Development of Virtual Skill Trainers and Their Validation Study Analysis Using Machine Learning

Seema Shedage
University of Central Arkansas
sshedage1@uca.edu

Jake Farmer
University of Central Arkansas
jfarmer2@uca.edu

Doga Demirel
Florida Polytechnic University
ddemirel@floridapoly.edu

Tansel Halic
University of Central Arkansas
tanselh@uca.edu

Sinan Kockara
University of Central Arkansas
skockara@uca.edu

Venkata S. Arikatla
Kitware Inc.
sreekanth.arikatla@kitware.com

Kevin Sexton
University of Arkansas for Medical Sciences
ksexton@uams.edu

Shahryar Ahmadi
University of Arkansas for Medical Sciences
sahmadi@uams.edu

ABSTRACT

Minimally invasive skills assessment is important in developing competent surgical simulators and executing reliable skills evaluation [9]. Arthroscopy and Laparoscopy surgeries are considered Minimally Invasive Surgeries (MIS). In MIS, the surgeon operates through small incisions with specialized narrow instruments, fiberoptic lights, and a monitor. Arthroscopy surgery is used to diagnose and treat joints problems, and Laparoscopic procedures are performed on the abdominal cavity. Due to non-natural hand-eye coordination, narrow field-of-view, and limited instrument control, MIS training is challenging to master. We are analyzing two simulators' data, Virtual Arthroscopic Tear Diagnosis and Evaluation Platform (VATDEP) and Gentleness Simulator. Both simulators went through the validation studies with human subjects. We recorded simulation data during the validation studies, such as tool motion, position, and task time. Recorded data went through the data preprocessing; after the data cleaning, we extracted the recoded data features and normalized them. Normalized features were used to input various machine learning algorithms, including K-nearest neighbor (KNN), Support vector machine (SVM), and Logistic regression (LR). The average accuracy was evaluated through k-fold cross-validation. The proposed methods validated using 10 subjects (5 experts, 5 novices) for the VATDEP simulator. 23 subjects (4 experts and 19 novices) for the Gentleness Simulator. The result shows a significant difference between the expert and novice population with the $p < 0.05$ using the Mann-Whitney U-test. The VATDEP simulator's classification algorithms' average accuracy is 74% and 80% for the Gentleness Simulator. The results show that the normalized features and with KNN, SVM, and LR classifiers can provide accurate classification of experts and novices. The evaluation technique proposed in this study can develop surgical training by providing appropriate feedback to trainees to evaluate proficiency.

ICISDM 2021, May 27–29, 2021, Silicon Valley, CA, USA

ACM ISBN 978-1-4503-8954-9/21/05…$15.00
https://doi.org/10.1145/3471287.3471296

CCS CONCEPTS

• **Human-centered computing** → Human-computer interaction (HCI); Interactive systems and tools, Computing methodologies; Modeling, and simulation; Simulation evaluation..

KEYWORDS

Machine Learning, Virtual Reality Simulator, Validation Study

ACM Reference Format:
Seema Shedage, Jake Farmer, Doga Demirel, Tansel Halic, Sinan Kockara, Venkata S. Arikatla, Kevin Sexton, and Shahryar Ahmadi. 2021. Development of Virtual Skill Trainers and Their Validation Study Analysis Using Machine Learning. In *2021 the 5th International Conference on Information System and Data Mining (ICISDM 2021), May 27–29, 2021, Silicon Valley, CA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3471287.3471296

1 INTRODUCTION

Minimally invasive surgeries (MIS) popularity, also known as minimal access, has risen widely and has become a universally accepted surgical technique [12]. Unlike traditional open surgery, where operations are performed through a large incision on the body, surgical procedures in MIS only require tiny incisions (e.g., pencil size) onto the patient's skin. In MIS, the entire procedure is performed with instruments and a fiber-optic camera inserted through the incisions. In arthroscopy, the surgeons need to assess and approach the rotator cuff from several different angles to delineate the tear pattern and then repair it anatomically fully. However, due to unintuitive hand-eye coordination, narrow and confined field-of-view, and confined space for instrument control, arthroscopy training is challenging to master.

An additional issue with MIS is the surgeon's gentleness while handling the tissue during the surgery. During the surgery, to avoid tissue damages, a medical practitioner must provide gentle handling of tissue. Tissue injuries can cause blood loss and long hospital recovery time, dead space, and meticulous hemostasis [3]. Gentle handling of tissues is turn out to be a vital metric to assess the beginners in MIS. Some institutes, such as the American Board of Surgery,

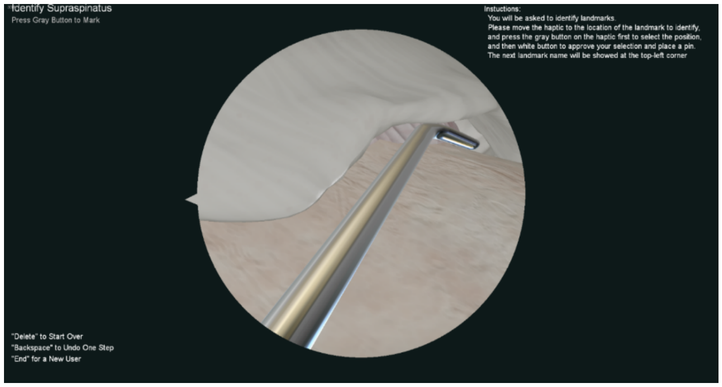

Figure 1: VATDEP Tear Diagnostic Task with Probe and Tear.

have already added gentleness assessments before assigning the new physician's certification [13, 14]. As gentleness is a primary determinant of surgical skill, and surgical skill is a strong predictor of surgical outcomes [2], a need for a way to safely practice and maintain gentleness with tissue is needed. Due to the lesser amount of force sensors, some attempts lack the appropriate measure of gentleness [8] or measured the force and acceleration on only one hand [10], or only measured one specific motion [7]. The benefits of virtual reality (VR) simulators can vary, such as easy accessibility with objective, quantitative feedback [14]. Standard training methods for arthroscopic and laparoscopic surgeries are subjective and require assessment tools [14], especially for gentleness. Therefore, gentleness assessment can benefit from a VR training environment [1]. Therefore, we have designed and developed Virtual Arthroscopic Tear Diagnosis and Evaluation Platform (VATDEP) and Gentleness simulators under experts' surgeon's guidance to measure a surgeon's gentleness and additional for arthroscopic rotator cuff repair training. Using simulation data from these simulators, we have classified surgeons and medical students into two groups and assessed these simulators' validity in an institutional setting.

2 STUDY DESIGN

Gentleness Simulator, as well as the VATDEP simulator [4], was created using Software Framework for Multimodal Interactive Simulations (SoFMIS) [6], a highly customizable, multithreaded simulation framework. SoFMIS allows for a quick and modular approach to creating visually realistic simulations with easy integration for haptic devices [11], external interfaces such as Arduino or any data acquisition (DAQ) devices, EMS22 sensors [15], and simplistic data recording. This framework allows for easy extension of modules within it by merely extending classes and providing a custom implementation to fit the developer's current needs. Many aspects of the simulation, such as rendering, physics simulation, object properties, and scene environments, are encapsulated, and in some respects, abstracted from the developer for simplicity's sake. Both studies are discussed in detail in sections 2.1 and 2.2.

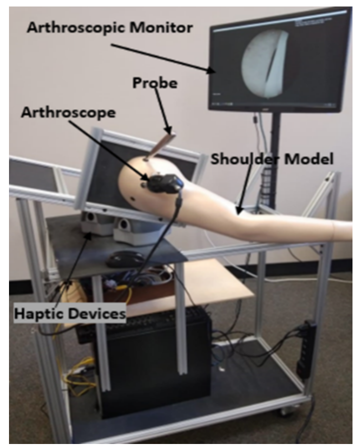

Figure 2: VATDEP Simulator Cart to Generate the Surgery Experience

2.1 Virtual Arthroscopic Tear Diagnosis and Evaluation Platform

VATDEP was used for evaluating subjects' anatomical knowledge and arthroscopic navigation skills for the rotator cuff tears. The rotator cuff contains muscles and tendons that surround the shoulder joint. During the VATDEP validation study, a total number of 10 attendees participated. The first task was the landmark identification task. In the landmark identification task, subjects had to identify major anatomical structures such as ligaments, tendons, and muscles within the shoulder anatomy. PBR is used to achieve a high level of visual realism. Figure 1 shows an initial view of the screen and a close up of the probe and tear.

VATDEP integrates simple arthroscopic tools into the simulator's physical portion, using an arthroscope and probe handle to give the subject a familiar setup used in the operating room. Figure 2 shows a customized cart's final design with an attached screen and human shoulder model to have real surgery experience.

The cart has been made with 80/20®of aluminum supporting two standard patient positioning techniques for shoulder arthroscopy, beach chair, and lateral decubitus positions 0. The cart also contains

Figure 3: Double Grasper Task Through the Different Stages

a mannequin human shoulder with widely used portal locations where instruments and arthroscopes are introduced. The portal locations are identified with our expert surgeon collaborator.

2.2 Gentleness Simulator

The gentleness simulator has two unique simulation scenarios to evaluate the gentleness for tissue manipulation and instrument handling, especially for laparoscopic surgeries; the first scenario is to use a tennis racket, and the second scenario is a double grasper task. (See section 2.2.1 and section 2.2.2)

2.2.1 Double Grasper Task. In this task, the subject needs to use both hands to handover the balloon from the right side of the task to the left side using graspers. If the grasper's jaws are tightly closed while transferring the balloon, the excessive force would be applied to the balloon and cause shaking during the transfer. Due to the shaking, the balloon can pop, and the task must be restarted. The subjects have an additional challenge introduced by passing the balloon through an archway in the middle of the scene to complete the transfer task. If the subject does not use reasonable force to grasp the soft body, it could get loose from the grasper and drop. This state is recorded as a penalty, and also the task completion time could increase. Figure 3 shows the scenes and phases for the Double Grasper Task.

2.2.2 Tennis Racket Task. In this task, the subject needs to tap the soft balloon to keep the balloon between two virtual planes slightly separated in the up direction.

As the subject performs the task during the validation study, motions regarding the instruments and the balloon are recorded in real-time. We used the Phantom Omni devices for instruments and haptic force feedback. Nvidia's PhysX software development kit was used to improve the soft body realism of the simulator. Soft body and haptic interaction forces are computed based on the normal to the tennis racket's head, allowing a realistic feeling over the surface area. This reaction force adds challenge to the task, as the subject must overcome the balloon's weight. Otherwise, they can likely fail to maintain the desired height. The soft body also holds the characteristic of popping if the applied force is greater than the threshold value. Figure 4 shows the tennis racket task.

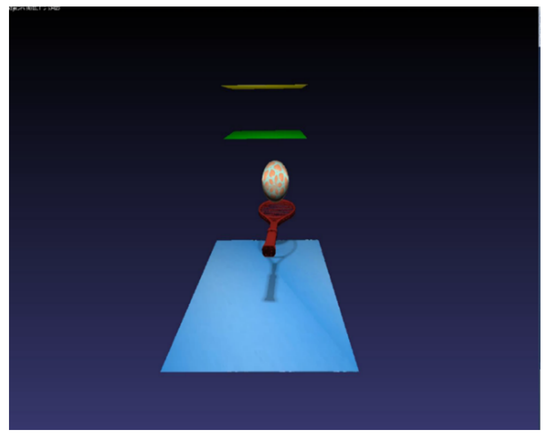

Figure 4: Soft body and Tennis Task

3 VALIDATION STUDY

We validated our simulator with human subjects at the University of Arkansas for Medical Science (UAMS). The study involved a total of 10 participants for VATDEP and 23 participants for Gentleness Simulator. Before the beginning of a designated task, the subjects were given a pre-questionnaire with questions about age, level of experience, hand dominance, etc. Following the questionnaire, the users were verbally told about the task. Each subject was given a certain amount of time to get familiar with the task and study environment. Then the subjects were expected to complete the tasks without any particular time limit. Once the tasks were completed, the subjects were given a post-questionnaire with questions regarding the realism and effectiveness of the simulator and open-ended questions for feedback on the simulator.

3.1 Data sets

For each participant, we recorded instrument positions in virtual scene distance units (mm), forces exerted forces on the anatomical structures in newton, arthroscope motions, virtual pin locations placed on anatomical structures and distance to the actual locations of anatomical structures, velocity(mm/s), and acceleration (mm/s2) of the tools (e.g., probe and shaver), shaved regions, shaving speed, and arthroscope motions, Time to complete the task and its sub-tasks (e.g., Time for each landmark) for VATDEP simulator. In the Gentleness Simulator, we recorded graspers' position, force in newton, soft body position, and task time. We also use data post-processing to compute measures for additional metrics [2]. These include average, mean, median velocity, jerk, turning angle in instrument motions, path length, soft body motion, etc. We subdivide the subjects' data into two expertise levels: expert and

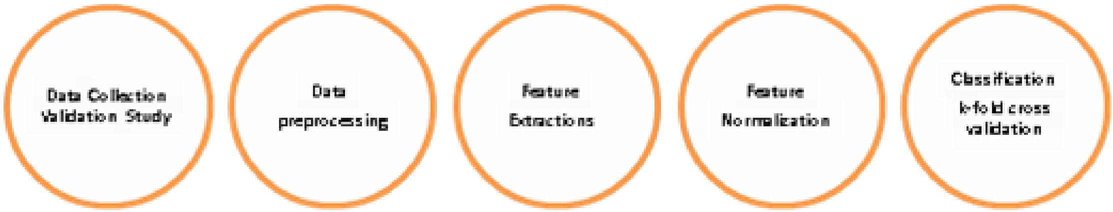

Figure 5: Data Analysis Flow Chart

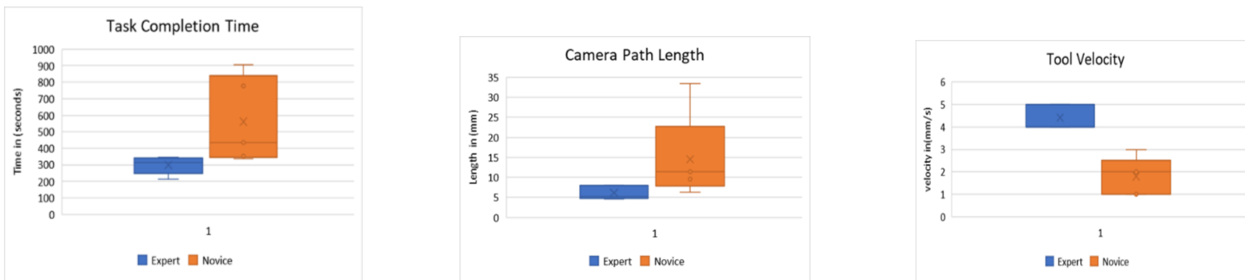

Figure 6: VATDEP Descriptive Statistical Analysis

novices. Since the number of the number subjects is limited, we classify them based on the number of years attending Post Graduate Year (PGY) 1-3 and (PGY) 4-5 for VATDEP. A total of 4 surgeons and 19 residents tested the gentleness simulator.

4 METHODOLOGY

In order to simplify the data analysis, reduce the computation time, and increase the applicability of the well-studied classification algorithms, we extracted the features such as velocity, acceleration, the path length from the raw data. Afterward, we applied the Mann-Whitney U test to determine the most significant difference between both groups and to find the distinction among the expertise level. The overall steps of the selection process can be seen in Figure 5

We also performed the descriptive statistical analysis on respected features to determine the difference between expert and novice. Figures 6 and 7 show the descriptive statistical analysis with features such as completion time, Camera Path Length, Tool Velocity, Double Grasper task time, Left grasper path length, and right grasper velocity.

We normalized the data set to compare the features between each other and eliminate the bias due to data saturation. This is to improve the accuracy of the analysis (e.g., removing the large skew in data). We used SciKit Learn and its preprocessing modules for normalization and machine learning [16]. SciKit Learn is an open source python package that includes various state of art classification, regression and clustering algorithms. The package includes fundamental data pre-processing functionalities such as binarizers, normalization, data transformation. The package has also post-processing operations for data reduction and visualization. We employed three normalization methods: standard scaler, min-max, and absolute value. In the end, we used K-Nearest Neighbors (n = 2), Logistic Regression, and SVM with both linear and radial basis function (RBF) kernels and considered the precision, recall, F1 score, and the average accuracy.

5 RESULTS AND ANALYSIS

Based on the post-questionnaire survey for the VATDEP, the expert group rated the 3D anatomical models' realism as 4.2 out of 5 on average. Realistic visual rendering was scored 4.4 out of 5, which is critical for the landmark identification task, and the anatomical correctness of the tear was noted as 4.2 out of 5. In the classification algorithms, we saw significant accuracy in the K-Nearest Neighbors algorithm. We achieved up to 81% correct classification on the PGY 4-5 and up to 89% classification accuracy on the PGY 1-3, which can be seen in Figure 8

In the Gentleness simulator, we have achieved the average lowest accuracy of 75% using the SVM linear kernel and the highest accuracy of 87% using the Logistic Regression algorithm. Our results demonstrated clear, distinguishable groups amongst subjects with the features: Curvature (p-0.010859), Force (p-0.03643), Soft body Path Length (p-0.0368409), Time to Complete (p-0.04136), and Turning Angle (p-0.02068). Figure 9, figure 10, and figure 11 show the classification results for the Gentleness simulator. We found a significant difference between both expert and novice population using Mann–Whitney U test for Curvature, Force, Soft body-path length, Time to complete the task, and turning angle features with the p-value less than 0.05. In content validity, the expert group rated the task's difficulty with the mean of 4.5 out of 5. Usefulness in learning hand-eye coordination and usefulness in learning ambidexterity expert stated mean of 4.5. Therefore, results prove that the gentleness simulator task can be useful for a learning purpose, according to experts [5].

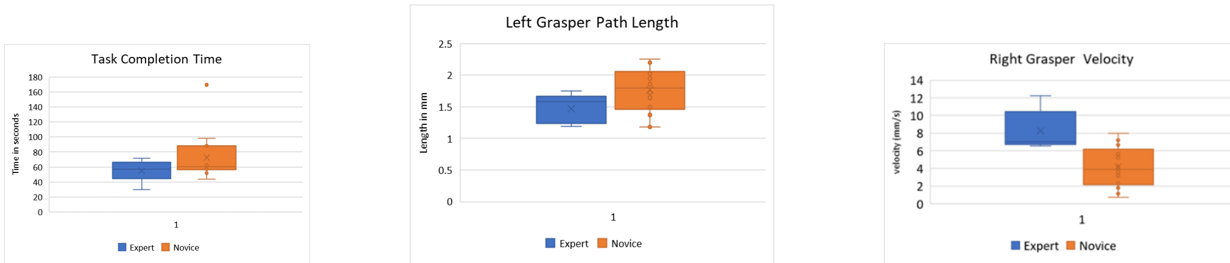

Figure 7: Double Grasper Task Descriptive Statistical Analysis

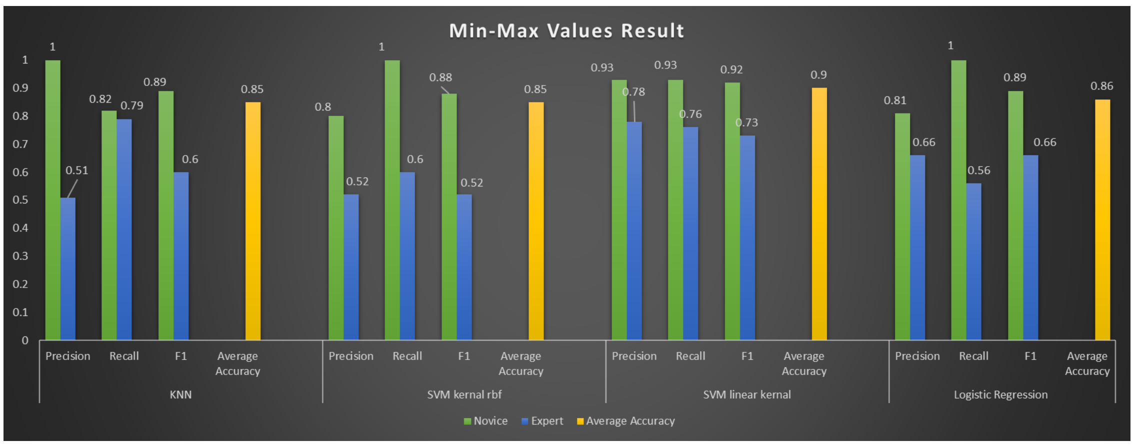

Figure 8: VATDEP Simulator Classification Result Using Min-Max

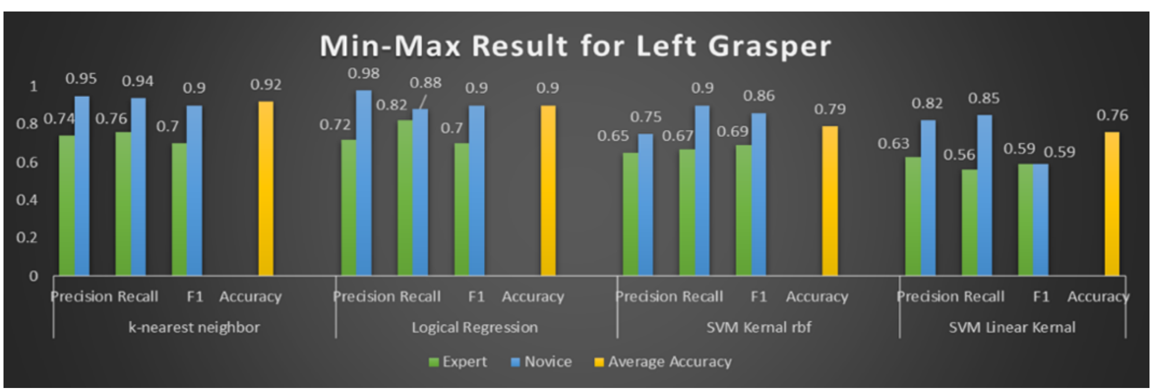

Figure 9: Left grasper simulator classification result using Min-Max

6 CONCLUSION

We recruited a total of 33 subjects, 23 of which were for the Gentleness Simulator, and 10 were for the VATDEP, ranging in experience from medical students to attending surgeons with several years of experience. We have shown that with data collected from our simulators, we can distinguish between expert and novice surgeons based on their skill levels, as well as a multitude of other factors such as gentleness and hand-eye coordination. Using a classification algorithm with Sci Kit Learn, we were able to show a distinction between two groups from both simulators. We also assessed the content validity of both simulators; experts rated the task difficulty with the mean of 4.5 out of 5. This shows our simulator can be useful for training purposes.

ACKNOWLEDGMENTS

This project was made possible by the Arkansas INBRE program, supported by a grant from the National Institute of General Medical Sciences, (NIGMS), P20 GM103429 from the National Institutes of Health (NIH). This project was also supported

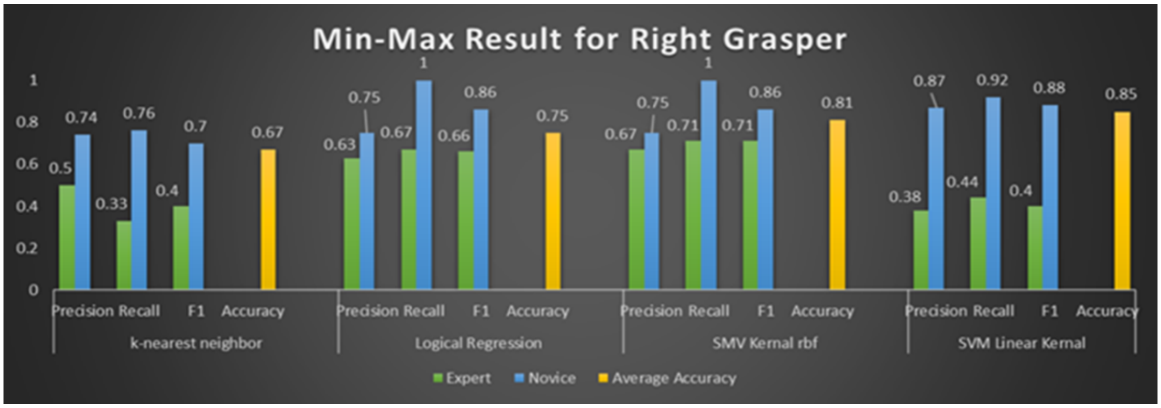

Figure 10: Right grasper simulator classification result using Min-Max

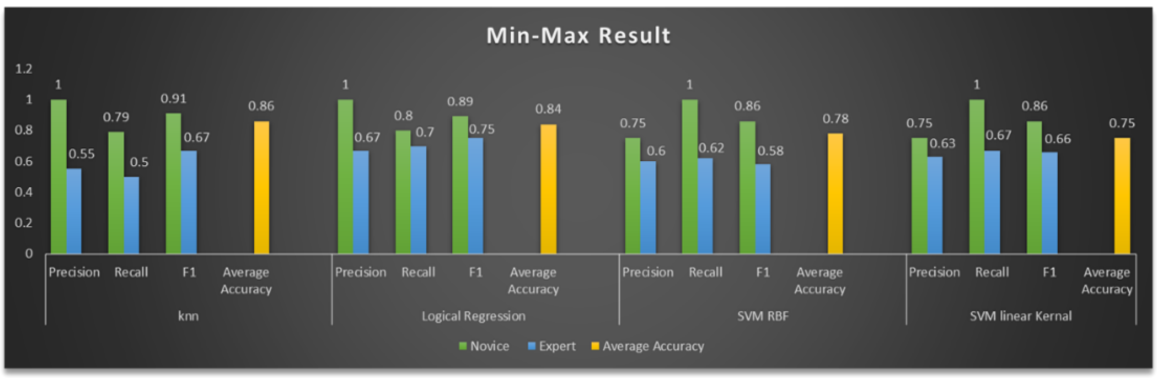

Figure 11: Tennis racket classification result using Min-Max

by NIH/NIAMS R44AR075481-01, NIH/NCI 5R01CA197491 and NIH/NHLBI NIH/NIBIB 1R01EB025241, R56EB026490.

REFERENCES

[1] Richard L. Angelo, Richard KN Ryu, Robert A. Pedowitz, William Beach, Joseph Burns, Julie Dodds, Larry Field, Mark Getelman, Rhett Hobgood, and Louis McIntyre. 2015. A proficiency-based progression training curriculum coupled with a model simulator results in the acquisition of a superior arthroscopic Bankart skill set. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 31, 10 (2015), 1854–1871.

[2] John D. Birkmeyer, Jonathan F. Finks, Amanda O'reilly, Mary Oerline, Arthur M. Carlin, Andre R. Nunn, Justin Dimick, Mousumi Banerjee, and Nancy JO Birkmeyer. 2013. Surgical skill and complication rates after bariatric surgery. *New England Journal of Medicine* 369, 15 (2013), 1434–1442.

[3] John L. Cameron. 1997. William Stewart Halsted. Our surgical heritage. *Annals of surgery* 225, 5 (1997), 445.

[4] Doga Demirel, Alexander Yu, Seth Cooper-Baer, Aditya Dendukuri, Tansel Halic, Sinan Kockara, Nizamettin Kockara, and Shahryar Ahmadi. 2017. A hierarchical task analysis of shoulder arthroscopy for a virtual arthroscopic tear diagnosis and evaluation platform (VATDEP). *The International Journal of Medical Robotics and Computer Assisted Surgery* 13, 3 (2017), e1799. DOI:https://doi.org/10.1002/rcs.1799

[5] Jake Farmer, Doga Demirel, Recep Erol, Daniel Ahmadi, Tansel Halic, Sinan Kockara, Venkata S. Arikatla, Kevin Sexton, and Shahryar Ahmadi. 2020. Systematic approach for content and construct validation: Case studies for arthroscopy and laparoscopy. *The International Journal of Medical Robotics and Computer Assisted Surgery* 16, 4 (2020), e2105.

[6] Tansel Halic, Sreekanth A. Venkata, Ganesh Sankaranarayanan, Zhonghua Lu, Woojin Ahn, and Suvranu De. 2011. A software framework for multimodal interactive simulations (SoFMIS). *Stud Health Technol Inform* 163, (2011), 213–217.

[7] Félix LAMATA13, Maria ANTOLIN, Samuel RODRÍGUEZ, and Alfonso OLTRA. 2006. Study of Laparoscopic Forces Perception for Defining Simulation Fidelity. *Medicine Meets Virtual Reality 14: Accelerating Change in Healthcare: Next Medical Toolkit* 119, (2006), 288.

[8] Thomas Mackel, Jacob Rosen, and Carla Pugh. 2006. Data mining of the E-pelvis simulator database: a quest for a generalized algorithm for objectively assessing medical skill. *Studies in Health Technology and Informatics* 119, (2006), 355–360.

[9] Behnaz Poursartip, Marie-Eve LeBel, Laura C. McCracken, Abelardo Escoto, Rajni V. Patel, Michael D. Naish, and Ana Luisa Trejos. 2017. Energy-based metrics for arthroscopic skills assessment. *Sensors* 17, 8 (2017), 1808.

[10] Alexis Sánchez, Omaira Rodríguez, Renata Sánchez, Gustavo Benítez, Romina Pena, Oriana Salamo, and Valentina Baez. 2014. Laparoscopic surgery skills evaluation: analysis based on accelerometers. *JSLS: Journal of the Society of Laparoendoscopic Surgeons* 18, 4 (2014).

[11] 2016. Touch. *3D Systems*. Retrieved February 14, 2021 from https://www.3dsystems.com/haptics-devices/touch

[12] Minimally Invasive Surgery. Retrieved from https://www.mayoclinic.org/tests-procedures/minimally-invasive-surgery/about/pac-20384771

[13] American Board of Surgery | www.absurgery.org. Retrieved April 10, 2019 from http://www.absurgery.org/default.jsp?index

[14] Fundamentals of Laparoscopic Surgery. *Fundamentals of Laparoscopic Surgery*. Retrieved April 10, 2019 from https://www.flsprogram.org/

[15] EMS22A30-C28-LS6 Bourns Inc. | Sensors, Transducers | DigiKey. Retrieved February 14, 2021 from https://www.digikey.com/en/products/detail/bourns-inc/EMS22A30-C28-LS6/3693519

[16] 6.3. Preprocessing data — scikit-learn 0.24.1 documentation. Retrieved February 14, 2021 from https://scikit-learn.org/stable/modules/preprocessing.html