This is a post-peer-review, pre-copyedit version of an article published in Artificial Intelligence in Medicine. The final authenticated version is available online at: http://dx.doi.org/10.1016/j.artmed.2017.03.007.

Automatic matching of surgeries to predict surgeons' next actions

Germain Forestier[a,b,*], François Petitjean[b], Laurent Riffaud[c,d], Pierre Jannin[c]

[a] *MIPS, University of Haute-Alsace, Mulhouse, France*
[b] *Faculty of Information Technology, Monash University, Melbourne, Australia*
[c] *INSERM MediCIS, Unit U1099 LTSI, University of Rennes 1, Rennes, France*
[d] *Department of Neurosurgery, Pontchaillou University Hospital, Rennes, France*

Abstract

Objective. More than half a million surgeries are performed every day worldwide, which makes surgery one of the most important component of global health care. In this context, the objective of this paper is to introduce a new method for the prediction of the possible next task that a surgeon is going to perform during surgery.
Material and Method. We formulate the problem as finding the optimal registration of a partial sequence to a complete reference sequence of surgical activities. We propose an efficient algorithm to find the optimal partial alignment and a prediction system using maximum a posteriori probability estimation and filtering. We also introduce a weighting scheme allowing to improve the predictions by taking into account the relative similarity between the current surgery and a set of pre-recorded surgeries.
Results. Our method is evaluated on two types of neurosurgical procedures: lumbar disc herniation removal and anterior cervical discectomy. Results show that our method outperformed the state of the art by predicting the next task that the surgeon will perform with 95% accuracy.
Conclusions. This work shows that, even from the low-level description of surgeries and without other sources of information, it is often possible to predict the next surgical task when the conditions are consistent with the previously recorded surgeries. We also showed that our method is able to assess when there is actually a large divergence between the predictions and decide that it is not reasonable to make a prediction.

Keywords: Temporal Analysis, Dynamic Time Warping, Surgical Process Modelling, Surgery

1. Introduction

In the USA alone, 1,000 new surgeries will have started within the next 10 minutes. This highlights how central surgeries have become for global health care. To support and assist surgical teams, Operating Rooms (ORs) have undergone tremendous changes. One of the targeted goals is the development of *context-aware* systems [1] that continuously monitor the activities performed in the ORs in order to provide an accurate and reliable support. The key challenge in developing these new methods is to process the data coming from sensors and real-time detection systems, in order to provide useful information and support decision making. This is extremely challenging because OR environments are very diverse, surgical interventions are very variable with specific patients, and different surgeons might have different levels of expertise. The richness and complexity of the data that is collected calls for new artificial intelligence methods [2] to support pre-, peri- and post-surgery (before, during and after). In this context, predictive data mining techniques [3] have long proven to be extremely relevant.

The field of Surgical Process Modeling (SPM) [4] targets the development of new methods that leverage from OR activities monitoring. In this field, several methods have already been proposed to automatically detect surgical activities. These methods rely either on manual annotations by an observer [5, 6] or on sensors present in the OR (*e.g.*, camera) [7, 8]. For example, the task performed by a surgeon can be automatically inferred by combining RFID chips on instruments (for identification) with accelerometers [9].

With the richness of the data comes the difficulty of analysing it, because of its complexity. For example, two surgeons performing the same surgery on the same patient might exhibit a very different course of specific actions, while being surgically very similar: they might use the same technique, have the same patient outcome, etc. However, from the low-level point of view (the sequence of low-level tasks like *cut*, *suture*, etc.), these surgeries will look very different from each other.

*Corresponding author – MIPS - Universite de Haute Alsace, 12 rue des freres Lumiere, 68093 Mulhouse, France, Tel.:+33 3 8933 6963, Fax.:+33 3 8942 3282

Email addresses: germain.forestier@uha.fr (Germain Forestier), francois.petitjean@monash.edu (François Petitjean), laurent.riffaud@chu-rennes.fr (Laurent Riffaud), pierre.jannin@univ-rennes1.fr (Pierre Jannin)

Extracting useful high-level knowledge from this low-level data has been one of the research themes targeted by the field of SPM [4, 10]: the objective is to understand surgeries to improve the quality of care. The above-mentioned sensors capture the surgical tasks performed in real-time, which opens the door to using artificial intelligence methods to provide real-time information to the surgical team.

This paper tackles the prediction of possible surgeons' subsequent actions, using low-level information alone. Predicting surgeons' possible next actions is critical for OR management: it can be used to provide useful real-time information to the surgical team (*e.g.*, nurses, anesthetist, junior surgeon), while allowing the surgeon to focus on more demanding tasks. For example, the nurses will be able to prepare the tool that is going to be used next, thus ensuring a smooth transition between the activities of the surgeon. Because predicting the next surgical task is central, such a prediction system will also be a keystone to the development of many other systems. For example, while the relative importance of the different factors that cause surgical error is unknown [11], technical skills acquisition are shown to correlate with a reduction of patient complications [12]. Thus, performing the right action at the right moment in surgery can greatly influence patient outcome. A study on patterns of technical error among surgical malpractice [13] highlighted that most technical errors occur in routine operations with experienced surgeons. One of the recommendation of the study is to focus surgical safety research on improving decision-making and performance in routine operations. This is why working on systems helping the surgeon to take action-oriented decisions is critical in the OR.

The data captured in the OR have a specific granularity level. A granularity level is defined as the level of abstraction at which the surgical procedure is described. MacKenzie et al. [14] were the first to propose a model of the surgical procedure that consists of different levels of granularity: the procedure, the step, the substep, the task, the subtask and the motion. Later, Lalys and Jannin [4] introduced a terminology consisting of phases defined as the major types of events occurring during surgery. Each phase is composed of several steps. A step is considered to be a sequence of activities used to achieve a surgical objective. The data used in this paper captures the activity of both hands for three different elements: *used instrument*, *performed action* and *targeted anatomical structure* [15]. Learning to predict the next activity of the surgeon from such low-level information is extremely challenging, because the next surgical action depends upon high-level information (such as phase of the surgery, technique used, patient-specific information, so-far reaction of the patient to the surgery, etc.), while a surgery is represented by a series of actions like "*cut* the *skin* with a *scalpel*".

Intuitively, our approach matches the on-going surgery to every surgery of a reference set of surgeries, and uses the next actions that have been performed in the reference set of surgeries to draw a prediction about the next action that will be performed in the current surgery. Our proposed approach includes the three following features:

1. **Optimal registration of a partial surgery:** We propose a new method to optimally register the on-going surgery (partial surgery) to any complete pre-recorded surgery. Our approach is based on the Dynamic Time Warping similarity measure [16], which is consistent with surgical processes [5].
2. **Voting for high-confidence prediction:** Using the optimally registered reference set of surgeries, we use voting to draw a high-confidence prediction about the next action that is going to be performed by the surgeon.
3. **Detecting when to predict with high-confidence:** Using the agreement rate among multiple predictors, we are able to detect when to perform a prediction and when it is not possible to draw an accurate prediction.
4. **Weighting the prediction according to sequence similarity:** Using the relative similarity between the on-going surgery and the set of pre-recorded surgery as weights, we are able to improve the prediction accuracy by giving more importance to similar surgical behaviors.

Our framework was assessed using two clinical datasets of lumbar disc herniation surgeries (LDH) and anterior cervical disectomy surgeries (ACD). The first dataset contains 24 LDH surgeries performed by multiple surgeons and was recorded at the Neurosurgery Department of a first site, named site A. The second dataset contains 18 ACD surgeries and was recorded at the Neurosurgery Department of a second site, named site B. We show that our method outperformed the state of the art on both datasets by providing a prediction with a 95% accuracy more than 85% of the times.

This article is an extended version of the article that was presented at the 15th Conference on Artificial Intelligence in Medicine in Europe [17]. In this extended version, we have improved the time complexity of the main algorithm from $\Theta(l \cdot k)$ to $O(l \cdot k)$ and improved our methodology with a new weighting technique for the predictions. We have also extended the validation of our work, by studying and comparing our method on a second clinical dataset. This paper is organized as follows. Section 2 introduces the related work in both surgical process modeling and prediction systems in health care and surgery. In Section 3, we present our method for high-confidence prediction of the next surgical activity that is going to be performed. In Section 4, we present experiments conducted to demonstrate the quality and performance of our approach compared to the state of the art. Finally, we conclude this work and describe future research in Section 5.

2. Related work

In this section, we briefly introduce the existing methods to record and recognize surgical activities (section 2.1), initial required step in our application. We then present some examples of prediction systems in health care and for surgery application (section 2.2).

2.1. Automatic recording of surgical activities

In the field of Surgical Process Modeling (SPM) [4], several methods have already been proposed to automatically detect surgical activities. These methods generally rely either on manual annotations by an observer [18, 6, 5] or on sensors present in the OR (*e.g.*, camera, motion detector, etc.) [19, 7, 8]. The data recorded by an observer are generally more accurate but are tedious to acquire. On the other hand, data recorded automatically using sensors scale up easily but are more prone to errors in the detection. In the following, we review principal efforts in developing methods for the automatic detection of surgeons activities.

Meissner et al. [9] showed recently that the tasks performed by a surgeon can be automatically inferred by combining RFID chips on instruments (for identification) with accelerometers. In this system, a hierarchical recognition model was used to detect instrument recognition and to infer the performed surgical action. Hidden Markov models (HMM) were then used to generate probability distributions over activities. Padoy et al. [19] also used HMM to combine low-level signals recorded in the OR. These low-level informations were combined with high-level information such as predefined phases to detect actions and to trigger events. Similarly, Bouarfa et al. [20] also used HMM with a pre-processing on the input sensor data in order to improve the detection of high-level surgical tasks.

Automatic detection of surgical activities can also be performed from videos. For example, Lalys et al. [21] proposed a framework for the automatic recognition of high-level surgical tasks using microscope videos analysis. The system was applied to cataract surgeries and combines computer vision techniques with time series analysis. SVM classifier was also considered by Lalys et al. [7] to detect phases and low-level surgical tasks using cameras in pituitary surgery.

Detecting the high-level surgical phases has also received some interest, as they provide a useful information about the current stage of the surgery, which could in turn be used to refine other models. For example, Bardram et al. [22] proposed a system using embedded and body-worn sensors data to train a decision tree in order to predict surgical phases. They studied sensor significance in order to identity the most important features for surgical phase prediction. More recently, Stauder et al. [23] used Random Forest (*i.e.*, a combination of decision trees) to predict surgical phases from sensors measurement.

Surgical robots represent a new source of data that is under study to detect the activities of surgeons. In this case, the movement of the robots, like the trajectories of the instruments, are used to infer the current activity of the surgeon. Significant amount of work has been devoted to the segmentation of surgical tasks into more detailed gestures [24, 25]. For example, Despinoy et al. [26] proposed a system to segment kinematic data from robotic training sessions. The goal there is to decompose the stream of kinematic data coming from a robot recording into individual gestures. The system is then used for training purposes. Note that phases and surgical activities are not the only interesting information to analyze. For example, Franke et al. [8] proposed a system to predict intervention time from low-level surgical activities.

All the above-mentioned sensors capture the surgical tasks performed in real-time, which opens the door to using artificial intelligence methods to provide real-time information to the surgical team. In this paper, we proposed to go beyond theses detection systems and we assume that the activities of the surgeons are detected. In this context, we address the problem of predicting the activities that the surgeon is going to perform in the future. Prediction systems play a crucial role in healthcare as presented in the following section.

2.2. Prediction systems in healthcare and surgery

Learning to predict the future from past observations is one of the key components that make it possible to bring value to the massive data stores that have been collected in medicine. For example, the system proposed by Liu et al. [27] has already proven its usefulness to predict patient information (e.g. blood count panel) from patient electronic health records. This system uses a hierarchical dynamical system and two modeling approaches: linear dynamical systems and gaussian processes to support predictive inferences. Huang et al. [28] also developed a predictive system in order to predict variation in clinical processes. They construct an appropriateness measure model based on historical clinical cases to predict variations in future cases of clinical processes. Based on a set of clinical cases extracted from Electronic Patient Records (EPRs) and a set of medical interventions, the system is able to predict if a variation is likely to happen in a specific clinical case. More recently, Bermejo et al. [29] proposed a system to predict the output of questionnaires used to measure the level of anxiety or depression in caregivers of schizophrenia patients. The goal of this prediction system is to anticipate an appropriate treatment or advice for the family caregivers from Primary Care consultations. In the context of surgery, Franke et al. [8] proposed a system for intervention time prediction from surgical low-level tasks. A surgical process model optimized for time prediction was designed together with a prediction algorithm. The predictions were used to support intervention scheduling and resource management. In all these applications, the performed predictions are used by medical teams to improve the quality and the efficiency of healthcare. Process Modeling was also recently considered to analyze surgical workflows. For example, Neumann

et al. [30] studied workflow modeling languages that could be suitable to model surgical processes. This kind of modeling would make it possible to leverage recent research advances in this field, such as developing prediction systems using business process models [31]. Sequence analysis has also been investigated for predicting information from surgical datasets. For example, Forestier et al. [5] proposed a system based on Dynamic Time Warping (DTW) that can predict the location and the level of expertise of a surgeon from recordings of surgical activities. DTW was also investigated to identify the variability of surgical procedures [10]. In the present work, DTW is compared to uniform scaling [32] which is based the Euclidean approach. Uniform scaling performs a linear transformation that stretches or contracts sequences uniformly over the sequence (resulting in sequences of equal lengths).

3. High-confidence prediction of the next surgical activity

We detail our approach in this section. We start by presenting our method for optimal sub-sequence matching in Section 3.1. In Section 3.2, we then show how to construct a discriminative model of the to-be-performed surgical actions; we also detail how to decide upon the situations in which we believe that uncertainty is too high to draw a high-confidence prediction.

3.1. Optimal sub-sequence matching

Let $\mathbb{S} = \{S_1, \cdots, S_N\}$ be the reference set of N sequences (surgeries), $S = \langle s_1, \cdots, s_l \rangle$ be one sequence of this set (a complete surgery), and $S^\star = \langle s^\star_1, \cdots, s^\star_k \rangle$ be a partial sequence (the ongoing surgery). Let us denote $S_{1,l'}$ a sub-sequence $\langle s_1, \cdots, s_{l'} \rangle$ of S. Our objective is to find the sub-sequence $S' = S_{1,l'}$ so that the cost of optimally registering the partial sequence $S^\star$ onto the sub-sequence S' of the reference sequence S is minimal.

Finding the cost of an optimal registration of one sequence onto another has been studied by the literature. The Dynamic Time Warping (DTW) similarity measure [16] makes it possible to find the optimal alignment of two sequences (and thus register them) in $\Theta(l_1 \cdot l_2)$ operations (with l_1 and l_2 the respective lengths of the realigned sequences), with regard to some scoring function. The consistency of this measure has been demonstrated for surgical processes in [5, 6] and is often used in medical applications [33].

In this section, we introduce a new objective function for finding the sub-sequence S' that best matches $S^\star$, and introduce a new algorithm, based on DTW, that can find S' in $O(k \cdot l)$ operations only (where k is the length of the prefix sequence to be match against another full sequence of length l).

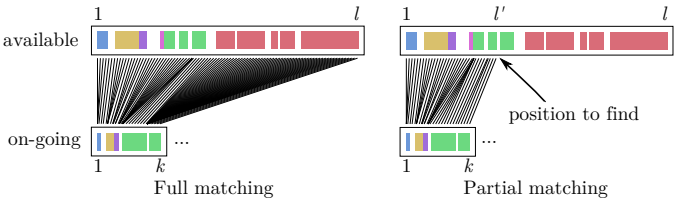


Figure 1: Illustration of the difference between a full (left) and partial (right) matching.

3.1.1. Sequence of surgical activities

$S = \langle s_1, \cdots, s_l \rangle$ is a sequence of surgical activities s_i represented by three elements: an action, an anatomical structure and an instrument, *e.g.*, (cut, skin, scalpel). In the following, we use the term *activity* to refer to these three elements. To be able to compare and align two sequences, it is mandatory to have a way to compare two single surgical activities. The simplest approach uses a binary distance which equals zero if two activities are identical, and at least one of their elements is different. However, this is not flexible as cutting the skin with a scalpel or with a surgical knife would be considered as doing two different activities. Consequently, we designed a specific metric that takes into account each element separately. In this paper, as we focus on predicting the activities of the right hand, we weighted each element of each activities by 1/3 (action, anatomical structure, instrument). It makes it possible to have a gradual evaluation of the similarity between two surgical activities, such as when the surgeon is targeting the same structure with two different instruments. More advanced metric can be designed if both hands are used or if the use of the microscope is considered [6].

3.1.2. Objective function

Our goal is to find the matching point l' in S that minimizes the optimal alignment between $S^\star$ and the subsequence $S_{1,l'}$:

$$match(S^\star, S) = \underset{1 \leqslant l' \leqslant l}{\arg\min}\ \mathrm{DTW}(S^\star, S_{1,l'}) \qquad (1)$$

Figure 1 presents the intuition about our objective function, compared to DTW's one.

Note that compared to other existing alignment technique like Smith-Waterman [34] there is no additive penalty for duplicating or skipping elements in DTW. We chose DTW (and its ability to sometime stretch subsequences) as we wanted to partially reduce the importance of actions' durations, but not to the point where we would only consider sequencing. Note also that the first element of both sequences have to be part of the resulting alignment, while it is not mandatory in Smith-Waterman. Figure 2 presents the trend of this objective function versus the value taken by l' on an example.

3.1.3. Efficient algorithm

An exhaustive search among all the possible matching points for l' will take $\Theta(\frac{l \cdot (l+1)}{2} \cdot k) = \Theta(l^2 \cdot k)$ oper-

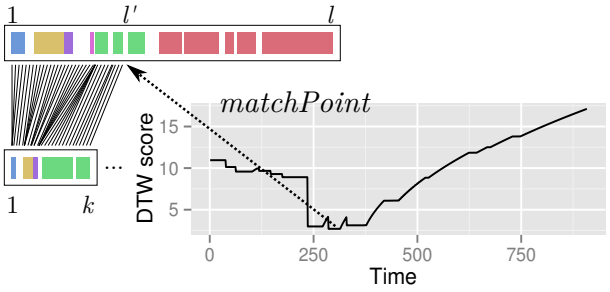


Figure 2: Illustration of the *matchPoint* resulting of the partial matching.

ations. Such a cubic complexity with the length of the matched sequences is incompatible with real-time matching, because a typical surgical procedure will often have more than 10,000 elements.

We now show how to modify the Dynamic Time Warping (DTW) algorithm to obtain an exact solution in $O(l \cdot k)$ operations without sacrificing the soundness of the process. There are three ideas in this algorithm:

1. Noticing that for all l', $DTW(S^\star, S_{1,l'})$ can be computed directly from the warping matrix constructed from $DTW(S^\star, S)$, by looking at the last column of each line of the matrix.
2. Remembering the smallest value of this last column.
3. Early abandoning the computation of the warping matrix if all values on one line are greater than $DTW(S^\star, S_{1,l'})$ for the best l' so far.

Put together, the two first elements make it possible to take the complexity from $\Theta(l^2 \cdot k)$ to $\Theta(l \cdot k)$, while the last element takes it to $O(l \cdot k)$. Note that with 10,000 elements, the difference in the complexity corresponds to an algorithm running more than 4 orders of magnitude faster than the naive solution. Our solution is presented in Algorithm 1.

The algorithm combines two main elements. First, we remember the score of DTW between each prefix of $S^\star$; this is done using the last **else if** statement (line 14). Second, we make sure that there is a possibility for $DTW(S_{1,l_2}, S^\star)$ to be lower than $DTW(S_{1,l'}, S^\star)$ with $S_{1,l'}$ be the best prefix of S found so far (and $l' < l_2$). We then use the fact that the value of $m[k, l']$ for any l' is greater than at least one cell in $m[\_, l'-1]$. It follows that at least one cell in the previous column has to be smaller than the current best alignment (for which the score is $m[k, l']$). We can then stop if this condition is not met.

Note that although this algorithm can be further optimized depending on δ (*i.e.*, the distance function between elements of the sequences), we chose here to give the algorithm for the general case. Furthermore, this adaptation of the algorithm did no alter the properties of optimality of DTW.

Algorithm 1 Optimal sub-sequence matching

Require: $S^\star = \langle s^\star_1, \cdots, s^\star_k \rangle$
Require: $S = \langle s_1, \cdots, s_l \rangle$
Let δ be a similarity between the elements of the sequences
Let $m[k, l]$ be a matrix storing partial costs
Let $l' \leftarrow 1$ be the matching point to find
1: $m[1,1] \leftarrow \delta(s^\star_1, s_1)$
2: **for** $i \leftarrow 2$ to k **do** $\{m[i,1] \leftarrow m[i-1,1] + \delta(s^\star_i, s_1)\}$
3: **for** $j \leftarrow 2$ to l **do** $\{m[1,j] \leftarrow m[1,j-1] + \delta(s^\star_1, s_j)\}$
4: **for** $j \leftarrow 2$ to l **do**
5: $\quad continue \leftarrow false$
6: $\quad$ **for** $i \leftarrow 2$ to k **do**
7: $\quad\quad m[i,j] \leftarrow \delta(s^\star_i, s_j) + \min(m[i-1,j], m[i,j-1], m[i-1,j-1])$
8: $\quad\quad$ **if** $not(continue) \wedge m[i,j] < m[k,l']$ **then**
9: $\quad\quad\quad continue \leftarrow true$
10: $\quad\quad$ **end if**
11: $\quad$ **end for**
12: $\quad$ **if** $not(continue)$ **then**
13: $\quad\quad$ **return** l'
14: $\quad$ **else if** $m[k,j] < m[k,l']$
15: $\quad\quad l' \leftarrow j$
16: $\quad$ **end if**
17: **end for**
18: **return** l'

3.2 A voting approach to draw high-confidence predictions

Our method uses the proposed optimal sub-sequence matching to draw predictions about the next surgical activity that will be performed. We will use the optimal sub-sequence matching from the on-going surgery $S^\star$ to every sequence S_i of $\mathbb{S}$. We can then use this information to draw a probability distribution $\hat{p}_{\text{next}}$ over the next possible state of the current surgery. More formally, the maximum likelihood estimate $\hat{p}_{\text{next}}$ for the next activity to be s given the previous activities $S^\star$ is:

$$\hat{p}_{\text{next}}(s|S^\star) = \frac{|\{S(match(S^\star, S) + 1) = s\}_{S \in \mathbb{S}}|}{|\mathbb{S}|} \quad (2)$$

Finally, we draw a prediction from the maximum a posteriori estimate of $\hat{p}_{\text{next}}$ using a majority vote [35], *i.e.*, select s for which $\hat{p}_{\text{next}}(s|S^\star) \geqslant 0.5$. We consider that if more than half of the predicted actions are similar, the system should perform a prediction. The goal of the majority vote is to perform a prediction only when a certain level of confidence is reached. In order to ensure and confer high-confidence to the system, we do not draw a prediction if no s obtains a majority or in case of ties. Note that $\hat{p}_{\text{next}}(s|S^\star)$ can be seen as an agreement rate on the prediction: a high value indicates an important agreement amongst the recorded surgeries about the next action that is going to be performed and conversely. The threshold on the agreement rate (0.5 in this paper) can be tuned according to need for a system performing very accurate

predictions but in limited number or a large number of predictions with an increased probability of errors. We have developed a web application [1] to allow the reader to try this prediction system easily. An open-source standalone implementation of the method is accessible at the same URL.

3.3. Weighting the predictions

Depending on the number of available recorded surgeries (*i.e.*, the size of $\mathbb{S}$), the number of predictions to combine for each action can be important. In this case, not all of the surgeries previously recorded will share a significant amount of common features with the ongoing surgery (S). Additionally, as we target real time prediction, our system has to stay highly efficient regardless of the number of surgeries in $\mathbb{S}$. Consequently, it is desirable to have a way to select the surgery from $\mathbb{S}$ and not use all of them for the prediction. For example, if we have 3 surgery in our set and the results of the prediction is action A for the first one and action B for the two others, the system will predict action B, as it has the majority. However, if we take into account the similarity between the ongoing surgery and the set of surgeries, it might be more accurate to predict A if the first surgery is highly similar to the ongoing surgery while the two others are far from to the ongoing surgery.

To cope with these issues, we propose to weigh the predictions provided by each partial alignment. The prediction scheme proposed in Equation (2) gives the same importance to each prediction. We propose to include an additional term to balance the influence of each prediction. Intuitively, we want to give more importance to the predictions provided by sequences which are highly similar to the ongoing surgery. To evaluate this similarity, we propose to rely on the DTW score given by the partial alignment performed by Algorithm 1. Indeed, the matrix m contains the minimal cost of the partial alignment along with the *matchPoint*. Thus, the *quality* of the partial alignment can be evaluated using this score. Using this process, we ensure that if the activities performed so far in the ongoing surgery are highly similar to the activities of a recorded surgery s, the prediction coming from s will have an important weight:

$$\hat{p}_{\text{next}}(s|S^\star) \propto \sum_{S\in\mathbb{S}} 1_{S(match(S^\star,S)+1)=s} \cdot e^{-\lambda \cdot DTW(S^\star, S_{1,match(S^\star,S)})} \tag{3}$$

Where $S_{1,match(S^\star,S)} = \langle S_1, \cdots, S_{match(S^\star,S)} \rangle$ is the subsequence of S up to the matching point (included) and $1_{S(match(S^\star,S)+1)=s}$ is an indicator function that takes value one when its subscripted condition is true and zero otherwise. We do not detail the normalization factor for readability. The λ parameter is used to control the importance given to the score in combining the results. A high λ value will give more importance to highly similar sequences. In the experiment, λ was set to 1 as the lengths of the sequences were limited. We then normalize the weight by dividing the sum for each next activity by the total sum for all next possible activities. As proposed with the majority voting scheme, a threshold can be used to perform a prediction only when a high confidence is reached. In the following, we keep the threshold of 0.5.

4. Experiments

4.1. Clinical data

We evaluated our framework on two datasets composed of two types of surgical procedures: Lumbar Disc Herniation (LDH) surgery and Anterior Cervical Discectomy (ACD) surgery. Figure 3 presents an extract of the LDH surgeries and Figure 4 an extract of the ACD surgeries. The legends illustrate the most common actions in the respective dataset. The white spaces correspond to times when the surgeon was not performing any action.

The LDH dataset is composed of 24 lumbar disc herniation surgeries recorded at the Neurosurgery Department of site A. Surgeries contain on average 680 actions. For this surgery, the list of actions is: *cut, coagulate, hold, dissect, install, remove, irrigate, sew, swab* and *drill*. The list of anatomical structures is: *skin, fascia, muscle, vertebra, ligament, duramater, nerveroot* and *disc*. And the list of surgical instruments is: *scalpel, scissors, dissectors, rongeurs, hooks, high-speed drill, suction tube, needle-holders, saline solution, retractors* and *forceps*. As all triples are not present (some triples of action, instrument, anatomical structure are irrelevant), our dataset contains only 108 different activities.

The surgeries involved 10 male and 14 female patients, with a median age of 52 years. These were exclusively patients with newly diagnosed disc hernia, no patient had undergone previous lumbar spine surgery which could increase surgical difficulties due to fibrosis. These lumbar disc surgeries are divided into three main steps: (1) approach of the disc, (2) discectomy and (3) closure. The herniated disc was approached via a posterior intermyolamar route. The patients were operated on by five junior and five senior surgeons. Senior surgeons have performed at least a hundred removals of lumbar disc herniation. All the junior surgeons have passed more than two years of their residency program but have only performed a few removals of lumbar disc herniation. We focused on the closure phase, because it allowed us to ensure that the main surgeon was the one operating (for a junior surgery, his or her senior sometimes takes over the surgery). We only considered the recording of the right hand which is the most active body part used to perform the most important activities.

The ACD dataset is composed of 18 anterior cervical discectomy surgeries recorded at Neurosurgery Department of site B. Surgeries contain on average 511 actions.

[1] http://germain-forestier.info/src/aiim2017/

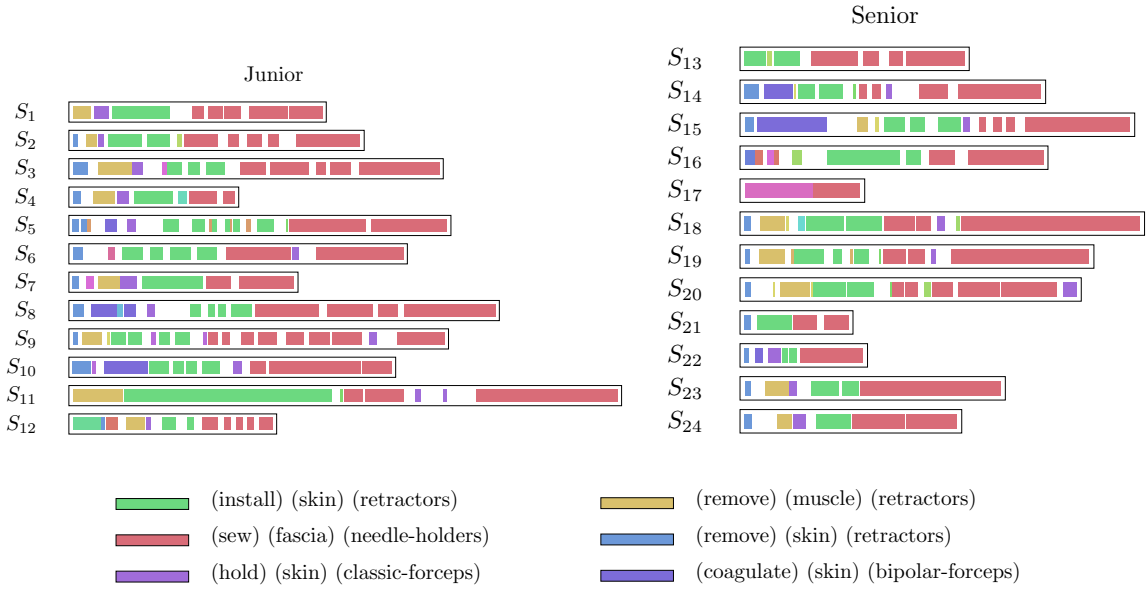


Figure 3: An extract of the first dataset of 24 LDH surgeries used for the experiments and the legend for the six most frequent actions.

For this surgery, the list of actions is: *cut, swab, sew, coagulate, install, dissect, irrigate, drill, remove* and *hold*. The list of anatomical structures is: *muscle, vertebra, skin, fascia, disc* and *ligament*. And the list of surgical instruments is: *scalpel, scissors, dissectors, rongeurs, hooks, high-speed drill, suction tube, needle-holders, saline solution, retractors, curettes, bonewax, drainage, drape, fluoroscopy* and *forceps*. The dataset contains 82 different activities. During this procedure, a cervical disc can be removed through an anterior approach. This means that surgery was done through the front of the neck as opposed to the back of the neck. A 1-level ACD surgical procedure can be decomposed into four major phases, whereas a fifth one may be necessary. These four phases are: the approach, the discectomy, the arthrodesis, and the closure phases. An additional phase of hemostasis may be mandatory in certain cases. The patients were operated on by two expert and two intermediate surgeons (refereed as junior in the following). As for the LDH dataset, we also focused on the closure phase and the right hand activities.

While it is impossible that this limited dataset could represent all types of LDH and ACD surgeries, the surgeons involved in this study indicated that these interventions contain typical surgical behavior for these types of surgeries.

4.2. Methodology

For both datasets, we compared three configurations: using procedures performed by the senior only, by the junior only and using all surgeries. Our aim was to observe the influence of the available surgeries (training data) on the quality of the predictions that were drawn. A leave-one-out cross-validation approach was used for each configuration: we selected one surgery out of the set of surgeries, and used it as the on-going intervention (this surgery was then removed from the set of reference surgeries). The left-out surgery was used to test our predictions, as if it was progressively discovered. Predictions were made every 5% of the progression of the intervention. Note that depending of the total duration of the intervention, the 5% can represent different durations. This choice has been made to perform the same number of predictions, regardless of the total duration of the intervention. We could then compare every prediction with the *actual* activity of the surgery. Every surgery was in turn considered as the on-going intervention.

We evaluated our system using the precision $\mathcal{P}$ (*i.e.*, number of good predictions / total number of predictions) and the recall $\mathcal{R}$ (*i.e.*, number of predictions / total number of expected predictions). We also used the F-measure $\mathcal{F}$ (harmonic mean between prediction and recall) to provide an overall evaluation. We compared the results of our method to the ones of the Euclidean state-of-the-art method [32]. We used the exact same process, but replaced the optimal sub-sequence matching with uniform scaling [32]. Uniform scaling performs a linear transformation that linearly stretches or contracts sequences uniformly over the whole sequence. We decided to use uniform scaling as a competitor for our approach, because the Euclidean distance is often used as the competitor of DTW to motivate the need for using a time warping approach.

4.3. Results on Lumbar Disc Herniation (LDH) dataset

Figure 5 presents the general results for the Lumbar Disc Herniation (LDH) dataset on the three configurations (Junior+Senior, Junior, Senior). We compared both

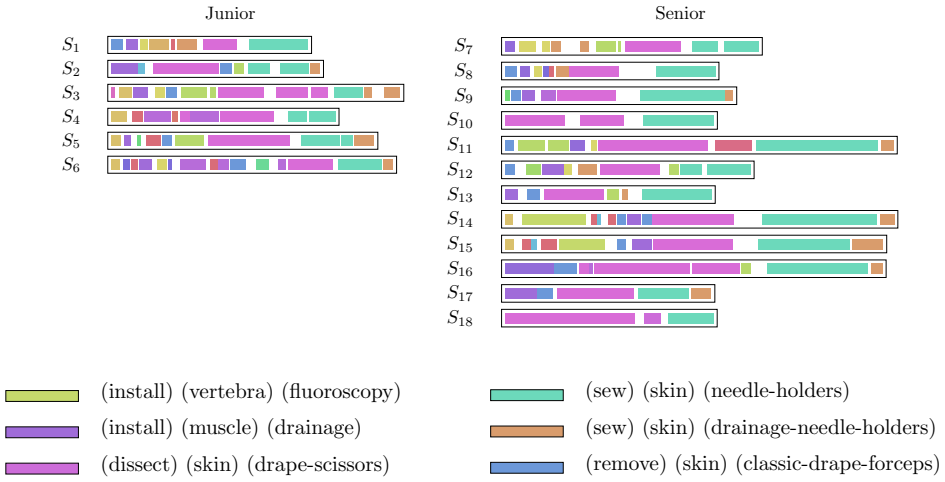


Figure 4: An extract of the second dataset of 18 ACD surgeries used for the experiments and the legend for the six most frequent actions.

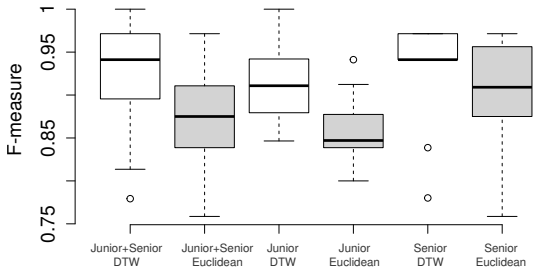


Figure 5: Results on the three configurations (Junior+Senior, Junior, Senior) for the two methods (DTW in white, Euclidean in gray) for the LDH dataset.

methods in terms of F-measure. We can see that our time-warping approach outperformed the state-of-the-art Euclidean approach, regardless of the considered configuration. The compact dispersion of the results for the senior case (as compared to the junior case) suggests that seniors have a more homogeneous behavior than junior surgeons, which is consistent with previous studies comparing junior and senior practices [5, 36, 37]. This trend was also observed in previous works using sequence of surgical activities [10, 5, 6]. This result also illustrates the influence of the set of available recordings in the quality of the prediction. Even though mixing all the surgeries together provided very good results, the best results were obtained for senior surgeons, whose surgical practice is usually more standardized and homogeneous. This supports our intuition that the more dedicated the training data is to the operating surgeon, the more accurate the predictions will be.

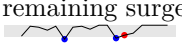
Table 1 details the prediction results for each of the 24 surgeries (using the 23 remaining surgeries as the training set). A sparkline (*e.g.*, ) presents, for each sequence, the evolution of the agreement rate among the

Table 1: Detailed results for every surgery of the LDH dataset; results with $\mathcal{F} \geqslant .9$ are shown in boldface.

Junior

Surg.	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	Agreement
S_1	1.000	0.944	**0.971**	
S_2	0.938	0.889	**0.913**	
S_3	0.846	0.722	0.779	
S_4	1.000	0.944	**0.971**	
S_5	0.941	0.944	**0.943**	
S_6	1.000	0.944	**0.971**	
S_7	0.882	0.944	**0.912**	
S_8	0.941	0.944	**0.943**	
S_9	0.875	0.889	0.882	
S_{10}	1.000	0.833	**0.909**	
S_{11}	1.000	1.000	**1.000**	
S_{12}	0.929	0.778	0.847	

Senior

Surg.	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	Agreement
S_{13}	1.000	0.889	**0.941**	
S_{14}	0.941	0.944	**0.943**	
S_{15}	1.000	0.833	**0.909**	
S_{16}	1.000	0.722	0.839	
S_{17}	1.000	0.833	**0.900**	
S_{18}	1.000	0.889	**0.941**	
S_{19}	1.000	0.944	**0.971**	
S_{20}	0.933	0.833	0.881	
S_{21}	1.000	0.944	**0.971**	
S_{22}	0.941	0.944	**0.943**	
S_{23}	1.000	1.000	**1.000**	
S_{24}	0.75	0.889	0.814	

predictions over the course of the surgery. The gray rectangle represents the interval $(0.5, 1]$ for which a majority is obtained. The blue dots represent the cases where our system did not provide a prediction (because no majority was obtained), while the red dots represent the inaccurate predictions. All other elements on the lines correspond to cases where our method predicted the next task accurately. The precision of our system is very high: in more than half of the surgeries, no mistake was every committed. Overall, our systems exhibits an average precision of 95%: our predictions do not eventuate 5% of the times only.

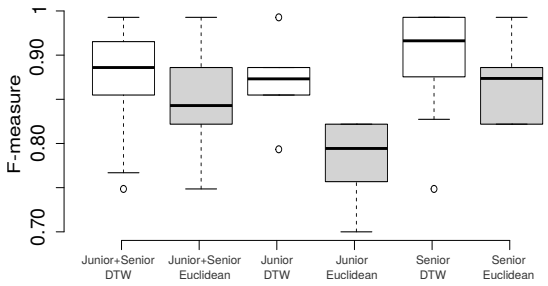


Figure 6: Results on the three configurations (Junior+Senior, Junior, Senior) for the two methods (DTW in white, Euclidean in gray) for the ACD dataset.

Moreover, our system provides a prediction 89% of the times (recall). This means that for the vast majority of cases, an agreement can be reached and a decision made. Furthermore, the consistency of our voting procedure is confirmed: for all the cases where the MAP (maximum a posteriori) estimate was below the majority threshold, and for which we thus did not provide a prediction (*i.e.*, blue dots in Table 1 – *Agreement* column), the MAP estimate was actually wrong. This confirms the relevance of our approach, by showing that we actually do not provide a prediction when no reliable choice can be made from the training set. This corresponds to the case where not enough similarity can be found between the on-going surgery and the reference set, which can be the case if specific activities are required during surgery. The highest number of errors were committed in S_{24} with a sequence of four wrong predictions in a row. This corresponds to the green activity in Figure 3, where the surgeon installed the retractors on the skin without stopping, while all the other surgeries exhibited several pauses. Finally, every prediction was made in less than 200 ms, which is compatible with real-time prediction in the OR.

4.4. Results on Anterior Cervical Discectomy (ACD) dataset

Figure 6 presents the general results for the Anterior Cervical Discectomy (ACD) dataset on the three configurations (Junior+Senior, Junior, Senior). We compared both methods in terms of F-measure. In this second dataset, our approach also outperformed the state-of-the-art Euclidean approach, regardless of the considered configuration. The Euclidean approach provided particularly poor results for the junior case. This can be explained by the limited amount of data available for this configuration, as only six sequences were available. The result is that it makes it difficult to find highly similar sequences in the reference set. Conversely, our method is able to operate non-linear distortions of the time axis, and hence to absorb some time variations; we posit that this help reduce the variance of the error for our system. Our approach, even with this limited amount of data managed to provide very good results with a precision of 87%. Similarly to the results obtained with the LHD dataset, the best results were obtained for the senior case, where behaviours are more homogeneous. When combining both junior and senior surgeries, the results of the Euclidean approach increased to a precision of 77% but our method still performed the best results with the precision reaching 93%.

Table 2: Detailed results for every surgery of the ACD dataset; results with $\mathcal{F} \geqslant .9$ are shown in boldface.

Junior				
Surg.	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	Agreement
S_1	1.000	0,778	0,870	
S_2	0,846	0,722	0,779	
S_3	0,900	0,556	0,687	
S_4	0,824	0,944	0,880	
S_5	1.000	0,722	0,839	
S_6	1.000	0,500	0,667	
Senior				
Surg.	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$	Agreement
S_7	1.000	0,556	0,714	
S_8	1.000	0,556	0,714	
S_9	1.000	0,833	**0,909**	
S_{10}	1.000	0,778	0,875	
S_{11}	1.000	0,778	0,875	
S_{12}	0,867	0,833	0,850	
S_{13}	1.000	0,667	0,800	
S_{14}	1.000	0,667	0,800	
S_{15}	1.000	0,611	0,759	
S_{16}	0,941	0,944	**0,943**	
S_{17}	1.000	0,778	0,875	
S_{18}	1.000	0,889	**0,941**	

Table 2 details the prediction results for every one of the 18 surgeries (using the 17 remaining surgeries as the training set). The precision of our system is very high: in 13 out of 18 surgeries, all of our predictions were correct. The overall precision is of 93% for this dataset. For example, there are only two senior surgeries S_{14} and S_{15} that used the activity "install, vertebra, fluoroscopy" (olive green in Figure 4) at the beginning of the phase. This makes a prediction difficult because we cannot obtain a majority vote. In this situation, our system did not perform prediction; this can be seen with the series of blue dots for S_{14} and S_{15} in Table 2. Our method is able to consistently automatically decide when *not* to providing a prediction, when it assesses that the current conditions are too dependent upon the current surgery. One can note that we currently consider all actions to be of equivalent importance in the computation of the accuracy. While it would be interesting to know if the system is able to predict the most important actions, the level of importance of a single action in the context of an entire surgery is still very difficult to assess.

4.5. Results with prediction weighting

Note that in the previous experiments, we assumed that all surgeries were relevant to make the predictions. This could be an issue for larger datasets because, intuitively, we would like to discard the surgeries that 'look' too different to the one being acquired. This is the basis for our weighting scheme: we construct a model where the prediction is influenced mostly by its close neighborhood (in our metric space). In some sense, this allows us to construct a model that is local and specific to the target

sequence. We present below our first results; we believe that the interest for this method will grow as the size of the datasets increases.

As presented in Section 3.3, an alternative to majority voting is the weighting of the prediction according the DTW score obtained from the partial matching. To evaluate this weighting scheme, we performed experiments on the LDH dataset (the largest available) where we replaced majority voting by prediction weighting. In this experiment, the λ parameter (see Equation (2)) was set to 1. Note that this parameter controls for the size and spread of the neighborhood that we want to consider as relevant for the predictions. Table 3 presents the results obtained on the LDH dataset. For the three configurations, using the weight instead of majority voting enabled an increase of the precision of the predictions. The amount of the increase is limited (between 1 to 2%) but allowed for each configuration to go beyond 96% of correct predictions. For the configuration with both junior and senior sequences, the recall value is higher when using prediction weights compared to using majority voting, which indicates that the prediction are both more precise and that we are able to predict the next task more often. The high recall values for the Euclidean method have to be balanced by their low precision resulting in lower F-measure for all configurations. These results confirm the relevance of weighting the predictions according to the similarity between the ongoing surgery and the reference surgeries. Furthermore, we believe that such technique will be more and more consistent as the size of the dataset increases, because there will compulsorily be surgeries that do not resemble the target one (and which we would then want to discard).

5. Conclusion

This work shows that it is possible to predict the next surgical task accurately. Our predictions are drawn from a low-level description of surgeries, without other source of information, and assume that the current surgery is consistent with the ones constituting the training set. Our contributions include (1) a definition of the objective function for the registration of a partial sequence to a complete reference sequence, (2) an efficient algorithm, based on DTW, to optimally minimize the above-mentioned objective function and (3) a prediction system that combines our optimal sub-sequence matching with MAP estimation and filtering. We also showed that our method is able to assess when the predictions are inconsistent and decide that it is not reasonable to make a prediction. Finally, we also introduced a new way for weighting the prediction to take into account the similarity between the ongoing surgery and the reference surgeries. Experiments on two datasets have shown that our method outperforms the state of the art and provides a prediction with high accuracy.

The fact that predicting surgical tasks is so central to the new generation of computer assisted surgery systems naturally opens up a number of clinical applications. We have mentioned in the introduction how this information can help ensuring a smooth running of the surgical procedure. Another application concerns the training of junior surgeons, where our system could be integrated in a simulation environment in order to provide help and feedback to the junior surgeon [38]. Our system could, on demand, provide a warning to the surgeon about his or her deviation from the standard practice of his or her colleagues. In future work, we want to validate this method on a more important dataset ($>$ 300 surgeries) and use our recent work on Dynamic Time Warping [39, 40] to improve the predictions. Furthermore, one possible extension of our system would be to use transition probabilities between surgical actions as a priori knowledge to refine and filter the predictions. We will also investigate the use of contextual information, like patient vital signs, in the prediction method.

Supplementary materials

The source code for the proposed method is available at http://germain-forestier.info/src/aiim2017/ (Accessed: January 2017)

Acknowledgments

This work was supported by the Australian Research Council under award DE170100037. This material is based upon work supported by the Air Force Office of Scientific Research, Asian Office of Aerospace Research and Development (AOARD) under award number FA2386-16-1-4023.

References

[1] Bricon-Souf N, Conchon E. Context awareness for medical applications. Medical Applications of Artificial Intelligence 2013;:355.

[2] Patel VL, Shortliffe EH, Stefanelli M, Szolovits P, Berthold MR, Bellazzi R, et al. The coming of age of artificial intelligence in medicine. Artificial Intelligence in Medicine 2009;46(1):5–17.

[3] Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. International Journal of Medical Informatics 2008;77(2):81–97.

[4] Lalys F, Jannin P. Surgical process modelling: a review. International Journal of Computer Assisted Radiology and Surgery 2013;8(5):1–17.

[5] Forestier G, Lalys F, Riffaud L, Collins DL, Meixensberger J, Wassef SN, et al. Multi-site study of surgical practice in neurosurgery based on surgical process models. Journal of Biomedical Informatics 2013;46(5):822 –9.

[6] Forestier G, Lalys F, Riffaud L, Trelhu B, Jannin P. Classification of surgical processes using dynamic time warping. Journal of Biomedical Informatics 2012;45(2):255–64.

[7] Lalys F, Riffaud L, Morandi X, Jannin P. Automatic phases recognition in pituitary surgeries by microscope images classification. In: Information Processing in Computer-Assisted Interventions; vol. 6135 of *LNCS*. Springer; 2010, p. 34–44.

[8] Franke S, Meixensberger J, Neumuth T. Intervention time prediction from surgical low-level tasks. Journal of Biomedical Informatics 2013;46(1):152 –9.

[9] Meißner C, Meixensberger J, Pretschner A, Neumuth T. Sensor-based surgical activity recognition in unconstrained environments. Minimally Invasive Therapy & Allied Technologies 2014;(0):1–8.

Table 3: Results of LDH dataset for the different configurations with and without prediction weighting.

Configuration	Method	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}$
Junior+Senior	DTW	0,955 (0,064)	0,894 (0,077)	0,921 (0,057)
	Euclidean	0,798 (0,090)	**0,995** (0,016)	0,883 (0,054)
	Weighted DTW	**0,964** (0,061)	0,896 (0,082)	**0,926 (0,056)**
Junior	DTW	0,948 (0,064)	0,894 (0,077)	**0,918** (0,052)
	Euclidean	0,769 (0,074)	**0,977** (0,037)	0,858 (0,042)
	Weighted DTW	**0,963** (0,063)	0,856 (0,093)	0,903 (0,059)
Senior	DTW	0,968 (0,077)	0,898 (0,070)	0,929 (0,059)
	Euclidean	0,828 (0,101)	**0,995** (0,016)	0,901 (0,062)
	Weighted DTW	**0,970** (0,073)	0,912 (0,069)	**0,937** (0,052)

[10] Forestier G, Petitjean F, Riffaud L, Jannin P. Non-linear temporal scaling of surgical processes. Artificial Intelligence in Medicine 2014;62(3):143 –52.

[11] Rogers SO, Gawande AA, Kwaan M, Puopolo AL, Yoon C, Brennan TA, et al. Analysis of surgical errors in closed malpractice claims at 4 liability insurers. Surgery 2006;140(1):25–33.

[12] Dlouhy BJ, Rao RC. Surgical skill and complication rates after bariatric surgery. The New England Journal of Medicine 2014;370(3):285–.

[13] Regenbogen SE, Greenberg CC, Studdert DM, Lipsitz SR, Zinner MJ, Gawande AA. Patterns of technical error among surgical malpractice claims: an analysis of strategies to prevent injury to surgical patients. Annals of surgery 2007;246(5).

[14] MacKenzie L, Ibbotson J, Cao C, Lomax A. Hierarchical decomposition of laparoscopic surgery: a human factors approach to investigating the operating room environment. Minimally Invasive Therapy & Allied Technologies 2001;10(3):121–7.

[15] Mehta N, Haluck R, Frecker M, Snyder A. Sequence and task analysis of instrument use in common laparoscopic procedures. Surgical endoscopy 2002;16(2):280–5.

[16] Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech and Signal Processing 1978;26(1):43–9.

[17] Forestier G, Petitjean F, Riffaud L, Jannin P. Optimal subsequence matching for the automatic prediction of surgical tasks. In: AIME 15th Conference on Artificial Intelligence in Medicine; vol. 9105. Springer; 2015, p. 123–32.

[18] Riffaud L, Neumuth T, Morandi X, Trantakis C, Meixensberger J, Burgert O, et al. Recording of surgical processes: a study comparing senior and junior neurosurgeons during lumbar disc herniation surgery. Neurosurgery 2010;67:325–32.

[19] Padoy N, Blum T, Ahmadi SA, Feussner H, Berger MO, Navab N. Statistical modeling and recognition of surgical workflow. Medical Image Analysis 2012;16(3):632–41.

[20] Bouarfa L, Jonker PP, Dankelman J. Discovery of high-level tasks in the operating room. Journal of Biomedical Informatics 2011;44(3):455–62.

[21] Lalys F, Riffaud L, Bouget D, Jannin P. A framework for the recognition of high-level surgical tasks from video images for cataract surgeries. IEEE Transactions on Biomedical Engineering 2012;59(4):966–76.

[22] Bardram JE, Doryab A, Jensen RM, Lange PM, Nielsen KL, Petersen ST. Phase recognition during surgical procedures using embedded and body-worn sensors. In: IEEE International Conference on Pervasive Computing and Communications. 2011,.

[23] Stauder R, Okur A, Peter L, Schneider A, Kranzfelder M, Feussner H, et al. Random forests for phase detection in surgical workflow analysis. In: Information Processing in Computer-Assisted Interventions. Springer; 2014, p. 148–57.

[24] Reiley CE, Hager GD. Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009. Springer; 2009, p. 435–42.

[25] Reiley CE, Hager GD. Decomposition of robotic surgical tasks: an analysis of subtasks and their correlation to skill. In: M2CAI Workshop, Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009. 2009,.

[26] Despinoy F, Bouget D, Forestier G, Penet C, Zemiti N, Poignet P, et al. Unsupervised trajectory segmentation for surgical gesture recognition in robotic training. IEEE Transactions on Biomedical Engineering 2015;.

[27] Liu Z, Hauskrecht M. Clinical time series prediction with a hierarchical dynamical system. In: Conference on Artificial Intelligence in Medicine. Springer; 2013, p. 227–37.

[28] Huang Z, Lu X, Gan C, Duan H. Variation prediction in clinical processes. In: Conference on Artificial Intelligence in Medicine. Springer; 2011, p. 286–95.

[29] Bermejo P, Lucas M, Rodríguez-Montes JA, Tárraga PJ, Lucas J, Gámez JA, et al. Single-and multi-label prediction of burden on families of schizophrenia patients. In: Conference on Artificial Intelligence in Medicine. Springer; 2013, p. 115–24.

[30] Neumann J, Rockstroh M, Vinz S, Neumuth T. Surgical workflow and process modeling. In: M2CAI Workshop, Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015. 2015,.

[31] Becker J, Breuker D, Delfmann P, Matzner M. Designing and implementing a framework for event-based predictive modelling of business processes. In: EMISA. 2014, p. 71–84.

[32] Yankov D, Keogh E, Medina J, Chiu B, Zordan V. Detecting time series motifs under uniform scaling. In: International Conference on Knowledge Discovery and Data mining. ACM; 2007, p. 844–53.

[33] Tormene P, Bartolo M, De Nunzio AM, Fecchio F, Quaglini S, Tassorelli C, et al. Estimation of human trunk movements by wearable strain sensors and improvement of sensor's placement on intelligent biomedical clothes. Biomedical engineering Online 2012;11(1):95.

[34] Smith TF, Waterman MS. Identification of common molecular subsequences. Journal of molecular biology 1981;147(1):195–7.

[35] Kittler J, Hatef M, Duin RP, Matas J. On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence 1998;20(3):226–39.

[36] Schumann S, Bühligen U, Neumuth T, et al. Distance measures for surgical process models. Methods Inf Med 2013;52(5):422–31.

[37] Pavlidis I, Tsiamyrtzis P, Shastri D, Wesley A, Zhou Y, Lindner P, et al. Fast by nature-how stress patterns define human experience and performance in dexterous tasks. Scientific Reports 2012;2.

[38] Zhou Y, Bailey J, Ioannou I, Wijewickrema S, O'Leary S, Kennedy G. Pattern-based real-time feedback for a temporal bone simulator. In: Symposium on Virtual Reality Software and Technology. ACM; 2013, p. 7–16.

[39] Petitjean F, Forestier G, Webb G, Nicholson A, Chen Y, Keogh E. Dynamic Time Warping averaging of time series allows faster and more accurate classification. In: IEEE International Conference on Data Mining. 2014, p. 470–9.

[40] Petitjean F, Forestier G, Webb GI, Nicholson AE, Chen Y, Keogh E. Faster and more accurate classification of time series by exploiting a novel Dynamic Time Warping averaging algorithm. Knowledge and Information Systems 2016;47(1):1–26.