# Simulation of the Evolution of a Virtual Patient's Physiological Status in the Operating Room: Application to Computer-assisted Anaesthesia Training

Hugo Boisaubert[1], Lucas Vincent[1], Corinne Lejus-Bourdeau[2,3] and Christine Sinoquet[1]
[1]*Research Laboratory of Digital Science in Nantes (LS2N) / UMR CNRS 6004, University of Nantes, 2 rue de la Houssinière, Nantes, France*
[2]*Experimental Universitary Laboratory for Simulation in Intensive Care (LESiMU) in Nantes, 9 rue Bias Ricordeau, Nantes, France*
[3]*Department of Anaesthesia and Intensive Care, Nantes University Hospital, 1 place Alexis Ricordeau, Nantes, France*

Keywords: Computer-assisted Medical Training, Virtual Patient, Operating Room, Anaesthesia, Reactive Scenario, Simulation, Prediction, Case-based Reasoning, Data Mining, Pattern Recognition, E-health Record, Event Trace, Multivariate Time Series.

Abstract: Half a million surgeries are performed every day around the world, which places safety and quality at the heart of global health issues. In this context, we introduce a novel approach, SVP-OR (Simulation of Virtual Patient at the Operating Room), designed for digital training support. For this purpose, we must evolve the physiological parameters of a virtual patient submitted to the actions of a user (trainee), and of a virtual medical team. We formulate the problem as a case-based reasoning approach in which (i) we identify real patients whose anaesthetic profiles show a region similar to the recent history of the virtual patient and (ii) we predict the near future of the virtual patient (a multivariate time series) from the multivariate time series of the most similar real patients. The first contribution in this paper consists in the design of a contextualized multidimensional pattern recognition approach. Our second contribution is the development of a generic framework based on the concept of contextualized multidimensional pattern, to predict the evolution of the virtual patient. In a third contribution, we instantiate our framework, and we evaluate and compare the realism of two predictive strategies.

## 1 INTRODUCTION

Over the world, the network of simulation platforms in intensive care, hosted by university hospitals, aims at training medical interns and nurses, as well as more experienced physicians. In particular, the Experimental Universitary Laboratory for Simulation in Intensive Care (LESiMU) in Nantes offers training in seven medical specialities, including anaesthesia. For this purpose, LESiMU uses high-fidelity patient simulators, namely mannequins, with immersion of the trainees in a full-scale interprofessional medical team. Currently, the training scenarios are written in advance, the trainer makes himself evolve the physiological parameters of the mannequin "by hand", in response to the actions of the medical team (including those of the trainee immersed in this team). There is therefore little variability in the scenarios.

Our work is **motivated** by a real need expressed by LESiMU. In order to improve the safety and quality of intraoperative care, the trainers at LESiMU wish to **vary the diversity of the scenarios** to be proposed to anaesthesia interns and nurse anesthetists, in initial training, as well as to more experienced practitioners. We proposed to develop a digitally assisted modality, based on the database of anaesthetic profiles recorded by Nantes University Hospital since 2004. In this modality, we wish the trainee to run a software program consisting in a generator of reactive scenarios. The application is parameterized with the surgery of interest, together with the age, gender, weight and medical history (*e.g.*, diabetes) of the virtual patient to be simulated. The other members of the medical team are simulated in a very simple way (icons carrying out actions and transmitting information, on the screen of trainee's laptop). Serving this training objective also responds in the long run to the need for



Boisaubert, H., Vincent, L., Lejus-Bourdeau, C. and Sinoquet, C.
Simulation of the Evolution of a Virtual Patient's Physiological Status in the Operating Room: Application to Computer-assisted Anaesthesia Training.
DOI: 10.5220/0010981000003123
In *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 5: HEALTHINF*, pages 228-239
ISBN: 978-989-758-552-4; ISSN: 2184-4305

anticipation through prediction, to estimate a risk at the operating room. This objective is inherent to the emerging paradigm of personalized medicine.

To generate reactive scenarios, one has to predict the evolution of the physiological parameters of the virtual patient after an action has been triggered by the medical team. More than 500,000 anaesthetic profiles have been collected since 2004, at Nantes University Hospital. Such anaesthetic data include a multivariate time series and an action trace. The multivariate time series describes the evolution of the physiological parameters (*e.g.*, blood pressure, oxygen saturation, heart rate) of a real patient, during their monitoring in the operating room. Hereinafter, we will also use the term ”variables” to refer to the physiological parameters. The action trace records the actions triggered by the medical team. Given a cohort of patients operated on for the same surgery, and in the same ”class” as the virtual patient to be simulated (age, gender, weight and medical history), the time series and action traces of the cohort contain knowledge to predict the evolution of the virtual patient after a given action has been triggered (For instance, a class may correspond to a 25-35 year old male weighting between 75 and 85 kg and with no medical antecedents). We can capture this knowledge envisaging either a modelling approach or a data mining approach.

In this paper, we present a case-based reasoning (CBR) approach. We consider the available appropriate cohort of real patients, that is patients with same surgery, age, gender, weight and medical history as the virtual patient to be simulated. We call **reactive simulation scenario** the dynamical process in which a user (the trainee) makes evolve a virtual patient. At the end of the reactive scenario, we obtain a multivariate time series together with the trace of actions carried out by both trainee and medical team until surgery completion.

In a case-based reasoning framework, we have to identify real patients whose historical data show a region similar to the recent history of the virtual patient. The historical data of a patient embraces a multivariate time series and an action trace. We have to handle two pattern recognition problems, one where the query is a sequence of time-stamped actions, the other one where the query is a multivariate time series. Finally, for a reactive simulation scenario, we must meet a real-time constraint.

In this paper, our first contribution consists in the design of a contextualized multidimensional pattern recognition approach, to identify real patients most similar in some region of their anaesthetic profiles to the simulated patient. Our second contribution is the design of a generic method to predict the evolution of the virtual patient, based on the concept of contextualized multidimensional pattern. In a third contribution, we instantiate our generic framework, and we describe, evaluate and compare the realism of two predictive strategies.

The remainder of the paper is organized as follows. Section 2 briefly mentions some related work. Section 3 describes the two main components of our CBR approach, that is (i) a contextualized pattern recognition method and (ii) a prediction strategy of short-term evolution for the virtual patient. This section ends with the description of the general CBR-based algorithm SVP-OR (Simulation of Virtual Patient at the Operating Room). Section 4 presents an evaluation of two instantiations of SVP-OR. Section 5 concludes and opens up future directions of research.

From now on, we denote $p$ the number of variables described in any anaesthetic profile of the patient cohort.

## 2 RELATED WORK

The motivation for our work is to implement short-term prediction on the fly of multivariate time series, in order to handle reactive scenarios simulating a surgical operation. Although other works in the literature may be related to our proposal, no such approach has ever been proposed.

These recent years, **digital training assistance** has entered the hospital. The range of assistance offered extends from the provision of scenarios to be replayed on computer from a selection of real scenarios, to the immersion in a virtual operating room (OR) (Nagendran et al., 2013; Qi et al., 2021).

Beside manual annotation, artificial intelligence has much to contribute to knowledge extraction from OR data. **Machine learning** techniques allow the automatic detection and prediction of surgical activities, for instance using Hidden Markov models (Meißner et al., 2014) or random forests (Stauder et al., 2014). In particular, deep learning can recognize various types of surgical procedures from videos (Khalid et al., 2020).

In the field of **data mining**, the literature reports work on the exploitation of low-level surgical tasks to predict the possible surgeons' subsequent tasks (Forestier et al., 2017), as well as interventional time (Franke et al., 2013). Works in the line of (Erdogan and Tarhan, 2018) apply process mining techniques to logs obtained in a healthcare context, with a focus on process discovery, conformance checking and process enhancement.

Much work has been done on multivariate **time series forecasting** (see for instance the recent survey in (Liu et al., 2021)). In our application, medical actions may exert synergistic or antagonistic effects on some of the patient's physiological parameters. Under these conditions, designing and training a prediction model based on the joint modelling of time series and event traces are highly challenging.

The case-based reasoning approach described in this paper relies on the concept of local similarity between real patients and a digital patient. A number of works reported in the literature deal with the notion of **similarity between patients**. Importantly, the massive and systematic use of electronic medical data has laid the foundation for personalized medicine. In this field, various definitions of global and local similarities have therefore been proposed; they leverage a variety of components of patient data, to apply clustering or classifying on patients, such as in (Ng et al., 2015; Brown, 2016; Parimbelli et al., 2018; Wang et al., 2019; Fang et al., 2021). However, in such works, prediction is focused on risk, survival time, for instance, and not on the dynamic behaviour of physiological parameters, as required by our application.

Among various artificial intelligence tracks, **case-based reasoning** can be considered as a form of similarity-based or analogical reasoning. Case-based reasoning (CBR) has been used extensively for diagnosis, classification, recommendation and therapy planning in medicine. The reader is directed to (Choudhury and Begum, 2016), specifically pages 138 to 140, for illustrations. The reader is also referred to (Goel and Diaz-Agudo, 2017) for a recent comprehensive overview on the developments in the field of CBR. So far, few works have exploited knowledge on patients' temporal data in the CBR framework. For instance, the work reported in (Ganzinger et al., 2019) organizes patient temporal data in a time-graph structure to calculate temporal similarities of disease progress among patients. In this case, the temporal data consist of medical events. In (Sha et al., 2016), the authors adapt a temporal similarity measure to the case of irregularly measured data. The recent work in (Mülâyim and Arcos, 2020) tackles CBR in presence of millions of patients' longitudinal records. This approach implements an anytime lazy k-nearest neighbor (kNN) algorithm by avoiding unnecessary neighbor assessments. For the situations when this speed-up may not suffice, the CBR approach can be interrupted earlier and it returns best-so-far kNNs. The positioning of our proposal is **disruptive** in the field of CBR: on the one hand, our approach allows the prediction of temporal subseries instead of more classical outcomes; on the other hand, the prediction task is iterated throughout a reactive scenario, in response to external sollicitations (medical actions).

# 3 CASE-BASED REASONING APPROACH

To make evolve the virtual patient during a surgery, we simulate their multivariate time series between two consecutive action triggers, and we iterate this process until the end of the surgery.

We propose a Contextualized Multidimensional Pattern Retrieval approach which relies on two tasks: (i) identification of real patients most similar to the virtual patient, in the cohort of patients in the same class as the virtual patient (Section 3.2); (ii) prediction of the near future of virtual patient from the previous real patients (Section 3.3). This second task is based on three subtasks : off-line annotation of the cohort of patients (Subsection 3.2.1), similarity search for action traces (Subsection 3.2.2), similarity search for time series (Subsection 3.2.3).

## 3.1 Contextualized Multidimensional Pattern

Our case-based reasoning (CBR) approach relies on Contextualized Multidimensional Pattern retrieval, in a set of multivariate time series each annotated with a trace of time-stamped triggers of actions.

We consider that (i) each variable is influenced by one action at least, (ii) an action may impact several variables, (iii) a variable impacted by several actions is subject to the combined effect of these actions, initiated at different time-steps or not.

The objective is to predict the evolution of the virtual patient subsequent to the triggering of an action A. The recent history of the virtual patient provides the context of the action triggering. Such context includes the "most recently triggered" actions (C and B on Figure 1), together with the recent evolution of the virtual patient's variables, prior the initiation of action A. On the example of Figure 1, the actions of the context are C, B and A, and a contextual time window is defined, that extends from the triggering of action C to that of action A. A **Contextualized Multidimensional Pattern** (CMP) can therefore be defined for the virtual patient. It is composed of (i) the restriction of the virtual patient's multivariate time series to the contextual time window and of (ii) the sequence of actions in the context, annotated with the durations between two successive triggers of actions. The latter sequence of intertwined actions and durations is called **action-signature** in the remainder of the paper. The number of actions in the action-signature, further referred to as *nba*, is specified through the expertise of anaesthesiologists. For instance, if *nba* equals 4, "D <10> C <35> B <72> A" represents an action-signature in which the time intervals between two successive triggers are respectively 10, 35 and 72 number of time-steps.

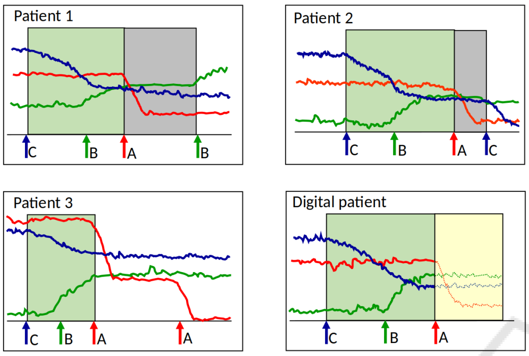


Figure 1: Concept of contextualized multidimensional pattern and use to simulate the evolution of the virtual patient. The mutivariate time series is composed of the three univariate time series respectively associated to the variables represented in red, green and blue. In this example, actions A, B and C respectively exert an impact on the variables represented in red, green and blue. These impacts are respectively a sharp decline, a slow growth, and a slow decline (until stabilization in all three cases). Patients 1, 2 and 3 are real patients. The multivariate time series of the four patients represented have all been left-truncated, as previous actions are assumed to have no effect on the variables beyond a certain duration. In this toy example, real patients 1, 2 and 3 have each their action-signature similar to that of the virtual patient. The mutivariate time series of patients 1 and 2 are supposed to be sufficiently similar to that of the virtual patient. We use the multivariate time series of patients 1 and 2 (gray rectangles truncated on the right when the action following A is triggered) to predict short-term evolution of the virtual patient (yellow rectangle).

## 3.2 Identifying Real Patients Most Similar to the Virtual Patient

To note, we only have to examine the real patients on targeted regions of their event traces and on multivariate time subseries. However, real-time identification of patients most similar to the virtual patient is challenging: (i) at the hospital, the available cohort of patients sharing the same characteristics as the virtual patient (surgery, age, gender, weight, pathological history) may be large, (ii) we deal with multivariate time series.

The current CMP of the virtual patient will be further referred to as the query. The number of actions in a CMP is a constant set by the training managers (*i.e.*, anaesthesia experts). This number is denoted *nba*, and includes the last action (*act*) triggered by the user (*i.e.*, the trainee).

### 3.2.1 Off-line Annotation

The first subtask is off-line and is dedicated to the annotation of the cohort. Based on this annotation, at each novel query, we wish to make available an index whose entries are actions. Such index will allow to efficiently retrieve real action-signatures (ASs) similar to the AS of the virtual patient.

### 3.2.2 Similarity for Action Traces

As already mentioned, the second (on-line) subtask in CMP retrieval involves index-based AS recognition. We add flexibility in two ways. Thanks to experts, we are able to categorize actions: for instance, it may be possible to replace an anaesthetic by another. Real patients found similar to the virtual patient (VP) are expected to have been anaesthetized with the same product as the VP, but product replacement is allowed. For small cohorts, this relaxation is crucial to raise the probability to identify patients similar to the VP. On the other hand, we recall that ASs are time-stamped. AS recognition must therefore take into account durations between any two consecutive actions. Again, relaxation is necessary. This time, for more realism, it is wise to gradually penalize time deviations from the VP's action-signature: we care that pairs of consecutive actions matched between VP and real patient are not too different for the most recent pairs; in constrast, we accept a lower similarity in time intervals if the actions are older. For a given relaxation percentage $\Delta$, the deviation allowed gradually decreases as follows:

$$\mathcal{D}ev_{time}(i) = \Delta\left(1 - \frac{i-1}{nba-1}\right), \tag{1}$$

where $i$ stands for the rank of the pair of consecutive actions under examination in the action-signature. For instance, in action-signature "D <10> C <35> B <72> A", the rank of (C,B) is 2. When applied to an action-signature of 4 actions and $\Delta = 5\%$, Eq. (1) yields the three following relaxation percentages: 5%, 3.3%, 1.6%. In practice, specifying parameter $\Delta$ will automatically discard real patients showing deviation from the VP, at rank $i$, above the corresponding threshold $\mathcal{D}ev_{time}(i)$.

### 3.2.3 Similarity for Time Series

Downstream AS recognition, the third substask consists in the retrieval of patients with multivariate time series similar to those of the virtual patient, and is far more challenging. Computing a multivariate dissimilarity measure from the outset, for instance a multivariate Dynamic Time Warping (DTW) distance is expensive, even for a number $p$ of variables equal to 3. As an alternative, we propose to rank the real patients relying on $p$ univariate dissimilarities: for each of the $p$ variables in our problem, we compute the dissimilarity between the univariate time series of virtual patient and real patient. We recall that we only have to compute these dissimilarities for restrictions of time series to the contextual window defined by the current CMP.

We denote $D$ a dissimilarity measure dedicated to univariate time series comparison. Subsection 4.2 will briefly indicate how we chose $D$ in the context of our application. We average the $p$ univariate $D$ measures computed for the $p$ variables:

$$\overline{D} = \frac{1}{p}\sum_{i=1}^{p} D_i, \tag{2}$$

with $D_i$ the dissimilarity obtained using $D$ for $i^{th}$ variable. Thereinafter, $\overline{D}$ will be referred to as the multivariate dissimilarity.

On this basis, the similarity score used to rank the real patients (already selected based on ASs) combines three similarity scores:

$$\mathcal{S}im_{AS+TS} = \kappa_1 \mathcal{S}im_{AS_{match}} + \kappa_2 \mathcal{S}im_{AS_{time}} + \kappa_3 \mathcal{S}im_{TS}, \tag{3}$$

in which $\mathcal{S}im_{AS_{match}}$ rewards pairwise matching of actions in query AS and real patient's AS, $\mathcal{S}im_{AS_{time}}$ rewards pairwise matching of time intervals in the previous ASs, and $\mathcal{S}im_{TS}$ is a normalized similarity score for time series. These three scores are defined as follows:

$$\mathcal{S}im_{AS_{match}} = \frac{1}{nba}\sum_{i=1}^{nba} sim(ASa_i^{(VP)}, ASa_i^{(RP)}), \tag{4}$$

where $ASa_i^{(VP)}$ and $ASa_i^{(RP)}$ respectively stand for $i^{th}$ action in the action-signature of virtual patient (VP) and $i^{th}$ action in the action-signature of real patient (RP), and $sim(s,s') = 1$ if $s = s'$ and 0 otherwise.

$$\mathcal{S}im_{AS_{time}} = \frac{\varepsilon_{max} - \varepsilon}{\varepsilon_{max}}, \tag{5}$$

$$\varepsilon = \sum_{i=1}^{nba-1} |ASt_i^{(VP)} - ASt_i^{(RP)}|, \tag{6}$$

where $ASt_i^{(X)}$ denotes the time interval between $i^{th}$ and $i+1^{th}$ actions in the action-signature of patient $X$, $X$ being either the VP or a real patient, and $\varepsilon_{max}$ is the largest value of $\varepsilon$ over all AS-selected patients examined.

$$\mathcal{S}im_{TS} = \frac{\overline{D}_{max} - \overline{D}}{\overline{D}_{max}}, \tag{7}$$

where $\mathcal{S}im_{TS}$ is equal to 1 for the query, as the dissimilarity of the query against itself is null, and $\mathcal{S}im_{TS}$ equals 0 for the real patient with largest multivariate dissimilarity $\overline{D}_{max}$, amongst all AS-selected patients examined.

### 3.3 Prediction Task

This prediction task is not straightforward, since (i) the time windows of the similar patients are not equal (gray rectangles in Figure 1), (ii) we do not know to which extent we must predict the evolution of the VP, (iii) the shapes of the univariate time series in real patients might be similar to that of the VP but with discordances in variable values at the triggering of the action of interest (action A in Figure 1). In the following, the prediction window of a similar patient is defined as the time window left- and right-bounded by the action just triggered and the next action's trigger. In Figure 1, such prediction windows are represented in gray rectangles. For simplicity, the action just triggered will be referred to "action A" in this section.

To simulate the multivariate time series over the largest possible time interval, we propose a strategy that iterates through the ordered right bounds $r_1, r_2, \cdots r_{n_s}$ of the prediction windows of the $n_s$ similar patients, considering these patients in parallel. At iteration $i$, for each variable $v$, we generate a (univariate) time series fragment by averaging over the univariate time subseries of the similar patients whose prediction windows extend after right bound $r_i$. Thus, at first iteration, we average over all subseries left-bounded by action A's trigger and right-bounded by $r_1$ ; at second iteration, we average over all remaining subseries that can be left-bounded by $r_1$ and right-bounded by $r_2$, and so on. The principle is shown in Figure 2 (a). The $n_s$ fragments thus obtained are concatenated to produce $\mathcal{U}_v^+$. Afterwards, for each variable $v$, the novel predicted univariate time series $\mathcal{U}_v^+$ is connected to the corresponding time series $\mathcal{U}_v$ generated so far for the virtual patient. The connection generally requires shifting along the y-axis, to start $\mathcal{U}_v^+$ from the value where $\mathcal{U}_v$ ended in. The principle is shown in Figure 2 (b). When repeated over the $p$ variables, this process makes evolve the $p$ physiological parameters of the virtual patient for some time after an action is triggered.

Importantly, in the scenario that is being played out, the next action after action A might be triggered by the trainee beyond the farthest time-step we were able to predict for. In this case, in the first version of our reactive scenario generator, we generate stable univariate time series to wait until the next action is triggered by the trainee or a timeout is reached.

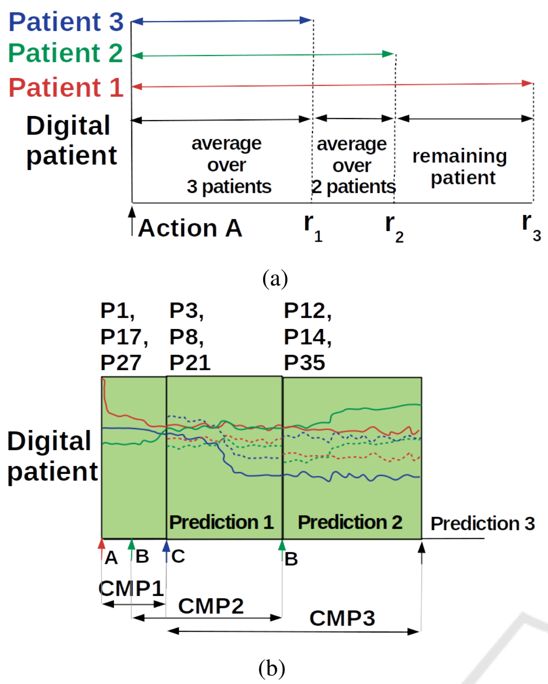


Figure 2: Incremental generation of the virtual patient's multivariate time series, from the real patients' multivariate time series, at action A triggering. In this example, $n_s = 3$ (number of real patients most similar to virtual patient), $p = 3$ (number of variables) and $nba = 3$ (number of actions in the Contextualized Multidimensional Pattern). (a) Averaging over less and less real patients for one variable produces a set of $n_s = 3$ univariate fragments. These fragments are concatenated to yield a synthetic time series running from action A to the farthest right bound ($r_3$). (b) The Contextualized Multidimensional Pattern $CMP_i$ drives Prediction$_i$. For example, in the prediction frame Prediction2, P12, P14 and P35 are the real patients showing highest similarity with the CMP2 of the virtual patient. The process explained in (a) provides $p = 3$ univariate fragments (dotted lines, one line per variable). A shift along the y-axis allows to concatenate each of these fragments with the corresponding univariate time series, in the multivariate time series simulated so far (solid lines).

## 3.4 Sketch of Algorithm

Algorithm 1 presents the main lines of our approach.

In line 1, a cohort of real patients is built: it is composed of the patients who shared the same characteristics $\mathcal{C}har$ (age, gender, weight, medical history) as the virtual patient when they underwent surgery $\mathcal{S}urg$; besides, the cohort is indexed by sequences of *nba* actions, for fastest similarity search of action-signatures in the patients' action traces. To start the simulation corresponding to the first $nba-1$ actions, we proceed as follows: at the trigger of the first action of the scenario, for instance action A's trigger, we select $n_s$ patients in $\mathcal{C}ohort$ who underwent action A,

---

Algorithm 1: Simulation of virtual patient.

**FUNCTION play_reactive_scenario($\mathcal{S}urg$, $\mathcal{C}har$, $nba$, $\Delta_\ell$, $\Delta_m$, $\Delta_h$, $n_s$)**
**INPUT:**
- $\mathcal{S}urg$, surgery of interest
- $\mathcal{C}har$, characteristics of virtual patient (VP)
- *nba*, number of actions in contextualized multidimensional pattern (CMP)
- $\Delta_\ell$, $\Delta_m$, $\Delta_h$, respectively low, medium and relaxation percentages to guide three CMP retrievals in parallel
- $n_s$, number of real patients most similar to virtual patient

**OUTPUT:**
- $\mathcal{T}$, event trace of VP
- $\mathcal{M}$, multivariate time series of VP

**VARIABLES:**
- $\mathcal{C}ohort$, cohort of real patients sharing same surgery $\mathcal{S}urg$ and characteristics $\mathcal{C}har$ as virtual patient
- $\mathcal{I}ndex$, indexation of $\mathcal{C}ohort$ by sequences of *nba* actions
- $CMP = (\mathcal{T}^*, \mathcal{M}^*)$, contextualized multidimensional pattern of VP, with
  - *e_time*, time ellapsed between the penultimate action and the current action *act* triggered by user
  - $\mathcal{T}^*$, action-signature of VP, of size *nba*, including last action *act*
  - $\mathcal{M}^*$, restriction of $\mathcal{M}$ to the current contextual time window defined by $\mathcal{T}^*$

1: $(\mathcal{C}ohort, \mathcal{I}ndex) \leftarrow$ **select_preannotated_cohort** $(\mathcal{S}urg, \mathcal{C}har)$
2: $(\mathcal{T}, \mathcal{M}) \leftarrow$ **initialize_digital_patient**$(nba-1)$

3: **while** (scenario is not over)
4: $(act, e\_time) \leftarrow$ **wait_for_action**()
5: **if** no action is triggered **then break end if**

6: $(\mathcal{T}^*, \mathcal{M}^*) \leftarrow$ **build_current_cmp**$(\mathcal{T}, \mathcal{M}, nba-1, act, e\_time)$
7: $\mathcal{P}at \leftarrow$ **identify_similar_patients**$(\mathcal{T}^*, \mathcal{M}^*, \Delta_\ell, \Delta_m, \Delta_h, n_s)$
8: $\mathcal{M} \leftarrow$ **grow_multivariate_time_series**$(\mathcal{M}, \mathcal{P}at)$
9: $\mathcal{T} \leftarrow$ **grow_action_trace**$(\mathcal{T}, e\_time, act)$
10: **end while**

---

and we predict virtual patient's evolution from them. At the occurrence of the second action of the scenario, say action B, the action-signature (AS) is "A <duration> B", and the prediction is made using this AS. Up to the occurrence of $nba-1^{th}$ action, we can only operate following this initialization mode (line 2). At the end of the initialization, an action trace $\mathcal{T}$ containing $nba-1$ actions has been grown and the corresponding multivariate time series $\mathcal{M}$ has been simulated for the digital patient.

From $nba^{th}$ action occurrence onwards (line 3 to 10), we successively use ASs of full length (*i.e.*, with *nba* actions). In line 6, once a novel action *act* has been triggered, the AS $\mathcal{T}^*$ is constructed by (i) considering the $nba-1$ latest actions in $\mathcal{T}$ and the corresponding time intervals, (ii) adding the time interval between penultimate action and current action *act*, *e_time*, (iii) and adding the current action *act*. The multivariate time subseries $\mathcal{M}^*$ corresponding to the temporal scope of $\mathcal{T}^*$ is then obtained. It is a suffix of the multivariate time series $\mathcal{T}$ simulated so far. Together with $\mathcal{T}^*$, $\mathcal{M}^*$ forms the current CMP.

In line 7, we identify the $n_s$ real patients most similar to the virtual patient. We recall that all patients with ASs similar to the digital patient's AS $\mathcal{T}^*$, given a specified threshold $\Delta$, are first identified (Subsection 3.2.2). To note, this identification is accelerated thanks to the off-line annotation mentioned in Subsection 3.2.1. Then, the patients thus obtained are ranked considering the score in Eq. (3), which takes into account the similarity with $\mathcal{T}^*$ and the similarity with subseries $\mathcal{M}^*$ (Subsection 3.2.3). If threshold $\Delta$ (see Subsection 3.2.2) is two low, we run the risk of not detecting any similar patient, and of having to restart a search with a higher $\Delta$ value. Therefore, we designed a procedure where three searches are run in parallel, with different $\Delta$ values. This way, we expect to obtain $n_s$ similar patients.

In line 8, these $n_s$ patients are used to predict the evolution of the digital patient as from current action $act$. The multivariate time series $\mathcal{M}$ is grown accordingly. The action trace $\mathcal{T}$ of the virtual patient is updated in line 9.

This iterative simulation stops when no more action is triggered.

# 4 APPLICATION

The SVP-OR framework has been implemented in Python 3.10.0. Subsection 4.1 first describes the data used to evaluate the SVP-OR approach. Then, Subsection 4.2 provides some information about the comprehensive comparative analysis of 8 dissimilarity measures which has been conducted separately to identify the most relevant measure for our application. Next, Subsection 4.3 describes two instantiations of SVP-OR. In Subsection 4.4, we finally evaluate these two instantiations.

## 4.1 Data

In the current part of our research project, the SVP-OR approach has been evaluated through realistic data for reasons of health data protection. These realistic data are produced by the DBLBS data generator that we have implemented, based on the expertise of anaesthesiologists. In a nutshell, this generator runs two steps.

The first step is driven by a grammar that provides a hierarchical description of a surgery, for a targeted patient (age, gender, weight, medical history). When fed with a user-specified number of patients $n_p$, this first step produces as many action traces. Expert knowledge is required to construct such grammar. In our work, we focused on inguinal hernia operation under laparoscopy, for a 30 year old male patient, with weight 80 and with no medical history. Figure 3 provides a simplified view of this surgery.

Based on additional expert knowledge supplemented by consultation of the literature, we were able to draw rules about the impacts of the actions on the physiological parameters (growth, decay, no impact), the delay, duration and intensity of these impacts. When fed with these rules, a number $n_p$ of action traces produced by first step, and initial values set for the parameters of each simulated patient, the second step produces $n_p$ multivariate time series. In this first version of the DBLBS generator, the number $p$ of variables considered equals 4 (heart rate; diastolic, systolic and average blood pressures). Thus, we generated 1000 simulated patients, that is 1000 realistic multivariate time series of length in the order of 200 time-steps, with measure points spaced every 30 s.

An incremental process based on expert feedback allowed a first validation of the DBLBS generator. Furthermore, we conducted a statistical analysis to compare the time series of real patients against the time series of patients simulated through DBLBS. Between 2002 and 2019, 418 patients underwent laparoscopic surgery with prosthetics setting, for an inguinal hernia, at Nantes University Hospital. Amongst these, 170 offered a complete anaesthetic profile. The corresponding patients are referred to as the reference cohort hereafter.

In a nutshell, we evidenced that for each variable, once the effect of the initial values has been erased through differencing, the empirical distribution functions of reference and simulated cohorts are close. Differencing is the operation that produces the times series $\{X_t - X_{t-1}\}_{t=2,\ldots,n}$ from the original time series $\{X_t\}_{t=1,\ldots,n}$ of length $n$. The empirical distribution function is an estimate of the cumulative distribution function that generated the values in a given sample. Moreover, we observed that the autocorrelation functions of the two cohorts are close. Autocorrelation is the correlation between a time series variable and its lagged version in time. Due to space limitations, these results are not detailed here.

## 4.2 Choice of Dissimilarity Measure

To select a dissimilarity measure $D$ (see Subsection 3.2.3) appropriate to our application, we conducted the extensive comparative study of 8 pre-selected univariate dissimilarities. The details of this study will be published elsewhere. In the rest of this section, we just mention some information about the context

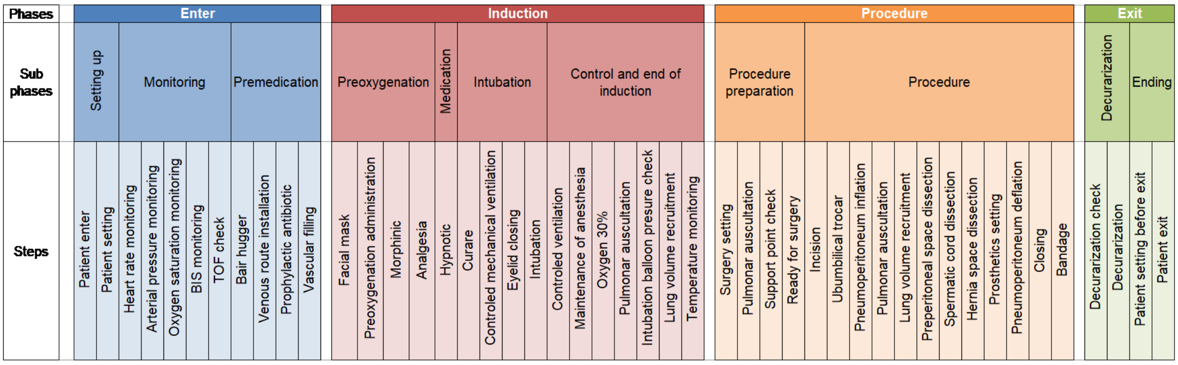

Figure 3: Surgery for the laparoscopic inguinal hernia operation with prosthetics setting.

of this study. Finally, we indicate which dissimilarity measure was chosen for our application.

Several studies are reported in the literature, that compare performances or properties of various similarity measures applied to time series. However, none of these studies deals with short time series, as required by our application. In a nutshell, we used 28 data sets in our experiments. Twenty-six data sets were from the UCR repository (Dau et al., 2018). The $27^{th}$ data set was simulated using the PHMC-LAR (Partially Hidden Markov Chain Linear AutoRegressive) model described in (Dama and Sinoquet, 2021): the 4-state PHMC-LAR model with autoregressive order 2 used is described in the Section 7.1.1 of this latter document. The $28^{th}$ data set was generated through the DBLBS generator we have designed, based on expert specification (see Subsection 4.1).

In each time series of these 28 data sets, subseries of length 10, 20 and 30 time-steps were drawn at random. The aim was to check whether our conclusions held for the three lengths. We perturbed each such subseries using 6 types of perturbations (for instance, by increasing amplitude at each time-step, depending on initial amplitude). We computed the dissimilarities between originals and perturbed versions for the 8 dissimilarity measures.

Table 1 briefly describes the 8 pre-selected univariate dissimilarities.

Our experiments allowed us to discard MPDist and the tsfresh-based measure. Moreover, we checked that DTW variants are relevant in presence of very short to short time series, as is the case in the SVP-OR framework. This result is not trivial. Amongst these variants, DTWdtai was chosen for its greatest speed.

We were compelled to run a comparative analysis focused on short time series, to fill a gap in the literature. Interestingly, other researchers and practitioners can rely on our results to apply some framework similar to SVP-OR: the DTW measure is appropriate.

Table 1: The 8 dissimilarity measures compared. DTW: Dynamic Time Warping.

| measure | abbreviation | library |
|---|---|---|
| **classical DTW** | DTWc | pyts |
| global alignement of two time series (Sakoe and Chiba, 1971) | | |
| **fast exact DTW** | DTWdtai | dtaidistance |
| exact calculus of DTW implemented in Cython language | | |
| **fast DTW** | DTWf | pyts |
| whole alignment of the two time series by repeatedly aligning subseries of the former (Salvador and Chan, 2007) | | |
| **Sakoe-Chiba DTW** | DTWsc | pyts |
| similarity assessment between the two time series limited within a warping path region, the Sakoe-Chiba band, defined as a region of same width along the diagonal of the alignment matrix (Sakoe and Chiba, 1978) | | |
| **Itakura DTW** | DTWi | pyts |
| warping path region in the shape of a parallelogram, defined through two parameters controlling the maximum warping width and the warping extent increase from beginning of warping path to maximum warping width (Itakura, 1975) | | |
| **multiscale DTW** | DTWm | pyts |
| downsampling of the two time series, optimal path thus obtained projected on the original scale, to be used as the warping path region (Müller et al., 2006) | | |
| **MPDist** | MPDist | MatrixProfile |
| feature-based dissimilarity measure (Gharghabi et al., 2018) | | |
| **transformed-based measure** | | tsfresh |
| Euclidian distance applied to the numerical representations of the two time series, obtained through the tsfresh feature extraction package (Christ et al., 2018) | | |

## 4.3 The Two Instantiations of the SVP-OR Algorithm

On expert advice, we set *nba* at 4 (see Subsection 3.1). We ran three similarity searches in parallel: we chose $\Delta_\ell = 1\%$, $\Delta_m = 5\%$ and $\Delta_h = 10\%$ (see Subsection 3.4).

To rank the real patients, we recall that score $\mathcal{S}im_{AS+TS}$ (Eq. 3) is a weighted combination of similarity scores on event traces and time series. We fixed the weights as follows: $\kappa_1 = \kappa_2 = 1$ (event traces), $\kappa_3 = 2$ (time series).

We instantiated the framework presented in Sec-

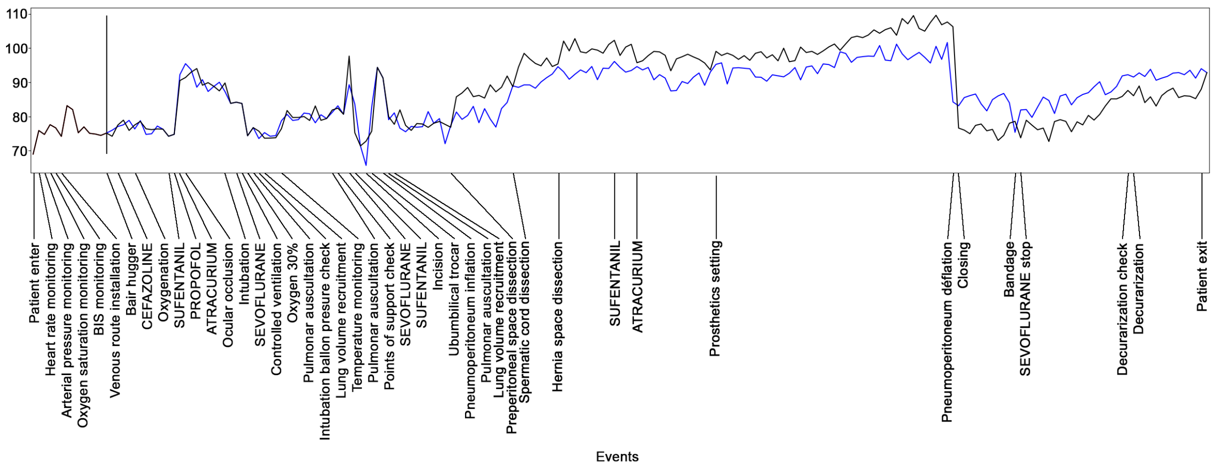


Figure 4: Comparison of real and simulated time series obtained through strategy #1 (most similar patient). The time series describe heart rate. Simulation has been performed as from the action consisting in the installation of a Bair Hugger (heating blanket) (black vertical line). Medical actions relevant to anaesthesia are indicated in capital letters. In black: real time series. In blue: simulated time series.

tion 3. The two variants considered, #1 and #2, differ by the value of $n_s$. We recall that $n_s$ denotes the number of real patients most similar to the virtual patient, used to predict short-term evolution of the latter (see Algorithm 1, line 8). #1 relies on the most similar patient, whereas #2 uses the 10 most similar patients.

### 4.4 Assessing the Realism of Simulated Trajectories

In this work, our aim is to evaluate how simulated trajectories (*i.e.*, simulated multivariate time series) depart from real trajectories. In this evaluation, each simulated trajectory is paced by the sequence of medical actions (*i.e.*, scenario) of a real trajectory against which it is compared. In this Subsection, we first outline the evaluation protocol used. Then we present and discuss the results.

#### 4.4.1 Evaluation Protocol

To assess the realism of the trajectories provided by each of variants #1 and #2, we implemented the following protocol. For $i^{th}$ patient in data set DBLBS ($1 \leq i \leq 1000$), we selected at random a time-step $\tau_i$ where an action was performed. We then simulated the rest of the surgery from time-step $\tau_i$, which means that we enforced the user to trigger the same acts of anaesthesia as in the real scenario, and at the same time-steps.

To note, under the normal conditions of use of our generator, the user is a human being. However, for the assessment purpose, we modified the interface of our generator: we enforced the generator to play the role of the user (trainee), following the real scenario of $i^{th}$ patient as from time-step $\tau_i$, that is the generator triggers the same anaesthesia actions as in the real scenario. Importantly, this particular context of use of the generator eliminates the problem of having to trigger the actions that are not perpetrated by the trainee. In our context, the real scenario is simply fed to the modified interface of our generator, as from time-step $\tau_i$.

Figure 4 shows a real time series and the predicted subseries.

We introduce the following notations and definitions:

**Notation 1.** *For each patient i, we call* $\mathcal{M}^{(R)}_{\tau_i}$ *the real multivariate time series between time-step* $\tau_i$ *and the end of the real surgery.*

**Notation 2.** *For each patient i,* $\mathcal{M}^{(P)}_{\tau_{i,\#j}}$ *denotes the predicted multivariate time series as from* $\tau_i$ *to the end of the surgery, when variant* $\#j$ *is used for prediction.*

**Definition 1.** $\overline{D}_{i,\#j}$ *is defined as the dissimilarity* $\overline{D}$ *(see Eq. 2) computed between the multivariate time series* $\mathcal{M}^{(R)}_{\tau_i}$ *and* $\mathcal{M}^{(P)}_{\tau_{i,\#j}}$. *In other words,* $\overline{D}_{i,\#j}$ *measures the dissimilarity between the observed physiological evolution (real trajectory) and predicted physiological behavior (simulated trajectory) of* $i^{th}$ *real patient, when variant #j is applied. A similarity score is obtained through normalization:* $Sim\text{-}g_{i,\#j} = \frac{\overline{D}_{max,\#j} - \overline{D}_{i,\#j}}{\overline{D}_{max,\#j}}$ *where* $\overline{D}_{max,\#j}$ *is the largest value obtained across the 1000 patients.*

**Notation 3.** *The distribution of normalized similarities* $Sim\text{-}g_{i,\#j}$ *computed for the 1000 patients in data set DBLBS is denoted* $\mathcal{D}_{Sim-g_{\#j}}$.

For each variant $\#j \in \{\#1, \#2\}$, we computed the distribution of normalized similarities $\mathcal{D}_{Sim-g_{\#j}}$ over the 1000 patients of data set DBLBS.

#### 4.4.2 Results and Discussion

Figure 5 shows the boxplots for the distributions $\mathcal{D}_{Sim-g_{\#1}}$ and $\mathcal{D}_{Sim-g_{\#2}}$. We conclude that for the surgery of interest, the time series generated by strategy #2 are less similar to the real time series (average around 0.78) than those generated through strategy #1 (average around 0.82). Besides, the simulation from 10 real patients (#2) creates more variability than the simulation from a single real patient (#1). This result is not trivial as variabilily could arise from the continual change of the single patient in strategy #1.

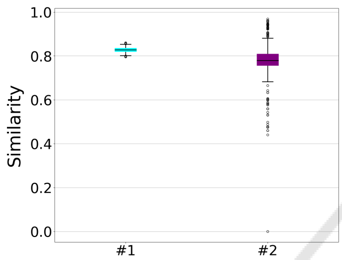


Figure 5: Prediction abilities of strategies #1 (most similar patient) and #2 (10 most similar patients). Each boxplot represents the distribution of similarities $\mathcal{D}_{Sim-g_{i,\#j}}$ between real and predicted multivariate time series of patient $i$, for the 1000 realistic patients of data set DBLBS, and for strategy $j$. For each patient, the prediction was made under the same left-truncated scenario (same sequence of actions) as in the real patient. The scenario was left-truncated at random over the 1000 patients, to assess the prediction ability along the whole surgery. The 1000 random truncations were the same for the two strategies.

We used a Wilcoxon / Mann-Whitney (WMW) test together with a Kolmogorov-Smirnov (KS) test to compare the distributions $\mathcal{D}_{Sim-g_{\#1}}$ and $\mathcal{D}_{Sim-g_{\#2}}$. We obtained a p-value equal to $4.69 \times 10^{-130}$ (WMW) and a p-value equal to $1.11 \times 10^{-251}$ (KS). We conclude that the distributions of time series simulated through #1 and #2 significantly differ.

The main conclusion to draw is that the two variants assessed produce trajectories sufficiently similar to real trajectories, when constrained by real scenarios. In the context of the application targeted, our simulation software will have to trigger the actions that are not perpetrated by the trainee. A synthetic representation of the steps, substeps, and possibly alternative steps will be constructed off-line for a given surgery. This representation will be leveraged on the fly each time an action by the medical team is plausible. In this case, assessing whether the simulated trajectories are still realistic seems far more challenging than in the present situation: it seems *a priori* that the only way to assess the realism of the simulation will require analysis by human experts.

## 5 CONCLUSION AND PERSPECTIVES

In this work, we have introduced a generalist framework, SVP-OR (Simulation of Virtual Patient at the Operating Room), which adopts a case-based approach to make evolve a virtual patient in response to the actions of a user and of a virtual medical team. The aim is to provide a digital training support for the user (trainee). SVP-OR lies on two corner stones, contextualized multidimensional pattern retrieval and short-term prediction from real patients' histories.

The ability to handle scenarios responsive to external sollicitations offers possibilities that go beyond the training objective considered here. First, our CBR-based SVP-OR framework allows to simulate the evolution of a patient's physiological status in a simple but efficient way. Therefore, without the need to learn a complex model in machine learning, we are able to develop a digital twin for a patient undergoing a surgery. In this way, for instance, we can anticipate the risks in the operating room for patients with medical antecedents. A second notable potentiality is of primary importance for developing research collaborations between University Hospitals and the outside world. Protected data cannot be easily shared with the outside world; a variant can be instantiated from our SVP-OR framework, to implement a simple but effective anonymization procedure capable of preserving the dependencies between the dynamics of variables, as well as between the variables and a trace of actions.

One of our next works will be to design a model to subsume all traces of actions ever observed for a given surgery. This synthetic representation will allow SVP-OR to automatically generate the actions of the medical team. Moreover, it will offer the way to identify inappropriate actions by the trainee.

A step further, we aim at integrating in the SVP-OR framework factors that may mislead the trainee. Such factors are for instance an abnormal situation, stress, interactions within the medical team that may be disruptive to the trainee.

## ACKNOWLEDGEMENTS

H. Boisaubert is supported by a PhD scholarship granted by the Pays de la Loire regional research project RFI OIC EXAN. The internship work of L. Vincent was supported by the FAME research cluster (Human Factors for Medical Technologies, NExT/ANR-16-IDEX-0007). All authors would like to thank F. Dama, PhD student, whose contribution in the development of the DBLBS generator is invaluable. The software development and execution of experiments were carried out at the CCIPL (Centre de Calcul Intensif des Pays de la Loire, Nantes, France).

## REFERENCES


Brown, S.-A. (2016). Patient similarity: emerging concepts in systems and precision medicine. *Frontiers in Physiology*, 7:561.

Choudhury, N. and Begum, S. (2016). A survey on case-based reasoning in medicine. *International Journal of Advanced Computer Science and Applications*, 7(8):132–136.

Christ, M., Braun, N., Neuffer, J., and Kempa-Liehr, A. (2018). Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). *Neurocomputing*, 307:72–77.

Dama, F. and Sinoquet, C. (2021). Partially Hidden Markov Chain Linear Autoregressive model: inference and forecasting. *arXiv:2102.12584*.

Dau, H., Keogh, E., Kamgar, K., Yeh, C.-C., Zhu, Y., Gharghabi, S., Ratanamahatana, C., Yanping, Hu, B., Begum, N., Bagnall, A., Mueen, A., Batista, G., and Hexagon-ML (2018). The UCR time series classification archive. https://www.cs.ucr.edu/ eamonn/time_series_data_2018.

Erdogan, T. and Tarhan, A. (2018). A goal-driven evaluation method based on process mining for healthcare processes. *Applied Sciences*, 8(6):894.

Fang, H., Tan, N., Tan, W., Oei, R., Lee, M., and Hsu, W. (2021). Patient similarity analytics for explainable clinical risk prediction. *BMC Medical Informatics and Decision Making*, 21(1):207.

Forestier, G., Petitjean, F., Riffaud, L., and Jannin, P. (2017). Automatic matching of surgeries to predict surgeons' next actions. *Artificial Intelligence in Medicine*, 81:3–11.

Franke, S., Meixensberger, J., and Neumuth, T. (2013). Intervention time prediction from surgical low-level tasks. *Journal of Biomedical Informatics*, 46(1):152–09.

Ganzinger, M., Schrodt, J., and Knaup-Gregori, P. (2019). A concept for graph-based temporal similarity of patient data. *Studies in Health Technology and Informatics*, 264:138–142.

Gharghabi, S., Imani, S., Bagnall, A., Darvishzadeh, A., and Keogh, E. (2018). Matrix Profile XII: MPdist: a novel time series distance measure to allow data mining in more challenging scenarios. In *IEEE International Conference on Data Mining (ICDM)*, pages 965–970.

Goel, A. and Diaz-Agudo, B. (2017). What's hot in case-based reasoning. In *Thirty-first AAAI Conference on Artificial Intelligence (AAAI)*, pages 5067–5069.

Itakura, F. (1975). Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 23(1):67–72.

Khalid, S., Goldenberg, M., Grantcharov, T., Taati, B., and Rudzicz, F. (2020). Evaluation of Deep Learning models for identifying surgical actions and measuring performance. JAMA Network Open 3(3):e201664.

Liu, Z., Zhu, Z., Gao, J., and Xu, C. (2021). Forecast methods for time series data: a survey. *IEEE Access*, 606–617:3091162.

Meißner, C., Meixensberger, J., Pretschner, A., and Neumuth, T. (2014). Sensor-based surgical activity recognition in unconstrained environments. *Minimally Invasive Therapy & Allied Technologies*, 23:198–205.

Mülâyim, M. and Arcos, J. (2020). Fast anytime retrieval with confidence in large-scale temporal case bases. *Knowledge-Based Systems*, 206:106374.

Müller, M., Mattes, H., and Kurth, F. (2006). An efficient multiscale approach to audio synchronization. In *International Conference on Music Information Retrieval (ISMIR)*, pages 192–197.

Nagendran, M., Gurusamy, K., Aggarwal, R., Loizidou, M., and Davidson, B. (2013). Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane Database of Systematic Reviews*, 8:CD00657.

Ng, K., Sun, J., Hu, J., and Wang, F. (2015). Personalized predictive modeling and risk factor identification using patient similarity. *MIA Joint Summits on Translational Science proceedings*, pages 132–136.

Parimbelli, E., Marini, S., Sacchi, L., and Bellazzi, R. (2018). Patient similarity for precision medicine: a systematic review. *Journal of Biomedical Informatics*, 83:87–96.

Qi, D., Ryason, A., Milef, N., Alfred, S., Abu-Nuwar, M., Kappus, M., De, S., and Jones, D. (2021). Virtual reality operating room with AI guidance: design and validation of a fire scenario. *Surgical Endoscopy*, 35(2):779–786.

Sakoe, H. and Chiba, S. (1971). A dynamic programming approach to continuous speech recognition. In *ICA, Paper 20 C13*.

Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1):43–49.

Salvador, S. and Chan, P. (2007). Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5):561–580.

Sha, Y., Venugopalan, J., and Wang, M. (2016). A novel temporal similarity measure for patients based on irregularly measured data in electronic health records. In *Seventh ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB)*, pages 337–344.

Stauder, R., Okur, A., Peter, L., Schneider, A., Kranzfelder, M., Feussner, H., and Navab, N. (2014). Random forests for phase detection in surgical workflow analysis. In *International Conference on Information Processing in Computer-Assisted Interventions*, pages 148–157.

Wang, N., Huang, Y., Liu, H., Fei, X., Wei, L., Zhao, X., and Chen, H. (2019). Measurement and application of patient similarity in personalized predictive modeling based on electronic medical records. *BioMedical Engineering OnLine*, 18:98.