Development of a machine learning model for automatic assessment of performance in virtual reality medical simulators

Lucas H. Sallaberry
Romero Tori
lucas.sallaberry@usp.br
tori@usp.br
Polytechnic School, University of São Paulo
São Paulo, Brazil

Fátima L. S. Nunes
fatima.nunes@usp.br
School of Arts, Sciences and Humanities, University of São Paulo
São Paulo, Brazil

ABSTRACT

Virtual reality simulators for medical training allow the student to learn and train medical procedures safely, before performing on a real patient. By collecting data from the trainees while they perform in the simulator, it is possible for the system to automatically give feedback about the student's performance in the procedure, in real time, making the training more independent, while also yielding information to the instructor that can be insightful regarding their student's learning. As studies using machine learning techniques to automatically assess performance in virtual reality simulators are scarce in the literature, this work proposes a model by comparing the results of different classifiers by using data collected in a dental simulator. The results so far have shown that the Naive Bayes classifier has achieved a lower Accuracy in most cases when compared to the Support Vector Machine, Multi-layer Perceptron and Random Forest classifiers.

CCS CONCEPTS

• **Applied computing** → *Health informatics*; • **Computing methodologies** → **Interactive simulation**; • **Human-centered computing** → *Virtual reality*.

KEYWORDS

Performance assessment, force feedback, dental simulator.

1 INTRODUCTION

Virtual Reality (VR) simulators for medical training can allow the student to learn and train a procedure safely, multiple times, without loss of material [4]. This kind of system can also assess the student's performance automatically, without the need of an instructor present, allowing for a more independent training experience.

The VIDA Odonto VR simulator is a system developed in a partnership between the Laboratory of Interactable Technologies (Interlab - EPUSP), Laboratory of Computer Applications for Health Care (LApIS - EACH/USP) e Laboratory of Simulation and Training (LaSiT - FOB/USP), currently focused on the teaching and training of the Inferior Alveolar Nerve Block (IANB) procedure, in a realistic virtual environment that is safe for the student's practice [5].

The objective of this MSc dissertation is to develop a model to automatically assess trainee's performance in VR medical simulators by using the data collected during training, and is related to the topics "Machine learning for VR" and "User studies and evaluation".

2 MOTIVATION

VR simulators are capable of collecting data related to the user's performance in the simulated procedure, such as the movements made with the medical instrument, without the need of additional peripherals. It is possible to use this data to implement models that can assess the student's performance automatically during the simulation, and even point out the aspects of the procedure that need more attention. The benefits of this approach are both for the student, by providing a more independent training, but also for the instructor, by providing useful, objective information about the performance that can be hard to examine by observation, such as velocity and precision. Peripherals, such as haptic devices that simulate force feedback, can also provide data that could only be obtained through sensors in a real simulation setting, such as the force used during the procedure.

3 RELATED WORK

We conducted a systematic review of the literature regarding studies that assessed the performance of participants in VR medical simulators automatically. Most studies used statistical techniques to assess performance (87%), such as the one presented by Brewin *et al.* [3], 29% used performance features to calculate a Score related to participant's performance, such as the study in [1], and Performance Curve was also a technique found in 21% of the studies, such as [2], where the participant repeats the same procedure multiple times in the simulator, and the features are plotted on each attempt, creating a curve that reflects how the performance changed through the trials. Machine Learning (ML) techniques were found in only four studies (6%), showing that this type of technique is underexplored within this category of application.

4 PROPOSED SOLUTION

The model has been developed for a VR dental simulator [5] as a proof of concept. In previously conducted experiments, the data from inexperienced and experienced participants in the IANB procedure was collected while they performed on the VIDA Odonto simulator, and this data will be used to train ML algorithms that can evaluate new learners.

Figure 1 shows the architecture of the proposed model. The data collected in the simulator consisted of the position (x, y and z in the virtual environment) and orientation (the syringe rotation in relation to the z, y and x axis, in this order). These values were sampled 60 times per second during the IANB procedure simulation. Features related to the performance of the user were then calculated

from this data, and algorithms for feature selection/fusion were applied to the dataset separately. The resulting features were used to train the algorithms. Once the best model is selected, it can then be integrated into the simulator so new trainees can have feedback on their performance automatically, as they train in it.
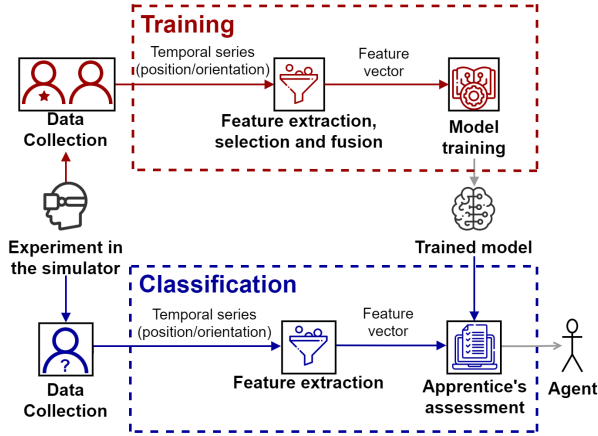

Figure 1: Proposed model architecture.

5 PRELIMINARY RESULTS

In a systematic review previously conducted, we gathered the performance metrics from the studies and found that 63% of the studies used the "Time taken for the whole procedure", which was the most used performance metric across studies, followed by "Instrument path length", which was used in 30% of the studies.

Using the data collected in the VIDA Odonto simulator from 43 inexperienced participants and 112 experienced participants in the IANB procedure, we calculated several features related to their performance in the procedure. Three different datasets were used to train the models: with the original features, with features selected using ReliefF (a filter type algorithm that selects features based on how well they can distinguish between elements of different classes) and with features calculated using Principal Component Analysis (PCA). For each dataset, the algorithms Support Vector Machine (SVM), Random Forest (RF), Multi-layer Perceptron (MLP) and Naive Bayes (NB) were trained.

The results shown in Table 1 are the mean Accuracy between folds on a ten folds stratified cross validation for the configuration of hyperparameters that achieved the highest mean Accuracy for each combination of algorithm (RF, NB, MLP and SVM) and dataset (Original, ReliefF and PCA). The NB classifier achieved the lowest Accuracies overall with the Original and PCA datasets, while the combination SVM/ReliefF achieved the highest Accuracy.

Additionally, an instructor evaluated the trajectories collected by the simulator during the experiments, and scored each of them. From the scores, we are training the algorithms to identify how trainees performed, based on the instructor assessment.

6 CONCLUSIONS

Although many studies in the literature attempt to assess the performance of users in VR simulators, only few of them apply ML

Table 1: Highest mean (standard deviation) Accuracy obtained from ten fold cross validation for each algorithm/dataset combination.

	RF	NB	MLP	SVM
Original	**0.75 (0.08)**	0.48 (0.14)	0.72 (0.10)	0.73 (0.02)
PCA	0.69 (0.15)	0.54 (0.13)	0.72 (0.09)	0.72 (0.02)
ReliefF	0.70 (0.11)	**0.71 (0.08)**	**0.73 (0.12)**	**0.76 (0.07)**

techniques for this proposal. From results obtained with different ML techniques and datasets, we found that the NB classifier had some of the worst Accuracies overall, while the best Accuracy was achieved with the combination SVM/ReliefF. In the next step we will evaluate new ML techniques and other manners of segmenting the dataset in order to develop a model for automatic assessment of performance to be integrated into the VIDA Odonto simulator.

ACKNOWLEDGMENTS

REFERENCES

[1] Kashif Akhtar, Kapil Sugand, Matthew Sperrin, Justin Cobb, Nigel Standfield, and Chinmay Gupte. 2015. Training Safer Orthopedic Surgeons: Construct Validation of a Virtual-Reality Simulator for Hip Fracture Surgery. *Acta Orthopaedica* 86, 5 (Sept. 2015), 616–621.

[2] S. M. B. I. Botden, I. H. J. T. de Hingh, and J. J. Jakimowicz. 2009. Suturing Training in Augmented Reality: Gaining Proficiency in Suturing Skills Faster. *Surgical Endoscopy* 23, 9 (Sept. 2009), 2131–2137.

[3] James Brewin, Tim Nedas, Ben Challacombe, Oussama Elhage, Jonas Keisu, and Prokar Dasgupta. 2010. Face, Content and Construct Validation of the First Virtual Reality Laparoscopic Nephrectomy Simulator. *BJU International* 106, 6 (Sept. 2010), 850–854.

[4] Cleber Gimenez Correa, Romero Tori, and Fatima L.S. Nunes. 2013. Haptic Simulation for Virtual Training in Application of Dental Anesthesia. In *2013 XV Symposium on Virtual and Augmented Reality*. IEEE, Cuiabá - Mato Grosso, Brazil, 63–72.

[5] Romero Tori, Gustavo Ziyu Wang, Lucas Henna Sallaberry, Allan Amaral Tori, Elen Collaço de Oliveira, and Maria Aparecida A. M. Machado. 2018. VIDA ODONTO: Ambiente de Realidade Virtual Para Treinamento Odontológico. *Revista Brasileira de Informática na Educação* 26, 02 (May 2018), 80–101.