



**HAL**  
open science

## Skill assessment of an epidural anesthesia using the PeriSIM simulator

Thibault Senac, Arnaud Lelevé, Richard Moreau, Laurent Krähenbühl,  
Florent Sigwalt, Christian Bauer

► **To cite this version:**

Thibault Senac, Arnaud Lelevé, Richard Moreau, Laurent Krähenbühl, Florent Sigwalt, et al.. Skill assessment of an epidural anesthesia using the PeriSIM simulator. *IEEE Transactions on Medical Robotics and Bionics*, 2021, 3 (1), pp.106-114. 10.1109/TMRB.2020.3048247 . hal-03092054

**HAL Id: hal-03092054**

**<https://hal.science/hal-03092054>**

Submitted on 24 Feb 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Skill assessment of an epidural anesthesia using the PeriSIM simulator

Thibault SÉNAC\*, Arnaud LELEVÉ\*, Richard MOREAU\*, Laurent KRÄHENBÜHL\*, Florent SIGWALT†, Christian BAUER†‡

\*University of Lyon, INSA Lyon, École Centrale de Lyon, UMR5005 CNRS, Ampère, F-69621, Villeurbanne, France,

†Département d'Anesthésie-Réanimation, Hôpital de la Croix Rousse, Hospices Civils de Lyon, Lyon, France,

‡Centre Lyonnais d'Enseignement par la Simulation en Santé, SAMSEI, Université Claude Bernard Lyon 1, Lyon, France

**Abstract**—Currently, the training in the lumbar epidural anesthesia procedure, a very common procedure, is mainly performed on patients under the supervision of experts. The reason for this situation is that the simulators that exist on the market are not realistic enough to efficiently learn this particularly haptic-sense-based gesture. To avoid this ethical issue and provide an efficient, repeatable, and affordable way of safely learning this gesture, the Ampère laboratory designed PeriSIM: a haptic training computer-based simulator. In this article, the trainees' skill assessment is tackled. This work presents various machine learning approaches to the evaluation of trainees and introduces a complete skill assessment algorithm to use in conjunction with the PeriSIM simulator. This work also aims to determine the important variables of an lumbar epidural anesthesia procedure for objective assessment purpose.

## I. INTRODUCTION

### A. Medical context

Lumbar epidural anesthesia is a commonly used, yet complex, medical gestures mainly performed during childbirth delivery. To perform these gesture the anesthetists use a Tuohy needle paired with a syringe filled with a physiologic saline solution. The procedure aims to get the needle tip into the epidural space. This epidural space is located near the spinal cord in the lumbar area. To reach this area, the needle has to be inserted between two vertebrae as shown in Fig.1. To precisely know the position of the needle tip, anesthetists exert pressure on the plunger of the syringe. As the injected fluid spreads differently according to the crossed tissues, this technique provides decent haptic feedback about the nature of the tissues located at the needle tip,

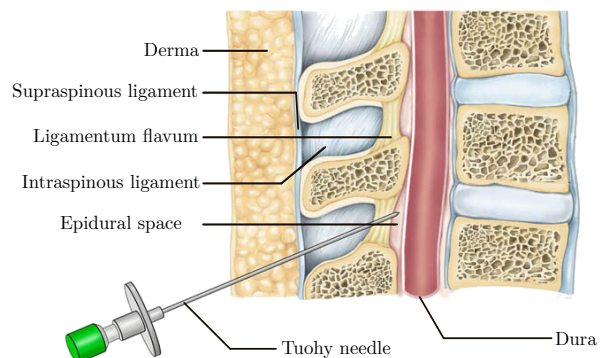


Figure 1: Insertion of a Tuohy needle in the epidural space

and thus the needle tip position. The latter sequentially crosses the following physical layers: derma, supraspinous ligament, intraspinous ligament, and ligamentum flavum. It then reaches the epidural space. As it is filled with fat and vessels, once the needle tip enters it, the fluid is injected into effortlessly. The anesthetists then experience what is called the *Loss Of Resistance* (LOR) principle.

The difficulty of the procedure lies in the fact that it is performed blindly while the epidural space is only 4 to 6mm wide on average. When the needle goes too far, it opens a breach in the *dura mater* which may lead to medical complications. Moreover, the needle insertion might require some force depending on the patient anatomy. Thus, the lack of visual cues, the necessary force level, and the required precision all contribute to the strong difficulty of the procedure. It then results in a long learning curve: studies indicate that up to 90 tries are required to reach 80% accuracy [1]. Moreover,

a common statement from public institutes such as the French H.A.S. [2] (*Haute Autorité de la Santé*) prescribes to "never do the first time on a patient", stressing further the need for a precise and reliable training mean for students having to learn the gesture.

### B. Simulation and skill assessment

The need for a reliable and precise training facility for the lumbar epidural anesthesia has brought some proposals such as the ones introduced by Manoharan *et al.* [3], Magill *et al.* [4], Hiemenz *et al.* [5] or Thao *et al.* [6]. Recent improvements of numeric technology allows to develop more complete simulators. Some of them are commercially available [7], [8] and others are still in development or open-source [9], [10]. No epidural simulator uses two technologies (haptic and pneumatic) to simulate both tissue resistance and syringe loss of resistance. The Ampère laboratory designed a simulator [11], [12] featuring the simulation of the LOR principle with a higher degree of precision than ever. Compared to the commercial ones and ones found in the literature, our simulator allows to train for needle insertion and LOR principle at the same time, to reproduce different kind of patients and to record different data (position, velocity, acceleration, force, etc.) during the process. This paper introduces results obtained with this simulator.

The main requirements for this simulator originate from an article by Vaughan *et al.* [1] stating some of the requirements of a "perfect" lumbar epidural simulator. Among these, this article covers the need for an automatic evaluation provided by the simulator. As this system is dedicated to hands-on training, the objective evaluation requirement is crucial. Usually, skill assessment techniques aim to provide a reliable and objective indication of the skills of a trainee. In the medical domain, one can cite as "gold standard" the Objective Structured Assessment of Technical Skills (OSATS) [13], [14] and the Global Operative Assessment of Laparoscopic Skills (GOALS) [15], [16] methods. Both have been tested in various studies [17], [18]. The former is dedicated to assessing skills in open surgery while the latter is aimed at assessing skills during laparoscopic procedures. Even though such evaluation scales are interesting, they were developed for medical procedures and cannot take advantage of the measurements that can be gathered using a computer-based haptic simulator.

Regarding the evaluation of gestures through measurements, several metrics can be highlighted. One of the simplest metrics that is often used to assess skill may

be the Task Completion Time (TCT) [19]. It often highlights that a masterful gesture is completed far quicker than a novice gesture. Regarding the metrics specific to a gesture, one can cite for example the deviation from the optimal path [20], motion smoothness [21], motion economy [20], path length [21] or curvature and affine velocity [22]. These are some of the metrics that may be used to assess mastery of a gesture in a 3D space. Concerning the specific lumbar epidural anesthesia procedure, one of the most commonly used metrics is what is called "overshoot" [23] in medical jargon. This overshoot corresponds to the distance crossed between the time the needle tip enters the epidural space and the time the gesture is stopped.

In this study, we propose an evaluation of the skill level of the trainees based on the data measured during their tries. Rather than using a single metrics, we chose to gather as many relevant variables as possible to come up with an objective classification of the mastery of the gesture using different machine learning methods

## II. EXPERIMENTAL SETUP

In this section, we detail the simulator and the experimental setup we used to gather the measurements.

### A. Materials and simulator design

As the lumbar epidural procedure features two main dependent parts (the needle and the syringe), the simulator was designed accordingly. To reproduce the needle, a customized Haption Virtuose 6D is involved and the syringe is reproduced with a double-acting pneumatic cylinder [11].

Concerning the needle insertion, the simulation of the force feedback is based on a simplified yet realistic enough representation of the anatomy involved during the procedure. We based the force generation on the work of Okamura *et al.* [24]. We also included the force trajectories designed by Kuckenbecker *et al.* [25] to improve the feeling of bone contact. A more complete description of the simulator design is available in [11].

Concerning the syringe simulation we used a double-acting pneumatic cylinder (Airpel® M16D100D [26]), with a position sensor (MPS-064TSNU0 [27] manufactured by SICK®), two pressure sensors (Festo® EPRB-1 [28]), a Festo® MPYE-5-M5-010 B servovalve controlled by a dSPACE® board. Moreover, some mechanical parts were added to improve

the overall realism. The complete simulator is presented in Fig.2.

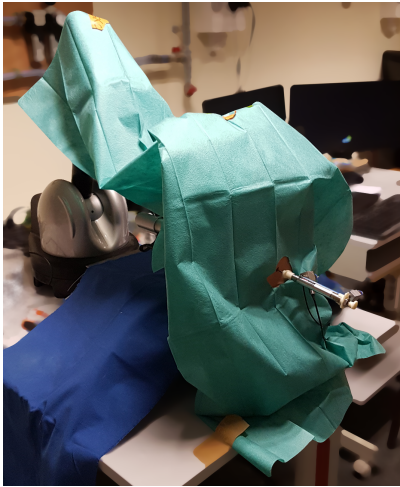


Figure 2: Complete simulator

### B. Experimental setup

Concerning the measurements, we gathered values from a total of twelve experts and twelve novices. On the one hand, the novices that took part in the measurement campaign were anesthesiology residents who were about to start their internship in a hospital. On the other hand, the experts were anesthetists who had performed at least 300 lumbar epidural anesthesia. Each person recorded twelve tries. These twelve tries corresponded to four standard pregnancy patients, four calcified elderly patients, and four obese patients. A patient is considered standard regarding the different measurements and resistant forces which are considered average. A calcified patient suffers from calcification of the ligaments thus resulting in higher cutting forces when the needle is cutting those ligaments. And the obese patient presents an important part of the fat in the tissues that increases the distance between the skin and the epidural space but also lowers the cutting force. The order of the twelve patients was set randomly and the type of patient was announced to the test subject before each test. Both novice and expert subjects got the opportunity to perform a couple of tries beforehand. This was a way for experts to get accustomed to the simulator and to see whether they needed to make any adjustments to their usual way of performing lumbar epidural anesthesia to comply with simulation constraints. For the novices, these first manipulations aimed at getting some pointers and advice on how to perform the procedure. Before

the novice trying the procedure for the first time, two senior anesthetists provided them with a lecture and some explanations concerning this procedure. The Fig.3 depicts the simulator in use.



Figure 3: Simulator in use

We did not record the test tries, but only the data from the twelve consecutive patients. For each try, the simulator records the position along the x, y and z axis of the handle of the electrical haptic interface, the number of insertions needed to perform the anesthesia, the direction of each insertion, the number of bone contacts, the time at which the needle reaches the epidural space, and the position of the plunger of the pneumatic cylinder. These measurements then serve as raw data to generate a total of 26 variables for each try.

## III. RESULTS AND ANALYSIS

### A. Global validation

In order to assess the realism of our simulator, a face validity questionnaire revealed a satisfactory overall score of 75% concerning the reproduction of the derma and the intraspinal ligament, 84% concerning the reproduction of the ligamentum flavum, 92% concerning the reproduction of the epidural space. Concerning the realism of the 3 different patients available (standard pregnancy patients, calcified elderly patients and obese patients), respectively 100%, 73% and 92% were very satisfied or satisfied.

Concerning the numerical values, we obtained similar results than Manoharan *et al.* [3] and Tran *et al.* [29].

### B. Data pre-processing

Before starting the data analysis, the data gathered has to be processed to be usable. In a first approach, we collected as many different variables as possible for each try:

- the position of the needle tip in three dimensions;
- forces applied to the tip of the needle in three dimensions;
- insertion angles, both pitch, and yaw;
- number of insertion;
- number of bone contacts;
- position of the needle tip at the end of the procedure;
- position of the piston of the pneumatic cylinder;
- pneumatic force produced by the pneumatic cylinder.

As most of these measurements are time-series, they constitute the base to define variables for skill assessment purpose. This represents a total of 26 variables:

- mean velocity of the needle tip along each axis;
- mean acceleration of the needle tip along each axis;
- variance of the velocity of the needle tip along each axis;
- variance of the acceleration of the needle tip along each axis;
- insertion angles (pitch and yaw);
- number of insertions;
- number of bone contacts;
- proportion of syringe emptied (which might be superior to 1)
- overshoot;
- mean velocity of the pneumatic cylinder piston;
- mean acceleration of the pneumatic cylinder piston;
- variance of the velocity of the pneumatic cylinder piston;
- variance of the acceleration of the pneumatic cylinder piston;
- success of the procedure (whether the needle punctured the *dura mater* or not).

Once these variables are computed for each try, one has to analyze them to produce some skill assessment. All

the work presented hereafter was achieved using machine learning methods provided by the software R.

### C. Data classification : a per variable approach

As a first approach to the skill assessment issue, data were studied and classified using each variable individually. We looked for trends in the data-set to set apart novices and experts. In that regard, some variables have already been identified as crucial concerning the lumbar epidural anesthesia procedure, the most prevalent being the overshoot ([30], [23]). Some other remarkable variables are the proportion of syringe emptied, variances of the velocity, and acceleration as a representation of motion smoothness and mean insertion velocity which can be linked to TCT. These variables are presented in Fig.4. Also, when faced with the skill assessment in general in surgical procedures, some variables are often highlighted such as acceleration [31] or TCT [19].

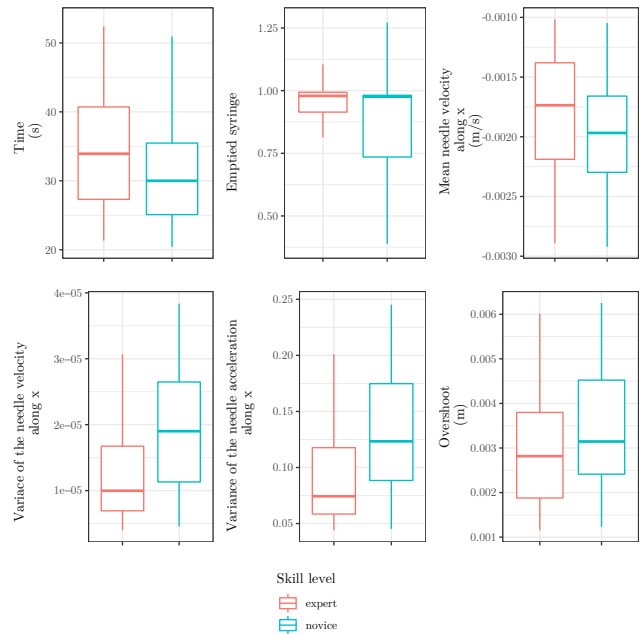


Figure 4: Remarkable variables from the data-set

These results clearly show a significant difference between novices and experts using the Wilcoxon test ( $p < 0.05$  in each case) on these variables and most of the results were expected the same. However, usually, the TCT tends to reduce with the increase in skill of the user and this does not seem to be the case here. Indeed the experts often prioritize security and precision instead of speed. That is why they tend to go at a slower and constant pace to be sure to reach the epidural space

with enough precision. Anyway, it seems that experience is taking into account possible serious adverse events and that success without harm is more important than speed. Zivkovic *et al.* [32] compared experts to novices in epidural anesthesia procedures and it shows that clinical experience was not correlated with procedure time.

With this mindset, it is possible to produce classifications based solely on the values of the aforementioned variables. The objective was to classify each try in two categories: the novices and the experts. To do so, the measures were divided into two parts: one training set containing two-thirds of the complete set of measures chosen at random and the last third to serve as the testing set. We then used an algorithm from `rpart` [33] (`rpart` method from `rpart` package) trained on the training set to produce a classification tree. Depending on the content of the training set, the classification tree does change. However, some variables often end up in the trees such as the overshoot for example. One of the trees obtained with this method is presented in Fig.5 as well as the corresponding classification performance. In the tree presented here, the overshoot is one of the deciding factors as well as velocity variances of the needle along the  $x$  and  $z$  axis. The  $x$  axis being the main insertion direction, the velocity variance along  $x$  is related to the smoothness of the needle trajectory. As expected experts tend to have a lower velocity variance along this axis. Also, as expected, the lower the overshoot, the most likely it is that the user is an expert. The  $z$  axis being the vertical axis, thus related to the inclination of the needle, `vzvar.haptic` (the variance of velocity along the  $z$ -axis) is related to the pitch of the insertion. The fact that experts tend to have a higher value on this variable may indicate that they tend to not stay on a horizontal plane while inserting the needle.

Unfortunately, this type of method does not provide sufficient performance in terms of classification rate as shown in Fig.6. Nevertheless, it does have some interests in a teaching environment. Indeed, it does highlight the important variables and thus provides some pointers to improve one's performance by showing which compartment to focus on. In this case, having a steady pace during the insertion appears to be a key factor of masterful lumbar epidural anesthesia.

#### D. Data classification : a more complex approach

1) *Cleaning the data:* Before implementing a more advanced classification method, the data can be "cleaned up" to remove any less important components. In that

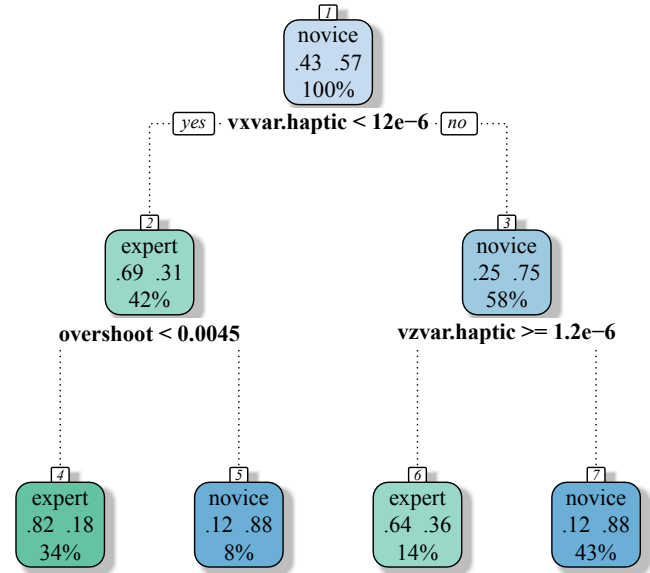


Figure 5: Example of classification tree obtained where `vxvar.haptic` is the velocity variance of the needle tip along  $x$  and `vzvar.haptic` the velocity variance of the needle tip along the  $z$  axis

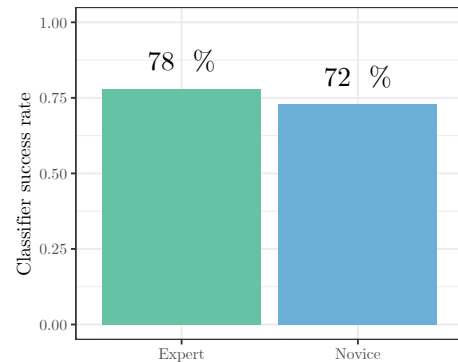


Figure 6: Classifier efficiency using the tree shown in Fig.5

regard, a Principal Components Analysis (PCA) is an unsupervised method used as a way to select the most relevant variables in the data-set. To do so, a PCA is applied and the most important components are selected using the PCA method from `FactoMineR` package [34]. The Fig.7 shows the obtained variance from the first components. To keep enough variance from the original data-set, five components are kept to rack up to more than 60% of the initial variance using the scree test [35].

From this, the initial variables are classified according to their importance in the construction of the first



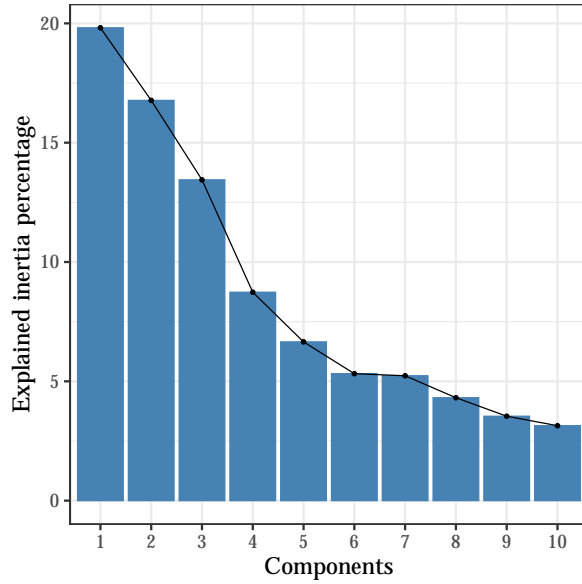


Figure 7: Total inertia represented by each component of the PCA

five components of the PCA, as shown in Fig.8. Knowing that some variables are crucial in the understanding of the outcome of the lumbar epidural anesthesia procedure, the different insertion angles, the number of insertions as well as the mean piston acceleration are taken out of the data-set as they do not seem to be impactful enough in terms of variability from one measurement to the next. In the end, the first five components account for almost 80% of the initial inertia once the undesirable components are taken out. The PCA result is shown in Fig.9.

Finally, to avoid the PCA to be lopsided toward a specific data point, the contribution of each observation is analyzed to seek out any abnormal ones. Table I displays the most important contributions across the five first components. We set a contribution threshold at 5 and gathered all observations that reach this threshold on any of the five first components of the PCA. Then we highlighted the highest in bold.

In this table, the observations taken out are :

- expert 5, try 1 because of its contribution toward the third component;
- expert 5, try 7 because of its contribution toward the third component.

We decided to keep the observations from novice 7 try 5 and 6 as the high contribution is located on the fifth component only. The presented observations, however,

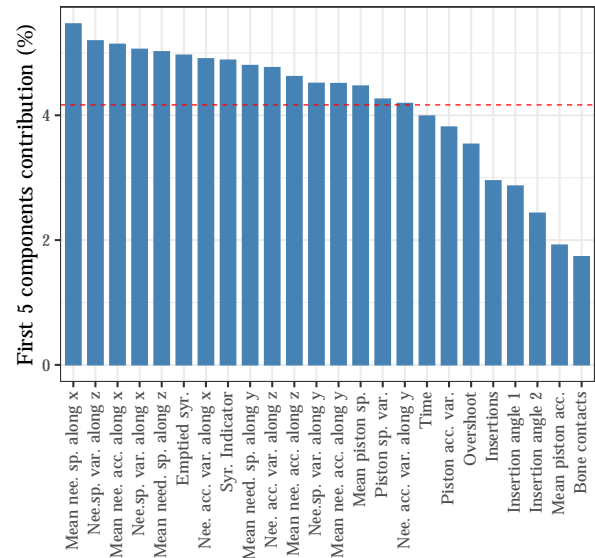


Figure 8: Sum of variables contributions to the first five dimensions of the PCA

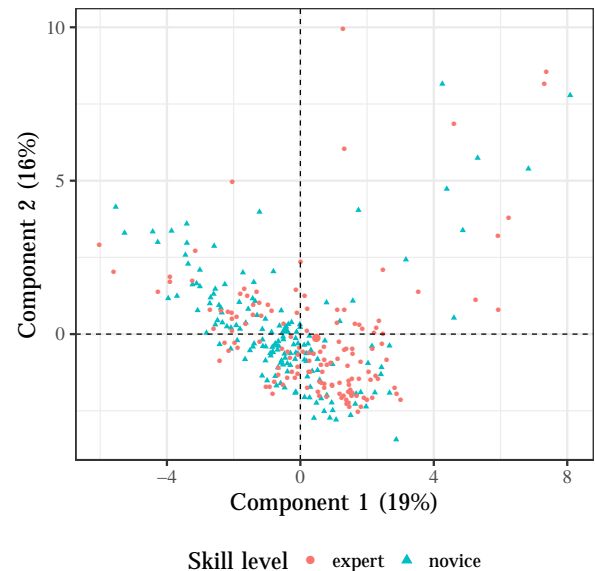


Figure 9: Visualization of the first two components of the PCA

Table I: Observations presenting high contributions to the five first variables

Observations	Co. 1	Co. 2	Co. 3	Co. 4	Co. 5
Nov. 7, try 5	0,31	11,83	0,17	7,39	<b>21,15</b>
Nov. 7, try 6	0,73	5,55	0,32	7,00	<b>18,18</b>
Nov. 9, try 3	0,85	5,90	0,64	0,05	0,01
Exp. 5, try 1	0,34	0,17	<b>25,30</b>	0,17	0,05
Exp. 5, try 4	0,98	8,67	0,08	0,43	0,99
Exp. 5, try 6	0,65	8,67	0,01	3,43	2,12
Exp. 5, try 7	0,35	0,29	<b>27,37</b>	1,55	0,05
Exp. 5, try 12	1,03	9,33	0,14	0,96	2,23

have been taken out of the data-set to establish the final version of the PCA used to classify the data. It has to be noted, that these two data points were re-introduced once the PCA had been computed. They were put apart only to get the best PCA result as possible.

2) *Implementing more advanced classification methods:* Once the data have been cleaned up using the method described above, another unsupervised method is used: t-SNE (t-distributed Stochastic Neighbor Embedding) [36] is applied to the data-set using the `Rtsne` method from `Rtsne` package [37] to get the result shown in Fig.10. The main objective of a t-SNE algorithm is to reorganize the data assuming it has a certain statistical distribution. By using this algorithm, similar data points were gathered, resulting in a more manageable data-set. With this algorithm, only the first three components were selected. As shown in Fig.10, there is a clear separation between groups of expert tries and novice tries, which is very encouraging. At that time the initial data-set was available in three different types:

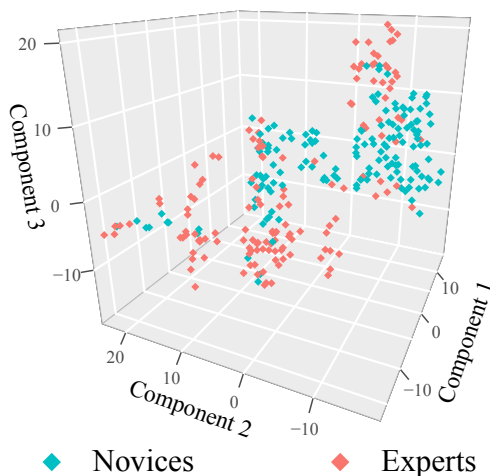


Figure 10: Representation of the data-set after a t-SNE analysis

- the "raw" measurements, once the non-significant variables were taken out;
- the PCA first five components;
- the t-SNE first three components.

From these results, three approaches were tested to try and find a successful enough classifier. The first tested classifier was a supervised method named: kNN algorithm (kNN method from `class` R package) [38]. For this algorithm, one only has to select the number of neighbors to consider. In this case, the optimal number of neighbors is selected for each version of the data-set looking at the success rate. The results are displayed in Fig.11. The kNN algorithm has the advantage of providing an estimation of its confidence for each ranking as it is possible to look at the number of neighbors of each category. As expected, this algorithm led to the best results using the first three components extracted by the t-SNE algorithm.

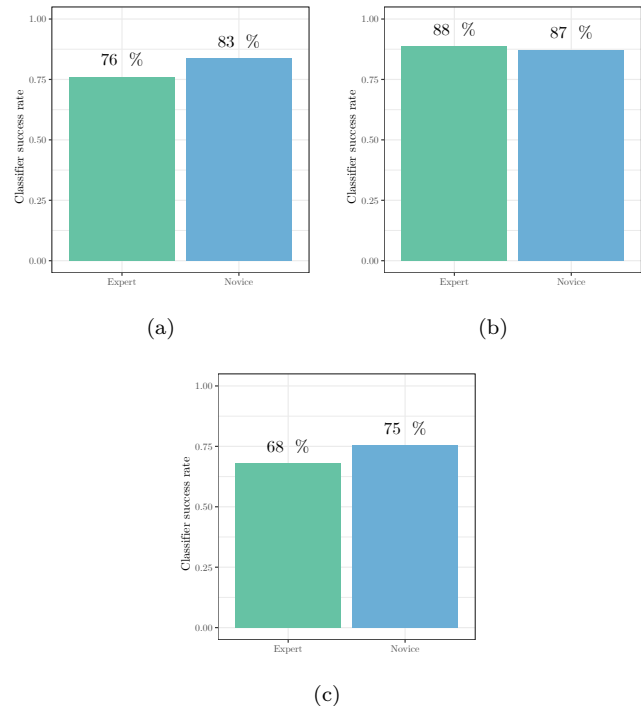


Figure 11: Classification success rates obtained using a kNN algorithm using: (a) PCA coordinates with  $k=16$ , (b) t-SNE coordinates using  $k=3$ , (c) raw data  $k=4$

Then another supervised method is used: SVM algorithm is implemented using the `SVM` method from the `e1071` package [39]. In this case, the most important part was to choose the most appropriate type of kernel



to delimit the different classes. Knowing the content of our data, it seemed interesting to use a radial kernel for the t-SNE version and a linear one was used for the rest. The success rate obtained using this type of algorithm is displayed in Fig.12. Once again it was possible to extract the confidence of the classification.

Finally, a third supervised method is implemented: basic neural network approach (`neuralnet` method from `neuralnet` package [40]). In this case, it was mostly a way of providing first-hand insight into this type of method that might need to be examined further to obtain better results. For each version of the data-set, the architecture (hidden layers) of the neural network was slightly modified using trial and error method to reach a decent enough result. These results are presented in Fig.13. The major drawback of this approach is that it does not provide more insight into the classification than the class itself. But once again, it was used to set a sort of basic benchmark to propose a more complete classification method.

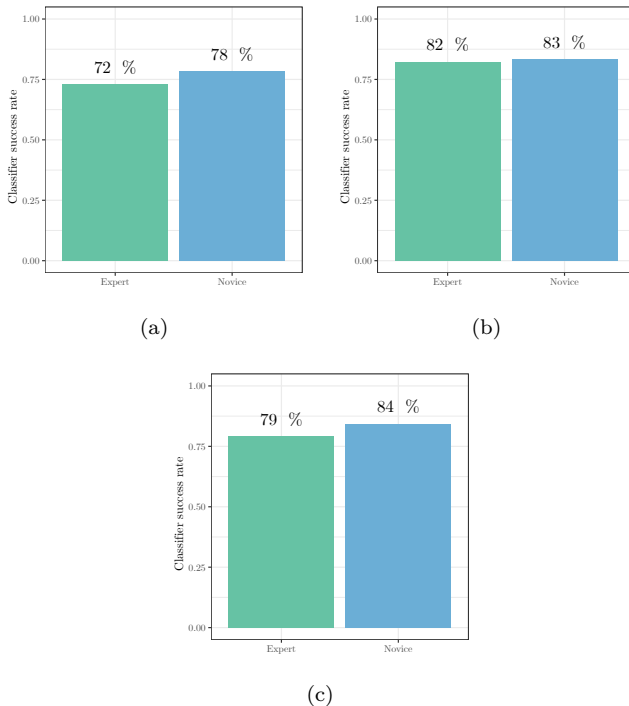


Figure 12: Classification success rates obtained using a SVM algorithm on : (a) PCA coordinates, (b) t-SNE coordinates, (c) raw data

These results are comparable to those obtained using an SVM approach or a kNN. If this may not be enough to be considered as the go-to method in this case. However, it seems that a more detailed approach to this

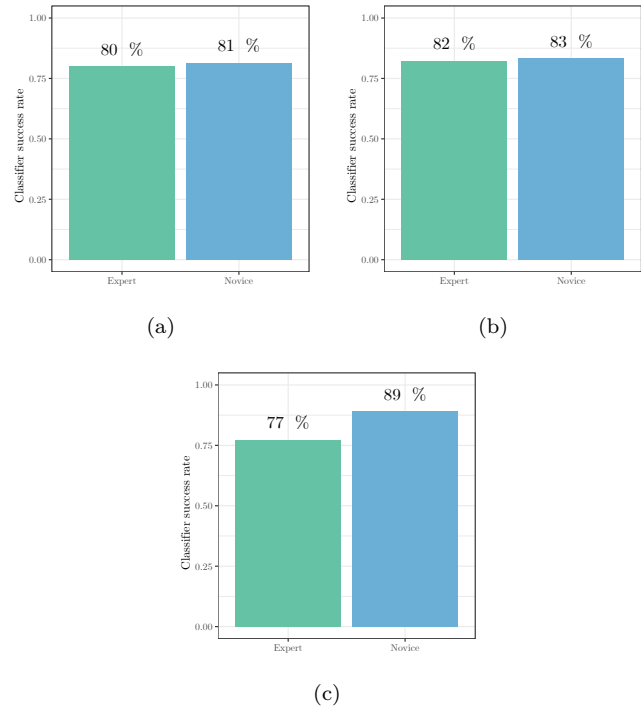


Figure 13: Classification success rates obtained using a neural network on: (a) PCA coordinates (hidden layers 6-4-1), (b) t-SNE coordinates (hidden layers 10-8-6-4-1), (c) raw data (hidden layers 10-8-6-4-1)

type of method may be a great source of improvement.

### E. Final results

The classification methods presented above are not sufficient to provide a general understanding of the skill of a person. Indeed, as it treats each trial separately, it does not inherently provide a general skill rating. To do so, it is possible to simply gather up the classifications of each person. Doing so allows us to put up a kind of skill rating by showing the proportion of expert rated attempts in all the trials of a person. The Table II and Table III shows the result of such a method on the data-set using the most effective classification algorithm described in the previous paragraph. These tables show the classifications obtained for each trial of a given subject in the first two columns. Then, the last two columns provide the final prediction and the confidence of this prediction. Whenever the algorithm is not able to conclude, for example when there are as many trials classified as expert and novice, the overall predicted ranking is undefined: a no-confidence index is issued.

Table II: Individual classification using data processed with a *t-SNE* algorithm and classed using a kNN with  $k = 3$

	Rankings		Evaluation	
	Novice	Expert	Pred.	Conf.
Novice 1	3	0	Novice	1
Novice 2	7	1	Novice	0,88
Novice 3	2	0	Novice	1
Novice 4	3	0	Novice	1
Novice 5	4	0	Novice	1
Novice 6	3	1	Novice	0,75
Novice 7	1	1	Undef.	-
Novice 8	4	0	Novice	1
Novice 9	5	2	Novice	0,71
Novice 10	2	0	Novice	1
Novice 11	5	0	Novice	1
Novice 12	4	0	Novice	1
Expert 1	0	2	Expert	1
Expert 2	1	3	Expert	0,75
Expert 3	0	3	Expert	1
Expert 4	0	4	Expert	1
Expert 5	1	3	Expert	0,75
Expert 6	0	2	Expert	1
Expert 7	0	2	Expert	1
Expert 8	0	5	Expert	1
Expert 9	3	3	Undef.	-
Expert 10	1	5	Expert	0,83
Expert 11	1	7	Expert	0,88
Expert 12	1	2	Expert	0,67

Table III: Individual classification using data processed with a PCA algorithm and classed using a kNN with  $k = 16$

	Rankings		Evaluation	
	Novice	Expert	Pred.	Conf.
Novice 1	3	0	Novice	1
Novice 2	4	4	Undef.	-
Novice 3	2	0	Novice	0,60
Novice 4	2	1	Novice	0,67
Novice 5	4	0	Novice	1
Novice 6	3	1	Novice	0,75
Novice 7	1	1	Undef.	-
Novice 8	3	1	Novice	0,75
Novice 9	5	2	Novice	0,71
Novice 10	2	0	Novice	1
Novice 11	5	0	Novice	1
Novice 12	4	0	Novice	1
Expert 1	0	2	Expert	1
Expert 2	2	2	Undef.	-
Expert 3	1	2	Expert	0,67
Expert 4	0	4	Expert	1
Expert 5	0	4	Expert	1
Expert 6	0	2	Expert	1
Expert 7	0	2	Expert	1
Expert 8	0	5	Expert	1
Expert 9	2	4	Expert	0,67
Expert 10	0	6	Expert	1
Expert 11	0	1	Expert	1
Expert 12	1	2	Expert	0,67

This methodology allows generating an overall skill rating for each person and drastically reducing the errors of the classifier. In this case, either using the kNN algorithm on the t-SNE coordinates or the PCA coordinates prevents classification errors. It only remains a few uncertain rating in each case.

#### IV. DISCUSSION

The three most common classification methods [41] have been implemented in order to compare expert and novice data recorded using the PeriSIM. All of them allow to distinguish experts and novices. As expected, their success rate depends on the method used to select the most relevant variables. This paper presents the different results of these methods according to the method of reduction used.

The results obtained using simple classification algorithms are quite encouraging. The proposed, overall skill rating algorithm allows getting a rather precise assessment of the skill of the user. There remain some uncertainties in the ranking provided by the algorithm. However, these uncertainties remain limited. Due to the relatively small test data-set, some users only count a few trials, which limits the accuracy of the algorithm. Indeed, raising the number of rated trials increases the precision of the ratings. Thus, more tries need to be recorded and rated using the data-set presented in this article as the training data-set.

Moreover, it might be interesting to investigate neural-network approaches deeper. Indeed, a simple and quick implementation of such a method already showed quite decent results. However, the lack of perspective on the result of the classification may be an important hurdle if such a classifier is used in a learning context.

Regarding the incorrect classifications, they might be caused by the specific trials being genuinely skillful or not. As each person performed all of his/her twelve trials in a row, novices may have increased their precision on some tries whereas some expert users may have taken the tries less seriously, thus resulting in some false ratings. This, however, does not take away the overall results obtained with this data-set.

In other possible improvements, it might be interesting to expand the algorithm to take into account the different types of patients provided by the simulator. It might be relevant to include different ratings, one for each patient type. However, with the current amount of recorded tries, dividing the data-set into three separate

ones might thin the data-set too much to be able to provide a precise enough ranking algorithm.

Another limit of this study is the lack of perspective about the use of such a simulator on the learning curve of the students. However, the presented results allow considering using this simulator in a teaching environment.

## V. CONCLUSION

In this paper, we proposed a simple yet but effective rating algorithm to assess the skill of a user when doing an lumbar epidural anesthesia procedure on the PeriSIM simulator. This haptic simulator has been developed previously in the Ampère Laboratory and aims at providing an efficient and affordable training tool for novice anesthetists. This study gathered a total of more than 200 data points and used them to design and test an objective and efficient skill assessment algorithm. The results obtained are encouraging and may require more recordings to deeply test the proficiency of the algorithm. Some changes might be undertaken to the type of algorithm used to classify the tries such as neural network approaches.

In conclusion, the proposed ranking algorithm may be summed up as presented in Fig.14. This complete algorithm can easily be improved using a different classification algorithm if needed.

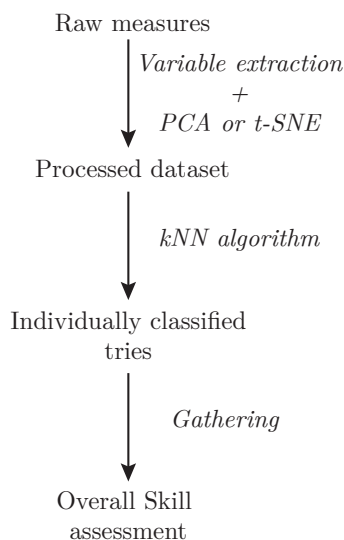


Figure 14: Summary of the skill assessment algorithm presented in this article for the epidural anesthesia procedure

## VI. ACKNOWLEDGMENT

Authors would like to thank the Neuro-immersion tech facility of the CRNL (Centre de Recherche en Neurosciences de Lyon) for their material support and the ANR (French National Research Agency) for financing SAMSEI project (ANR-11-IDFI-0034) under the supervision of Pr. X. Martin.

## REFERENCES

- [1] N. Vaughan, V. N. Dubey, M. Y. Wee, and R. Isaacs, "A review of epidural simulators: Where are we today?" *Medical Engineering & Physics*, vol. 35, no. 9, pp. 1235–1250, Sept. 2013. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S135045331300057X>
- [2] J.-C. Granry and M.-C. Moll, "Etat de l' art (national et international) en matiere de pratiques de simulation dans le domaine de la sante," Haute Autorite de la Sante (HAS), Tech. Rep., 2012.
- [3] V. Manoharan, D. van Gerwen, J. J. van den Dobbelen, and J. Dankelman, "Design and validation of an epidural needle insertion simulator with haptic feedback for training resident anaesthesiologists," in *Haptics Symposium (HAPTICS), 2012 IEEE*. IEEE, 2012, pp. 341–348.
- [4] J. C. Magill, M. F. Byl, M. F. Hinds, W. Agassounon, S. D. Pratt, and P. E. Hess, "A Novel Actuator for Simulation of Epidural Anesthesia and Other Needle Insertion Procedures," *Simulation in Healthcare*, vol. 5, no. 3, pp. 179–184, June 2010. [Online]. Available: <https://insights.ovid.com/crossref?an=01266021-201006000-00009>
- [5] L. Hiemenz, J. McDonald, D. Stredney, and D. Sessanna, "A physiologically valid simulator for training residents to perform an epidural block," in *Proceedings of the 1996 Fifteenth Southern Biomedical Engineering Conference*, Mar. 1996, pp. 170–173.
- [6] D. Thao, A. T. M, and S. M. A, "Development and Evaluation of an Epidural Injection Simulator with Force Feedback for Medical Training," *Studies in Health Technology and Informatics*, pp. 97–102, 2001. [Online]. Available: <http://www.medra.org/servlet/aliasResolver?alias=iospress&ISSN&ISSN=0926-9630&volume=81&spage=97>
- [7] Epimed, "Thoraxis neuraxial simulator," 2016. [Online]. Available: <http://www.epimed.com/products/thoraxis/>
- [8] NYSORA, "Epidural anesthesia block ultrasound simulator," <https://www.nysora.com/nysora-ultrasound-simulators-epidural-anesthesia-block/> (visited on 2020-12-15).
- [9] J. Bortman, Y. Baribeau, J. Jeganathan, Y. Amador, F. Mahmood, M. Shnider, M. Ahmed, P. Hess, and R. Matyal, "Improving clinical proficiency using a 3-dimensionally printed and patient-specific thoracic spine model as a haptic task trainer," *Regional Anesthesia & Pain Medicine*, vol. 43, no. 8, pp. 819–824, 2018. [Online]. Available: <https://rapm.bmj.com/content/43/8/819>
- [10] M. Han, A. A. Portnova, M. Lester, and M. Johnson, "A do-it-yourself 3d-printed thoracic spine model for anesthesia resident simulation," *PLOS ONE*, vol. 15, no. 3, pp. 1–10, 03 2020. [Online]. Available: <https://doi.org/10.1371/journal.pone.0228665>

- [11] T. Senac, A. Leleve, R. Moreau, L. Krahenbuhl, F. Sigwalt, C. Bauert, and Q. Rouby, "Designing an Accurate and Customizable Epidural Anesthesia Haptic Simulator," in 2019 International Conference on Robotics and Automation (ICRA). Montreal, QC, Canada: IEEE, May 2019, pp. 8353–8359. [Online]. Available: <https://ieeexplore.ieee.org/document/8794199/>
- [12] P.-J. Alès Roux, N. Herzig, A. Lelevé, R. Moreau, and C. Bauer, "3D Haptic Rendering of Tissues for Epidural Needle Insertion using an Electro-Pneumatic 7 Degrees of Freedom Device," in 2016 IEEE IROS, IEEE, Ed., no. 5175–5180. Daejeon, South Korea: IEEE, Oct. 2016. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01340723>
- [13] M. C. Vassiliou, L. S. Feldman, C. G. Andrew, S. Bergman, K. Leffondre, D. Stanbridge, and G. M. Fried, "A global assessment tool for evaluation of intraoperative laparoscopic skills," The American journal of surgery, vol. 190, no. 1, pp. 107–113, 2005.
- [14] "OSATS," <http://www.csats.com/osats> (visited on 2020-12-15).
- [15] J. A. Martin, G. Regehr, R. Reznick, H. Macrae, J. Murnaghan, C. Hutchison, and M. Brown, "Objective structured assessment of technical skill (OSATS) for surgical residents," British journal of surgery, vol. 84, no. 2, pp. 273–278, 1997.
- [16] "GOALS," <http://www.csats.com/goals> (visited on 2020-12-15).
- [17] H. Niitsu, N. Hirabayashi, M. Yoshimitsu, T. Mimura, J. Taomoto, Y. Sugiyama, S. Murakami, S. Saeki, H. Mukaida, and W. Takiyama, "Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room," Surgery Today, vol. 43, no. 3, pp. 271–275, Mar. 2013. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3574562/>
- [18] G. Sroka, L. S. Feldman, M. C. Vassiliou, P. A. Kaneva, R. Fayez, and G. M. Fried, "Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room: a randomized controlled trial," The American journal of surgery, vol. 199, no. 1, pp. 115–120, 2010.
- [19] S. K. MISHRA, A. GANPULE, A. KURIEN, V. MUTHU, and M. R. DESAI, "Task completion time: Objective tool for assessment of technical skills in laparoscopic simulator for urology trainees," Indian journal of urology, vol. 24, no. 1, pp. 35–38, Jan. 2008.
- [20] C. Feng, J. W. Rozenblit, and A. Hamilton, "Fuzzy Logic-Based Performance Assessment in the Virtual, Assistive Surgical Trainer (VAST)," in 15th Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems (ecbs 2008), Mar. 2008, pp. 203–209.
- [21] S. Cotin, N. Stylopoulos, M. Ottensmeyer, P. Neumann, D. Rattner, and S. Dawson, "Metrics for laparoscopic skills trainers: The weakest link!" in Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002, T. Dohi and R. Kikinis, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 35–43.
- [22] J. A. Cifuentes-Quintero, P. Boulanger, M. T. Pham, R. Moreau, and F. Prieto, "Automatic Gesture Analysis Using Constant Affine Velocity," in 36th IEEE EMBC, Chicago, IL, United States, Aug. 2014. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01076675>
- [23] A. M. Cometa, B. M. Lopez, T. Vasilopoulos, A. J. Destephens, A. Bigos, D. E. Lizzdas, N. Gravenstein, and S. Lampotang, "Does the technique for assessing loss of resistance alter the magnitude of epidural needle tip overshoot?" Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare, vol. 15, no. 3, pp. 154–159, June 2020.
- [24] A. Okamura, C. Simone, and M. O’Leary, "Force Modeling for Needle Insertion Into Soft Tissue," IEEE Transactions on Biomedical Engineering, vol. 51, no. 10, pp. 1707–1716, Oct. 2004. [Online]. Available: <http://ieeexplore.ieee.org/document/1337139/>
- [25] K. Kuchenbecker, J. Fiene, and G. Niemeyer, "Improving contact realism through event-based haptic feedback," IEEE Transactions on Visualization and Computer Graphics, vol. 12, no. 2, pp. 219–230, Mar. 2006. [Online]. Available: <http://ieeexplore.ieee.org/document/1580456/>
- [26] "Double-Acting Archives," <http://airpot.com/product-category/product-lines/pneumatic-actuation/airpel-anti-stiction-air-cylinders/metric-models/double-acting-metric-models/> (visited on 2020-12-15).
- [27] "MPS-064tsnu0 | Magnetic cylinder sensors SICK," <https://www.sick.com/ag/en/magnetic-cylinder-sensors/position-sensors/mps-t/mps-064tsnu0/p/p230381> (visited on 2020-12-15).
- [28] "Miniature Ratiometric Pressure Transducer | EPRB Series | TE Connectivity," <http://www.te.com/usa-en/product-CAT-PTT0007.html> (visited on 2020-12-15).
- [29] D. Tran, King-Wei Hor, A. Kamani, V. Lessoway, and R. Rohling, "Instrumentation of the Loss-of-Resistance Technique for Epidural Needle Insertion," IEEE Transactions on Biomedical Engineering, vol. 56, no. 3, pp. 820–827, Mar. 2009. [Online]. Available: <http://ieeexplore.ieee.org/document/4760227/>
- [30] L. van ADRICHEM, "Avoiding overshoot," DSPE Mikroniek, vol. 3, pp. 36–40, 2009.
- [31] A. GHASEMLOONIA, Y. MADDAHI, K. ZAREINIA, S. LAMA, J. C. DORT, and G. R. SUTHERLAND, "Surgical Skill Assessment Using Motion Quality and Smoothness," Journal of Surgical Education, vol. 74, no. 2, pp. 295–305, Mar. 2017. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1931720416302100>
- [32] N. Zivkovic, G. van Samkar, H. Hermanns, P. Lirk, M. W. Hollmann, J. J. van den Dobbelen, D. J. van Gerwen, and M. F. Stevens, "Face and construct validity of tu-delft epidural simulator and the value of real-time visualization," Regional Anesthesia & Pain Medicine, vol. 44, no. 3, pp. 298–302, 2019. [Online]. Available: <https://rapm.bmj.com/content/44/3/298>
- [33] T. THERNEAU, B. ATKINSON, B. RIPLEY, and M. B. RIPLEY, "Package ?rpart?" Available online: [cran.ma.ic.ac.uk/web/packages/rpart/rpart.pdf](http://cran.ma.ic.ac.uk/web/packages/rpart/rpart.pdf) (accessed on 20 April 2016), 2015.
- [34] F. HUSSON, S. LE, and J. PAGES, Exploratory multivariate analysis by example using R, second edition ed., ser. CRC computer science and data analysis series. Boca Raton: CRC Press, 2017.
- [35] R. B. CATTELL, "The Scree Test For The Number Of Factors," Multivariate Behavioral Research, vol. 1, no. 2, pp. 245–276, Apr. 1966. [Online]. Available: [https://doi.org/10.1207/s15327906mbr0102\\_10](https://doi.org/10.1207/s15327906mbr0102_10)

- [36] L. v. d. MAATEN and GEOFFREY Hinton, “Visualizing Data using t-SNE,” Journal of Machine Learning Research, no. 9, 2008.
- [37] J. KRIJTHE and LAURENS van der Maaten (Author of original C++, “Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation,” Nov. 2018. [Online]. Available: <https://CRAN.R-project.org/package=Rtsne>
- [38] B. RIPLEY and W. VENABLES, “class: Functions for Classification,” Jan. 2019. [Online]. Available: <https://CRAN.R-project.org/package=class>
- [39] D. MEYER, E. DIMITRIADOU, K. HORNIK, A. WEINGESSEL, F. LEISCH, t. l. C++-code), and t. l. C++-code), “e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien,” June 2019. [Online]. Available: <https://CRAN.R-project.org/package=e1071>
- [40] F. Günther and S. Fritsch, “neuralnet: Training of Neural Networks,” The R Journal, vol. 2, no. 1, pp. 30–38, 2010. [Online]. Available: <https://doi.org/10.32614/RJ-2010-006>
- [41] R. Dias, A. Gupta, and S. Yule, “Using machine learning to assess physician competence: A systematic review,” Acad Med., vol. 94, no. 3, pp. 427–439, Mar. 2019.