, metacarpophalangeal joint; and IPJ, interphalangeal joint. Fingers nomenclature: 1, thumb; 2, index; 3, middle; 4, ring; and 5, little.

Table II shows the confusion matrices that summarize the performances obtained by classifying each single task repetition as performed by an expert or by a novice using the best classifier/feature vector couple for each task. The results obtained considering the full-set of sensors and a subset of them (see hereafter in this section) are shown together in Tables I and II for comparison. According to these results, trial-based classification achieved an error rate of 0% for both tasks, all the 180 trials being assigned to the correct experience class. Subject-based classification successfully evaluated at least nine out of the ten trials for each subject for both tasks, and consequently, each user was matched with the correct experienced/inexperienced class.

Moreover, the sensor selection procedure identified 11 sensor outputs out of 20 as the most significant (see Fig. 4 for reference). Fig. 5 reports the results of the classification performances using the selected sensor subset. Error rates were organized in a $[41 \times 3]$ table, and test for

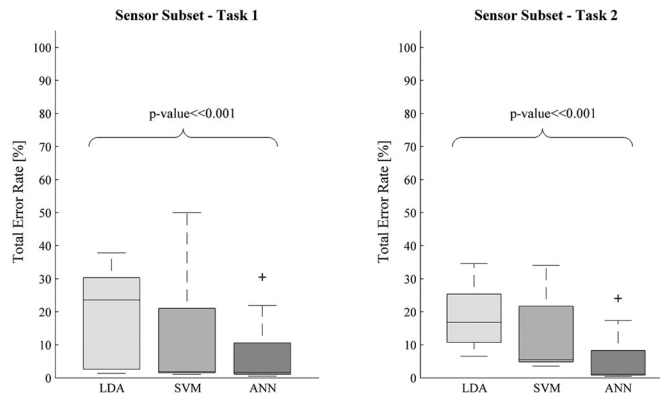


Fig. 5. Box plots of the classification performances using the selected sensor subset: total error rates are shown for each classifier and for both tasks (task 1: SIS, task 2: SRS). Friedman test results are reported as p -values (Alpha value was set to $p = 0.05$).

normality showed that at least two column data out of three were not coming from a continuous distribution ($p < 0.05$) for both the task datasets. Friedman test results and the post-hoc analysis demonstrated that the classifiers performed differently ($p < 0.05$), with significant differences among each pair of classifiers ($p < 0.017$).

The ANN performed better than SVM and LDA (see Fig. 5), and its relative best feature vector results are reported in Table I. Trial-based classification error rate was 1.67% for the first task and 1.11% for the second one, three and two trials out of 180 being wrongly classified, respectively, for the two tasks, as Table II shows. Even when considering a reduced number of sensors, subject-based classification correctly evaluated at least nine out of ten trials for each subject and allowed each subject to be correctly classified as an expert/novice.

Finally, the ANN achieved the lowest error rate and outperformed LDA and SVM in all the tests. ANN performance was not negatively affected by the variability of input training data resulting from differences in subjects' hand anatomy, level of experience, familiarity with surgical instrumentations, etc. This was largely due to the possibility of varying the number of neurons in each layer and of adapting weight of connections according to the input data. The SVM, by means of the kernel function, is characterized by a decision rule that is a simple linear function in the kernel space and this makes it more stable and characterized by low variance (variance reflects the sensitivity of the classifier to the training set used). LDA, instead, was less accurate than both SVM and ANN. This classifier is notoriously sensitive to variability of input data and to bad dimensionality of the training matrix, and this could have affected its performance in our specific analyses, despite the shrinkage regularization.

IV. CONCLUSION

This work describes a sensory-glove-based framework for the objective classification of technical surgical expertise during OS training. To the best of our knowledge, this is the first application of a sensory glove to the evaluation of OS training. Glove-based technology appears to be one of the most promising solutions to capture the complexity of hand movement assessment in OS training. Previous works already assessed surgical performances using a sensory glove [14], [35], [36], [37], but attempts were mainly focused on the MIS approach.

The proposed method adopts descriptive statistical parameters as features for the classification stage to develop an assessment tool available for a wide range of surgical contexts, removing the need of specific

sensor- or task-based metrics. A further advantage of this method is its low computational complexity.

In conclusion, the developed glove-based system can represent a feasible evaluation tool for OS training, contributing to facilitate the introduction of new technology-based skill assessments as new standards in surgical education.

ACKNOWLEDGMENT

The authors would like to thank all the volunteers who took part in this study, Prof. S. Seri (School of Life, Health Sciences, Aston University, Birmingham, U.K.) for overall revision, and Dr. P. Cavallo (The MathWorks Inc., Natick, MA, USA) for English Language revision of the manuscript.

REFERENCES

- [1] R. M. Satava, "The revolution in medical education—The role of simulation," *J. Graduate Med. Educ.*, vol. 1, no. 2, pp. 172–175, 2009.
- [2] P. Figert, A. Park, D. Witzke, and R. Schwartz, "Transfer of training in acquiring laparoscopic skills," *J. Amer. Coll. Surgeons*, vol. 193, no. 5, pp. 533–537, 2001.
- [3] C. E. Reiley, H. C. Lin, D. D. Yuh, and G. D. Hager, "Review of methods for objective surgical skill evaluation," *Surgical Endoscopy*, vol. 25, no. 2, pp. 356–366, 2011.
- [4] S. Viriyasiripong *et al.*, "Accelerometer measurement of head movement during laparoscopic surgery as a tool to evaluate skill development of surgeons," *J. Surgical Educ.*, vol. 73, no. 4, pp. 589–594, 2016.
- [5] H. Abboudi *et al.*, "Current status of validation for robotic surgery simulators—A systematic review," *BJU Int.*, vol. 111, no. 2, pp. 194–205, 2013.
- [6] J. Davies, M. Khatib, and F. Bello, "Open surgical simulation—A review," *J. Surgical Educ.*, vol. 70, no. 5, pp. 618–627, 2013.
- [7] R. K. Reznick and H. MacRae, "Teaching surgical skills—changes in the wind," *New Engl. J. Med.*, vol. 355, no. 25, pp. 2664–2669, 2006.
- [8] I. Oropesa *et al.*, "EVA: Laparoscopic instrument tracking based on endoscopic video analysis for psychomotor skills assessment," *Surgical Endoscopy*, vol. 27, no. 3, pp. 1029–1039, 2013.
- [9] M. K. Powers *et al.*, "Crowd-sourcing assessment of surgeon dissection of renal artery and vein during robotic partial nephrectomy: A novel approach for quantitative assessment of surgical performance," *J. Endourol.*, vol. 30, no. 4, pp. 447–452, 2016.
- [10] A. Sánchez *et al.*, "Laparoscopic surgery skills evaluation: Analysis based on accelerometers," *J. Soc. Laparoendoscopic Surgeons*, vol. 18, no. 4, pp. 1–5, 2014.
- [11] V. Datta, S. Mackay, M. Mandalia, and A. Darzi, "The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model," *J. Amer. Coll. Surgeons*, vol. 193, no. 5, pp. 479–85, 2001.
- [12] I. Oropesa *et al.*, "Methods and tools for objective assessment of psychomotor skills in laparoscopic surgery," *J. Surgical Res.*, vol. 171, no. 1, pp. 81–95, 2011.
- [13] M. K. Chmarra, S. Klein, J. C. F. de Winter, F.-W. Jansen, and J. Dankelman, "Objective classification of residents based on their psychomotor laparoscopic skills," *Surgical Endoscopy*, vol. 24, no. 5, pp. 1031–1039, 2010.
- [14] K. Kahol *et al.*, "Measuring movement expertise in surgical tasks," in *Proc. 14th ACM Int. Conf. Multimedia*, Santa Barbara, CA, USA, 2006, pp. 719–722.
- [15] Z. Lin *et al.*, "Objective skill evaluation for laparoscopic training based on motion analysis," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 977–985, Apr. 2013.
- [16] R. A. Watson, "Use of a machine learning algorithm to classify expertise: Analysis of hand motion patterns during a simulated surgical task," *Acad. Med.*, vol. 89, no. 8, pp. 1163–1167, 2014.
- [17] I. Oropesa *et al.*, "Supervised classification of psychomotor competence in minimally invasive surgery based on instruments motion analysis," *Surgical Endoscopy*, vol. 28, no. 2, pp. 657–670, 2014.
- [18] G. Costantini, G. Saggio, L. Sbernini, N. Di Lorenzo, and D. Casali, "Towards an objective tool for evaluating the surgical skill," in *Computational Intelligence*, J. J. Merelo, A. Rosa, J. M. Cadenas, A. Dourado, K. Madani, and J. Filipe, Eds. New York, NY, USA: Springer, 2016, pp. 325–335.

- [19] G. Saggio *et al.*, "Objective surgical skill assessment: An initial experience by means of a sensory glove paving the way to open surgery simulation?" *J. Surgical Educ.*, vol. 72, no. 5, pp. 910–917, 2015.
- [20] G. L. Santosuoso, G. Saggio, F. Sora, L. Sbernini, and N. Di Lorenzo, "Advanced algorithms for surgical gesture classification," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Florence, Italy, 2014, pp. 3596–3600.
- [21] G. Saggio, S. Bocchetti, C. A. Pinto, and G. Orenco, "Electronic interface and signal conditioning circuitry for data glove systems useful as 3D HMI tools for disabled persons," in *Proc. Int. Conf. Health Informat.*, Rome, Italy, 2011, pp. 248–253.
- [22] G. Saggio, F. Riillo, L. Sbernini, and L. R. Quitadamo, "Resistive flex sensors: A survey," *Smart Mater. Struct.*, vol. 25, no. 1, pp. 1–30, 2016.
- [23] G. Saggio, S. Bocchetti, G. Orenco, and C. A. Pinto, "Electronic interface and signal conditioning circuitry for data glove systems useful as 3D HMI tools for disabled persons," in *Proc. Int. Conf. Health Informat.*, Rome, Italy, 2011, pp. 248–253.
- [24] G. Saggio and C. A. Pinto, "Virtuality supports reality for e-health applications," in *Virtual Reality*, J.-J. Kim, Ed. Rijeka, Croatia: InTech, 2011, pp. 247–272.
- [25] L. Dipietro, A. M. Sabatini, and P. Dario, "A survey of glove-based systems and their applications," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 38, no. 4, pp. 461–482, Jul. 2008.
- [26] G. Heumer, H. Ben Amor, M. Weber, and B. Jung, "Grasp recognition with uncalibrated data gloves—A comparison of classification methods," in *Proc. IEEE Virtual Reality Conf.*, Charlotte, NC, USA, 2007, pp. 19–26.
- [27] D. Zhang, Y. Geng, X. Zhang, Y. Zhang, and G. Li, "Real-time performance of hand motion recognition using kinematic signals for impaired hand function training," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng.*, San Diego, CA, USA, 2013, pp. 339–342.
- [28] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, Third Quarter 2013.
- [29] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K. R. Müller, "Single-trial analysis and classification of ERP components," *Neuroimage*, vol. 56, no. 2, pp. 814–825, 2011.
- [30] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.
- [31] K. Levenberg, "A method for the solution of certain problems in least squares," *Quart. Appl. Math.*, vol. 5, pp. 164–168, 1944.
- [32] D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM J. Appl. Math.*, vol. 11, no. 2, pp. 431–441, 1963.
- [33] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," in *Proc. Int. Joint Conf. Neural Netw.*, San Diego, CA, USA, 1990, vol. 3, pp. 21–26.
- [34] I. T. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer-Verlag, 2002, p. 488.
- [35] F. Amirouche, J. R. Martin, M. Gonzalez, and L. Fergusson, "Experimental set-up and sensory glove interface for microsurgery," *Proc. Inst. Mech. Eng. H*, vol. 222, no. 1, pp. 89–99, 2008.
- [36] J. D. Lemos, A. M. Hernandez, G. Soto-Romero, and A. Valade, "Instrumented glove for in-hand movement tracking in neurosurgical simulation," in *Proc. IEEE Int. Conf. Electron., Circuits Syst.*, Marseille, France, 2014, pp. 686–689.
- [37] R. C. King, L. Atallah, B. Lo, and G.-Z. Yang, "Development of a wireless sensor glove for surgical skills assessment," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 5, pp. 673–679, Sep. 2009.