USE OF CLASSIFICATION ALGORITHMS IN DETERMINING SURGICAL SKILL LEVELS THROUGH SURGEONS' HAND MOVEMENT BEHAVIORS

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

ATILIM UNIVERSITY

BY

DAMLA TOPALLI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

THE DEPARTMENT OF SOFTWARE ENGINEERING

MARCH 2018

Approval of the Graduate School of Natural and Applied Sciences, Atılım University.

Prof. Dr. Ali Kara

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Ali Yazıcı

Head of Department

This is to certify that we have read the thesis **"Use of Classification Algorithms in Determining Surgical Skill Levels through Surgeons' Hand Movement Behaviors"** submitted by **"Damla Topallı"** and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

Assoc. Prof. Dr. Nergiz Ercil Çağıltay

Supervisor

Examining Committee Members

Prof. Dr. Ali Yazıcı

Assoc. Prof. Dr. Nergiz Ercil Çağıltay

Assoc. Prof. Dr. Hadi Hakan Maraş

Asst. Prof. Dr. Erol Özçelik

Asst. Prof. Dr. Yavuz İnal

Date: March 30, 2018

I declare and guarantee that all data, knowledge and information in this document has been obtained, processed and presented in accordance with academic rules and ethical conduct. Based on these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Damla Topallı

Signature:

ABSTRACT

USE OF CLASSIFICATION ALGORITHMS IN DETERMINING SURGICAL SKILL LEVELS THROUGH SURGEONS' HAND MOVEMENT BEHAVIORS

Topallı, Damla

Ph.D., Software Engineering Department

Supervisor: Assoc. Prof. Dr. Nergiz Ercil Çağıltay

March 2018, 105 pages

Today, endoscopic surgeries have become an alternative for open procedures whenever possible. In this technique, the surgeon performs the operation by using a camera and light source, called 'endoscope', and special operational tools in order to operate through small entry points. For such types of operations, surgeons are required to gain several skills, whose development needs hands-on practice in them which is a challenge in surgical education programs. Several technology-enhanced training environments have been developed to improve current surgical education programs. However, in order to better integrate these technologies into the traditional methods, it is critical to understand the skill levels and prepare appropriate content according to the trainees' requirements. In other words, the trainees' skill levels need to be assessed regularly for better preparing the content and the sequence of the training program according to their individual requirements. The current skill level assessment techniques are mainly based on expert observations which are criticized as expensive and subjective. In this respect, the present study aims to evaluate the surgical skills objectively by using hand movement metrics through computer-based simulation software in Neurosurgery. This study is conducted with 28 surgical residents who were considered as intermediate or novice in their education. The evaluations are mainly concentrated on the hand movements of the trainees on computer simulated surgical training software. Accordingly, first an estimation of skill levels of intermediate and novice surgeons by using classification methods through performance metrics is performed. Secondly, velocity-based hand metrics are calculated using the hand movement data for classifying intermediate and novice surgeons. After that, by adapting BIT algorithm, which is an open source eye-event classification algorithm, to the hand movement data, new hand movement event metrics are proposed. Through these metrics, the participants' eye and hand movement events are analyzed. Finally, the results of the classification by using these newly introduced metrics are presented. As a conclusion, this thesis study attempts to better classify the intermediate and novice surgical residents' skill levels through their hand-movement events. The results are very promising showing that the proposed metrics potentially improve the accuracy of the classification. The researchers believe that, in the future by using the performance metrics together with hand- and eye- movement events metrics in a combined manner, the level of the accuracy may even be improved.

Keywords: virtual simulation environment, surgical education, skill-based training, eye-hand coordination, hand movement event metrics, feature selection, classification

ÖZ

SINIFLANDIRMA ALGORİTMALARININ CERRAHLARIN EL HAREKETİ DAVRANIŞLARI ÜZERİNDEN BECERİ SEVİYELERİNİN TAHMİNİNDE KULLANIMI

Topallı, Damla

Doktora, Yazılım Mühendisliği Bölümü

Tez Yöneticisi: Doç. Dr. Nergiz Ercil Çağıltay

Mart 2018, 105 sayfa

Günümüzde endoskopik ameliyatlar uygulanması mümkün olduğunda açık ameliyat yerine tercih edilen bir alternatif haline gelmiştir. Bu operasyonlarda cerrah, 'endoskop' adı verilen bir kamera ve ışık kaynağı ve özel operasyonel araçları kullanarak işlemi gerçekleştirir. Bu tür ameliyatları gerçekleştirebilmek için cerrahların sürekli pratik yaparak gerekli becerileri kazanmaları gerekmektedir. Dolayısıyla bu becerilerin geliştirilmesi günümüzdeki eğitim programları açısından önemli bir hedeftir. Mevcut cerrahi eğitim programlarını iyileştirmek üzere çeşitli teknolojiler ile zenginleştirilmiş eğitim programları geliştirilmektedir. Ancak, bu teknolojileri geleneksel yöntemlere daha iyi entegre edebilmek için, cerrahların beceri düzeylerini anlamak ve gereksinimlerine göre uygun içerik hazırlamak önemlidir. Diğer bir deyişle, eğitim programının içeriğinin ve sırasının eğitim alan kişilerin bireysel ihtiyaçlarına uygun bir şekilde hazırlanması için, beceri düzeylerinin düzenli olarak değerlendirilmesi gerekmektedir. Mevcut beceri seviyesi değerlendirme teknikleri, çoğunlukla pahalı ve öznel olması nedeniyle eleştirilen uzman gözlemlerine dayanmaktadır. Bu bağlamda, bu çalışma, nöroşirürjide bilgisayar tabanlı simülasyon yazılımı ile el hareket ölçütlerini kullanarak cerrahi becerilerin objektif olarak değerlendirmesini amaçlamaktadır. Bu çalışma cerrahi eğitim alan 28 öğrenci ile gerçekleştirilmiştir. Değerlendirmeler temel olarak katılımcıların bilgisayara dayalı benzetim yazılımı üzerindeki el hareketleri esas alınarak gerçekleştirilmiştir. Buna göre, öncelikle benzetim tabanlı bir cerrahi eğitim yazılımı ortamından alınan performans ölçütleri kullanılarak öznitelikler çıkarılmış, çeşitli sınıflandırma algoritmaları ile orta ve acemi düzey cerrahların beceri düzeyleri tahmin edilmiştir. İkinci olarak benzetim ortamında el hareketlerine dayalı hız tabanlı ölçütler hesaplanmış, bu ölçütler orta ve acemi düzey cerrahları sınıflandırmak için kullanılmıştır. Daha sonra, açık kaynaklı bir göz hareketi sınıflandırması algoritması olan BIT algoritmasını, el hareketleri verisine uyarlayarak, yeni el hareketi ölçütleri önerilmiştir. Bu önerilen ölçütler ile, katılımcıların göz ve el hareketi verileri analiz edilerek, orta ve acemi düzey cerrahların el-göz davranışlarındaki farklılıklar anlaşılmıştır. Son olarak, bu çalışmada önerilen el hareket ölçütleri kullanılarak öznitelikler çıkarılmış ve sınıflandırma algoritmaları kullanılarak orta ve acemi düzey cerrahların beceri düzeyleri tahmin edilmiştir. Sonuç olarak, bu çalışmada, el hareketi verilerinden elde edilen ölçütler kullanılarak, acemi ve orta düzeydeki cerrahların beceri seviyelerinin daha iyi anlaşılması hedeflenmiştir. Sonuçlar, önerilen özniteliklerin tahminlerin doğruluğunu potansiyel olarak arttırdığını göstermektedir. Araştırmacılar, gelecekte, el ve göz özniteliklerinin bir arada kullanılması ile performans değerlendirmelerinin doğruluk seviyesinin daha da iyileştirilebileceğine inanmaktadırlar.

Anahtar Kelimeler— sanal benzetim ortamları; cerrahi eğitim; beceriye dayalı eğitim; el-göz koordinasyonu; el hareketi ölçütleri; öznitelik seçimi; sınıflandırma

To My Family

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere appreciation to my supervisor Assoc. Prof. Dr. Nergiz Ercil Çağıltay for her continuous support, guidance, motivation and insight throughout my thesis study. She guides me to explore the importance of research; I learned a lot from her. I offer sincere thanks for her extensive knowledge, enthusiasm and patience during this period. I'm really proud of being one of her students.

I would like to express my appreciation to examination committee members; head of Software Engineering Department, Prof. Dr. Ali Yazıcı for his valuable comments and guidance during the PhD program; Assoc. Prof. Hadi Hakan Maraş for his valuable contributions throughout the research; and my thesis progress jury members Assoc. Prof. Dr. Erol Özçelik and Asst. Prof. Dr. Yavuz İnal for their valuable time, comments and suggestions, all helping to improve my thesis work. I am also grateful to Asst. Prof. Dr. Gül Tokdemir for precious suggestions and Dr. Güler Kalem for her support, sharing experiences and motivation.

Additionally, I would like to thank my friends; Seda Çamalan for her valuable support and Gülden Alaman, Hazan Dağlayan Sevim for their precious friendship and support, Batuhan Coşar and Furkan Kurtaran, for their support and motivation during this time, Bugra Sahinoglu, Ozan Can Acar, Tolga Üstünkök, and colleagues from Computer Science departments, Atilim University who voluntarily participate in this research, for their valuable time and contributions.

Last but not the least; I would like to express my appreciation to my dearest parents Selma Topallı and Mahmut Topallı for their endless love, support and patience. I owe my curiosity of this research field to them. They have always believed in me and encouraged me all through my life.

This study is conducted for endoneurosurgery education purposes as part of a project named Endo-neurosurgery Education (ECE: Tubitak 1001, Project No: 112K287) supported by TÜBİTAK 1001 for three years (2011 – 2014). I would like to thank the support of TÜBİTAK 1001 program for realizing this research. This study is conducted with the equipment in Atilim University Simulation Laboratory, established by Assoc. Prof. Dr. Nergiz Ercil Çağıltay. Finally, thanks to the members of ECE Project team and Emre Tuner for his support in SimLab, and the surgical residents of Hacettepe University Medical School, who willingly participated in this study, for their valuable time and contributions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ECE	-	Educational Computer-based-simulation Environment
MIS	-	Minimally Invasive Surgery
BIT	-	Binocular-Individual Threshold
FN	-	Fixation Number
FD	-	Fixation Duration
SN	-	Saccade Number
SSD	-	Stand-Still Duration
SSN	-	Stand-Still Number
MAE	-	Mean Absolute Error
RMSE	-	Root Mean Squared Error
RAE	-	Relative Absolute Error
RRSE	-	Root Relative Squared Error
TPR	-	True Positive Rate
FPR	-	False Positive Rate
ROC	-	Reciever Operating Characteristic
SVM	-	Support Vector Machine
PUK	-	Pearson VII Universal Kernel
KNN	-	K-Nearest Neighbors
LWL	-	Locally Weighted Learning
PCA	-	Principal Component Analysis
FOV	-	Field of View

CHAPTER 1

INTRODUCTION

Classification is an important concept that people use very often in their daily lives. Classification helps people to better recognize and group things. For instance the types of animals, plants and everything can be better organized and studied through classification. In order to classify objects and things their common behaviors and features are detected and based on these rules the classification is implemented. In the world of computer-based analysis, classification techniques have been used to automatically differentiate the events, things and objects. Several classification algorithms have been developed to improve the accuracy level of the classification that can be used in several fields. Recently, the classification algorithms are being used to analyze data and create support in different types of information systems, such as music mood detection (Bhat, Amith, Prasad, & Mohan, 2014), face recognition (Larrain, Bernhard, Mery, & Bowyer, 2017), predicting diabetes (Kaur & Chhabra, 2014) and heart disease (Kumar & Sahoo, 2015).

With the technological developments in the field of surgery, minimally invasive surgical (MIS) techniques have become the standard of surgical care for many patients. With this technology, surgical fields now incorporate video imaging with the help of camera and miniature surgical devices rather than fingers (Minna Silvennoinen, Helfenstein, Ruoranen, & Saariluoma, 2012). Yet, despite its several benefits for better treatment of patients, it is known that MIS operation techniques are difficult to learn and more than 30 procedures for the learning curves have been reported (Dankelman, Grimbergen, & Stassen, 2007; Moore & Bennett, 1995). The number of procedures required can be even increased to 100 procedures depending on the type of the operation (Lehmann et al., 2005). Schreuder (2011) mentioned in his study that a problem is the pressure on reducing the operation time in order to be not only cost-effective, but also ethically considered (Schreuder, Oei, Maas, Borleffs, & Schijven, 2011). Studies show that MIS is more demanding and requires more preparation than open surgery (Gomoll, O'toole, Czarnecki, & Warner, 2007; Panait, Bell, Roberts, & Duffy, 2008; Tuijthof et al., 2010). For this reason, strengthening and training of the required skills is essential for surgical candidates in the early training phase of MIS. Traditional skills do not guarantee success in MIS (Tuijthof et al., 2010; Waxberg, Goodell, Avgerinos, Schwaitzberg, & Cao, 2004). Hence, each individual needs to master the surgical skill before operating on actual patients (Schreuder et al., 2011). A large part of surgical skills are gained in the operating room while operating on patients (Dankelman & Di Lorenzo, 2005). Hence, developing these skills before entering an operating room provides a more focused and efficient performance, reduces the time spent in the operating room, and enhances patient safety (Waxberg et al., 2004).

Unlike open surgery, MIS is, by nature, a technique that is very suitable for simulation based training (Schreuder et al., 2011). The specific psychomotor skills and eye–hand coordination needed for this type of surgery can be developed easily through simulation (Derossis et al., 1998; Grantcharov, Bardram, Funch-Jensen, & Rosenberg, 2003). For skills training, box trainers or computer-enhanced trainers may be used, but in the past decade, new virtual reality (VR) trainers have been introduced for minimally invasive techniques and purposes. Nowadays, simulation training, often enhanced by VR techniques, is used for a wide range of training purposes: laparoscopy (Gurusamy, Aggarwal, Palanivelu, & Davidson, 2008), robot-assisted surgery (Kenney, Wszolek, Gould, Libertino, & Moinzadeh, 2009), endoscopy (Bittner, Mellinger, Imam, Schade, & MacFadyen, 2010), cystoscopy (Schout et al., 2010), hysteroscopy (Bajka et al., 2010), and intervention radiology (Ahmed et al., 2010).

Earlier studies show that it is possible to transfer the skills learned on a simulator to real operations, leading to fewer errors and shorter operating time (Larsen et al., 2009; Thijssen & Schijven, 2010). Recently, e-learning programs and ‘‘serious games’’ for MIS embedded in training curricula, step-by-step approaches encouraging the making and solving of mistakes, and a diversity of storylines have been introduced (Verdaasdonk et al., 2009). However, simulator training cannot stand on its own and needs to be part of a training curriculum. The way the surgical simulator is used in a particular teaching curriculum determines its validity for the cause. Studies show that proficiency-based skill training lead to fewer errors in the operating room and reduce the operating time. In that concern, several simulation systems have been developed to further support the surgical education (Robb, Aharon, & Cameron, 1997; Rudman et al., 1998). However, studies in the literature show that the research conducted for the surgery education is very limited when it comes to technical skills assessment (Derossis, DaRosa, Dutta, & Dunnington, 2000). According to Andersen, simulation applications that allow the assessment and learning of expert intra-operative judgment should include the following outline:

- Cognitive task analysis (CTA) of the operative steps and potential points of risk for each surgical procedure.
- The ability to detect the situational awareness of the performer and the options considered to avoid error at critical steps.
- An assessment (scoring) of options considered or attempted.
- Immediate evaluation feedback to inform improved performance.
- A program of deliberate practice in which progressively more challenging scenarios can be introduced, based on the trainee’s demonstrated skills (Andersen, 2012).

In that concern, the classification methods may improve the surgical education programs by providing more objective and cheaper assessment based on the data collected through computer-simulation software.

This study attempts to better understand the skill differences of intermediate and novice surgical residents through their hand movements. For this, through haptic user interfaces the hand movement’s data of the surgical residents are collected while they were performing tasks that are defined in computer-simulated endoscopic surgery scenarios. For this purpose, four scenarios have been used. This study is conducted by 28 surgical residents who are considered as intermediate or novice in their education. The data collected from such computer-based simulation environments can be grouped as performance data such as the task accuracy and the task duration to successfully perform each task. In addition to these metrics some behavioral data also can be collected from these environments. For instance, the eye-behaviors of participants while they are performing surgical tasks can be collected and analyzed. Additionally, the data about hand behaviors can also be collected. Based on the collected data, first an estimation of skill levels of intermediate and novice surgeons by using classification methods through performance metrics taken from the computer simulation-based surgical training environment is performed. Secondly, classification is performed through velocity-based hand metrics. These metrics are calculated by getting data from the simulation software that is developed by using the Unity game engine (Unity, 2017). In the literature, there is no study that gets the velocity-based metrics from such game engines. By using this approach, a more standardized and easier method is proposed for calculating velocity-based metrics. Besides these attempts, by adapting Binocular-Individual Threshold (BIT) algorithm, which is an open source eye-movement event algorithm to the hand movement data, new features are proposed. Using these features, the participants' eye and hand movement events are analyzed, and another classification attempt has been conducted.

This thesis is organized as follows: Chapter 2 describes the research methodology of the study. Chapter 3 presents the estimation of skill levels of intermediate and novice surgeons by using classification methods through performance metrics. Chapter 4 covers the results of classification through velocity-based hand metrics. Chapter 5 presents the differences of intermediate and novice surgeons based on eye and hand movement events, in which new features are proposed by adapting BIT algorithm. Chapter 6 covers the results of classification using newly proposed hand metrics. Chapter 7 details the discussion and conclusion of the study, and finally limitations and future work is provided in Chapter 8.

CHAPTER 2

METHODOLOGY

Today, the availability of objective metrics of surgical performance is considered to be critical for training surgeons and evaluating their performance. In this respect, virtual reality simulators provide objective assessment of surgical skill levels. These simulators are able to measure the performance parameters of the surgeons' in surgical simulation environment providing insights about their skill levels. Accordingly, the aim of this study is to understand the skill differences of intermediate and novice surgical residents through their hand movements. The study is conducted with 28 surgical residents. Their performances while performing four scenarios of simulated surgical tasks were monitored and metrics based on their performance and behaviors such as eye and hand events for each hand were recorded. Based on the collected data, first skill levels of intermediate and novice surgeons are estimated by using classification methods through performance metrics. Later, velocity-based hand metrics were calculated using the hand movement data - the coordinate points gathered from simulation- based system- in Unity (Unity, 2017). These features are then used for classifying intermediate and novice surgical residents. Additionally, participants' eye and hand movements were analyzed and new hand features were proposed according to the level of smoothness of movement. In this study, an open source eye movement classification algorithm, namely BIT, is adapted to classify the hand movement of surgical residents. Finally, these newly introduced hand metrics are used to estimate the skill levels of surgical residents. The research procedure of this thesis is given in Figure 2.1 below.

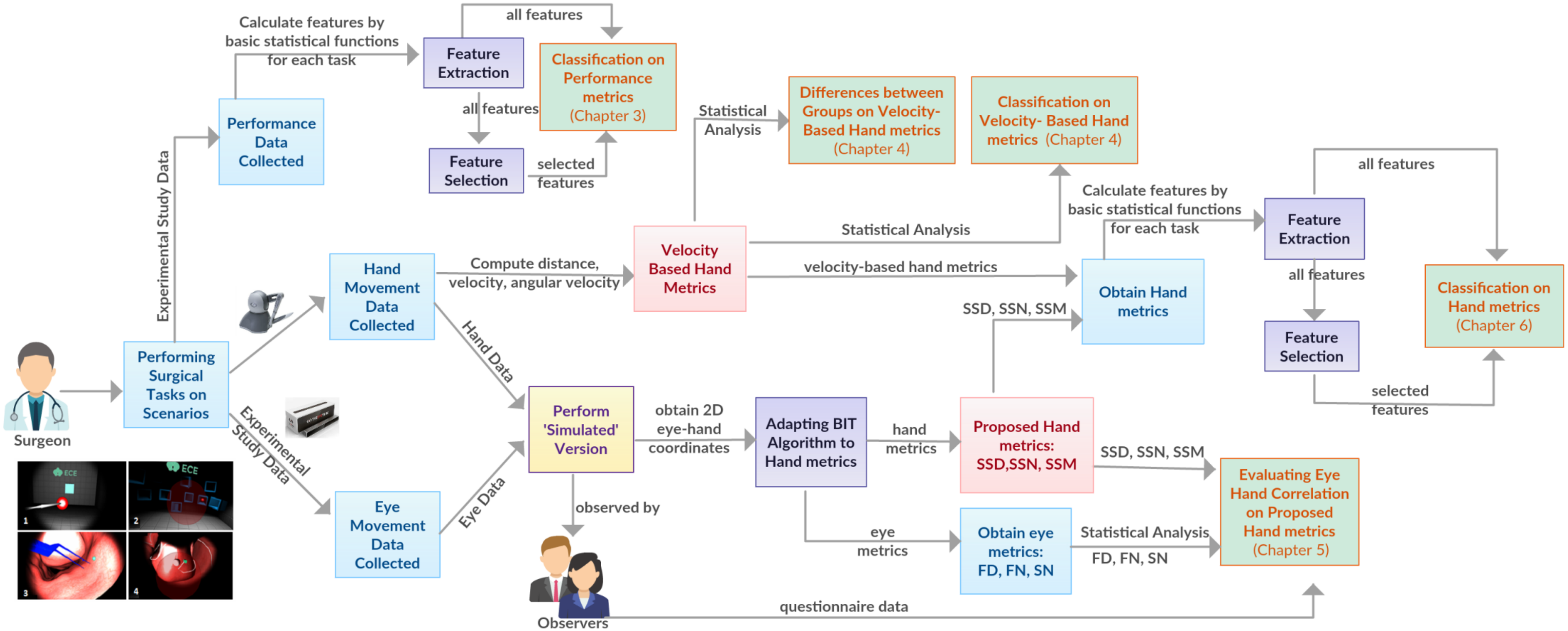


Figure 2. 1 Research Procedure of the Thesis

2.1 Participants

A total of 28 surgical residents (21 doctors and 7 interns) from the Department of Neurosurgery (12 participants) or Otolaryngology (ENT) (9 participants) from the Hacettepe Medical School in Ankara, Turkey, participated in this study. There were two skill level groups of participants. Among those, 16 participants (3 female) were novices whose average age was 26.94 (SD = 6.57) working as research assistants in the neurosurgery and ENT departments. On the average, they had observed 12.69 (SD = 16.21) and assisted in 5.75 (SD = 13.48) surgeries. On the other hand, 12 participants (1 female) were intermediates whose average age was 30.15 (SD = 2.15). On the average, they had observed 56.53 (SD = 31.97) and assisted in 34.85 (SD = 30.17) surgeries. On the average, the intermediate group had performed 16.00 (SD = 17.69) operations as surgeons. Detailed information about the participants is given in Table 2.1.

Table 2. 1 Information about Participants

		Gender		**Endoscopic Surgical Expertise**		
Skill Level	**Age**	**F**	**M**	**Observed**	**Assisted**	**Performed**
Intermediate	30.15	1	11	56.53	34.85	16.00
Novice	26.94	3	13	12.69	5.75	0.00

In this thesis, the categorization of skill levels is defined based on Silvennoinen et al.'s study (M Silvennoinen, Mecklin, Saariluoma, & Antikainen, 2009). Accordingly, participants who have operated at least one endoscopic surgery are considered as 'intermediate', whereas others who have observed and assisted in endoscopic operations, but have not performed any surgeries by themselves are considered as 'novice'.

2.2 Scenarios

Four scenarios (prepared in single-handed and both-handed versions) were used for the surgical training process used in the experimental study. Two of them (Scenarios 1 and 2) were prepared for the purpose of practicing general skills, such as learning the usage of the surgical tools with an endoscope and gaining depth-perception in a simulated 3D environment. The other two scenarios were closer to the operational procedures, using the simulated anatomical model. Specifically, in the both-hands version of these scenarios, it is aimed to improve right- and left-hand coordination and eye-hand coordination skills.

Scenario-1: Moving the Ball into the Box

'Moving the Ball into the Box' scenario for single-hand and both-hands (tool and the light source with camera) are shown in Figure 2.2-A and B, respectively. In this scenario, each participant is asked to approach the red ball with the haptic device, catch it, and then move it into the green box. The position of the ball and the box changes randomly in each task. The participant must complete this process successfully, which includes 10 tasks within the allocated period. If the process is not completed within 10 seconds, the ball and the box disappear.

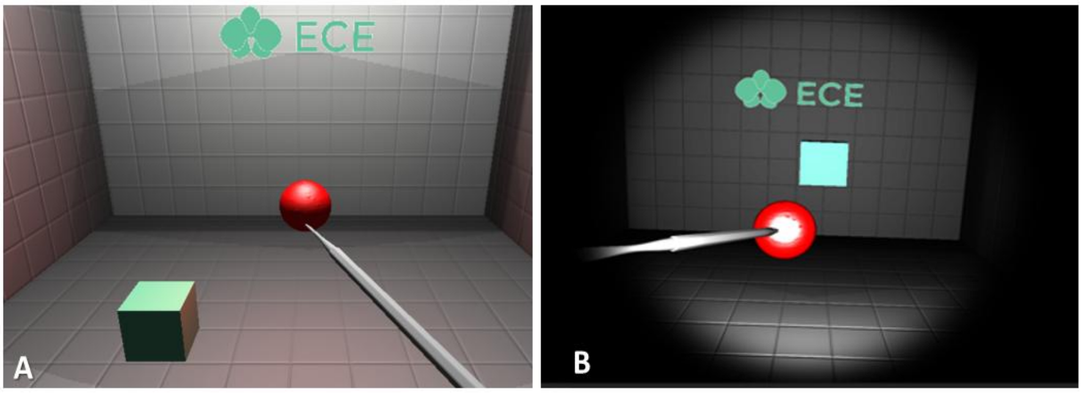


Figure 2. 2 Moving the Ball into the Box (Scenario-1)

Scenario 2: Catching the Balls in boxes with an Endoscope

The layout of this scenario is prepared for single-hand and both-hand versions as shown in Figure 2.3-A and B, respectively. In this scenario, participants should catch the red balls in the cubes with the camera as the tool using their dominant hand in the single-hand version. Similarly, in the both-hand version, the camera is used as the tool controlled by their dominant-hand, whereas the light is used as the endoscope and controlled by the non-dominant hand. In order to catch the red balls, it is important to approach the ball from the right angle by using the camera within the given time period of time, which is 10 seconds. In this scenario, ten balls appear one after another in different cubes randomly on the scene.

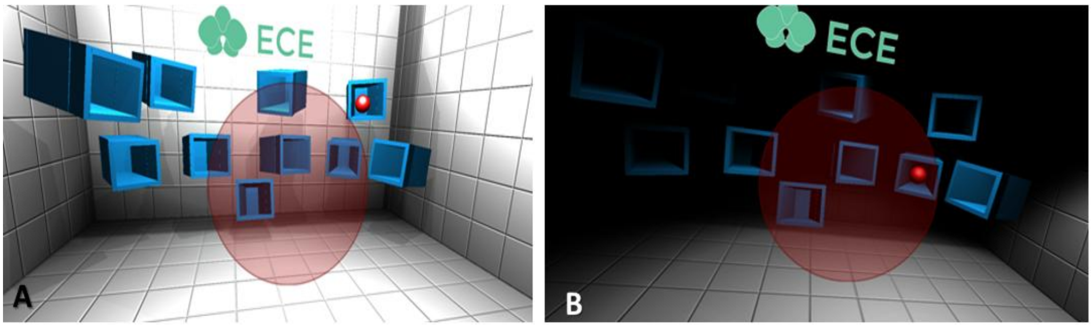


Figure 2. 3 Catching the Objects in boxes with Endoscope (Scenario-2)

Scenario-3: Clearing the Nose

‘Clearing the Nose’ scenario for single-hand and both-hand (tool and the light source and camera) versions are shown in Figure 2.4-A and B, respectively. In this scenario, the participant must remove the green ball-like objects, which are spread through the nose model. In the single-hand version, the camera acts as the tool and the participant removes the objects. The camera is used as the light source and the cautery model as the tool to collect the objects in the both-hand version. In case of a collision i.e. if the haptic device touches the tissue, it provides a force feedback that feels like the device is pushed back in the hands of the user.

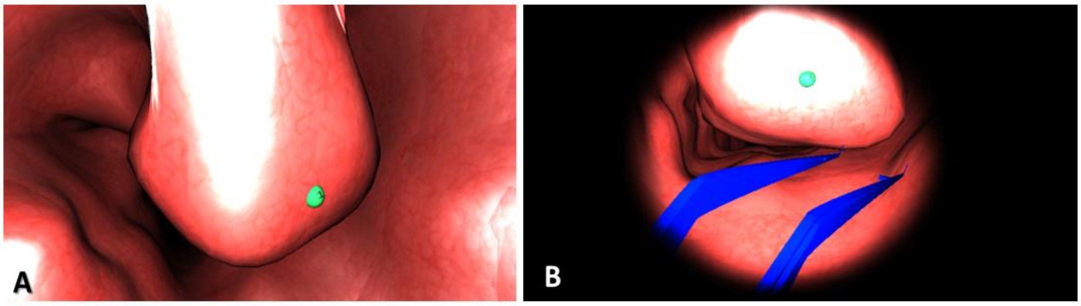


Figure 2. 4 Clearing the Nose Scenario (Scenario-3)

Scenario-4: Following the Ball with an Endoscope

In this scenario, the participant should move the white ball to follow a path starting from the yellow node until reaching the green node on the simulated anatomical nose model (See Figure 2.5-A: single-handed and B: both-handed) representing a higher fidelity. In order to move the white ball, the participant has to stay within the right angle and distance, which is approved by the focus area turning green as in Figure 2.5-A. Otherwise, the focus area remains red, indicating that the participant cannot move the ball in the scene until approaching it with a right angle.

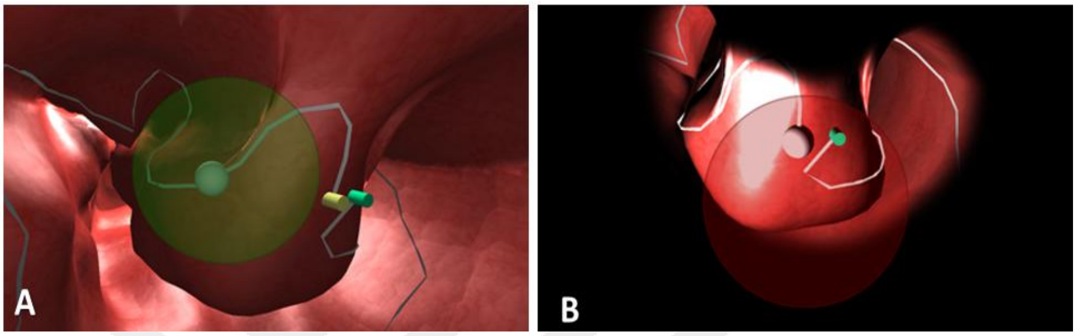


Figure 2. 5 Following the Ball with an Endoscope (Scenario-4)

Specifically in both-handed version, the coordinated movement of both hands controlling the camera and the light source is required to complete the task successfully.

2.3 Metrics in Scenarios

All scenarios are designed and developed as ten repetitive tasks, except Scenario-4, which has fifteen tasks. The metrics recorded automatically by the computer for each scenario appeared in Table 2.2.

Table 2. 2 Metrics Recorded for each Scenario

	Scenario							
	I		**II**		**III**		**IV**	
Metrics	DH/NH	BH	DH/NH	BH	DH/NH	BH	DH/NH	BH
Time	✓	✓	✓	✓	✓	✓	✓	✓
Distance	✓	✓	✓	✓	✓	✓	✓	✓
Camera distance		✓		✓		✓		✓
Accuracy	✓	✓	✓	✓	✓	✓	✓	✓
Catch time					✓			
Time in Error						✓	✓	✓
Distance in Error							✓	✓
Deviation count							✓	✓

For each specific scenario, a total of eight metrics were collected while performing the tasks in the simulation environment. The three common metrics among all include; ‘time’ which is the total time period in seconds for completing each task, ‘distance’ which is total distance covered by the haptic device as the tool and ‘accuracy’ which determines if the task is performed thoroughly in the given time period. As stated before, each task has to be performed successfully within the given time. In the both-hand scenarios, along with the tool distance, the camera distance covered by the haptic device as camera, is also recorded. ‘Catch time’ is measured only for Scenario-2, which is the partial task time to catch the green ball. ‘Time in Error’ is the duration in which the focus area remains red representing the idle time, ‘Distance in Error’ is the distance travelled while the focus area is still red, and ‘deviation count’ is the number of collisions with the tissue, and only measured in Scenario-4.

2.4 Apparatus

In order to control the force applied on the tissue, in theory, a surgeon would like to feel the force, position and other tactile information generated by the instrument (Westebring–Van Der Putten, Goossens, Jakimowicz, & Dankelman, 2008). In MIS, and in comparison to open surgery, the sense of touch is limited, hence, surgeons must rely more on the feeling of net forces resulting from tool-tissue interactions (Basdogan et al., 2004). For this reason, haptic devices can be integrated into training simulations for MIS procedures (Basdogan et al., 2004) in order to provide a similar real-world practice and realistic sense of touch. Accordingly, in this study, a mid-range professional ‘Geomagic Touch’ haptic device is used as seen in Figure 2.6.

Figure 2. 6 Experimental Setup through Haptic Devices

For the present work, using this device, the authors obtained one hundred data points per second representing 3D coordinates of hand motion for each task in each scenario. Additionally, the eye-gaze data of the participants is collected with a 60 Hz eye-tracking device, the Eye Tribe (The Eye Tribe, 2014). This tool is used to track the user’s eye movements and calculate the on-screen gaze coordinates.

2.5 Experimental Procedure

At the beginning of the experiment, the participants were asked to fill out a questionnaire including their demographic information, dominant-hand and experience level (i.e. years in the department, number of operations observed, assisted, and performed). After that, a brief instructional video was shown and oral explanations were given to the participants about the experimental procedure. Each participant was asked to perform an introductory level scenario, named “Using a Haptic Device”, shown in Figure 2.7, with the aim to train the participants for the use of the haptic device since most participants had never been involved in such experimental studies before.

Figure 2. 7 Using Haptic Device Scenario

Four scenarios were developed for surgical training purposes as a part of an Educational Computer-based-simulation Environment (ECE) project. All of these scenarios were implemented in Unity 3D with C#. Unity provides a high fidelity, low cost and easy to use simulation environment (Craighead, Burke, & Murphy, 2008). Each scenario was prepared in two versions: single-handed and both-handed. Accordingly, each scenario was performed under three different hand conditions (i.e. dominant hand, non-dominant hand, and both-hands). The experimental study was started with single-handed tasks, after completing all the tasks of that scenario, continued with both-handed tasks. In order to eliminate the order effect, half of the participants started the experiment with their dominant hands, and the other half with non-dominant hands. In both-hand tasks, the surgical tool is controlled by the dominant hand, whereas the camera (as endoscope) is controlled by the non-dominant hand. During the experimental study, the performance data for each task and the eye and hand movements of the surgical residents were collected using haptic devices.

2.6 Data Analysis Methods

In this study, all of the scenarios were implemented in Unity 3D using C#. Through this computer-based simulation environment, performance and hand movement data of the participants' is collected using haptic devices. Additionally, new velocity-based features (i.e. distance, velocity and angular velocity) were calculated using the 3D coordinates of position and rotation vectors of both-hands in Unity 3D environment. The features for classification were extracted using MATLAB (R2015a). The feature selection, cross validation and classification processes were performed within WEKA machine learning suite (Witten, Frank, Hall, & Pal, 2016).

Moreover, to understand the differences between intermediate and novices a statistical data analysis was performed on SPSS (version 21; IBM Corporation). In this study, two different statistical analyses were performed using SPSS. Firstly, the difference between the intermediate and novice groups based on their hand movement metrics was analyzed using Mann-Whitney U test. This is a non-parametric test alternative to the independent-samples t-test used when the sample size is lower than 30 and the normality assumptions were violated. Secondly, binary logistic regression analysis was performed to estimate the skill levels of surgeons based on their hand movement metrics.

CHAPTER 3

ESTIMATING SKILL LEVELS USING CLASSIFICATION ALGORITHMS THROUGH PERFORMANCE METRICS

In this section, current methods for surgical skill assessment were discussed through experience level definition and classifications found in the literature in order to provide a standard view of surgical skill levels. The results of classification on four scenarios based on performance metrics such as time, distance, catch time and accuracy were presented.

3.1 Defining Surgical Skill Levels

Previous studies show that there are different classifications of experience levels such as dividing the groups into two levels, such as beginner and experienced; three levels experienced, intermediate and novice; or more than three levels. Details of such two-level, three-level, or other classifications appear in what follows.

2 levels (Novice and Expert)

Aggarwal et al.'s study (2006) aims to establish and validate a VR (virtual reality) simulator curriculum to provide an evidence-based approach for laparoscopic training program (Rajesh Aggarwal, Grantcharov, Moorthy, Hance, & Darzi, 2006). In that study, the minimally invasive VR simulator (MIST-VR) has 12 abstract laparoscopic tasks, each at 3 graduated levels of difficulty (easy, medium, and hard). These tasks are assigned to two different groups, twenty medical students (novices) and ten experienced laparoscopic surgeons (experts) in order to highlight the performance differences between two groups. The expert criterion was defined as having performed more than 100 laparoscopic cholecystectomies. The performance difference between the two groups was measured by the time taken to perform the tasks, the path length, and the number of errors for each hand (Rajesh Aggarwal et al., 2006).

3 levels (Novice, Intermediate and Expert)

Based on the literature findings, generally the participants in such studies are divided into three experience levels; novice, intermediate and expert. For instance, Shetty et al. (2012) claims that camera handling and navigation are essential skills in laparoscopic surgery, and that there are no standardized VR-based camera navigation curricula available (Shetty et al., 2012). For this reason, in order to improve technical performance in the operating room, virtual reality (VR) simulation may be useful to develop camera skills for novices. Accordingly, an experimental study is conducted to understand the effect of VR simulation strategies for laparoscopic training. The participants are divided into three groups novice, intermediate, or advanced, based on the number of the surgeries they have assisted or performed: novice (performed or assisted in less than 10 laparoscopic surgeries), intermediate (performed or assisted in 10 to 100 laparoscopic surgeries) and advanced (performed more than 100 laparoscopic surgeries) (Shetty et al., 2012). Similar to this work, Srivastava et al. (2004) also conducted an experimental study in the field of shoulder arthroscopy operations, where the participants were divided into three groups based on their hands-on experience, such as novice (no hands-on experience), intermediate (performed or assisted in 1 to 50 shoulder arthroscopies) and expert (performed or assisted in more than 50 shoulder arthroscopies) groups (Srivastava et al., 2004).

Another study conducted by Schreuder et al. (2009) consists of both basic skills and gynecologic procedural simulations to compare the performance among three groups; novice (no hands-on laparoscopic experience), intermediate (performed 10 to 75 laparoscopic procedures), and expert (performed more than 100 laparoscopic procedures), similarly based on the number of laparoscopic procedures performed (Schreuder, van Dongen, Roeleveld, Schijven, & Broeders, 2009). A similar surgical level classification was performed to analyze the learning rate for laparoscopic skills on a virtual reality training system (Grantcharov et al., 2003). In Grantcharov et al.’s

(2003) study, the assessment of laparoscopic skills is carried out in terms of time, errors, and the economy of hand movements in the simulator with the aim to define the ability of the simulator to differentiate between surgeons with different laparoscopic experience, again based on the number of laparoscopic surgeries performed: beginner (fewer than 10 laparoscopic procedures), intermediate (performed 15 to 80 laparoscopic procedures) and master (performed more than 100 laparoscopic procedures) (Grantcharov et al., 2003). In Korndorffer et al.'s (2005) study, it is aimed to differentiate between participants at different levels of experience by using the completion time as the performance metric (Korndorffer et al., 2005). The data were collected from participants including novice (medical students), intermediate (post graduate 2-4 years), and advanced (expert surgeons) groups (Korndorffer et al., 2005). Similarly, Scott et al.'s study (2001) documents the laparoscopic performance of three groups of subjects with different levels of experience over the curricula, based on task completion time (Scott et al., 2001). Additionally, the relationship between task completion time and the number of practice repetitions was also examined (Scott et al., 2001).

More than 3 levels

Balik et al.'s study (2010) grouped surgeons into four levels where the operative experience was represented by the individuals' number of previous surgical procedures carried out (Balik et al., 2010). Four skill levels were determined as follows: Level 1: the first 60 procedures, Level 2: 61 to 120 procedures, Level 3: 121 to 180 procedures, and Level 4: more than 180 procedures (Balik et al., 2010). Another study conducted by Windsor, Diener & Zoha (2008) has defined five levels: undergraduate tertiary students, medical students, novice surgical trainees, advanced surgical trainees and experienced laparoscopic surgeons (Windsor, Diener, & Zoha, 2008).

Similarly, Silvennoinen et al. (2009) also categorizes expertise and skill levels in minimal invasive surgery into five groups, namely 'beginner' (merely has non-specific knowledge of a domain), 'novice' (just started to develop the elementary knowledge assumed in the domain), 'intermediate' (already extended his/her knowledge above the beginner level), 'sub expert' (a medical specialist being capable of solving clinical problems outside their domain of expertise), and ‘expert’ (has specialized knowledge of the subdomain) (M Silvennoinen et al., 2009).

3.2 Assessing Surgical Skill Levels

Recent developments in MIS have resulted in an increased interest in objective assessment methods for surgical skills. The observation process carried out by training specialists is a subjective method by nature. For this reason, certain objective measures have been proposed to improve the quality of assessment, such as objective structured assessment of technical skills (OSATS), a method for testing specific operative skills in surgical trainees (Cagiltay, Ozcelik, Sengul, & Berker, 2017; Martin et al., 1997). Van Hove et al. (2010) conducted a systematic search for studies addressing the validity and reliability of the methods for objective skills assessment within surgery and gynaecology (Van Hove, Tuijthof, Verdaasdonk, Stassen, & Dankelman, 2010). It is reported that OSATS and Virtual reality simulators have been studied most, OSATS can be accepted as a “gold standard” for objective skills assessment purposes (Van Hove et al., 2010). However, it is also stated that the use of OSATS in the actual operating theatre is not as frequent, leading to doubts whether it can distinguish between different levels of performance in such real scenarios (Van Hove et al., 2010). Moorthy et al. (2003) also reported certain constraints, such as resources and time to find supervising surgeons to observe and evaluate the performance of trainees as a drawback of OSATS assessments (Moorthy, Munz, Sarker, & Darzi, 2003).

Based on the findings in the literature, it is seen that no standard method is defined to determine the experience levels of surgeons. In general, only the number of operations performed in a specific surgical procedure is considered as a cutoff point while determining skill levels. However, the number of operations performed does not necessarily imply that a surgeon can operate professionally without flaws (Darzi, Smith, & Taffinder, 1999). Also, assessment may be easier in the surgical skills training laboratory than in the theatre, where the tasks should be organized carefully for realistic surgical practice considering the essential requirements of feasibility, reliability, and validity (Darzi et al., 1999).

Different from OSATS, virtual reality simulators are widely used as a practice-based method in surgical skill assessment, such as Minimally Invasive Surgical Trainer – Virtual Reality (MIST-VR), a subject performs a task holding two standard laparoscopic instruments whose movements are electronically tracked, recorded and evaluated. The system provides low-level analysis of the positions, forces, and times recorded during training to assess surgical skill (Darzi & Mackay, 2002; Lin, Shafran, Yuh, & Hager, 2006) also defined as a useful objective assessment tool for evaluating the psychomotor skills of senior, junior, and novice laparoscopists (Gallagher, Richie, McClure, & McGuigan, 2001).

There are other studies for objective skill level assessment using classification methods. For instance, Chmarra et al. (2010) conducted a research with 31 gynecologic surgical residents, 10 were ‘experienced’ who performed more than a hundred , 10 were ‘intermediate’ who performed 10 to 100 laparoscopic surgical procedures, and 11 were ‘novice’- medical students who had no prior experience in laparoscopic surgery (M. K. Chmarra, Klein, de Winter, Jansen, & Dankelman, 2010). Considering six metrics: total time, path length, depth perception, motion smoothness, angular area, and volume, the study uses Linear Discriminant Analysis (LDA) method is used for classification, estimating the skill levels with an accuracy of 74.2% (M. K. Chmarra et al., 2010).

3.3 Materials and Methods

In order to estimate the skill levels, the performance data of the participants is obtained while working on four scenarios, and in three different hand conditions, namely dominant hand (DH), non-dominant hand (NH) and both-hand (BH). There were three main stages performed as the method of this chapter: feature extraction, feature selection and evaluating classifiers, as given in Figure 3.1.

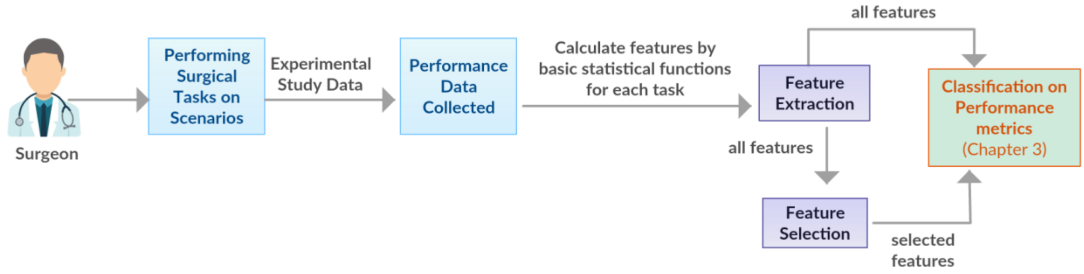


Figure 3. 1 Methodology for Classification on Performance Metrics

In the first stage, the features were extracted using performance metrics such as time, distance, camera distance, catch time, deviation count and accuracy for each scenario. In addition to these metrics, basic statistical functions were used to extract some features such as sum, min, max, mean, standard deviation and variance for each task in the scenarios. Number of all extracted features for all hand conditions in four scenarios was given in Table 3.1 below.

Table 3. 1 Number of Features Extracted (All Features) for Scenarios

	# of Features		
Scenarios	**DH**	**NH**	**BH**
I	55	55	72
II	55	55	55*
III	72	72	89
IV	136	136	158

** In Scenario-2 under both-hand condition, there occurred a machine failure problem while gathering hand data for the distance of tool as controlled by the dominant-hand. Hence, features related to tool distance cannot be included in the classification process.*

After this stage, feature selection methods were implemented to improve the performance of classifiers and the accuracy of the classification process. Three methods, namely reliefF- an instance-based supervised approach for ranking attributes (Urbanowicz, Meeker, LaCava, Olson, & Moore, 2017; Y. Zhang, Ding, & Li, 2008); wrapper subset evaluation, similarly a supervised method, using naïve Bayes classifier along with cross-validation (Witten et al., 2016) and PCA (Principle Component Analysis) - a well-known unsupervised approach for dimensionality reduction, constructing new features from the original ones (Goswami & Chakrabarti, 2014)- were compared in order to find the best accuracy. After evaluating these three methods on Scenario-1 under all hand conditions, reliefF is found to be the best method which gives higher accuracy in classification results (see Appendix A). Number of features after the selection process was given in Table 3.2 below.

Table 3. 2 Number of Selected Features for Scenarios

	# of Selected Features		
Scenarios	**DH**	**NH**	**BH**
I	6	10	19
II	11	22	18
III	26	14	28
IV	30	31	54

In the final stage, classifiers were trained and evaluated using 10-fold cross-validation. The feature selection (using three different methods) and classification processes were performed by using WEKA workbench (Witten et al., 2016). Brief explanations on these methods were provided in the following section.

3.3.1 Classification Methods and Algorithms

Classification is an important task in data mining, which can be simply explained as determining the category of a given record through a classification model to differentiate between instances belonging to different classes (Pan-Ning Tan, Michael Steinbach, & Vipin Kumar, 2006). Data classification process consists of two stages; learning and classification. At the first stage, classification model is constructed by using training data set at learning stage. After that, this constructed model is used for predicting the class for an unseen instance at classification stage (Han, Pei, & Kamber, 2011). Previous studies in the literature reported the top 10 data mining algorithms for classification process, such as SVM, kNN, AdaBoost and Naïve Bayes (Wu et al., 2008). Along with those, a total of 12 algorithms were used in this study.

3.3.1.1 Naive Bayes Method

Naive Bayes is a statistical classification method based on Bayes' probability theorem with the assumption of strong independence between the features (Murphy, 2006). Previous studies reported that this approach is easy to construct and interpret

(Wu et al., 2008), and effective in text classification and medical diagnosis (Rish, 2001).

3.3.1.2 Logistic / Simple Logistic Regression Models

These classifiers are also based on statistical approaches by building linear logistic regression models (Agarwal, Pandey, & Tiwari, 2012). Simple regression functions are acted as base learners for fitting the logistic models (Landwehr, Hall, & Frank, 2005).

3.3.1.3 Support Vector Machines (SVM)

The Support Vector Machine (SVM) is a popular machine leaning method for classification, non-parametric, supervised learning algorithm (Chang & Lin, 2011; Cortes & Vapnik, 1995). In WEKA environment, sequential minimal optimization (SMO) is implemented for training a support vector classifier (Witten et al., 2016). The SVM algorithm proposed a classification method which linearly classifies the optimal hyper-plane by mapping the input space into a high-dimensional feature space. This method can also perform nonlinear classification through "kernel trick"(Cortes & Vapnik, 1995). Three kernel functions were used in this study; namely polynomial, normalized polynomial and the Pearson VII universal kernel (PUK) functions (Üstün, Melssen, & Buydens, 2006) kernels. Based on the findings in the literature, several studies used SVM method to classify surgical levels based on hand motion patterns (Allen et al., 2010; Robert A Watson, 2014).

3.3.1.4 K- Nearest Neighbor (KNN)

KNN is a lazy learner algorithm, also referred as instance based learning approach which stores all the training data, and classify the unseen record based on a similarity measure. The similarity is calculated by distance metric, in this study by using Euclidean distance measure, and defining k closest points in order to perform classification (Steinbach & Tan, 2009; Wu et al., 2008). In this study, KNN is performed with different number of nearest neighbors, including k = (1, 3, 5, 7, 9). Determining the value of k is important in this method to correctly identify the class of an instance, where smaller k values are sensitive to the noise points and larger values leads to misclassification of other instances in the neighborhood.

3.3.1.5 Locally Weighted Learning (LWL)

Similar to KNN approach, LWL is also an instance-based algorithm which assign weights to each instance, and then use these weights for the prediction (Atkeson, Moore, & Schaal, 1997; Frank, Hall, & Pfahringer, 2002).

3.3.1.6 Boosting and Bagging

Boosting (i.e. Adaboost algorithm) and Bagging methods are two well known classification techniques to create ensemble models. A collection of individual classifiers were used to obtain the ensemble model to increase the accuracy of classifiers by voting the decisions of each individual classifiers in the ensemble (Dietterich, 2000; Witten et al., 2016).

3.3.1.7 Rule- based Algorithms

Rule-based classification approaches used a sequence of rules extracted from the training set for determining the class of an unseen instance. Rule-based models are constructed based on if-then rules. In this study, PART (a method based on obtaining rules from partial decision trees) and Jrip (RIPPER) methods were used as rule-based classifiers (Frank & Witten, 1998; Nguyen & Choi, 2008; Witten et al., 2016).

3.3.1.8 Decision Trees

Decision Trees are easy to use and efficient classification models, which are constructed by using information gain algorithm, i.e. placing the best discriminator attribute at the root node. For instance, J48, a decision tree based learner, which is the optimized version of the most popular tree classifier C4.5 (Nguyen & Choi, 2008; Quinlan, 2014). Random Forest is a type of ensemble classifier that uses decision trees as a based classifier, which is reported as a method performing faster training and being more stable (Chan & Paelinckx, 2008).

3.3.2 Cross Validation

Cross-Validation is a statistical method of evaluating and comparing learning algorithms by dividing data into two parts as training and test sets (Refaeilzadeh, Tang, & Liu, 2009). The most common form of cross validation is k-fold cross validation (Refaeilzadeh et al., 2009), where k is the number of partitions, named as folds. After that, k iterations of training and validation are performed by each time leaving out one partition as test set, and the other (k-1) partitions as training set. In this study, classifiers were trained and evaluated using 10-fold cross-validation.

3.4 Results

In this section, the results of the surgical experience level classification for all the four scenarios and three hand conditions are presented. The analyses are carried out using a total of 12 algorithms from Weka workbench (Witten et al., 2016). In what follows, the best accuracy results for estimating the skill levels appear for each scenario, by using all feature set and selected feature set with ReliefF method and the algorithm used for classification as in Table 3.3.

Table 3. 3 Best Accuracies and Related Classification Algorithm

	All Feature Set			**Selected Feature Set**		
Scenario	**DH**	**NH**	**BH**	**DH**	**NH**	**BH**
I	67.85 *LWL*	60.71 *SVM PolyKernel*	71.42 *LWL*	75.00 *LWL*	75.00 *SVM PolyKernel*	78.57 *KNN (k=9)*
II	64.28 *Jrip*	82.14 *LWL*	75.00 *KNN (k=1)*	67.85 *KNN (k=7)*	85.71 *AdaboostM1 SVM Puk*	75.00 *Logistic*
III	64.28 *KNN (k=7)*	82.14 *Simple Logistic*	57.14 *KNN (k=9)*	75.00 *KNN (k=3)*	85.71 *AdaboostM1*	76.92 *KNN (k=3)*
IV	78.57 *Jrip*	64.28 *SVM PolyKernel*	71.42 *KNN (k=3)*	82.14 *SVM NormPoly*	75.00 *SVM Puk*	78.57 *SVM PolyKernel*

LWL: Locally Weighted Learning; SVM: Support Vector Machine with polynomial, normalized polynomial and the Pearson VII universal kernel; KNN: K- Nearest Neighbor; Jrip: Ripper

As seen from the results in Table 3.3, the best accuracy result for estimating the skill levels of the participants in Scenario-1 was obtained in the both-hand condition using the selected features with an accuracy of 78.57% (22/28) with $KNN_{(k=9)}$ algorithm. Both for dominant and non-dominant hand conditions, the skill levels of surgeons were estimated with an accuracy of 75% (21/28).

In Scenario-2, the best accuracy result for estimating the skill levels of the participants was obtained in the non-dominant hand condition using the selected features with an accuracy of 85.71% (24/28) with AdaboostM1 and SVM algorithm using the Pearson VII function-based universal kernel. After that, in the both-hand condition, the best accuracy is found as 75% (21/28). In the dominant-hand condition, a best accuracy of 67.85% (19/28) is obtained using the selected feature set.
In Scenario-3, the best accuracy result for estimating the skill levels of the participants was obtained in non-dominant hand condition using the selected features with an accuracy of 85.71% (24/28) with AdaboostM1 algorithm. Under the dominant-hand condition, the accuracy skill level estimation is 75% (21/28), whereas an accuracy of 76.92% (20/26) was obtained in the both-hand condition.

In Scenario-4, the best accuracy result for estimating the skill levels of the participants was obtained in dominant hand condition using selected features with an accuracy of 82.14% (23/28) with SVM algorithm using normalized polynomial kernel. Later, in the both-hand condition, the best accuracy is found as 78.57% (22/28). Finally, in non-dominant hand condition, a best accuracy of 75% (21/28) is obtained. The detailed classification results for each scenario and the information about features selected were given in Appendices A and B, respectively.

As a result, considering all the scenarios and hand conditions, the best estimation result for this study is obtained in Scenario-2 and Scenario-3, both in the non-dominant hand condition, with the accuracy of %85.71, where 24 instances were correctly classified out of 28 as belonging to the either novice or intermediate group.

3.5 Discussions and Conclusions

The results of this study show that, by using classification algorithms, through some features collected by computer-simulated virtual surgical training environments with haptic interfaces, it is possible to estimate the intermediate and novice surgical skills with 79-86%. The non-dominant- and both-hand skills in three scenarios (1, 2, and 3) were more decisive for a better assessment of these skill levels. This study proposes an assessment approach by using certain features based on the performance data collected in computer-based endoscopic surgery training simulation environments. However, the proposed estimation accuracy can be even improved by developing additional features from more detailed data about the trainees such as hand behaviors, eye behaviors and their correlations.

Accordingly, for specific scenarios, by pre-defined threshold values, which can be calculated using such classification techniques alongside experimental data, the skill levels of trainees can be calculated. This information may help educators to improve their assessment of candidates. Additionally, this information can be used to create more adaptive training programs by regularly assessing the skill levels of the trainees through different scenarios developed for specific surgical procedures. In the future, training scenarios can be arranged with such threshold values and assessment modules to guide the trainees and support educators in the field.

As a conclusion, given the limited number of participants in this specific field of surgery, estimating the skill levels as novice and intermediate with the accuracy of 86% can be considered as a successful result. Nevertheless such accuracy can still be improved by using some other data. Additionally, when other skill levels, such as experts and the beginners, are included in such a study with a higher number of participants, the accuracy can be improved yet again. Hence, in the future, by applying such techniques, more standardized training programs can be developed for skill-based surgical education.

CHAPTER 4

VELOCITY- BASED HAND METRICS TO DIFFERENTIATE SURGICAL SKILL LEVELS

Several methods were reported in previous studies to assess the skill levels of surgeons. Traditionally, the skill assessment of surgical trainees is done through observation by a number of supervised surgeons, providing verbal feedback to the trainees. For instance, Adrales et al. (2003) conducted a study with 27 subjects having various years of experience, who performed on laparoscopic simulations. A task-specific checklist is used to assess participants including four skills on a 5-point scale; respect for the tissue, economy of movement, flow of the operation and spatial orientation (Adrales et al., 2003). It is revealed that years of experience directly correlated with the ratings given to these skills, whereas it is inversely correlated with the time needed to complete each procedure (Adrales et al., 2003). Additionally, it is also indicated that speed while performing tasks and quality of performance increased with surgical experience (Adrales et al., 2003).

With the advancements in technology in surgical domain regarding MIS procedures, there has been an increased interest in objective assessment methods for surgical skills. The observation process carried out by training specialists is a subjective method by nature. For this reason, certain objective measures have been proposed to improve the quality of assessment. For instance, ‘structured human grading’ can be considered as an extension to the traditional observation approach, with the aim to standardize the evaluation by using a rated-scale checklist (Reiley, Lin, Yuh, &

Hager, 2011). One of the well-known example of this method is objective structured assessment of technical skills (OSATS), which have been introduced for testing specific operative skills in surgical trainees (Cagiltay et al., 2017; Martin et al., 1997). Such methods can increase the standardization of surgical skill evaluation, however still some issues were reported about objectivity, time constraints and the need for further research on correlation between technical skill and patient outcome (Moorthy et al., 2003; Reiley et al., 2011).

Another method for objective skill assessment is to use virtual reality simulators, based on gathering metrics such as time, distance covered, accuracy of the task along with metrics related to hand motion such as position and speed of the surgical device. For instance, Felsher et al. (2005) presented a study to validate the ability of the simulator to distinguish experienced surgeons from the beginners using metrics time to reach target, percentage of visualized area and time spent in clear view, completion rate and overall efficiency, while performing on a colonoscopy simulation (Felsher et al., 2005). It is reported that experienced endoscopists reached the target more rapidly, visualized more of the surface with a higher proportion of time in clear view and in overall performance they were more efficient than novices (Felsher et al., 2005). Hence, the virtual simulation environments provide several measures to better understand the surgical performance and skill levels. However, in the literature, even there are some studies to better understand the hand-movement metrics taken from the virtual reality (VR) environments, studies attempt to use such metrics to classify the trainees according to their experience levels are limited. This study aims to focus on hand movement metrics to differentiate skill levels and to better understand the trainees' performances. The findings of the literature about hand metrics were summarized as below.

4.1 Metrics based on Hand Movement

Today, by using robotic surgical systems and medical simulators, it is possible to analyze surgical motion, to understand the differences in levels of technical skills in surgeons (Reiley et al., 2011). Reiley et al. (2011) reported the methods of tracking surgical motion and the available data-collection systems. It is reported that almost all systems measure tool motion (e.g., position and velocity) and some additionally measure the force of tracked instruments, where electromagnetic, mechanical, or optical systems can be used for motion tracking (M. Chmarra, Grimbergen, & Dankelman, 2007; Reiley et al., 2011). For instance, wearable devices such as 'Instrumented Gloves' shown in Figure 4.1, are used in hand movement analysis to gather hand dexterity metrics (Lemos, Hernandez, & Soto-Romero, 2017).

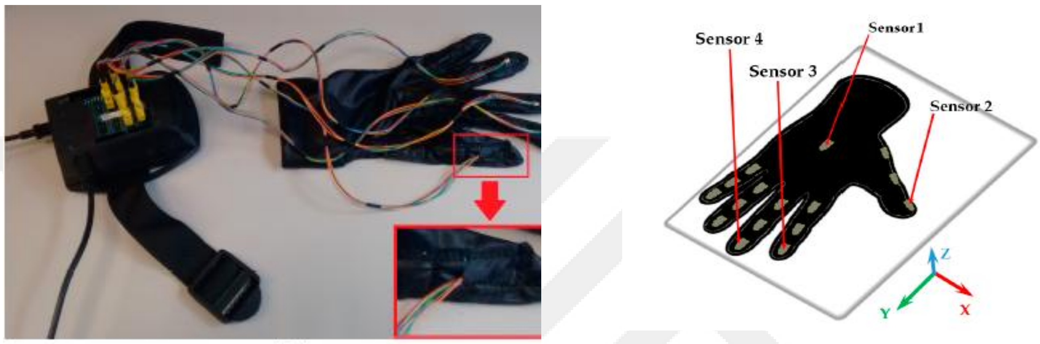


Figure 4. 1 Hardware components of Instrumented Glove (Lemos et al., 2017)

Using such systems, it is possible to track the motion of hands/instruments in a 3D environment to observe the differences among surgeons having different skill levels. In Bauernschmitt et al. (2006) study, force feedback is implemented into telemanipulated surgery using surgical instruments equipped with haptic feedback, in order the improve the telemanipulator systems presented in a previous study (Mitsuishi, Tomisaki, Yoshidome, Hashizume, & Fujiwara, 2000), where the feedback of force on surgical skills is evaluated dependent on different surgical experience (Bauernschmitt et al., 2006). Hence, another behavior difference between novices and experts can be reported as the applied forces on the haptic device used in the simulators. In that study, applied forces and speed of hand motion were recorded for heart surgeon participants. When the feedback of force in the dominant and non-dominant hand on surgical skills is evaluated, it is revealed that experienced surgeons worked with significantly less force in the non-dominant hand than the young surgeons (Bauernschmitt et al., 2006).

In Hofstad et al., (2013)'s study, there were three groups of surgeons; experts, intermediates and novices performing a labyrinth task. It is reported that motion analysis is a valid objective way for assessing psychomotor skills in MIS (Hofstad et al., 2013). Nine motion-related metrics derived from the position and the orientations of the instruments were collected (Hofstad et al., 2013):

- Time to complete each task
- Bimanual dexterity: ability to control to instruments at the same time)
- Path length: total movement of the tip of the instrument
- Angular length: total change in angle of the tip of the instrument
- Depth perception: total distance travelled
- Response orientation: total amount of instrument rotation around its axis
- Motion smoothness: total change in acceleration of the instrument
- Number of sub-movements
- Average velocity of the instrument measured in mm/s.

It is revealed that experts and intermediates performed significantly better than the novices in terms of time and parameters measuring the amount of instrument movement and experts had significantly better bimanual compared to the intermediates and novices as well as they have performed the task in a shorter instrument path length with the non-dominant hand than the intermediates (Hofstad et al., 2013). It is also reported that there was no difference according to the motion smoothness metric (M. Chmarra, Kolkman, Jansen, Grimbergen, & Dankelman, 2007; Hofstad et al., 2013; Maithel, Villegas, Stylopoulos, Dawson, & Jones, 2005). Maithel et al. (2005) conducted a study with 30 surgical residents, performing on Computer Enhanced Laparoscopic Training System (CELTS), generating an overall score using six metrics: time, depth perception, path length, response orientation and motion smoothness (Maithel et al., 2005). The results of the study showed that the senior residents performed significantly better than the junior residents overall on all parameters, except for motion smoothness (Maithel et al., 2005).

In Chmarra et al. (2010)'s study, it is aimed to define a method for surgical skill level classification, based on psychomotor laparoscopic skills alone. There were thirty one surgical residents distributed as: 10 experienced, 10 intermediate, and 11 novice, where each participant performed four tasks: pipe cleaner, rubber band beads, and circles. Six assessment metrics were used in the analysis extracted from the MIS tools motion: total time, path length, depth, motion smoothness, angular area, and volume. Using these metrics, the model able to classify 23 out of 31 cases correctly (M. K. Chmarra et al., 2010).

Pellen et al. (2009) reported that in their study, there were four groups as consultant, senior, junior and student among 160 participants. Motion analysis data was obtained using optical strips attached to the instruments and three motion related metrics were collected: time, path length (cumulative distance in mm. providing information about economy of movement) and smoothness (the cumulative number of instrument accelerations). The results imply that consultants outperformed students and juniors and seniors dissected faster, more efficiently and more accurately than juniors and students (Pellen, Horgan, Barton, & Attwood, 2009).

As a summary, earlier studies report that metrics such as total time, distance- path length, depth, hand motion smoothness, angular area, speed of motion and force/torque were collected while the surgeons performing on virtual reality surgical trainers, in order to objectively assess their surgical skill levels. Hence, it is understood that, metrics related to the hand movements of surgeons can be used for assessing surgical skill levels in surgical simulation environments.

4.2 Procedure of the Study

The aim of this study is to understand the behavior differences between intermediate and novice surgeons based on their hand movement data collected while they were performing tasks in both-handed simulation scenarios. Time and distance metrics were collected during the experiment by using haptic devices, then velocity and angular velocity metrics were calculated for both dominant- and non-dominant hands controlling the tool and the camera, respectively. The procedure for this chapter is given in Figure 4.2.

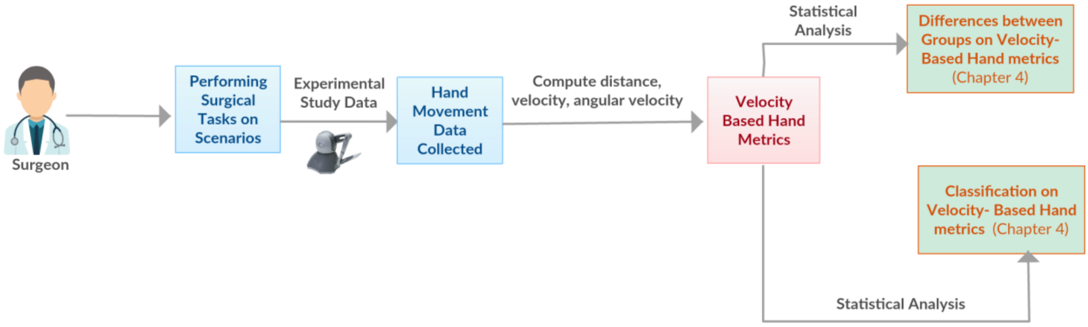


Figure 4. 2 Procedure for Analysis on Velocity- based Hand Metrics

A total of 28 surgical residents (21 doctors and 7 interns) from the Department of Neurosurgery (12 participants) or Otolaryngology (ENT) (9 participants) participated in this research. Four scenarios as described earlier were developed for surgical training purposes as a part of an Educational Computer-based-simulation Environment (ECE) project. All of these scenarios were implemented in Unity 3D (Unity, 2017) with C#. Unity provides a high fidelity, low cost and easy to use simulation environment (Craighead et al., 2008). Each scenario was prepared in two versions: single-handed and both-handed. Accordingly, each scenario was performed under three different hand conditions (i.e. dominant hand, non-dominant hand, and both-hand). The experimental study was started with single-handed tasks, after completing all the tasks of that scenario, continued with both-handed tasks. In this study, we specifically concentrated on both-handed scenarios to understand the participants' hand movement under dominant- and non-dominant-hand conditions. In order to understand the hand movement differences between groups following metrics were recorded using software developed in Unity 3D.

4.3 Hand Metrics extracted from Hand Movement Data

The hand movement data of the participants is recorded including the timestamp and 3D vector positions (coordinates) of both tool controlled by their dominant-hand and camera controlled by their non-dominant hand. A hundred data points per second recorded by the software as the hand coordinates for both hands with the help of the haptic devices. Among the collected data, the Euclidian distance formula (Formula 1) is used to compute the distance between the current position denoted by a 3D vector

A (x_a, y_a, z_a) and the next position denoted by a 3D vector B (x_b, y_b, z_b) of the tool and camera in Unity environment. The distance is measured in mm.

$$distance_{A,B} = \sqrt[2]{(x_a - x_b)^2 + (y_a - y_b)^2 + (z_a - z_b)^2} \qquad (1)$$

After that, the linear velocity (*V)* of the device (tool or camera) is measured by (Formula 2) in mm/s, where the change in the position is denoted by Δx:

$$V_{device} = \frac{X_{final} - X_{initial}}{time} = \frac{\Delta x}{t} \qquad (2)$$

Finally, the magnitude of angular velocity (ω) regarding the rotational speed of the device (change in the angle of the instrument, in the plane perpendicular to the instruments axis) in 3D environment is calculated by (Formula 3) where ΔΘ representing angular rotation in degrees occurs in a time Δt (in seconds).

$$\omega_{device} = \frac{\Delta\Theta}{\Delta t} \qquad (3)$$

Accordingly, performing on all four scenarios, hand metrics such as tool distance, tool velocity, tool angular velocity for the dominant hand; camera distance, camera velocity and camera angular velocity for the non-dominant hand were calculated and used for the analysis.

4.4 Results

The results of this study are provided in two sections. In the first section, the difference between the intermediate and novice groups based on their hand movement behaviors was analyzed using Mann-Whitney U test. In the second, binary logistic regression analysis was performed to estimate the skill levels of surgeons.

4.4.1 Differences between Intermediate and Novices on Hand Metrics

Since the normality assumptions were violated as assessed by Shapiro-Wilk's test ($p < .05$) for novices, and the sample size (n) was less than 30 for the groups, the Mann-Whitney U test was implemented as a non-parametric alternative to the independent-samples t-test to determine the differences between intermediate and novice hand movement metrics. For all scenarios, distributions of the metrics for intermediate and novice groups were similar, as assessed by visual inspection. It is seen that, considering the medians of all metrics, there were no statistically significantly difference between these groups for Scenario-1 and Scenario- 4, as given in Table 4.1. On the other hand, camera distance and angular velocity metrics in Scenario-2 and the camera velocity metric in Scenario-3 were significantly different. In Scenario-2, there occurred a machine failure problem while gathering hand data for the distance of tool as the camera controlled by the dominant-hand. In this case, also velocity, which is based on the distance covered, cannot be calculated. Hence, dominant-hand condition for Scenario-2 is omitted from the analysis.

Table 4. 1 Median values for hand metrics according to skill levels

		Scenario							
		I		**II**		**III**		**IV**	
Hand	**Metrics**	INT	NVC	INT	NVC	INT	NVC	INT	NVC
DH	Tool Distance (mm.)	67.43	67.60	-	-	109.41	136.39	44.96	44.66
	Tool Velocity	20.15	17.28	-	-	11.16	12.99	10.31	9.63
	Tool Angular Velocity	3.77	3.86	3.83	3.79	3.92	3.84	3.68	3.79
NH	Camera Distance (mm.)	17.88	16.23	***70.34***	***84.96****	63.47	53.36	56.01	59.74
	Camera Velocity	5.47	3.91	12.31	13.30	***6.37***	***4.35****	14.54	11.67
	Camera Angular Velocity	2.98	3.50	***3.68***	***3.81****	3.52	3.70	3.90	3.86

** Significance at the 0.05 level*

In Scenario-2, camera distance metric was statistically significantly different between intermediates (Mdn = 70.34) and novices (Mdn. = 84.96), U = 140, z = 2.043, p = 0.042. Additionally, camera angular velocity metric was also statistically significantly different between intermediates (Mdn = 3.68) and novices (Mdn. = 3.81), U = 139, z = 1.999, p = 0.047. These results show that, intermediates took less distance with camera, than that of novices. According to the angular velocity metric, representing the magnitude of the camera rotation is lower for intermediates compared to the novices. In other words, their economy of movement is better than that of the novices by considering the camera rotation and endoscope movements.

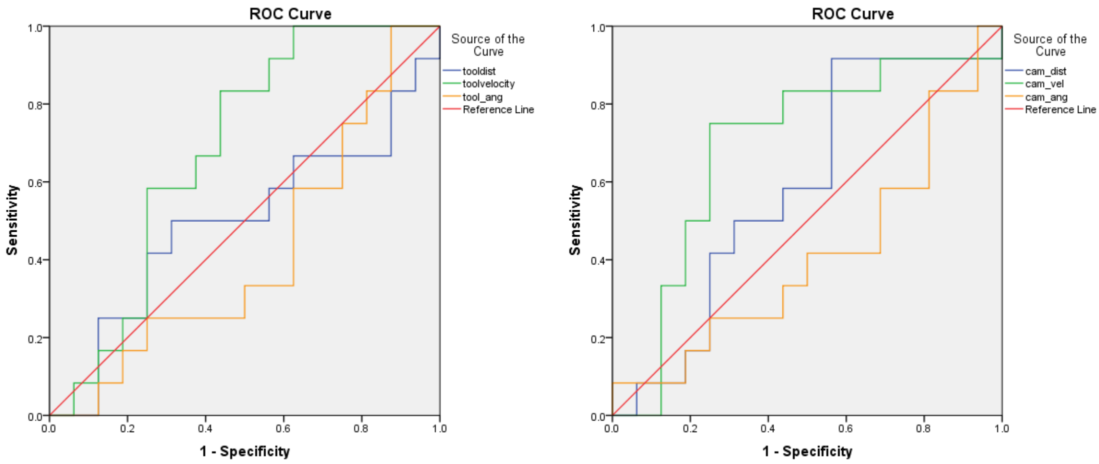
In Scenario-3, there was a statistically significantly difference on the camera velocity metric between intermediates (Mdn = 6.37) and novices (Mdn = 4.35), U = 44, z = -2.057, p = 0.041. The result shows that intermediates control the camera faster using their non-dominant hands, compared to the novices.

4.4.2 Estimating the Skill Levels of Surgeons

Considering four scenarios used in this study, a binomial logistic regression was performed to ascertain the effects of hand metrics; distance, velocity, and angular velocity of both the tool and the camera on the estimation of surgical skills levels of surgeons. The outliers detected was omitted from the analysis, where studentized residuals expected to have values within ±2 range, if they have been standardized in binary logistic regression (Christensen, 2006).

In this analysis, along with the overall classification accuracy (%), sensitivity, specificity, positive and negative predictive values were also calculated to assess the ability of a logistic regression model to correctly classify cases. All these measures were calculated based on a cut-off point of 0.5 (50%). This means that a participant with a predicted probability of the surgical skill expertise that is greater than or equal to 0.5 would be classified as an intermediate (since the positive actual state is given as 'intermediate' while performing the analysis) and novice, otherwise. However, instead of concentrating on one cut-off point only, all possible cut-off points in the data can be considered, and shows how each cut-off point changes the specificity and sensitivity of the test. A visual representation of this can be presented in Receiver Operating Characteristic (ROC) curve, which is a plot of TPR (sensitivity) along the y-axis versus FPR (1 – specificity) in x-axis (Hilbe, 2011; Pang-Ning Tan, Michael Steinbach, & Vipin Kumar, 2006). A model which is more close to the upper left corner of the ROC curve (TPR=1, FPR=0)(Pang-Ning Tan et al., 2006), indicates a better discrimination. In what follows, the results of logistic regression analysis were presented along with visual representations, ROC curve.

In Scenario-1, the logistic regression model was not statistically significant, $\chi2(7) = 7.188$, $p = .41$. The model explained 30.0% (Nagelkerke R^2) of the variance in surgical skill levels and correctly classified 75.0% of cases (intermediates with 66.7% and novices with 81.3%). The ROC Curve is presented for metrics regarding tool in Figure 4.3-A and camera in Figure 4.3-B for Scenario-1, respectively.



A **B**

Figure 4. 3 ROC Curves for metrics regarding tool and camera in Scenario-1

The area under the ROC curve for tool velocity was .68, 95%CI [.48, .88] and camera velocity was .69, 95% CI [.48, .90] as seen in Figure 3, representing a better discrimination compared to other measures. In Scenario-2, the logistic regression model was statistically significant, $\chi2(6) = 23.612$, $p < .0005$. The model explained 79.7% (Nagelkerke R^2) of the variance in surgical skill levels and correctly classified 88.5% of cases (intermediates with 83.3% and novices with 92.9%). The ROC Curve is presented for measures regarding camera given in Figure 4.4 for Scenario-2.

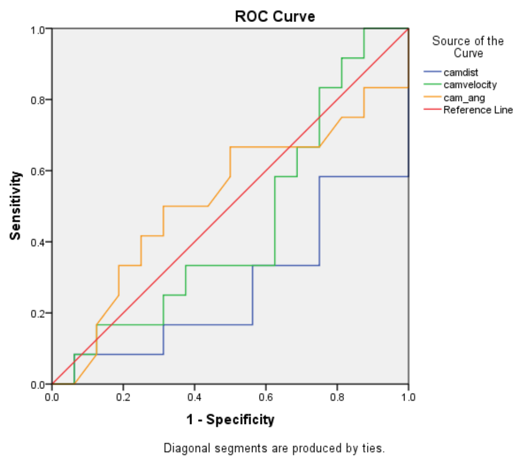


Figure 4. 4 ROC Curves for metrics regarding camera in Scenario-2

As seen in Figure 4.4, the area under the ROC curve for the angular velocity of camera was .52, 95%CI [.30, .75], representing a better discrimination compared to other metrics. In Scenario-3, the logistic regression model was not statistically significant, χ2(6) = 7.311, p = .29. The model explained 35.1% (Nagelkerke R^2) of the variance in surgical skill levels and correctly classified 75.0% of cases (intermediates with 72.7% and novices with 76.9%). The ROC Curve is presented for metrics regarding tool in Figure 4.5-A and camera in Figure 4.5-B for Scenario-3, respectively.

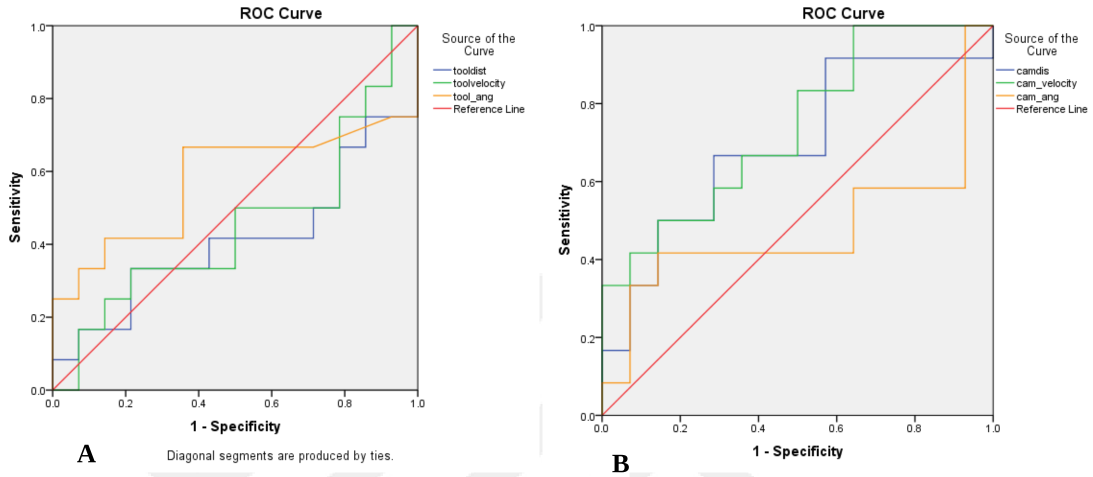


Figure 4. 5 ROC Curves for metrics regarding tool and camera in Scenario-3

The area under the ROC curve for the angular velocity of tool was .58, 95%CI [.33, .82] as shown in Figure 4.5-A. The area under the ROC curve for the velocity and distance metrics for the camera were .74, 95% CI [.55, .93] and .70, 95% CI [.48, .90], respectively, which is an acceptable level of discrimination as reported by a previous study (Hosmer Jr, Lemeshow, & Sturdivant, 2013).

In Scenario-4, the logistic regression model was also statistically significant, χ2(6) = 22.434, p = .001. The model explained 81.7% (Nagelkerke R^2) of the variance in surgical skill levels and correctly classified 91.7% of cases (intermediates with 90.0% and novices with 92.9%). The ROC Curve is presented for metrics regarding tool in Figure 4.6-A and camera in Figure 4.6-B for Scenario-4, respectively.

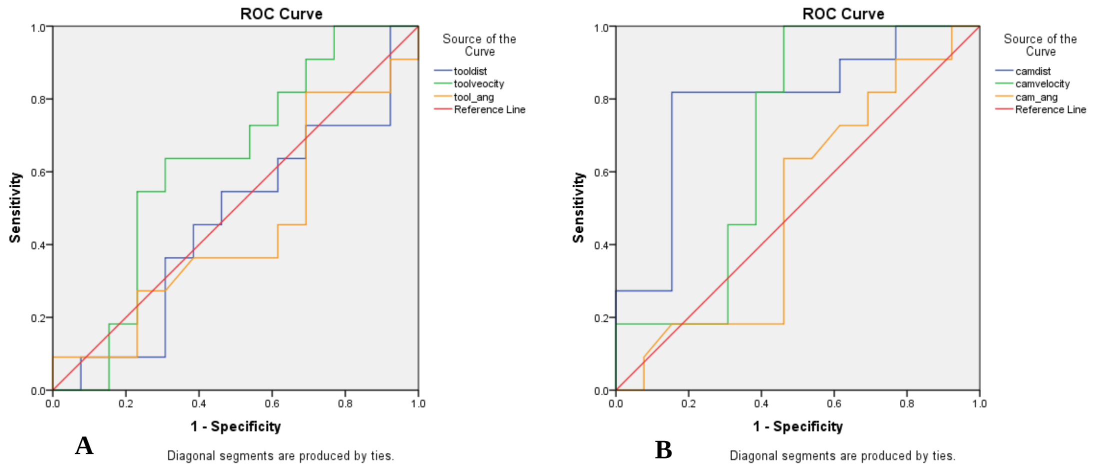


Figure 4. 6 ROC Curves for metrics regarding tool and camera in Scenario-4

The area under the ROC curve for the velocity of tool was .62, 95%CI [.39, .85]. The area under the ROC curve for the velocity and distance metrics for the camera were .70, 95% CI [.47, .92] and, .79, 95% CI [.60, .98], respectively, which is an acceptable level of discrimination (Hosmer Jr et al., 2013).

4.5 Discussions and Conclusions

The results of this study show that hand movement metrics such as distance, velocity and angular velocity can potentially be used as a metric for assessing surgical skill levels. According to the results of Mann-Whitney U test, there were no statistically significantly difference between these groups for Scenario-1 and Scenario- 4. On the other hand, camera distance and angular velocity measures in Scenario-2 and the camera velocity metric in Scenario-3 were significantly different. Firstly, the result implies that intermediates control the camera faster using their non-dominant hands, compared to the novices. This finding is supportive to the Adrales et al.'s study, reporting that task speed and quality of performance increased with surgical experience (Adrales et al., 2003). Pellen et al., (2009)'s findings also supportive that senior surgeons dissected faster, more efficiently and more accurately than juniors and student groups (Pellen et al., 2009).

Secondly, intermediates took lower distance with camera using their non-dominant hands, than that of novices. Finally, angular velocity metric (the magnitude of the camera rotation) is lower for intermediates compared to the novices. This finding is also supportive to Hofstad et al., (2013)'s study, reporting that experts and intermediates performed significantly better than the novices in terms of time and parameters measuring the amount of instrument movement and added that experts performed the task in a shorter instrument path length with the non-dominant hand than the intermediates (Hofstad et al., 2013). The results of binary logistic regression analysis to estimate the skill levels of surgeons show that, for both Scenario-1 and Scenario-3, the model correctly classified 75.0% of cases. The classification accuracy of model increases in Scenario-2 and Scenario-4, where the model correctly classified 88.5% and 91.7% cases, respectively. These results show that the level of difficulty of the tasks that are defined in each scenario, the fidelity level of the scenario, the order of the scenarios provided in the curriculum may all be influencing factors for the accuracy of the estimations. However the results also show that when the scenarios and educational programs are appropriately designed, it is possible to increase the accuracy of the skill level estimation above 90% which can be considered as acceptable to support the training programs.

After that, the area under the ROC curve was presented in order to visualize the measures in each scenario, to determine better discriminators in classification process compared to others (see Table 4.2).

Table 4. 2 Summary of Area under ROC Curve (AUC) Results for Scenarios

Hand Condition	**Metrics**	**Scenarios**			
		I	**II**	**III**	**IV**
DH	Tool Distance				
	Tool Velocity	.68			.62
	Tool Angular Velocity			.58	
NH	Camera Distance			.70	.79
	Camera Velocity	.69		.74	.70
	Camera Angular Velocity		.52		

As seen in Table 4.2, camera velocity is the metric, which results in a better discrimination in Scenarios 1, 3 and 4, followed by camera distance (Scenarios 3 and 4) and tool velocity (Scenarios 3 and 4). Angular velocity of both camera and tool were the metric also helping to differentiate between surgical skill levels in Scenarios 2 and 3, respectively. Additionally, it can be seen from Table 4.2 that non-dominant hand metrics result in a better discrimination compared to the dominant-hand metrics. These results are supportive to the earlier studies reporting that an experienced MIS surgeon can be distinguished from a less experienced one by the higher ability to control the instrument in the non-dominant hand and the higher degree of simultaneous (coordinated) movements of the two instruments (Hofstad et al., 2013).

As a conclusion, four main findings of this study can be summarized as below:

- Intermediates took less distance with camera, than that of novices.
- Angular velocity metric (the magnitude of the camera rotation) is lower for intermediates compared to the novices.

In other words, their economy of movement is better than that of the novices by considering the camera rotation and endoscope movements.

- Intermediates control the camera faster using their non-dominant hands, compared to the novices.
- Non-dominant hand metrics (distance, velocity and angular velocity of camera) results in a better discrimination compared to the dominant-hand metrics.

CHAPTER 5

PROPOSED HAND METRICS USING BIT ALGORITHM AND EYE-HAND COORDINATION ANALYSIS OF SURGEONS

Besides its several benefits, minimal invasive surgery (MIS) education has many challenges. Surgeons involved in these type of operations need to develop several skills, such as eye-hand coordination to mention an important one (Hernandez et al., 2004). As the location of the scene can only be observed through a monitor in MIS, mislocation can make it impossible for the surgeon to observe his/her hands as well as the operative scene simultaneously, making eye-hand coordination even more critical (Wentink, 2001). Since the operator is does not directly at his/her hand, but the monitor, in MIS the real-world depth perception tends to fade, causing mismatch problems in hand or tool movements in time and space (Batmaz, de Mathelin, & Dresp-Langley, 2017). Hence, eye-hand coordination becomes a critical issue to tackle for MIS purposes. Studies show that, during the training programs, surgeons can develop their skills to handle these problems. In order to support such programs, supportive technology with enhanced tools and methods is required. Tracking the hand movements is one of those technologies to better enhance the current training programs. In this respect, the analysis of operational tool trajectory has been suggested as an effective method for monitoring surgery training; also, several other hand-movement metrics have been proposed such as path length, motion smoothness, depth perception, response orientation and grasping (Stylopoulos & Vosburgh, 2007). However, detecting the location of a given object, such as the tool or hand, in a precise manner is important for each of these metrics (Helsen, Elliott, Starkes, & Ricker, 2000; Oropesa, Chmarra, et al., 2013). Tracking hand and instrument movements using markers, known as ‘motion analysis’, has been suggested by earlier studies as an alternative method in assessing the related skills by measuring the economy of movement (Datta, Chang, Mackay, & Darzi, 2002). Furthermore, it has been reported that motion analysis are useful tools to assess performance compared to the OSATS (Objective structured assessment of technical skill) and time alone (Hernandez et al., 2004). Additionally, in the literature video processing methods (Jiang, Zheng, & Atkins, 2015; Oropesa, Sánchez-González, et al., 2013) and motion tracking systems (Oropesa et al., 2011) have been proposed to detect the tool position in a precise way, giving rise to other practical concerns.

Today, computer-based simulation environments provide for several objective assessment capabilities which can be attained continuously during the training period and without any expert supervision (Ayodeji, Schijven, Jakimowicz, & Greve, 2007). However, there are still some questions that need to be answered to better integrate these metrics into the traditional curricula (Wilson et al., 2010). Besides, the calculation methods for many of these metrics are implicit and within the simulators, making them difficult to manipulate. Despite the fact that these metrics provide some insight into the relationship between eye and hand and their coordination, still there is a need improve our understanding how hand movements are guided and controlled by vision (Wilson et al., 2010).

Accordingly, in this chapter, some additional hand-movement metrics have been proposed using precise data taken from a computer-based simulation environment by using a haptic interface. Additionally, the surgical residents' eye and hand behaviors are analyzed considering eye-gaze and hand movement metrics.

5.1 Metrics for Behavior Differences of Surgeons

Effective and objective metrics are necessary in order to provide for proper feedback and continuous analysis of the psychomotor skills in MIS (Cagiltay et al., 2017; Oropesa et al., 2014). In that concern, in the literature several metrics have been proposed which can be grouped into two categories. The first group attempts to understand the performance of the trainees, focusing on task accuracy and duration.

Beside these performance metrics, others have also been proposed to as regards the behaviors of the trainees during the tasks, such as eye events (fixations and saccades) and hand movement metrics. The present work is mainly based on the eye-events and hand movement behaviors of the trainees.

5.1.1 Metrics Based on Eye Events

Colby and Goldberg (2000) have defined a calculation for eye events according to which, ‘saccade’ is the size and direction of the eye movement and a representation of visual space in parietal cortex to be remapped in advance of the eye movement; whereas ‘fixation duration’ is the period through which the representation of the visual scene in parietal cortex is stable (Colby & Goldberg, 1992).

5.1.2 Metrics Based on Hand Behaviors

Idle Time: Latko et al., (1997) provide some definitions for hand movements. According to them, when no regular exertions are detected, the hand activity is considered as idle, and when there is infrequent motion, it is considered as steady motion. Based on the frequency of the motion, they also propose ‘consistent conspicuous’ (long pauses or slow motions), slow steady motion, and rapid steady motion (Latko et al., 1997). Apart from these, one has to consider other studies conducted on motor behaviors in surgical skills which are based mainly on the path-length, the amount of time to complete a procedure, and idle time which also needs to be considered but has remained rather neglected (D'Angelo et al., 2015). Oropesa et al., (2011) define idle time as lack of movement of both hands representing the delay in motor planning or decision making (Oropesa et al., 2011).

Smoothness of hand function: One example of motion metrics is the smoothness of hand function. Oropesa et al. (2013) define motion this as abrupt changes in acceleration resulting in jerky movements of the instrument (Oropesa, Chmarra, et al., 2013). According to Mohamadipanah et al. (2016), further research is necessary to better understand the role and usage of psychomotor metrics, such as smoothness, to assess the performance during certain medical procedures (Mohamadipanah et al., 2016).

Working Space: This is proposed as a metrics for the economy of the area and economy of volume efficiency in MIS (Oropesa, Chmarra, et al., 2013), and is defined using an electromagnetic sensor to track the participants' hand movements and the summation of distances from the sensor's average spatial location (Mohamadipanah et al., 2016).

5.1.3 Eye- Hand Behavior Differences among Surgeons

In the literature, there are several studies focusing on the differences in the visual attention strategies adopted by experts and novices (Eivazi et al., 2012). As a means of measurement, eye tracker tools are used to record the eye movements of the surgeons during the operations. A study proposed by Law et al. (2004) states that visual information is important in surgeons' manipulative performance, such as in laparoscopic surgery (Law, Atkins, Kirkpatrick, & Lomax, 2004). Hence, comparing the differences in the surgeons' eye movement behavior in case of expert surgeons and novices may shed light on what can be incorporated into training as an innovative way of assessing skills (Law et al., 2004). In Law et al. (2004)'s study, eye movements are compared among 5 experts and 5 novices performing a one-handed aiming task on a computer-based laparoscopic surgery simulator , reporting that I experts were quicker and made fewer errors than novices, that novices needed more visual feedback to complete the tasks, and that experts maintained eye gaze on the target while manipulating the tool whereas novices tracked the movement of the tool until it reached the target (Law et al., 2004).

Another study for visual attention conducted by Tien et al. (2010) compares the eye movements of 4 experts and 4 novices performing a simulated gall bladder removal task on a dummy patient with heartbeat and simulated vital signs displayed on a secondary monitor (Tien, Atkins, Zheng, & Swindells, 2010). It is reported that novices had difficulty concentrating on the surgical display in such a way that they could hardly observe the vital signs of the patient, whereas experts regularly glanced at the vital signs in order to observe the patient condition (Tien et al., 2010). In the context of visual attention, Shetty et.al (2012) report that one of the most essential skills in laparoscopic surgery is camera handling and navigation (Shetty et al., 2012). Surgeons rely on camera operators for the purpose of visualization of the operative field. In that study, the behavior differences were analyzed among the participants' in three groups (novice, intermediate, and advanced) on camera navigation, coordination and target visualization modules of a laparoscopic simulation training environment (Shetty et al., 2012), with the final outcome that I novices required significantly more repetitions to complete the camera navigation module compared to the other groups (Shetty et al., 2012).

In Richstone et al.'s (2010) study, the use of eye metrics is investigated to assess surgical skills objectively (Richstone et al., 2010). The study is carried out with twenty-one surgeons in both simulated and live operational settings. Using linear discriminate analysis (LDA) and nonlinear neural network analyses (NNA), they were able to correctly classify surgeons as expert or non-expert with 91.9% and 92.9% accuracy, respectively, based on their complex eye and pupillary movements (Richstone et al., 2010).

Cao, MacKenzie and Payandeh (1996) presented a task and motion analysis in endoscopic surgery by one expert and five novice surgeons (Cao, MacKenzie, & Payandeh, 1996). It is found that novices' movements are much slower than the expert, also requiring more motion attempts in total for the tasks to be completed (Cao et al., 1996).

5.2 Methods and BIT Algorithm

The aim of this research is to assess the relationship between eye-gaze and hand-movement metrics in order to understand the behavior differences of intermediate and novice surgeons in a simulation-based endoscopic surgery environment. For this purpose, BIT algorithm is a fully automatic, velocity-based algorithm to determine fixations (i.e., fixation duration and fixation number) and saccades from the eye data, using direction-, eye-, task- and individual- specific thresholds (van der Lans, Wedel, & Pieters, 2011). Accordingly, in this study, BIT algorithm is used to identify the fixation duration, fixation number and the saccades of the eye-gaze data. The algorithm is also implemented to the hand movement data collected within a surgical simulation environment to develop new hand features. The source code of the algorithm is available on the authors' website using MATLAB (van der Lans et al., 2011).

5.2.1 Eye Metrics

Fixation identification is a statistical description of the monitored eye-gaze behaviors (Salvucci & Goldberg, 2000). In this study, using the BIT algorithm, three metrics namely Fixation Duration (FD), Fixation Number (FN) and Saccade Number (SN) are identified. These specific eye movement events, fixation and saccades, are defined below.

- A *fixation* eye movement can be defined as the "pauses over informative regions of interests" (Salvucci & Goldberg, 2000). Salvucci and Goldberg (2000) reported that fixations are generally determined in the range of 200-400 ms considering the duration, whereas in another study it is reported as 260-330 ms (van der Lans et al., 2011).
- *Saccades* are the "rapid eye movements between fixations" (Salvucci & Goldberg, 2000). Since velocity-based algorithms use velocity thresholds while identifying fixations and saccades, it is reported that saccades have higher velocities compared to fixations (Salvucci & Goldberg, 2000) and that they contribute to a variety of ocular motor behaviors, such as the reflexive movements towards novel stimuli or gaze shifts for any learned tasks (Leigh & Kennard, 2004).

In the literature, it can be seen that different algorithms include different constant thresholds to define fixations and saccades. In the present study, BIT algorithm is chosen because of its ability to automatically specify the threshold values, leading to a variety of thresholds depending on the task and individual.

5.2.2 Proposed Hand Metrics

Based on the definition of eye movement events, a similar statement can be made for hand movement events. In the present work, two hand metrics are introduced to identify the hand movements of the participants as explained below:

- *'Stand Still'* metric is proposed as the period when the hand movement remains within a very small range and lower velocity for some time. In other words, the stand-still measure determines the 'idle state' of the hand movement. By running the BIT algorithm, such events can be classified into '*Stand Still Duration*' (SSD) and '*Stand Still Number*' (SSN) for hand movements.
- In this respect, the '*Sudden Sharp Movement' (SSM)* metric is also proposed to identify very fast, sharp hand movements while performing any given task.

5.2.3 Participants

During the experimental study, some participants' eye and hand data cannot be recorded due to some machine failure problems. The participants without eye or hand data recorded should be omitted from the analysis, since the aim is to understand the correlation between eye and hand data as pair. Hence, a total of 15 out of 28 surgical residents (10 doctors and 5 interns) from the Department of Neurosurgery (6 participants) or Otolaryngology (ENT) (4 participants) from Hacettepe Medical School in Ankara, Turkey considered for this study.

Among these participants, 10 (1 female) were novices whose average age was 25.6 (SD = 3.62) and worked as research assistants in the neurosurgery and ENT departments. On average, they had observed around 8 (SD = 10.17) and assisted in 1 (SD = 13.48) surgeries. In the other group, however, 5 participants (1 female) were intermediates whose average age was 28.40 (SD = 1.52) and who had observed 52 (SD = 32.71) and assisted in around 40 (SD = 29.12) surgeries. On average, the intermediate group had performed around 24 (SD = 20.38) operations as surgeons.

Detailed information about these participants according to their skill level is given in Table 5.1.

Table 5. 1 Information about Participants in Eye-Hand Coordination Research

		Gender		Endoscopic Surgical Expertise		
Skill Level	**Age**	**F**	**M**	**Observed**	**Assisted**	**Performed**
Intermediate	28.4	1	4	52.0	39.6	23.8
Novice	25.6	1	9	8.2	1.0	0.0

5.2.4 Procedure

In this research, the research procedure mainly consisted of two stages. First, an experimental study was conducted in a laboratory with the participants. Performance (i.e., time, distance, camera distance and accuracy), eye-gaze and the hand movement data of the participants were recorded while performing the tasks.

After that, the recorded data is given as an input to the ‘simulated’ version of these scenarios using Unity. In order to simulate those, scenarios using a camera and a tool should be considered, where we can observe the participants’ performance through the camera. The procedure for this chapter is shown in Figure 5.1.

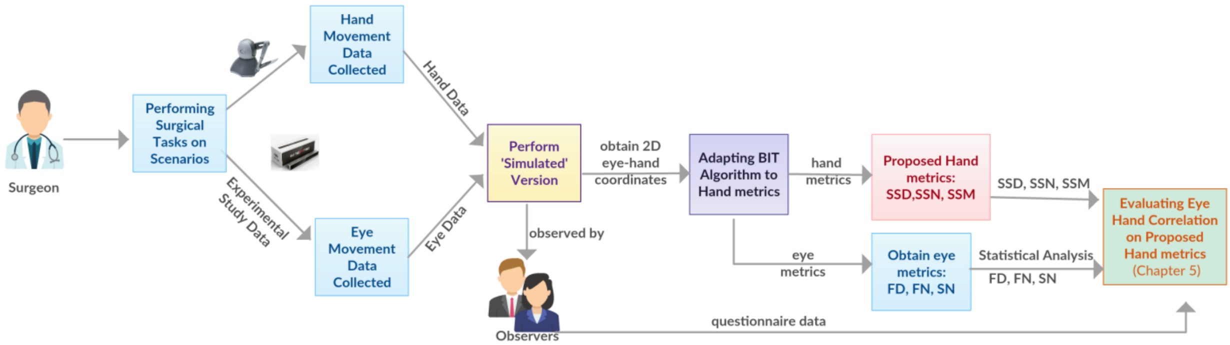


Figure 5. 1 Procedure for the Analysis on Proposed Hand Metrics

Experimental Study in Laboratory

At the beginning of the experiment, the participants were asked to fill out a questionnaire including their demographic information, dominant-hand and experience level (i.e., years in the department, number of operations observed, assisted, and performed). After that, a brief instructional video was shown and briefing was provided about the procedure. Each participant was asked to perform these scenarios, using both their dominant and non-dominant hands. Specifically in both-hand scenarios, it is aimed to improve eye-hand coordination skills. Along with the performance data, the eye-gaze data (i.e., pupil size, fixation, raw and smoothed X, Y coordinates of both left and right eye) and hand movement data (i.e., tool and camera position, tool and camera rotation as 3D vectors) were collected during experimental study and stored.

Simulated Performance of Participants

The performance of each participant in both Scenario 1 and 3 are simulated in Unity environment in order to observe their performances individually and with focus on eye-hand coordination. For such approach scenarios under both-hand conditions, using a tool and a camera is appropriate since the observer is watching the performance through the view of the camera. This is the reason of choosing these two scenarios out of four. With the help of this simulation, the operation is monitored. A blue pointer is used – point at the end of the surgical tool - showing the hand movement of the participant which denotes the current location of the tool in 3D simulation environment and the 'eye' icon showing the point on the screen that the participant currently looking (Figure 5.2).

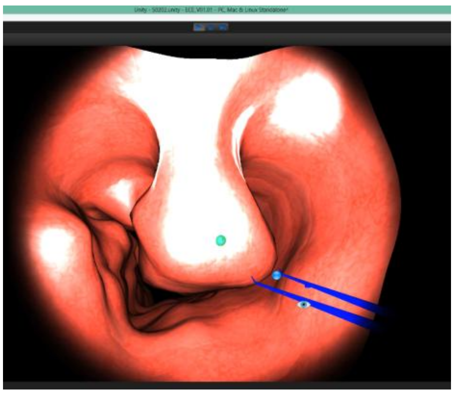


Figure 5. 2 Simulated Performance in Scenario-3

Additionally, the simulated performance enables observers to identify tissue-contact, left-right hand coordination and the eye-hand coordination of participants. In this research design, the observation data were gathered through questionnaire (see Appendix C - D) to understand such behavior differences between novice and intermediate groups. Five researchers, all of whom are graduate students in the field of engineering, monitor the participants' performances as observers in Scenario-3, which is an environment similar to the operational procedures. During the experiment, eye-gaze coordinates were gathered in a top-left oriented 2D coordinate system. The screen resolution is 1920 x 1080 pixels, in other words, the horizontal field of view (x-coordinate) is 1920 pixels, whereas the vertical field of view (y-coordinate) is 1080 pixels. The field of view (FOV) of the camera, from the left-perspective for Scenario-3 can be seen in Figure 5.3.

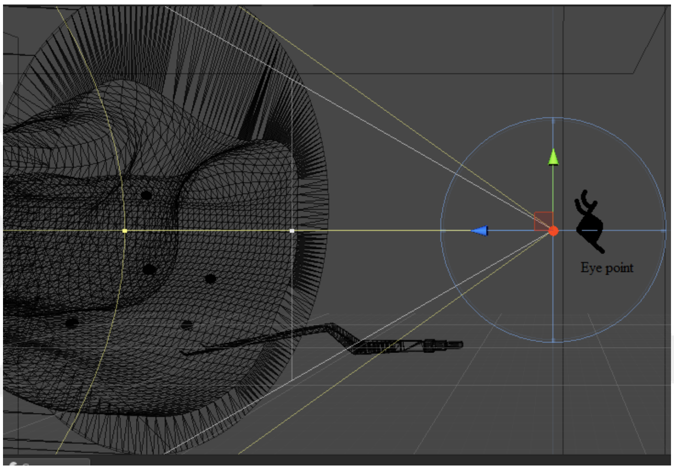


Figure 5. 3 Camera's FOV in Simulation Environment

However, hand movement coordinates, regarded as the position of tool and camera in the scenarios, were represented as 3D vectors. The origins for eye O_{eye} and for hand O_{hand} coordinates have been represented in a 2D scene and appear in Figure 5.4-A and B.

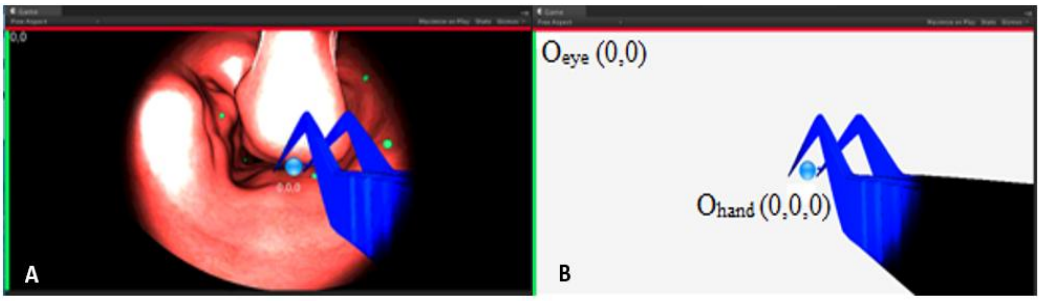


Figure 5. 4 Eye-Gaze and Hand Movement Coordinates in Scenario-3

Hence, in the simulated version, a conversion of hand data from a 3D vector to 2D on-screen coordinates is needed. This conversion is done using 'world point' to 'screen point' transformation in Unity environment, where, the bottom-left of the screen is represented as origin (0, 0) and the right-top coordinate is (1920, 1080) in 2D environment.

Using this approach, the raw coordinates for both eye-gaze and hand movement data were obtained as the output of the simulated version of the scenarios, which is then given as input to the BIT Algorithm. Then, the classification process of the eye and hand movements as either fixation or saccade events were identified by running BIT Algorithm, separately for the eye data and the hand data. The output of the BIT algorithm performed on the eye data is the FD, FN and SN metrics, whereas SSD, SSN and SSM metrics were obtained as the outputs from the hand data.

5.3 Results

The results of this study are provided in three sections. In the first one, the eye-hand correlation results for scenario-1 and in the second section for Scenario-3 are presented according to the novice and intermediate participants. In the last section, analysis of the questionnaire data is presented. A correlation analysis is performed to assess the relationship between eye-gaze and hand movement metrics considering three pairs: the fixation duration of the eye-gaze (FD) and stand still duration of the hand movement (SSD), the fixation number of the eye-gaze (FN) and stand still number of the hand movement (SSN), and saccade number of the eye-gaze (SN) and sudden-sharp movement (SSM) for the hand-movement, where the correlation coefficient r is calculated. The value of $|r|$ from 0.1 to 0.3 represents a small correlation, from 0.3 to 0.5 represents a moderate correlation, and larger than 0.5 shows a strong correlation as reported by Cohen (Cohen, 1988; Laerd Statistics, 2017).

5.3.1 Eye- Hand Correlation Results for Scenario-1

The results of eye-hand correlation for the two groups (intermediate and novice) of participants in Scenario-1 are given in Table 5.2 for both the eye metrics (FD, FN, and SN) and the hand metrics (SSD, SSM and SSN).

Table 5. 2 Descriptive Statistics for Eye-Hand Measures in Scenario-1

Skill Levels	Eye Metrics						Hand Metrics					
	FD		FN		SN		SSD		SSN		SSM	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Intermediate	121.53	19.05	12.15	1.91	20.80	13.46	92.99	10.25	9.30	1.02	55.80	37.48
Novice	112.70	26.55	11.26	2.65	28.60	32.46	101.99	15.04	10.19	1.50	55.10	15.82

Eye-Hand Correlation for Intermediates

A Pearson's product-moment correlation was run to assess the relationship between eye-gaze and hand movement metrics, (FD- SSD and FN- SSN) and saccades (SN-SSM) for intermediates. For all of these three pairs, the preliminary analyses showed the relationship to be linear with both variables normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there were no outliers. Also, there was a strong negative correlation between FD – SSD and FN- SSN metrics in intermediates, $r = -.836$ and $r = -.837$, respectively. On the other hand, a strong positive correlation existed between saccade measures SN and SSM in the intermediate group, $r = .755$ (Table 3).

Eye-Hand Correlation for Novices

A Pearson's product-moment correlation was run to assess the relationship between eye-gaze and hand movement metrics, and fixations (FD- SSD and FN- SSN) and saccades (SN- SSM) for novices. For the first two pairs related to fixation, preliminary analyses showed the relationship to be linear with both variables normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there were no outliers. There was a moderate positive correlation for both FD- SSD and FN- SSN metrics among the novice participants, $r = .448$. However, not all variables were normally distributed for saccade metrics SN and SSM, as assessed by Shapiro-Wilk's test ($p < .05$). Accordingly, a Spearman's rank-order correlation was run to assess the relationship between the saccade number of eye-gaze data and the sudden sharp movements of hand data. The results show a strong positive correlation for SN and SSM metrics, $r_s = .590$ (Table 5.3).

Table 5. 3 Eye- Hand Correlation Results for Scenario-1

	Eye- Hand Metrics		
Skill Level	**FD - SSD**	**FN - SSN**	**SN - SSM**
Intermediate	-.836 *Strong negative*	-.837 *Strong negative*	.755 *Strong positive*
Novice	.448 *Moderate positive*	.448 *Moderate positive*	.590 *Strong positive*

5.3.2 Eye- Hand Correlation Results for Scenario-3

In what follows, the results of eye-hand correlation are provided for the two groups in Scenario-3. Descriptive statistics for eye metrics (FD, FN, and SN) and hand metrics (SSD, SSM, SSN) also appear in Table 5.4.

Table 5. 4 Descriptive Statistics for Eye- Hand metrics in Scenario-3

	Eye Metrics						**Hand Metrics**					
	FD		**FN**		**SN**		**SSD**		**SSN**		**SSM**	
Skill Levels	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Intermediate	151.25	30.48	15.12	3.04	41.40	23.58	121.37	16.60	12.15	1.65	395.00	143.63
Novice	133.42	42.49	13.34	4.25	91.90	82.30	118.28	15.13	11.84	1.52	486.00	143.86

Eye-Hand Correlation for Intermediates

A Spearman's rank-order correlation was run to assess the relationship between FD-SSD and FN- SSN metrics since not all variables were normally distributed, as assessed by Shapiro-Wilk's test ($p < .05$). There was a strong negative correlation for the pairs related to fixation, FD- SSD and FN- SSN metrics. In other words, an increase in eye fixation duration and fixation number was strongly correlated with a decrease in the hand stand still duration and stand still number in intermediates, $r_s = -.900$, $p < .05$ (Table 5). However, both variables of saccade metrics, SN and SSM, were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). Pearson's correlation was run to assess the relationship between the saccade number of eye-gaze data and sudden sharp movements of hand data. There was also a strong positive correlation for SN and SSM metrics, $r = .846$ (Table 5).

Eye-Hand Correlation for Novices

A Pearson's product-moment correlation was run to assess the relationship between eye-gaze and hand movement metrics, the fixation (FD- SSD and FN- SSN) and saccades (SN- SSM) for novices. For the first two pairs related to fixation, preliminary analyses showed the relationship to be linear with both variables normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there were no outliers. There was a moderate negative correlation for both FD- SSD and FN- SSN metrics in novices, $r = -.443$ and $-.441$, respectively. However, not all variables were normally distributed for saccade metrics SN and SSM, as assessed by Shapiro-Wilk's test ($p < .05$). Accordingly, a Spearman's rank-order correlation was run to assess the relationship between the saccade number of eye-gaze data and sudden sharp movements of hand data. The results show a small positive correlation for SN and SSM metrics for novices, $r_s = .06$ (Table 5.5).

Table 5. 5 Eye- Hand Correlation Results for Scenario-3

	Eye- Hand Metrics		
Skill Level	**FD - SSD**	**FN – SSN**	**SN - SSM**
Intermediate	-.900* *Strong negative*	-.900* *Strong negative*	.846 *Strong positive*
Novice	-.443 *Moderate negative*	-.441 *Moderate negative*	.06 *Small positive*

**correlation is significant at the 0.05 level*

5.3.3 Analyzing the Questionnaire Data

Five researchers (1 female), all of whom are graduate students in the field of engineering; monitor the participants' performances as observers in Scenario-3 in order to assess the left-right hand coordination and eye-hand coordination. The items seen in Table 6 were asked to the participants to answer which they should choose one of the five alternatives (1: Strongly Disagree, 5: Strongly agree) in a likertscale-type questionnaire. 11 out of 15 participants (3 intermediates, 8 novices) were evaluated using the questionnaire data. The descriptive results appear in Table 5.6.

Table 5. 6 Descriptive Results for Questionnaire Analysis of the Observers

	Intermediate		Novice	
Questionnaire Item	M	SD	M	SD
Participant shows developed depth perception skills in a 3D environment	3.33	.42	2.12	.34
Participant shows developed skills on hand-eye coordination	3.53	.30	1.92	.52

A Mann-Whitney U test was run to determine if there were any differences in the scores for the given expressions (see Table 6) between intermediate and novice groups. The distributions of the scores for intermediates and novices were not similar among all and as assessed through visual inspection. According to the results, considering these expressions, the 3D depth perception and eye-hand coordination skills of intermediates (mean rank = 10.00) were significantly higher than novices (mean rank = 4.50), U=0, z= -2.461, p=.012.

5.4 Discussions and Conclusions

The findings of this chapter are discussed below, under three sections based on fixation and saccade metrics regarding to hand movements.

Eye-hand Coordination related to Fixation Metrics in Scenario-1:

The results of this study show that all three eye-gaze and hand movement metrics are strongly correlated for the intermediates, indicating that their eye-hand coordination skills are improved. On the other hand, fixation metrics (FD-SSD and FN-SSN) are moderately correlated for novices, which mean that they require eye-hand skills improvements. In other words, the intermediate participants' eye-hand coordination skills are better improved compared to the novices. These results also show that in Scenario-1, when the fixation duration (FD) and fixation number (FN) of intermediates increase, their stand still duration (SSD) and stand still number (SSN) decreases. The increase in the fixation duration can be observed because of an increase in their concentration while performing the tasks. In such cases, the stand still duration regarding the idle state of hand movements decreases, resulting in serial movements without discontinuation. In other words when the concentration of experts' increases, their hand movements become smoother leading to fewer occurrences of stand-still.

Different from intermediates, in Scenario-1, an increase in the fixation duration (FD) and fixation number (FN) of eye-gaze among the novice participants correlates with an increase in their hand movement metrics, stand still duration (SSD) and stand still number (SSN). At this stage, it can be inferred that the novices' hand movements become more stable when their fixation of eye increases. This may be because of their eye-gaze behaviors, which is reported in the previous studies as the experts maintained eye gaze on the target while manipulating the tool, whereas novices tracked the movement of the tool until it reached the target (Law et al., 2004).

Eye-hand Coordination related to Fixation Metrics in Scenario-3:

In Scenario-3, a significantly strong negative correlation is found between eye fixations (FD and FN) and hand movements (SSN and SSD) of intermediates; whereas this was strong negative in Scenario-1, implying that their hand-eye coordination is improved in Scenario-3. This I can be because of their learning through the scenarios or the higher fidelity level of Scenario-3 and, as such, remains to be studied further. On the other hand, novices performed better in terms of eye-hand correlation, where the strength of the correlation remains the same as moderate, but the direction changes from positive to negative. This result indicates that when the eye fixation increases, their hand movements also become smoother. This is an indication that their hand-eye movement coordination is improved in Scenario-2 after practicing in Scenario -1.

Eye-hand Coordination related to Saccade Metric:

A strong correlation occurs between SN and SSM in Scenario-1 and 3, where in Scenario-3 its correlation coefficient is slightly increased for both groups. This result shows that, when there is a saccade in their eye movements, a sudden sharp movement also occurs in their hand movements. However, the results show that, in Scenario-3, for the novice group a considerably smaller correlation exists, which means lower correlation between their saccade numbers (SN) and sudden sharp movement (SSM). This may also show the need of their hand-eye skill improvements. Similarly, the results obtained based on the questionnaire data considering the observers’ evaluations on the skill levels of the intermediates in terms of their 3D depth perception, eye- hand coordination and left-right hand coordination are reported as higher than that of the novices.

CHAPTER 6

ESTIMATING SKILL LEVELS USING CLASSIFICATION ALGORITHMS THROUGH HAND MOVEMENT METRICS

Recently, the use of simulation techniques in medical education is an emerging topic in surgical training process. During this training, assessing skill levels of trainees is a critical issue for better evaluation and guidance. Previous studies reported that the assessment methods were subjective in the field and MIS requires specific surgical skills to be assessed objectively (Cagiltay & Berker, 2018; M. Chmarra, Grimbergen, et al., 2007). Hence, there is an increasing need for cost-effective methods of objective evaluation of skill levels. Accordingly, several metrics have been proposed for objective skill assessment that can be used through box trainers or virtual reality simulators. Based on the categorization presented by Fried and Feldman, metrics have been classified into 2 groups as efficiency and quality metrics (Fried & Feldman, 2008; Oropesa et al., 2011). Efficiency metrics were obtained by the use of tracking devices, including motion-derived and force-derived metrics (Oropesa et al., 2011). It is also reported that mostly used motion efficiency metrics were time, path length and economy of movement, but sometimes there is no significant difference observed due to the changing factors such as task difficulty (R Aggarwal et al., 2009; Eriksen & Grantcharov, 2005; Oropesa et al., 2011; Sánchez-Peralta et al., 2010; A. Zhang, Hünerbein, Dai, Schlag, & Beller, 2008). Hence, speed (Eriksen & Grantcharov, 2005; Yamaguchi et al., 2007; A. Zhang et al., 2008) and motion smoothness (Pellen et al., 2009; Sokollik, Gross, & Buess, 2004; Van Sickle, McClusky III, Gallagher, & Smith, 2005) metrics were also proposed as a discriminator factor of such metrics (Oropesa et al., 2011). Other motion-related metrics are reported as depth, angular area, volume and spatial perception, which are rarely considered in studies due to difficult validation (Oropesa et al., 2011). These studies show that, by using motion analysis, an improvement during training can be occurred (Sokollik et al., 2004; Torkington, Smith, Rees, & Darzi, 2001). It is also reported that basic psychomotor skills can be assessed by motion analysis (M. Chmarra, Grimbergen, et al., 2007).

6.1 Related Work

Based on the findings in the literature, there are several studies providing solutions for automating the objective assessment of surgical expertise using hand-related metrics by performing classification. For instance, Watson (2014) hypothesized that machine learning approach can be used to improve prediction of surgical expertise using hand motion patterns of two groups with different levels of laparoscopic proficiency, experts and novices (Robert A Watson, 2014). It is reported that, 14 experts and 10 surgical residents as novices participated in their study, and based on their hand motion patterns, using SVM method the accuracy of 83% , and Lempel–Ziv (LZ) metric the accuracy of 70% was obtained (Lempel & Ziv, 1976; Robert A Watson, 2014). Allen et al. (2010) also used SVM method to classify 4 experts and 26 novices, based on the 3D position and orientation of the tool. Four metrics were gathered such as time, path length, volume and applied force while performing three surgical tasks. The average accuracy for these three tasks were reported as 91.6% (Allen et al., 2010).

A previous study conducted by Overby and Watson (2014) also investigates the hand-motion pattern differences between Fundamentals of Laparoscopic Surgery (FLS) certified and non-certified surgeons, where the hand motion data collected through micro-electromechanical gyroscope tracking devices worn on both hands during a laparoscopic procedure (Overby & Watson, 2014). The results of that study indicates that the complexity of hand-motion patterns increased with higher surgical experience or grade (Overby & Watson, 2014), added that experts use simpler subtask motifs in more complex and dense patterns (Overby & Watson, 2014; Robert Anthony Watson, 2012, 2013).

Another study conducted by Chmarra et al. (2010) uses Linear Discriminant Analysis (LDA) method to classify residents' level as belonging to one of the three groups; experienced, intermediates and novices, based on psychomotor laparoscopic skills, using six motion analysis metrics: were total time, path length, depth, motion smoothness, angular area, and volume (M. K. Chmarra et al., 2010). The accuracy of 74% obtained in that study (M. K. Chmarra et al., 2010). Varadarajan et al (2009) also uses LDA method based on Hidden Markov Models (HMM) analysis, to differentiate between 8 surgeons as experts, intermediates and novices using kinematic data with 78 motion variables (Varadarajan, Reiley, Lin, Khudanpur, & Hager, 2009). The results show an accuracy of 87% (Varadarajan et al., 2009).

6.2 Material and Methods

In this study, hand movement metrics (i.e. distance, velocity, angular velocity of tool and camera, presented in Chapter 4 and SSD, SSN, SSM presented in Chapter 5) in Scenario-1 and 3, were used in order to estimate the skill levels. The research procedure for this chapter is given in Figure 6.1. Three main stages of classification performed in the given order: feature extraction, feature selection and classifier evaluation to find the best accuracy. In the first stage, the features were extracted using hand movement metrics and, basic statistical functions were used to calculate some features for each task in the scenarios. Accordingly, a total of 101 features extracted for both scenarios using MATLAB. After that, one of a well-known feature selection method, ReliefF, was implemented to improve the performance of classifiers and the accuracy of the classification process. As a result of this process, 34 features were selected for Scenario-1, whereas 42 features were selected for Scenario-3, as seen in Table 6.1.

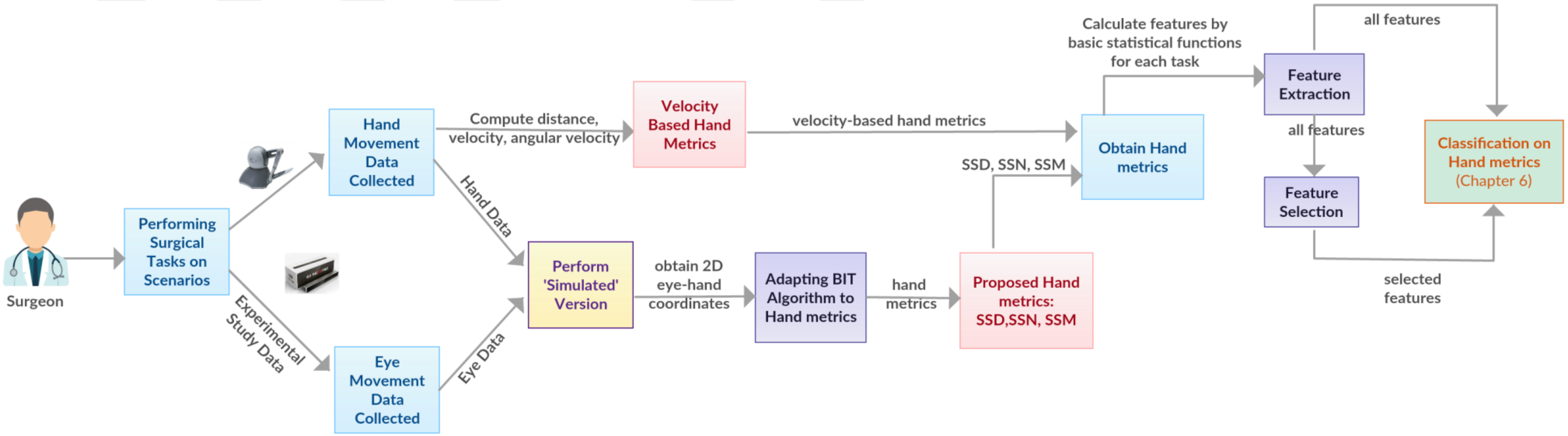


Figure 6. 1 Procedure for Classification on Proposed Hand Metrics

Table 6. 1 Selected Hand Features Set for Scenarios

Scenario	Selected Feature Set
I	Distance_1, 3, 4, max, std, var, Camera distance_ 3, 5, 6, 9, 10, median, min, Tool velocity_1, 2, 7, 10, min, Camera velocity_1, 5, 6, 9, 10, max, mean, median, min, sum, Camera angular velocity_max, median, SSD_median, sum, SSN_ median, sum
III	Dominant_hand, Distance_6,7,9, max, min, std, var, Camera distance_4,6,9,10,mean, median, min, sum, Tool velocity_6,7, std, var, Tool angular velocity_mean, Camera velocity_1, 4, 5, 6, 7, 9, 10, max, mean, median, min, sum, Camera angular velocity_ min, sum, SSN_ max, mean, median, std, var, SSM_mean

In Table 6.1, selected features were presented with related task id's and statistical functions. For instance, 'Distance_1, 3, 4, max, std, var' representing the features of tool distances regarding to the tasks 1, 3, 4, maximum tool distance, standard deviation and variance of the tool distances among a total of ten tasks.

Finally, classifiers were trained and evaluated using 10-fold cross-validation. The feature selection (using three different methods) and classification processes (using a total of 12 algorithms) were performed by using WEKA workbench (Witten et al., 2016). Data is collected from 28 and 26 participants (2 participants hand movement data cannot be recorded for both-hand condition in Scenario-3, due to a machine failure), while performing on Scenario-1 and 3, respectively.

6.3 Results

The results of this study are provided in two sections. In the first section, the classification results based on hand metrics for Scenario-1 and in Scenario-3 are presented according to the novice and intermediate participants. The detailed classification results for each scenario were given in Appendix E. In the second section, analysis of the questionnaire data is presented.

6.3.1 Results for Estimating Skill Levels on Hand Metrics

In Scenario-1, the best accuracy result for estimating the skill levels of the participants was obtained with an accuracy of 85.71% (24/28) with $KNN_{(k=1)}$ algorithm, whereas an accuracy of 80.76% (21/ 26) was obtained in Scenario-3 with $KNN_{(k=3)}$ algorithm, as seen in Table 6.2.

Table 6. 2 Classification Results based on Hand Features

	Best Accuracy (%)	
Scenario	All Features	Selected Features
I	78.57 *AdaBoost*	85.71 $KNN_{k=1}$
III	76.92 $KNN_{k=3}$	80.76 $KNN_{k=3}$

6.3.2 Analyzing the Questionnaire Data

Five researchers (1 female), all of whom are graduate students in the field of engineering, monitor the participants' performances as observers in Scenario-3. In order to assess participants' skill levels, their performances considering four skills, such as respect for the tissue, economy of movement, flow of operation and spatial orientation were evaluated which they should choose one of the five alternatives (from 0 to 4) in a questionnaire (See Appendix D, Part III). The questions were taken from the rating scale used in Adrales et al.'s study (Adrales et al., 2003). In this study, 11 out of 28 participants (3 intermediates, 8 novices) were selected and evaluated using the questionnaire data. The descriptive results appear in Table 6.3.

Table 6. 3 Descriptives for Participants' Skill Level Assessment in Questionnaire

	Intermediate		**Novice**	
Questionnaire Item	**M**	**SD**	**M**	**SD**
Clinical Judgement- Respect for Tissue	1.53	.11	.45	.32
Dexterity (economy of movement)	1.93	.30	.47	.32
Serial /Simultaneous Complexity (Flow of operation)	1.53	.30	.60	.40
Spatial Orientation	1.53	.30	.35	.36

A Mann-Whitney U test was run to determine if there were any differences in the scores for the given expressions (see Table 6.3) between intermediate and novice groups. The distributions of the scores for intermediates and novices were not similar among all and as assessed through visual inspection. According to the results, considering the respect for tissue, economy of movement and spatial orientation skills, intermediates (mean rank = 10.00) performed significantly better than novices (mean rank = 4.50), p=.012. Similarly, for the skill ‘serial/ simultaneous complexity’ representing the flow of operation, intermediates (mean rank = 9.83) performed better than novices (mean rank = 4.56), U=.5, z = -2.380, p=.017.

6.4 Conclusions

The results of this study show that, by using classification algorithms, through hand movement metrics collected by computer-simulated virtual surgical training environments with haptic interfaces, it is possible to estimate the intermediate and novice surgical skills with 81-86%. Using these metrics, the accuracies have been improved when compared to the results of classification based on performance metrics presented in Chapter 3. Also this result slightly improved the accuracy of 83% presented in a previous study using SVM algorithm (Robert A Watson, 2014). Hence, it can be concluded that metrics relying on hand behaviors of surgical residents can provide better insights about their skill levels in surgical education programs.

Additionally, based on the observers’ evaluations in questionnaire data analysis, there occurs a significant difference between the intermediate and novice surgeons’ skills related to their hand movement behaviors such as respect for the tissue, economy of movement, flow of operation and spatial orientation. The results indicate that intermediates were more respectful to the tissue, with higher economy of movement and more serial flow of operation and spatial orientation, without getting lost while performing the tasks in the simulation environment compared to the novices.

As a conclusion, hand movement metrics collected through simulation-based environments can be used for objectively assessing the skill levels of trainees in skill-based surgical education.

CHAPTER 7

DISCUSSIONS AND CONCLUSION

In this study, in order to automatically assess the skill levels of surgical residents through computer-simulation software environments, new features named the “stand still” metric which reflects the period when the hand movement remains within a very small range and lower velocity for some time. This metric is calculated as ‘*Stand Still Duration*’ (SSD) and ‘*Stand Still Number*’ (SSN) for hand movements. Additionally, the ‘*Sudden Sharp Movement’ (SSM)* metric is also proposed to identify very fast, sharp hand movements while performing any given task and calculated through BIT algorithm. The classification of the participants as novice and intermediate is performed on four scenarios by using performance data, velocity based metrics and proposed hand metrics. The summary of the results are provided in Table 7.1. The calculations for the proposed hand metrics cannot be performed for the scenarios 2 and 4. As the design of these scenarios do not include an operation tool, tracking the dominant hand tool position was not possible for these scenarios. Additionally, the calculations are only implemented for both-hand condition for the proposed metrics.

Table 7. 1 Summary of Classification Results in Both Hand Condition

Scenarios	Performance Metrics	Velocity Metrics	Velocity and Proposed Hand Metrics
I	78.57	75.00	85.71
II	75.00	88.50	-
III	76.92	75.00	80.76
IV	78.57	91.70	-

As seen from Table 7.1, by including the velocity based metrics and proposed hand metrics the accuracy of the estimation is improved from 79 to 86 for Scenario-1 and from 71 to 81 for Scenario-3. These accuracies for classifying intermediate and novice skill levels can be considered as successful when compared to the similar previous studies classifying expert and novice skill levels as given in Table 7.2.

Table 7. 2 Classification Results of Previous Studies

Accuracy	Reference	Method	# of Participants	# of Skill Levels	Skill Levels
83.0	(Robert A Watson, 2014)	SVM	24	2	Expert, novice
70.0	(Robert A Watson, 2014)	LZ-Metric	24	2	Expert, novice
91.6	(Allen et al., 2010)	SVM	30	2	Expert, novice
80.3	(Allen et al., 2010)	Z-Score Norm.	30	2	Expert, novice
74.0	(M. K. Chmarra et al., 2010)	LDA	31	3	Expert, intermediate, novice

SVM: Support Vector Machine; LZ: Lempel–Ziv metric; Z-score Normalization; LDA:Linear Discriminant Analysis method

These findings imply that, in future, by including experts in such analyses, these accuracies may be even improved since differentiating between experts and novices are easier compared to distinguish intermediate and novices.

Another finding of our study is that the intermediate-level participants have a higher degree of eye-hand coordination skills compared to the novices. The increase in intermediate level participants' visual concentration leads to smoother hand movements. In similar situations, the novices' hand movements have become more stable with the increasing fixation of eye. For both intermediate and novice participants, when there is a saccade in their eye movements, a sudden sharp movement also occurs in their hand movements. Additionally, the results also imply that after the first round of practice, all participants' eye-hand coordination skills are improved.

The major contribution of this thesis is evaluating the impact of hand movement metrics on objective surgical skill assessment. For this purpose, new hand movement metrics are proposed by adapting an open source eye-movement classification BIT algorithm in computer-based simulation software through haptic interface. It can be inferred from the literature findings that hand movements of surgical residents were tracked by using wearable devices such as instrumented gloves including sensors. However, in our study, the hand movement data were gathered using haptic devices in a surgical simulation environment, which may lead to a more realistic practice in training and assessment.

As a conclusion, this study show that by using the velocity-based features with the performance data and the proposed hand metrics, the accuracy of the classification can be improved. The researchers believe that by including the eye-movements events to these results it can be even more improved. These results are very promising that in the future, the computer-based simulation software can adapt the classification algorithms to continuously assess the skill levels of the endoscopic surgery trainees and adapt the content of the training material and the provided order according to their individual requirements. Such an approach would help the educators to better integrate these tools in their training programs and to get support on the assessment process of the trainees.

CHAPTER 8

LIMITATIONS AND FUTURE WORK

This study was conducted with 28 participants because, in general, the number of surgeons in the neurosurgery and ENT departments is very limited. However, in future, it may be possible to validate the results of this study by a larger number of participants, having different skill levels. Additionally, in this study, the number of scenarios used in the analysis is limited due to the requirement of considering only both-handed scenarios including a surgical tool and a camera. Hence, in the future, the experimental studies should be conducted with more scenarios, having different difficulty levels in a surgical training curriculum. Finally, the results of this study can be further validated by other additional measures such as the applied forces on the haptic device used in the simulators, in order to increase the classification accuracy of surgical skill levels.

Recently, using the machine learning techniques in surgical training constitutes a new area of research. This current study provides a surgical simulation system infrastructure to gather the hand movements' data of experts obtained by special tactile devices. Accordingly, an intelligent support system can be created by examining the hand motion patterns of experienced surgeons to better guide novices in a surgical education program. In this manner, it is also possible to monitor trainees' progress in the education program individually. As a future work, by using these techniques, cost-efficient educational tools can be developed in order to improve education efficiency and patient safety.

REFERENCES

Adrales, G. L., Park, A. E., Chu, U. B., Witzke, D. B., Donnelly, M. B., Hoskins, J. D., et al. (2003). A valid method of laparoscopic simulation training and competence assessment1, 2. *journal of surgical research, 114*(2), 156-162.

Agarwal, S., Pandey, G., & Tiwari, M. (2012). Data mining in education: data classification and decision tree approach. *International Journal of e-Education, e-Business, e-Management and e-Learning, 2*(2), 140.

Aggarwal, R., Crochet, P., Dias, A., Misra, A., Ziprin, P., & Darzi, A. (2009). Development of a virtual reality training curriculum for laparoscopic cholecystectomy. *British Journal of Surgery, 96*(9), 1086-1093.

Aggarwal, R., Grantcharov, T., Moorthy, K., Hance, J., & Darzi, A. (2006). A competency-based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill. *The American journal of surgery, 191*(1), 128-133.

Ahmed, B. A., Connolly, B. L., Shroff, P., Chong, A. L., Gordon, C., Grant, R., et al. (2010). Cumulative effective doses from radiologic procedures for pediatric oncology patients. *Pediatrics, 126*(4), e851-e858.

Allen, B., Nistor, V., Dutson, E., Carman, G., Lewis, C., & Faloutsos, P. (2010). Support vector machines improve the accuracy of evaluation for the performance of laparoscopic training tasks. *Surgical endoscopy, 24*(1), 170.

Andersen, D. K. (2012). How can educators use simulation applications to teach and assess surgical judgment? *Academic Medicine, 87*(7), 934-941.

Atkeson, C. G., Moore, A. W., & Schaal, S. (1997). Locally weighted learning for control *Lazy learning* (pp. 75-113): Springer.

Ayodeji, I., Schijven, M., Jakimowicz, J., & Greve, J. (2007). Face validation of the Simbionix LAP Mentor virtual reality training module and its applicability in the surgical curriculum. *Surgical endoscopy, 21*(9), 1641-1649.

Bajka, B. H., Clarke, J. M., Topping, D. L., Cobiac, L., Abeywardena, M. Y., & Patten, G. S. (2010). Butyrylated starch increases large bowel butyrate levels and lowers colonic smooth muscle contractility in rats. *Nutrition research, 30*(6), 427-434.

Balik, E., Asoglu, O., Saglam, S., Yamaner, S., Akyuz, A., Buyukuncu, Y., et al. (2010). Effects of surgical laparoscopic experience on the short-term postoperative outcome of rectal cancer: results of a high volume single center institution. *Surgical Laparoscopy Endoscopy & Percutaneous Techniques, 20*(2), 93-99.

Basdogan, C., De, S., Kim, J., Muniyandi, M., Kim, H., & Srinivasan, M. A. (2004). Haptics in minimally invasive surgical simulation and training. *IEEE computer graphics and applications, 24*(2), 56-64.

Batmaz, A. U., de Mathelin, M., & Dresp-Langley, B. (2017). Seeing virtual while acting real: Visual display and strategy effects on the time and precision of eye-hand coordination. *PloS one, 12*(8), e0183789.

Bauernschmitt, R., Braun, E. U., Knoll, A., Mayer, H., Nagy, I., & Lange, R. (2006). *Implementation of force feedback into telemanipulated surgery: Assessment of surgical experience.* Paper presented at the Computers in Cardiology, 2006.

Bhat, A. S., Amith, V., Prasad, N. S., & Mohan, D. M. (2014). *An efficient classification algorithm for music mood detection in Western and Hindi music using audio feature extraction.* Paper presented at the Signal and Image Processing (ICSIP), 2014 Fifth International Conference on.

Bittner, J. G., Mellinger, J. D., Imam, T., Schade, R. R., & MacFadyen, B. V. (2010). Face and construct validity of a computer-based virtual reality simulator for ERCP. *Gastrointestinal endoscopy, 71*(2), 357-364.

Cagiltay, N. E., & Berker, M. (2018). Technology-enhanced surgical education: attitudes and perceptions of the endoscopic surgery community in Turkey. *BMJ Simulation and Technology Enhanced Learning*, bmjstel-2017-000238.

Cagiltay, N. E., Ozcelik, E., Sengul, G., & Berker, M. (2017). Construct and face validity of the educational computer-based environment (ECE) assessment scenarios for basic endoneurosurgery skills. *Surgical endoscopy, 31*(11), 4485-4495.

Cao, C., MacKenzie, C., & Payandeh, S. (1996). *Task and motion analyses in endoscopic surgery.* Paper presented at the Proceedings ASME Dynamic Systems and Control Division.

Chan, J. C.-W., & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment, 112*(6), 2999-3011.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST), 2*(3), 27.

Chmarra, M., Grimbergen, C., & Dankelman, J. (2007). Systems for tracking minimally invasive surgical instruments. *Minimally Invasive Therapy & Allied Technologies, 16*(6), 328-340.

Chmarra, M., Kolkman, W., Jansen, F., Grimbergen, C., & Dankelman, J. (2007). The influence of experience and camera holding on laparoscopic instrument movements measured with the TrEndo tracking system. *Surgical endoscopy, 21*(11), 2069-2075.

Chmarra, M. K., Klein, S., de Winter, J. C., Jansen, F.-W., & Dankelman, J. (2010). Objective classification of residents based on their psychomotor laparoscopic skills. *Surgical endoscopy, 24*(5), 1031-1039.

Christensen, R. (2006). *Log-linear models and logistic regression*: Springer Science & Business Media.

Cohen, J. (1988). Set correlation and contingency tables. *Applied Psychological Measurement, 12*(4), 425-434.

Colby, C., & Goldberg, M. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science, 255*(5040), 90-92.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297.

Craighead, J., Burke, J., & Murphy, R. (2008). *Using the unity game engine to develop sarge: a case study.* Paper presented at the Proceedings of the 2008

Simulation Workshop at the International Conference on Intelligent Robots and Systems (IROS 2008).
D'Angelo, A.-L. D., Rutherford, D. N., Ray, R. D., Laufer, S., Kwan, C., Cohen, E. R., et al. (2015). Idle time: an underdeveloped performance metric for assessing surgical skill. *The American journal of surgery, 209*(4), 645-651.
Dankelman, J., & Di Lorenzo, N. (2005). Surgical training and simulation: Taylor & Francis.
Dankelman, J., Grimbergen, C. K. A., & Stassen, H. G. (2007). New technologies supporting surgical intervenltions and training of surgical skills-a look at projects in europe supporting minimally invasive techniques. *IEEE Engineering in Medicine and Biology Magazine, 26*(3), 47-52.
Darzi, A., & Mackay, S. (2002). Skills assessment of surgeons. *Surgery, 131*(2), 121-124.
Darzi, A., Smith, S., & Taffinder, N. (1999). Assessing operative skill: needs to become more objective. *BMJ: British Medical Journal, 318*(7188), 887.
Datta, V., Chang, A., Mackay, S., & Darzi, A. (2002). The relationship between motion analysis and surgical technical assessments. *The American journal of surgery, 184*(1), 70-73.
Derossis, A. M., DaRosa, D. A., Dutta, S., & Dunnington, G. L. (2000). A ten-year analysis of surgical education research. *The American journal of surgery, 180*(1), 58-61.
Derossis, A. M., Fried, G. M., Abrahamowicz, M., Sigman, H. H., Barkun, J. S., & Meakins, J. L. (1998). Development of a model for training and evaluation of laparoscopic skills 1. *The American journal of surgery, 175*(6), 482-487.
Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine learning, 40*(2), 139-157.
Eivazi, S., Bednarik, R., Tukiainen, M., von und zu Fraunberg, M., Leinonen, V., & Jääskeläinen, J. E. (2012). *Gaze behaviour of expert and novice microneurosurgeons differs during observations of tumor removal recordings.* Paper presented at the Proceedings of the Symposium on Eye Tracking Research and Applications.
Eriksen, J., & Grantcharov, T. (2005). Objective assessment of laparoscopic skills using a virtual reality stimulator. *Surgical Endoscopy And Other Interventional Techniques, 19*(9), 1216-1219.
Felsher, J. J., Olesevich, M., Farres, H., Rosen, M., Fanning, A., Dunkin, B. J., et al. (2005). Validation of a flexible endoscopy simulator. *The American journal of surgery, 189*(4), 497-500.
Frank, E., Hall, M., & Pfahringer, B. (2002). *Locally weighted naive bayes.* Paper presented at the Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence.
Frank, E., & Witten, I. H. (1998). Generating accurate rule sets without global optimization.
Fried, G. M., & Feldman, L. S. (2008). Objective assessment of technical performance. *World journal of surgery, 32*(2), 156-160.
Gallagher, A. G., Richie, K., McClure, N., & McGuigan, J. (2001). Objective psychomotor skills assessment of experienced, junior, and novice

laparoscopists with virtual reality. *World journal of surgery, 25*(11), 1478-1483.
Gomoll, A. H., O'toole, R. V., Czarnecki, J., & Warner, J. J. (2007). Surgical experience correlates with performance on a virtual reality simulator for shoulder arthroscopy. *The American journal of sports medicine, 35*(6), 883-888.
Goswami, S., & Chakrabarti, A. (2014). Feature selection: A practitioner view. *International Journal of Information Technology and Computer Science (IJITCS), 6*(11), 66.
Grantcharov, T. P., Bardram, L., Funch-Jensen, P., & Rosenberg, J. (2003). Learning curves and impact of previous operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills. *The American journal of surgery, 185*(2), 146-149.
Gurusamy, K., Aggarwal, R., Palanivelu, L., & Davidson, B. (2008). Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. *British Journal of Surgery, 95*(9), 1088-1097.
Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*: Elsevier.
Helsen, W. F., Elliott, D., Starkes, J. L., & Ricker, K. L. (2000). Coupling of eye, finger, elbow, and shoulder movements during manual aiming. *Journal of motor behavior, 32*(3), 241-248.
Hernandez, J., Bann, S., Munz, Y., Moorthy, K., Datta, V., Martin, S., et al. (2004). Qualitative and quantitative analysis of the learning curve of a simulated surgical task on the da Vinci system. *Surgical Endoscopy And Other Interventional Techniques, 18*(3), 372-378.
Hilbe, J. M. (2011). Logistic regression *International Encyclopedia of Statistical Science* (pp. 755-758): Springer.
Hofstad, E. F., Våpenstad, C., Chmarra, M. K., Langø, T., Kuhry, E., & Mårvik, R. (2013). A study of psychomotor skills in minimally invasive surgery: what differentiates expert and nonexpert performance. *Surgical endoscopy, 27*(3), 854-863.
Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398): John Wiley & Sons.
Jiang, X., Zheng, B., & Atkins, M. S. (2015). Video processing to locate the tooltip position in surgical eye–hand coordination tasks. *Surgical innovation, 22*(3), 285-293.
Kaur, G., & Chhabra, A. (2014). Improved J48 classification algorithm for the prediction of diabetes. *International Journal of Computer Applications, 98*(22).
Kenney, P. A., Wszolek, M. F., Gould, J. J., Libertino, J. A., & Moinzadeh, A. (2009). Face, content, and construct validity of dV-trainer, a novel virtual reality simulator for robotic surgery. *Urology, 73*(6), 1288-1292.
Korndorffer, J. R., Clayton, J. L., Tesfay, S. T., Brunner, W. C., Sierra, R., Dunne, J. B., et al. (2005). Multicenter construct validity for southwestern laparoscopic videotrainer stations. *journal of surgical research, 128*(1), 114-119.

Kumar, S., & Sahoo, G. (2015). Classification of heart disease using naive bayes and genetic algorithm *Computational Intelligence in Data Mining-Volume 2* (pp. 269-282): Springer.

Laerd Statistics. (2017). Pearson's product-moment correlation using SPSS Statistics. Statistical tutorials and software guides. . from https://statistics.laerd.com/

Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine learning, 59*(1-2), 161-205.

Larrain, T., Bernhard, J. S., Mery, D., & Bowyer, K. W. (2017). Face recognition using sparse fingerprint classification algorithm. *IEEE Transactions on Information Forensics and Security, 12*(7), 1646-1657.

Larsen, C. R., Soerensen, J. L., Grantcharov, T. P., Dalsgaard, T., Schouenborg, L., Ottosen, C., et al. (2009). Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. *Bmj, 338*, b1802.

Latko, W. A., Armstrong, T. J., Foulke, J. A., Herrin, G. D., Rabourn, R. A., & Ulin, S. S. (1997). Development and evaluation of an observational method for assessing repetition in hand tasks. *American Industrial Hygiene Association Journal, 58*(4), 278-285.

Law, B., Atkins, M. S., Kirkpatrick, A. E., & Lomax, A. J. (2004). *Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment.* Paper presented at the Proceedings of the 2004 symposium on Eye tracking research & applications.

Lehmann, K. S., Ritz, J. P., Maass, H., Cakmak, H. K., Kuehnapfel, U. G., Germer, C. T., et al. (2005). A prospective randomized study to test the transfer of basic psychomotor skills from virtual reality to physical reality in a comparable training setting. *Annals of surgery, 241*(3), 442.

Leigh, R. J., & Kennard, C. (2004). Using saccades as a research tool in the clinical neurosciences. *Brain, 127*(3), 460-477.

Lemos, J. D., Hernandez, A. M., & Soto-Romero, G. (2017). An instrumented glove to assess manual dexterity in simulation-based neurosurgical education. *Sensors, 17*(5), 988.

Lempel, A., & Ziv, J. (1976). On the complexity of finite sequences. *IEEE Transactions on information theory, 22*(1), 75-81.

Lin, H. C., Shafran, I., Yuh, D., & Hager, G. D. (2006). Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions. *Computer Aided Surgery, 11*(5), 220-230.

Maithel, S., Villegas, L., Stylopoulos, N., Dawson, S., & Jones, D. (2005). Simulated laparoscopy using a head-mounted display vs traditional video monitor: an assessment of performance and muscle fatigue. *Surgical Endoscopy And Other Interventional Techniques, 19*(3), 406-411.

Martin, J., Regehr, G., Reznick, R., Macrae, H., Murnaghan, J., Hutchison, C., et al. (1997). Objective structured assessment of technical skill (OSATS) for surgical residents. *British Journal of Surgery, 84*(2), 273-278.

Mitsuishi, M., Tomisaki, S., Yoshidome, T., Hashizume, H., & Fujiwara, K. (2000). *Tele-micro-surgery system with intelligent user interface.* Paper presented at the Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on.

Mohamadipanah, H., Parthiban, C., Law, K., Nathwani, J., Maulson, L., DiMarco, S., et al. (2016). *Hand smoothness in laparoscopic surgery correlates to psychomotor skills in virtual reality*. Paper presented at the Wearable and Implantable Body Sensor Networks (BSN), 2016 IEEE 13th International Conference on.
Moore, M. J., & Bennett, C. L. (1995). The learning curve for laparoscopic cholecystectomy. *The American journal of surgery, 170*(1), 55-59.
Moorthy, K., Munz, Y., Sarker, S. K., & Darzi, A. (2003). Objective assessment of technical skills in surgery. *BMJ: British Medical Journal, 327*(7422), 1032.
Murphy, K. P. (2006). Naive bayes classifiers. *University of British Columbia, 18*.
Nguyen, H. A., & Choi, D. (2008). *Application of data mining to network intrusion detection: classifier selection model.* Paper presented at the Asia-Pacific Network Operations and Management Symposium.
Oropesa, I., Chmarra, M. K., Sánchez-González, P., Lamata, P., Rodrigues, S. P., Enciso, S., et al. (2013). Relevance of motion-related assessment metrics in laparoscopic surgery. *Surgical innovation, 20*(3), 299-312.
Oropesa, I., Sánchez-González, P., Chmarra, M. K., Lamata, P., Fernández, A., Sánchez-Margallo, J. A., et al. (2013). EVA: laparoscopic instrument tracking based on endoscopic video analysis for psychomotor skills assessment. *Surgical endoscopy, 27*(3), 1029-1039.
Oropesa, I., Sánchez-González, P., Chmarra, M. K., Lamata, P., Pérez-Rodríguez, R., Jansen, F. W., et al. (2014). Supervised classification of psychomotor competence in minimally invasive surgery based on instruments motion analysis. *Surgical endoscopy, 28*(2), 657-670.
Oropesa, I., Sánchez-González, P., Lamata, P., Chmarra, M. K., Pagador, J. B., Sánchez-Margallo, J. A., et al. (2011). Methods and tools for objective assessment of psychomotor skills in laparoscopic surgery. *journal of surgical research, 171*(1), e81-e95.
Overby, D. W., & Watson, R. A. (2014). Hand motion patterns of Fundamentals of Laparoscopic Surgery certified and noncertified surgeons. *The American journal of surgery, 207*(2), 226-230.
Panait, L., Bell, R. L., Roberts, K. E., & Duffy, A. J. (2008). Designing and validating a customized virtual reality-based laparoscopic skills curriculum. *Journal of surgical education, 65*(6), 413-417.
Pellen, M. G., Horgan, L. F., Barton, J. R., & Attwood, S. E. (2009). Construct validity of the ProMIS laparoscopic simulator. *Surgical endoscopy, 23*(1), 130-139.
Quinlan, J. R. (2014). *C4. 5: programs for machine learning*: Elsevier.
Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation *Encyclopedia of database systems* (pp. 532-538): Springer.
Reiley, C. E., Lin, H. C., Yuh, D. D., & Hager, G. D. (2011). Review of methods for objective surgical skill evaluation. *Surgical endoscopy, 25*(2), 356-366.
Richstone, L., Schwartz, M. J., Seideman, C., Cadeddu, J., Marshall, S., & Kavoussi, L. R. (2010). Eye metrics as an objective assessment of surgical skill. *Annals of surgery, 252*(1), 177-182.
Rish, I. (2001). *An empirical study of the naive Bayes classifier.* Paper presented at the IJCAI 2001 workshop on empirical methods in artificial intelligence.

Robb, R. A., Aharon, S., & Cameron, B. M. (1997). Patient-specific anatomic models from three dimensional medical image data for clinical applications in surgery and endoscopy. *Journal of digital imaging, 10*(1), 31-35.
Rudman, D. T., Stredney, D., Sessanna, D., Yagel, R., Crawfis, R., Heskamp, D., et al. (1998). Functional endoscopic sinus surgery training simulator. *The Laryngoscope, 108*(11), 1643-1647.
Salvucci, D. D., & Goldberg, J. H. (2000). *Identifying fixations and saccades in eye-tracking protocols.* Paper presented at the Proceedings of the 2000 symposium on Eye tracking research & applications.
Sánchez-Peralta, L. F., Sánchez-Margallo, F. M., Moyano-Cuevas, J. L., Pagador, J. B., Enciso-Sanz, S., Sánchez-González, P., et al. (2010). Construct and face validity of SINERGIA laparoscopic virtual reality simulator. *International journal of computer assisted radiology and surgery, 5*(4), 307-315.
Schout, B., Muijtjens, A. M., Hendrikx, A. J., Ananias, H. J., Dolmans, V. E., Scherpbier, A. J., et al. (2010). Acquisition of flexible cystoscopy skills on a virtual reality simulator by experts and novices. *BJU international, 105*(2), 234-239.
Schreuder, H. W., Oei, G., Maas, M., Borleffs, J. C., & Schijven, M. P. (2011). Implementation of simulation in surgical practice: minimally invasive surgery has taken the lead: the Dutch experience. *Medical teacher, 33*(2), 105-115.
Schreuder, H. W., van Dongen, K. W., Roeleveld, S. J., Schijven, M. P., & Broeders, I. A. (2009). Face and construct validity of virtual reality simulation of laparoscopic gynecologic surgery. *American Journal of Obstetrics & Gynecology, 200*(5), 540. e541-540. e548.
Scott, D. J., Young, W. N., Tesfay, S. T., Frawley, W. H., Rege, R. V., & Jones, D. B. (2001). Laparoscopic skills training. *The American journal of surgery, 182*(2), 137-142.
Shetty, S., Panait, L., Baranoski, J., Dudrick, S. J., Bell, R. L., Roberts, K. E., et al. (2012). Construct and face validity of a virtual reality–based camera navigation curriculum. *journal of surgical research, 177*(2), 191-195.
Silvennoinen, M., Helfenstein, S., Ruoranen, M., & Saariluoma, P. (2012). Learning basic surgical skills through simulator training. *Instructional Science, 40*(5), 769-783.
Silvennoinen, M., Mecklin, J.-P., Saariluoma, P., & Antikainen, T. (2009). Expertise and skill in minimally invasive surgery. *Scandinavian Journal of Surgery, 98*(4), 209-213.
Sokollik, C., Gross, J., & Buess, G. (2004). New model for skills assessment and training progress in minimally invasive surgery. *Surgical Endoscopy And Other Interventional Techniques, 18*(3), 495-500.
Srivastava, S., Youngblood, P. L., Rawn, C., Hariri, S., Heinrichs, W., & Ladd, A. L. (2004). Initial evaluation of a shoulder arthroscopy simulator: establishing construct validity. *Journal of shoulder and elbow surgery, 13*(2), 196-205.
Steinbach, M., & Tan, P.-N. (2009). kNN: k-nearest neighbors. *The top ten algorithms in data mining,* 151-162.
Stylopoulos, N., & Vosburgh, K. G. (2007). Assessing technical skill in surgery and endoscopy: A set of metrics and an algorithm (C-PASS) to assess skills in surgical and endoscopic procedures. *Surgical innovation, 14*(2), 113-121.

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Classification: Alternative techniques* (Vol. 5).
Tan, P.-N., Steinbach, M., & Kumar, V. (2006). Classification: basic concepts, decision trees, and model evaluation. *Introduction to data mining, 1*, 145-205.
The Eye Tribe. (2014). Basics. from http://theeyetribe.com/dev.theeyetribe.com/dev.theeyetribe.com/general/index.html
Thijssen, A. S., & Schijven, M. P. (2010). Contemporary virtual reality laparoscopy simulators: quicksand or solid grounds for assessing surgical trainees? *The American journal of surgery, 199*(4), 529-541.
Tien, G., Atkins, M. S., Zheng, B., & Swindells, C. (2010). *Measuring situation awareness of surgeons in laparoscopic training.* Paper presented at the Proceedings of the 2010 symposium on eye-tracking research & applications.
Torkington, J., Smith, S., Rees, B., & Darzi, A. (2001). The role of the basic surgical skills course in the acquisition and retention of laparoscopic skill. *Surgical endoscopy, 15*(10), 1071-1075.
Tuijthof, G. J., Van Sterkenburg, M. N., Sierevelt, I. N., Van Oldenrijk, J., Van Dijk, C. N., & Kerkhoffs, G. M. (2010). First validation of the PASSPORT training environment for arthroscopic skills. *Knee Surgery, Sports Traumatology, Arthroscopy, 18*(2), 218-224.
Unity. (2017). Unity 2017: The world-leading creation engine. from https://unity3d.com/unity
Urbanowicz, R. J., Meeker, M., LaCava, W., Olson, R. S., & Moore, J. H. (2017). Relief-based feature selection: introduction and review. *arXiv preprint arXiv:1711.08421*.
Üstün, B., Melssen, W. J., & Buydens, L. M. (2006). Facilitating the application of support vector regression by using a universal Pearson VII function based kernel. *Chemometrics and Intelligent Laboratory Systems, 81*(1), 29-40.
van der Lans, R., Wedel, M., & Pieters, R. (2011). Defining eye-fixation sequences across individuals and tasks: the Binocular-Individual Threshold (BIT) algorithm. *Behavior Research Methods, 43*(1), 239-257.
Van Hove, P., Tuijthof, G., Verdaasdonk, E., Stassen, L., & Dankelman, J. (2010). Objective assessment of technical surgical skills. *BJS, 97*(7), 972-987.
Van Sickle, K., McClusky III, D., Gallagher, A., & Smith, C. (2005). Construct validation of the ProMIS simulator using a novel laparoscopic suturing task. *Surgical Endoscopy And Other Interventional Techniques, 19*(9), 1227-1231.
Varadarajan, B., Reiley, C., Lin, H., Khudanpur, S., & Hager, G. (2009). *Data-derived models for segmentation with application to surgical assessment and training.* Paper presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention.
Verdaasdonk, E., Dankelman, J., Schijven, M., Lange, J., Wentink, M., & Stassen, L. (2009). Serious gaming and voluntary laparoscopic skills training: a multicenter study. *Minimally Invasive Therapy & Allied Technologies, 18*(4), 232-238.
Watson, R. A. (2012). Computer-aided feedback of surgical knot tying using optical tracking. *Journal of surgical education, 69*(3), 306-310.

Watson, R. A. (2013). Quantification of surgical technique using an inertial measurement unit. *Simulation in Healthcare, 8*(3), 162-165.
Watson, R. A. (2014). Use of a machine learning algorithm to classify expertise: analysis of hand motion patterns during a simulated surgical task. *Academic Medicine, 89*(8), 1163-1167.
Waxberg, S. L., Goodell, K. H., Avgerinos, D. V., Schwaitzberg, S. D., & Cao, C. G. (2004). *Evaluation of physical versus virtual surgical training simulators.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
Wentink, B. (2001). Eye-hand coordination in laparoscopy-an overview of experiments and supporting aids. *Minimally Invasive Therapy & Allied Technologies, 10*(3), 155-162.
Westebring–Van Der Putten, E., Goossens, R., Jakimowicz, J., & Dankelman, J. (2008). Haptics in minimally invasive surgery–a review. *Minimally Invasive Therapy & Allied Technologies, 17*(1), 3-16.
Wilson, M., McGrath, J., Vine, S., Brewer, J., Defriend, D., & Masters, R. (2010). Psychomotor control in a virtual laparoscopic surgery training environment: gaze control parameters differentiate novices from experts. *Surgical endoscopy, 24*(10), 2458-2464.
Windsor, J. A., Diener, S., & Zoha, F. (2008). Learning style and laparoscopic experience in psychomotor skill performance using a virtual reality surgical simulator. *The American journal of surgery, 195*(6), 837-842.
Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann.
Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., et al. (2008). Top 10 algorithms in data mining. *Knowledge and information systems, 14*(1), 1-37.
Yamaguchi, S., Konishi, K., Yasunaga, T., Yoshida, D., Kinjo, N., Kobayashi, K., et al. (2007). Construct validity for eye–hand coordination skill on a virtual reality laparoscopic surgical simulator. *Surgical endoscopy, 21*(12), 2253-2257.
Zhang, A., Hünerbein, M., Dai, Y., Schlag, P. M., & Beller, S. (2008). Construct validity testing of a laparoscopic surgery simulator (Lap Mentor®). *Surgical endoscopy, 22*(6), 1440-1444.
Zhang, Y., Ding, C., & Li, T. (2008). Gene selection algorithm by combining reliefF and mRMR. *BMC genomics, 9*(2), S27.

APPENDICES

APPENDIX A. CLASSIFICATION RESULTS ON PERFORMANCE MEASURES

Table A.1. Classification Results of Scenario-1 – DH using All Features

Algorithm	Option	Accuracy	Correctly Classified	Kappa Statistics	MAE	RMSE	RAE(%)	RRSE(%)	TPR	FPR	Precision	Recall	F-Measure
LWL		67.85	19	0.33	0.38	0.48	77.35	97.55	0.58	0.25	0.64	0.58	0.609
SMO	Puk	57.14	16	0	0.42	0.65	86.63	131.12	0	0		0	
KNN	K=3	57.14	16	0.14	0.44	0.56	86.34	112.64	0.58	0.43	0.5	0.58	0.538
KNN	K=5	57.14	16	0.16	0.49	0.52	99.64	105.52	0.66	0.5	0.5	0.66	0.571
Regression		57.14	16	0	0.5	0.5	101.06	100.14	0	0		0	
JRip		57.14	16	0.08	0.46	0.55	93.27	110.84	0.33	0.25	0.5	0.33	0.401
SMO	Norm. Polykernel	53.57	15	-0.07	0.46	0.68	93.85	136.47	0	0.06	0	0	0
KNN	K=1	53.57	15	0.08	0.46	0.65	94.36	131.54	0.58	0.5	0.46	0.58	0.519
Bagging		53.57	15	-0.04	0.49	0.5	100.75	101.02	0.08	0.12	0.33	0.08	0.133
PART		53.57	15	0.04	0.46	0.62	93.94	124.97	0.41	0.37	0.45	0.41	0.435
Decision Tree J48		53.57	15	0.04	0.45	0.62	91.54	124.48	0.42	0.37	0.45	0.41	0.435
KNN	K=7	50	14	0.02	0.52	0.53	105.14	107.75	0.58	0.56	0.44	0.58	0.501
Random Forest		50	14	0	0.52	0.23	105.09	106.61	0.5	0.5	0.42	0.5	0.462
Naive Bayes		46.42	13	0	0.53	0.71	108.09	142.38	0.75	0.75	0.43	0.75	0.545
Logistic		46.42	13	-0.03	0.53	0.73	108.28	146.59	0.58	0.62	0.41	0.58	0.483
Simple Logistic		42.85	12	-0.14	0.51	0.55	103.46	111.76	0.41	0.56	0.35	0.41	0.385
AdaBoostM1		42.85	12	-0.19	0.55	0.68	111.56	136.25	0.25	0.43	0.3	0.25	0.273
SMO	PolyKernel	35.71	10	-0.28	0.64	0.8	129.94	160.59	0.33	0.62	0.28	0.33	0.308
KNN	K=9	32.14	9	-0.34	0.53	0.54	108.22	109.33	0.33	0.68	0.26	0.33	0.296
	Max Accuracy	**67.85**	**19**										**0.609**

Table A.2. Classification Results of Scenario-1 – DH using Selected Features by ReliefF Method

Algorithm	Option	Accuracy	Correctly Classified	Kappa Statistics	MAE	RMSE	RAE(%)	RRSE(%)	TPR	FPR	Precision	Recall	F-Measure
LWL		75	21	0.47	0.34	0.45	69.75	90.52	0.58	0.12	0.78	0.58	0.667
SMO	Norm. Polykernel	64.28	18	0.25	0.35	0.59	72.19	119.71	0.5	0.25	0.6	0.5	0.545
KNN	K=7	64.28	18	0.25	0.45	0.49	91.89	100.04	0.5	0.25	0.6	0.5	0.545
AdaBoostM1		64.28	18	0.25	0.36	0.5	73.9	100.19	0.5	0.25	0.6	0.5	0.545
Regression		64.28	18	0.25	0.47	0.51	96.91	103.24	0.5	0.25	0.6	0.5	0.545
JRip		64.28	18	0.25	0.41	0.51	84.81	102.76	0.5	0.25	0.6	0.5	0.545
Random Forest		64.28	18	0.27	0.43	0.5	88.84	100.27	0.58	0.31	0.58	0.58	0.583
Bagging		60.71	17	0.18	0.44	0.48	90.29	96.71	0.5	0.31	0.54	0.5	0.522
KNN	K=3	57.14	16	0.08	0.44	0.54	89.35	109.93	0.33	0.25	0.5	0.33	0.401
KNN	K=9	57.14	16	0.08	0.48	0.51	98.68	103.54	0.33	0.25	0.5	0.33	0.401
KNN	K=1	53.57	15	0.02	0.46	0.65	94.41	131.57	0.33	0.31	0.44	0.33	0.381
KNN	K=5	53.57	15	0.04	0.45	0.52	92.54	104.63	0.42	0.37	0.45	0.41	0.435
PART		53.57	15	0.02	0.48	0.62	97.49	125.5	0.33	0.31	0.44	0.33	0.381
Decision Tree J48		53.57	15	0.02	0.45	0.58	91.94	117.003	0.33	0.31	0.44	0.33	0.381
Logistic		50	14	-0.02	0.46	0.55	93.67	110.48	0.42	0.44	0.42	0.42	0.417
Simple Logistic		50	14	-0.02	0.47	0.51	96.05	102.28	0.42	0.44	0.42	0.42	0.417
SMO	PolyKernel	50	14	-0.02	0.5	0.7	101.06	141.63	0.42	0.44	0.42	0.42	0.417
SMO	Puk	50	14	-0.04	0.5	0.7	101.06	141.63	0.33	0.37	0.4	0.33	0.364
Naive Bayes		46.42	13	0	0.5	0.65	101.75	131.28	0.75	0.75	0.42	0.25	0.545
	Max Accuracy	**75**	**21**										**0.667**

Table A.3. Classification Results of Scenario-1 – DH using Selected Features by Wrapper Method

Algorithm	Option	Accuracy	Correctly Classified	Kappa Statistics	MAE	RMSE	RAE(%)	RRSE(%)	TPR	FPR	Precision	Recall	F-Measure
SMO	Puk	64.28	18	0.25	0.35	0.59	72.19	119.7	0.67	0.43	0.53	0.67	0.593
Decision Tree J48		64.28	18	0.27	0.37	0.54	75.15	109.06	0.58	0.31	0.58	0.58	0.583
KNN	K=3	60.71	17	0.22	0.41	0.51	84.66	101.23	0.67	0.43	0.53	0.67	0.593
KNN	K=5	57.14	16	0.16	0.48	0.52	98.22	104.71	0.67	0.5	0.5	0.67	0.571
Simple Logistic		53.57	15	0.04	0.46	0.52	94.62	105.4	0.41	0.37	0.45	0.41	0.435
SMO	Norm. Polykernel	53.57	15	0	0.46	0.68	93.85	136.47	0.25	0.25	0.43	0.25	0.316
KNN	K=1	53.57	15	0.09	0.46	0.65	94.38	131.56	0.67	0.56	0.47	0.67	0.552
KNN	K=7	53.57	15	0.09	0.46	0.49	94.95	98.78	0.67	0.56	0.47	0.67	0.552
PART		53.57	15	0.06	0.45	0.6	92.24	120.99	0.5	0.44	0.46	0.5	0.480
Random Forest		53.57	15	0.08	0.5	0.52	102.71	106.11	0.58	0.5	0.47	0.58	0.519
Naive Bayes		50	14	0.05	0.49	0.66	99.07	133.58	0.75	0.68	0.45	0.75	0.563
Logistic		50	14	0.02	0.51	0.71	102.88	141.94	0.58	0.56	0.43	0.58	0.500
KNN	K=9	46.42	13	-0.01	0.49	0.51	99.47	102.95	0.67	0.68	0.42	0.67	0.516
AdaBoostM1		46.42	13	-0.08	0.55	0.63	111.9	126.22	0.41	0.5	0.38	0.41	0.400
Regression		42.85	12	-0.19	0.52	0.55	104.84	110.03	0.25	0.43	0.3	0.25	0.273
SMO	PolyKernel	39.28	11	-0.22	0.61	0.77	122.72	156.07	0.33	0.56	0.31	0.33	0.320
JRip		39.28	11	-0.33	0.56	0.59	114.33	118	0	0.31	0	0	0
LWL		35.71	10	-0.23	0.54	0.61	110.09	122.08	0.5	0.75	0.33	0.5	0.400
Bagging		35.71	10	-0.34	0.52	0.53	105.83	107.31	0.16	0.5	0.2	0.16	0.182
	Max Accuracy	**64.28**	**18**										**0.593**

Table A.4. Classification Results of Scenario-1 – DH using Selected Features by PCA Method

Algorithm	Option	Accuracy	Correctly Classified	Kappa Statistics	MAE	RMSE	RAE(%)	RRSE(%)	TPR	FPR	Precision	Recall	F-Measure
Naive Bayes		57.14	16	0.16	0.47	0.53	94.4	105.38	0.67	0.5	0.5	0.67	0.571
SMO	PolyKernel	57.14	16	0	0.42	0.65	86.63	131.12	0	0		0	
SMO	Norm. Polykernel	57.14	16	0	0.42	0.65	86.63	131.12	0	0		0	
PART		57.14	16	0	0.49	0.49	99.91	100.06	0	0		0	
Decision Tree J48		57.14	16	0	0.49	0.49	99.91	100.06	0	0		0	0
SMO	Puk	53.57	15	-0.07	0.46	0.68	93.85	136.48	0	0.06	0	0	0
KNN	K=1	53.57	15	0.08	0.46	0.65	94.33	131.53	0.58	0.5	0.47	0.58	0.519
KNN	K=5	53.57	15	0.08	0.5	0.53	101.07	106.85	0.58	0.5	0.47	0.58	0.519
KNN	K=7	53.57	15	0.06	0.52	0.55	105.14	109.59	0.5	0.43	0.46	0.5	0.480
AdaBoostM1		50	14	-0.08	0.54	0.6	109.89	120.35	0.17	0.25	0.33	0.17	0.222
Regression		50	14	-0.13	0.51	0.51	102.29	102.61	0	0.125	0	0	0
Logistic		46.42	13	-0.13	0.49	0.52	99.96	103.54	0.25	0.37	0.33	0.25	0.286
KNN	K=3	46.42	13	-0.06	0.51	0.59	103.41	118.19	0.5	0.56	0.4	0.5	0.444
KNN	K=9	46.42	13	-0.06	0.5	0.52	101.06	103.64	0.5	0.56	0.4	0.5	0.444
Bagging		46.42	13	-0.129	0.52	0.54	105.99	107.93	0.25	0.37	0.33	0.25	0.286
Random Forest		46.42	13	-0.12	0.53	0.58	108.14	116.27	0.25	0.37	0.33	0.25	0.286
JRip		42.85	12	-0.24	0.56	0.59	113.88	118.5	0.08	0.31	0.16	0.08	0.111
Simple Logistic		39.28	11	-0.22	0.49	0.51	100.46	101.78	0.33	0.56	0.31	0.33	0.320
LWL		35.71	10	-0.28	0.58	0.63	117.41	126.53	0.33	0.62	0.29	0.33	0.308
	Max Accuracy	**57.14**	**16**										**0.571**

Table A.5. Classification Results of Scenario-1 – NH using All Features

Algorithm	Option	Accuracy	Correctly Classified	Kappa Statistics	MAE	RMSE	RAE(%)	RRSE(%)	TPR	FPR	Precision	Recall	F-Measure
SMO	PolyKernel	60.71	17	0.22	0.39	0.62	79.41	125.54	0.67	0.44	0.53	0.67	0.593
Logistic		60.71	17	0.21	0.39	0.63	79.44	125.54	0.58	0.37	0.54	0.58	0.560
Simple Logistic		60.71	17	0.20	0.46	0.57	93.39	113.80	0.58	0.37	0.54	0.58	0.560
Random Forest		60.71	17	0.19	0.50	0.52	100.27	104.67	0.50	0.31	0.54	0.50	0.522
KNN	K=7	57.14	16	0.14	0.52	0.56	105.15	112.68	0.58	0.44	0.50	0.58	0.538
Regression		57.14	16	0.00	0.50	0.50	101.06	100.15	0.00	0.00		0.00	
Bagging		53.57	15	-0.05	0.50	0.50	100.75	101.02	0.08	0.12	0.33	0.08	0.133
KNN	K=3	50	14	0.02	0.51	0.63	103.42	126.82	0.58	0.56	0.44	0.58	0.500
KNN	K=5	50	14	0.02	0.48	0.56	98.23	112.39	0.58	0.56	0.44	0.58	0.500
SMO	Norm. Polykernel	50	14	-0.14	0.50	0.70	101.07	141.63	0.00	0.12	0.00	0.00	0.000
SMO	Puk	50	14	-0.14	0.50	0.70	101.07	141.63	0.00	0.12	0.00	0.00	0.000
KNN	K=9	46.42	13	-0.01	0.51	0.53	103.45	107.05	0.67	0.69	0.42	0.67	0.516
Naive Bayes (NB)		46.42	13	-0.03	0.52	0.71	105.47	142.57	0.58	0.62	0.41	0.58	0.483
LWL		46.42	13	-0.03	0.53	0.62	106.89	124.79	0.58	0.62	0.41	0.58	0.483
AdaBoostM1		46.42	13	-0.10	0.52	0.65	105.01	129.81	0.33	0.44	0.36	0.33	0.348
Jrip		46.42	13	-0.13	0.55	0.61	111.60	123.28	0.25	0.37	0.33	0.25	0.286
KNN	K=1	42.85	12	-0.12	0.57	0.73	114.46	145.92	0.50	0.62	0.37	0.50	0.429
PART		39.28	11	-0.20	0.59	0.74	118.68	148.65	0.42	0.62	0.33	0.42	0.370
Decision Tree J48		35.71	10	-0.28	0.61	0.75	123.54	151.11	0.33	0.62	0.28	0.33	0.308
	Max Accuracy	**60.71**	**17**										**0.593**

Table A.6. Classification Results of Scenario-1 – NH using Selected Features by ReliefF Method

Algorithm	Option	Accuracy	Correctly Classified	Kappa Statistics	MAE	RMSE	RAE(%)	RRSE(%)	TPR	FPR	Precision	Recall	F-Measure
SMO	PolyKernel	75	21	0.49	0.25	0.50	50.53	100.15	0.75	0.25	0.69	0.75	0.720
Logistic		71.42	20	0.42	0.28	0.53	57.75	107.06	0.67	0.25	0.67	0.67	0.667
SMO	Puk	71.42	20	0.41	0.28	0.53	57.75	107.06	0.67	0.25	0.67	0.67	0.667
Regression		71.42	20	0.41	0.35	0.42	69.99	83.97	0.67	0.25	0.67	0.67	0.667
Simple Logistic		67.85	19	0.36	0.41	0.50	82.34	100.24	0.75	0.37	0.60	0.75	0.667
KNN	K=3	67.85	19	0.36	0.42	0.49	84.66	98.19	0.75	0.37	0.60	0.75	0.667
KNN	K=7	64.28	18	0.30	0.48	0.51	96.99	101.66	0.75	0.44	0.56	0.75	0.643
KNN	K=9	64.28	18	0.30	0.48	0.50	96.30	100.71	0.75	0.43	0.56	0.75	0.643
Bagging		64.28	18	0.25	0.47	0.51	94.97	102.89	0.50	0.25	0.60	0.50	0.545
KNN	K=5	60.71	17	0.21	0.45	0.49	92.54	99.74	0.58	0.37	0.54	0.58	0.560
Random Forest		60.71	17	0.2	0.42	0.50	85.25	100.59	0.58	0.37	0.53	0.58	0.560
Naive Bayes (NB)		57.14	16	0.17	0.43	0.63	87.31	127.16	0.75	0.56	0.50	0.75	0.600
LWL		57.14	16	0.16	0.43	0.55	88.09	110.33	0.67	0.50	0.50	0.67	0.571
KNN	K=1	57.14	16	0.14	0.43	0.63	87.70	126.42	0.58	0.44	0.50	0.58	0.538
AdaBoostM1		57.14	16	0.12	0.41	0.55	82.34	110.59	0.50	0.37	0.50	0.50	0.500
Jrip		53.57	15	0.06	0.49	0.56	100.13	112.62	0.50	0.44	0.46	0.50	0.480
SMO	Norm. Polykernel	53.57	15	-0.04	0.46	0.68	93.85	136.48	0.08	0.12	0.33	0.08	0.133
PART		50	14	0	0.49	0.65	100.00	130.37	0.50	0.50	0.42	0.50	0.462
Decision Tree J48		50	14	0.00	0.48	0.63	96.92	126.70	0.50	0.50	0.43	0.50	0.462
	Max Accuracy	**75**	**21**										**0.720**

Table A.7. Classification Results of Scenario-1 – BH using All Features

Algorithm	**Option**	**Accuracy**	**Correctly Classified**	**Kappa Statistics**	**MAE**	**RMSE**	**RAE(%)**	**RRSE(%)**	**TPR**	**FPR**	**Precision**	**Recall**	**F-Measure**
LWL		71.42	20	0.44	0.36	0.5	73.94	101.01	0.83	0.37	0.62	0.83	0.714
PART		71.42	20	0.42	0.31	0.52	63.01	104.48	0.75	0.31	0.64	0.75	0.692
Decision Tree J48		71.42	20	0.42	0.32	0.52	64.94	105.34	0.75	0.31	0.64	0.75	0.692
KNN	K=1	64.28	18	0.31	0.37	0.58	74.33	115.45	0.83	0.5	0.56	0.83	0.667
KNN	K=3	64.28	18	0.31	0.45	0.56	91.69	113.37	0.83	0.5	0.56	0.83	0.667
Logistic		60.71	17	0.22	0.39	0.63	79.81	125.55	0.67	0.43	0.53	0.67	0.593
SMO	PolyKernel	60.71	17	0.2	0.39	0.62	79.41	125.54	0.58	0.37	0.53	0.58	0.560
KNN	K=9	60.71	17	0.2	0.5	0.52	101.07	104.81	0.58	0.37	0.53	0.58	0.560
Jrip		60.71	17	0.18	0.44	0.56	89.3	113.12	0.5	0.31	0.54	0.5	0.522
AdaBoostM1		60.71	17	0.15	0.34	0.51	69.38	102.39	0.33	0.19	0.57	0.33	0.421
SMO	Puk	57.14	16	0	0.42	0.65	86.63	131.12	0	0		0	
Regression		57.14	16	0	0.5	0.5	101.07	100.15	0	0		0	
KNN	K=5	53.57	15	0.09	0.47	0.53	95.38	106.25	0.67	0.56	0.47	0.67	0.552
KNN	K=7	53.57	15	0.08	0.5	0.53	101.07	107.18	0.58	0.5	0.47	0.58	0.519
Simple Logistic		53.57	15	0.04	0.44	0.55	89.22	110.31	0.42	0.37	0.45	0.42	0.435
Random Forest		53.57	15	0.04	0.49	0.5	98.47	101.22	0.42	0.37	0.45	0.42	0.435
SMO	Norm. Polykernel	53.57	15	-0.02	0.46	0.68	93.85	136.48	0.16	0.18	0.4	0.16	0.235
Bagging		53.57	15	-0.04	0.49	0.5	100.75	101.02	0.08	0.12	0.33	0.08	0.133
Naive Bayes (NB)		46.42	13	-0.03	0.54	0.72	109.35	144.79	0.58	0.62	0.41	0.58	0.483
	Max Accuracy	**71.42**	**20**										**0.714**

Table A.8. Classification Results of Scenario-1 – BH using Selected Features by ReliefF Method

Algorithm	**Option**	**Accuracy**	**Correctly Classified**	**Kappa Statistics**	**MAE**	**RMSE**	**RAE(%)**	**RRSE(%)**	**TPR**	**FPR**	**Precision**	**Recall**	**F-Measure**
KNN	K=9	78.57	22	0.57	0.4	0.44	81.19	88.15	0.83	0.25	0.71	0.83	0.769
KNN	K=3	75	21	0.5	0.28	0.44	56.53	87.67	0.83	0.31	0.67	0.83	0.741
KNN	K=7	75	21	0.5	0.37	0.44	75.57	87.99	0.83	0.31	0.67	0.83	0.741
KNN	K=5	71.42	20	0.42	0.32	0.44	65.54	87.74	0.75	0.31	0.64	0.75	0.692
SMO	Norm. Polykernel	71.42	20	0.41	0.28	0.53	57.75	107.06	0.67	0.25	0.67	0.67	0.667
SMO	Puk	71.42	20	0.4	0.28	0.53	57.75	107.06	0.58	0.18	0.7	0.58	0.636
KNN	K=1	67.85	19	0.36	0.33	0.54	67.63	109.53	0.75	0.37	0.6	0.75	0.667
Logistic		67.85	19	0.35	0.31	0.56	64.3	112.43	0.67	0.31	0.61	0.67	0.640
SMO	PolyKernel	67.85	19	0.35	0.32	0.56	64.97	113.55	0.67	0.31	0.61	0.67	0.640
Random Forest		64.28	18	0.28	0.43	0.48	86.84	96.72	0.67	0.37	0.57	0.67	0.615
AdaBoostM1		64.28	18	0.27	0.33	0.5	67.17	100.82	0.58	0.31	0.58	0.58	0.583
Simple Logistic		60.71	17	0.22	0.4	0.54	81.21	108.09	0.67	0.43	0.53	0.67	0.593
Regression		60.71	17	0.22	0.44	0.54	89.21	107.42	0.67	0.44	0.53	0.67	0.593
LWL		57.14	16	0.14	0.47	0.6	94.21	121.09	0.58	0.44	0.5	0.58	0.583
Jrip		57.14	16	0.12	0.46	0.58	92.77	117.33	0.5	0.37	0.5	0.5	0.500
Decision Tree J48		57.14	16	0.12	0.44	0.62	89.25	125.2	0.5	0.37	0.5	0.5	0.500
Bagging		57.14	16	0.1	0.47	0.52	95.86	104.31	0.42	0.31	0.5	0.42	0.455
PART		53.57	15	0.04	0.46	0.64	93.87	127.57	0.42	0.37	0.45	0.42	0.435
Naive Bayes (NB)		50	14	0.02	0.46	0.64	93.06	127.99	0.58	0.56	0.44	0.58	0.500
	Max Accuracy	**78.57**	**22**										**0.769**

Table A.9. Classification Results of Scenario-2– All Hand Conditions using All Features

		DH			NH			BH		
Algorithm	**Option**	**Accuracy**	**Correctly Classified**	**F-Measure**	**Accuracy**	**Correctly Classified**	**F-Measure**	**Accuracy**	**Correctly Classified**	**F-Measure**
PART		64.28	18	0.444	**82.14**	**23**	**0.800**	67.85	19	0.640
Jrip		**64.28**	**18**	**0.500**	78.57	22	0.769	53.57	15	0.435
SVM	Puk	60.71	17	0.154	82.14	23	0.737	60.71	17	0.154
Decision Tree J48		60.71	17	0.421	82.14	23	0.800	67.85	19	0.640
Regression		57.14	16	?	57.14	16	?	57.14	16	?
Simple Logistic		53.57	15	0.435	78.57	22	0.769	57.14	16	0.455
KNN	K=3	53.57	15	0.435	64.28	18	0.615	46.42	13	0.348
KNN	K=7	53.57	15	0.435	60.71	17	0.621	46.42	13	0.444
KNN	K=9	53.57	15	0.480	64.28	18	0.643	50	14	0.462
Bagging		53.57	15	0.133	53.57	15	0.133	53.57	15	0.133
SVM	Normalized Polykernel	50	14	0.000	57.14	16	0.400	53.57	15	0.133
KNN	K=1	50	14	0.462	**71.42**	**20**	**0.667**	**75**	**21**	**0.720**
LWL		50	14	0.222	82.14	23	0.815	39.28	11	0.261
AdaBoost		50	14	0.300	75	21	0.720	53.57	15	0.381
Naive Bayes		46.42	13	0.545	71.42	20	0.692	53.57	15	0.552
KNN	K=5	46.42	13	0.400	60.71	17	0.621	42.86	12	0.385
Random Forest		46.42	13	0.348	78.57	22	0.750	67.85	19	0.571
Logistic		42.85	12	0.333	57.14	16	0.538	57.14	16	0.500
SVM	PolyKernel	42.85	12	0.333	64.28	18	0.615	64.28	18	0.545
	Max Accuracy	**64.28**	**18**	**0.545**	**82.14**	**23**	**0.815**	**75**	**21**	**0.720**

Table A.10. Classification Results of Scenario-2 – All Hand Conditions using Selected Features by ReliefF Method

		DH			NH			BH		
Algorithm	**Option**	**Accuracy**	**Correctly Classified**	**F-Measure**	**Accuracy**	**Correctly Classified**	**F-Measure**	**Accuracy**	**Correctly Classified**	**F-Measure**
KNN	K=7	**67.85**	**19**	**0.471**	78.57	22	0.769	64.28	18	0.583
SVM	Puk	67.85	19	0.400	**85.71**	**24**	**0.800**	67.85	19	0.526
KNN	K=5	67.85	19	0.400	64.28	18	0.643	60.71	17	0.476
KNN	K=9	64.28	18	0.444	75	21	0.741	67.85	19	0.640
SVM	PolyKernel	60.71	17	0.560	64.28	18	0.615	**71.42**	**20**	**0.714**
Simple Logistic		60.71	17	0.522	67.85	19	0.667	71.42	20	0.636
Regression		60.71	17	0.522	78.57	22	0.769	64.28	18	0.615
PART		60.71	17	0.522	82.14	23	0.800	67.85	19	0.640
SVM	Normalized Polykernel	60.71	17	0.353	60.71	17	0.421	64.28	18	0.545
Random Forest		57.14	16	0.538	78.57	22	0.750	67.85	19	0.609
Bagging		57.14	16	0.500	78.57	22	0.750	75	21	0.720
Decision Tree J48		57.14	16	0.500	82.14	23	0.800	67.85	19	0.640
Jrip		57.14	16	0.455	67.85	19	0.609	53.57	15	0.480
KNN	K=3	57.14	16	0.250	67.85	19	0.667	60.71	17	0.476
KNN	K=1	50	14	0.417	78.57	22	0.750	71.42	20	0.667
LWL		50	14	0.300	82.14	23	0.815	57.14	16	0.455
Naive Bayes		46.42	13	0.483	78.57	22	0.769	64.28	18	0.643
AdaBoost		42.85	12	0.385	85.71	24	0.833	64.28	18	0.545
Logistic		28.57	8	0.167	60.71	17	0.621	75	21	0.720
	Max Accuracy	**67.85**	**19**	**0.560**	**85.71**	**24**	**0.833**	**75**	**21**	**0.720**

Table A.11. Classification Results of Scenario-3 – DH using All Features

Algorithm	Option	Accuracy	Correctly Classified	Kappa Statistics	MAE	RMSE	RAE(%)	RRSE(%)	TPR	FPR	Precision	Recall	F-Measure
KNN	K=7	64.28	18	0.27	0.43	0.49	86.79	98.11	0.58	0.31	0.58	0.58	0.583
Random Forest		64.28	18	0.23	0.48	0.51	98.42	101.6	0.42	0.18	0.62	0.42	0.500
KNN	K=5	60.71	17	0.18	0.41	0.49	82.58	98.74	0.5	0.31	0.54	0.5	0.522
KNN	K=1	57.14	16	0.14	0.43	0.63	87.66	126.39	0.58	0.44	0.5	0.58	0.538
Decision Tree J48		57.14	16	0.12	0.44	0.62	89.87	125.78	0.5	0.37	0.5	0.5	0.500
SVM	Puk	57.14	16	0	0.43	0.65	86.63	131.12	0	0		0	
Regression		57.14	16	0	0.5	0.5	101.07	100.15	0	0		0	
KNN	K=3	53.57	15	0.04	0.45	0.56	91.68	112.02	0.42	0.37	0.45	0.42	0.435
PART		53.57	15	0.04	0.47	0.66	94.76	133.35	0.42	0.37	0.45	0.42	0.435
Naive Bayes		53.57	15	0.02	0.46	0.66	92.76	133.05	0.33	0.31	0.44	0.33	0.381
Bagging		53.57	15	-0.04	0.49	0.5	100.75	101.02	0.08	0.12	0.33	0.08	0.133
SVM	Norm. Polykernel	53.57	15	-0.07	0.46	0.68	93.85	136.48	0	0.06	0	0	0.000
Jrip		50	14	0	0.51	0.62	104.71	124.8	0.5	0.5	0.43	0.5	0.462
KNN	K=9	50	14	-0.11	0.46	0.49	92.32	98.85	0.08	0.18	0.25	0.08	0.125
SVM	PolyKernel	46.42	13	-0.08	0.53	0.73	108.29	146.6	0.42	0.5	0.38	0.42	0.400
AdaBoost		46.42	13	-0.08	0.53	0.67	108.39	134.74	0.42	0.5	0.38	0.42	0.400
LWL		42.85	12	-0.14	0.53	0.64	108.05	128.74	0.42	0.56	0.35	0.42	0.385
Simple Logistic		42.85	12	-0.16	0.56	0.65	112.55	130.73	0.33	0.5	0.33	0.33	0.333
Logistic		39.28	11	-0.2	0.61	0.78	122.68	156.02	0.42	0.62	0.33	0.42	0.370
	Max Accuracy	**64.28**	**18**										**0.583**

Table A.12. Classification Results of Scenario-3 – DH using Selected Features by ReliefF Method

Algorithm	Option	Accuracy	Correctly Classified	Kappa Statistics	MAE	RMSE	RAE(%)	RRSE(%)	TPR	FPR	Precision	Recall	F-Measure
Naive Bayes		64.28	18	0.27	0.41	0.59	83.36	117.43	0.58	0.31	0.58	0.58	0.583
Logistic		67.85	19	0.35	0.32	0.56	64.82	113.15	0.67	0.31	0.61	0.67	0.640
Simple Logistic		67.85	19	0.36	0.37	0.51	74.54	101.79	0.75	0.37	0.6	0.75	0.667
SVM	PolyKernel	57.14	16	0.12	0.42	0.65	86.63	131.12	0.5	0.37	0.5	0.5	0.500
SVM	Norm. Polykernel	57.14	16	0.06	0.42	0.65	86.63	131.12	0.25	0.18	0.5	0.25	0.333
SVM	Puk	57.14	16	0.08	0.42	0.65	86.63	131.12	0.33	0.25	0.5	0.33	0.400
KNN	K=1	67.85	19	0.37	0.33	0.55	67.61	109.52	0.83	0.44	0.59	0.83	0.690
KNN	K=3	75	21	0.5	0.37	0.53	75.27	106.8	0.83	0.31	0.66	0.83	0.741
KNN	K=5	71.42	20	0.44	0.41	0.5	82.58	100.96	0.83	0.37	0.62	0.83	0.714
KNN	K=7	67.85	19	0.36	0.4	0.47	81.69	94.94	0.75	0.37	0.6	0.75	0.667
KNN	K=9	50	14	-0.06	0.42	0.47	85.16	93.63	0.25	0.31	0.37	0.25	0.300
LWL		60.71	17	0.2	0.42	0.55	85.7	110.91	0.58	0.37	0.53	0.58	0.560
AdaBoost		46.42	13	-0.03	0.5	0.64	102.82	129.84	0.58	0.62	0.41	0.58	0.483
Bagging		50	14	-0.02	0.49	0.52	99.5	104.93	0.42	0.44	0.42	0.42	0.417
Regression		57.14	16	0.16	0.48	0.55	96.66	110.12	0.67	0.5	0.5	0.67	0.571
PART		60.71	17	0.15	0.4	0.61	81.43	122.73	0.33	0.18	0.57	0.33	0.421
Jrip		50	14	0	0.5	0.6	102.21	121.49	0.5	0.5	0.43	0.5	0.462
Decision Tree J48		67.85	19	0.32	0.36	0.56	73.33	111.43	0.5	0.18	0.67	0.5	0.571
Random Forest		60.71	17	0.22	0.45	0.5	91.68	101.15	0.67	0.44	0.53	0.67	0.593
	Max Accuracy	**75**	**21**										**0.741**

Table A.13. Classification Results of Scenario-3 – NH using All Features

Algorithm	Option	Accuracy	Correctly Classified	Kappa Statistics	MAE	RMSE	RAE(%)	RRSE(%)	TPR	FPR	Precision	Recall	F-Measure
Simple Logistic		82.14	23	0.63	0.31	0.43	61.95	86.43	0.83	0.19	0.77	0.83	0.800
Logistic		75	21	0.51	0.28	0.51	56.81	102.5	0.92	0.37	0.65	0.92	0.759
PART		75	21	0.48	0.27	0.47	53.93	95.05	0.67	0.18	0.73	0.67	0.696
Decision Tree J48		75	21	0.48	0.27	0.48	54.42	96.84	0.67	0.18	0.72	0.67	0.696
Random Forest		75	21	0.48	0.42	0.44	85.67	89.03	0.67	0.18	0.72	0.67	0.696
AdaBoost		75	21	0.47	0.26	0.45	52.51	91.27	0.58	0.12	0.78	0.58	0.667
LWL		71.42	20	0.42	0.34	0.51	69.38	101.61	0.75	0.31	0.64	0.75	0.692
Naive Bayes		64.28	18	0.3	0.36	0.59	72.98	118.68	0.75	0.43	0.56	0.75	0.643
Jrip		64.28	18	0.25	0.39	0.55	79.97	110.53	0.5	0.25	0.6	0.5	0.545
KNN	K=1	57.14	16	0.17	0.43	0.63	87.68	126.41	0.75	0.56	0.5	0.75	0.600
Regression		57.14	16	0	0.5	0.5	101.06	100.14	0	0		0	
KNN	K=7	53.57	15	0.11	0.48	0.5	98.01	100.11	0.75	0.62	0.47	0.75	0.581
KNN	K=5	53.57	15	0.09	0.46	0.49	92.54	98.61	0.67	0.56	0.47	0.67	0.552
KNN	K=3	53.57	15	0.08	0.46	0.53	94.03	107.25	0.58	0.5	0.47	0.58	0.519
SVM	PolyKernel	53.57	15	0.06	0.46	0.68	93.85	136.48	0.5	0.44	0.46	0.5	0.480
Bagging		53.57	15	-0.04	0.49	0.5	100.75	101.02	0.08	0.12	0.33	0.08	0.133
SVM	Puk	53.57	15	-0.07	0.46	0.68	93.85	136.48	0	0.06	0	0	0.000
SVM	Norm. Polykernel	46.42	13	-0.21	0.53	0.73	108.29	146.6	0	0.19	0	0	0.000
KNN	K=9	42.85	12	-0.14	0.5	0.51	101.86	102.39	0.42	0.56	0.36	0.42	0.385
	Max Accuracy	**82.14**	**23**										**0.800**

Table A.14. Classification Results of Scenario-3 – NH using Selected Features by ReliefF Method

Algorithm	**Option**	**Accuracy**	**Correctly Classified**	**Kappa Statistics**	**MAE**	**RMSE**	**RAE(%)**	**RRSE(%)**	**TPR**	**FPR**	**Precision**	**Recall**	**F-Measure**
AdaBoost		85.71	24	0.71	0.18	0.36	37.16	73.42	0.92	0.18	0.79	0.92	0.846
SVM	PolyKernel	82.14	23	0.64	0.18	0.42	36.09	84.64	0.92	0.25	0.73	0.92	0.815
Simple Logistic		82.14	23	0.64	0.31	0.41	62.97	82.95	0.83	0.18	0.77	0.83	0.800
Bagging		78.57	22	0.56	0.33	0.39	67.12	78.73	0.75	0.18	0.75	0.75	0.750
Logistic		78.57	22	0.55	0.21	0.46	44.24	92.77	0.67	0.12	0.8	0.67	0.727
Random Forest		75	21	0.49	0.3	0.37	61.79	74.53	0.75	0.25	0.69	0.75	0.720
PART		75	21	0.48	0.25	0.47	52.02	94.87	0.67	0.18	0.73	0.67	0.696
Decision Tree J48		75	21	0.48	0.26	0.48	54.42	96.84	0.67	0.18	0.73	0.67	0.696
KNN	K=7	71.42	20	0.44	0.4	0.44	81.69	87.42	0.83	0.37	0.62	0.83	0.714
LWL		71.42	20	0.43	0.34	0.51	69.67	101.36	0.75	0.31	0.64	0.75	0.692
Jrip		71.42	20	0.41	0.34	0.51	69.59	101.61	0.67	0.25	0.67	0.67	0.667
KNN	K=1	67.85	19	0.37	0.33	0.54	67.64	109.55	0.83	0.43	0.58	0.83	0.690
KNN	K=3	67.85	19	0.37	0.31	0.43	63.56	87.01	0.83	0.44	0.59	0.83	0.690
Naive Bayes		64.28	18	0.3	0.35	0.57	71.09	114.06	0.75	0.44	0.56	0.75	0.643
Regression		64.28	18	0.22	0.37	0.48	76.01	95.99	0.33	0.12	0.67	0.33	0.444
KNN	K=5	60.71	17	0.22	0.4	0.46	81.17	92.02	0.67	0.43	0.53	0.67	0.593
KNN	K=9	57.14	16	0.11	0.42	0.45	85.16	90.04	0.42	0.31	0.5	0.42	0.455
SVM	Puk	57.14	16	0	0.42	0.65	86.63	131.12	0	0		0	
SVM	Norm. Polykernel	53.57	15	-0.02	0.46	0.68	93.85	136.48	0.16	0.18	0.4	0.167	0.235
	Max Accuracy	**85.71**	**24**										**0.846**

Table A.15. Classification Results of Scenario-3 – BH using All Features

Algorithm	**Option**	**Accuracy**	**Correctly Classified**	**Kappa Statistics**	**MAE**	**RMSE**	**RAE(%)**	**RRSE(%)**	**TPR**	**FPR**	**Precision**	**Recall**	**F-Measure**
KNN	K=9	57.14	16	0.11	0.49	0.51	100.27	102.2	0.42	0.31	0.5	0.42	0.455
Random Forest		57.14	16	0.11	0.49	0.51	100.2	102.81	0.42	0.31	0.5	0.42	0.455
Simple Logistic		57.14	16	0.08	0.46	0.53	93.38	105.68	0.33	0.25	0.5	0.33	0.400
SVM	Norm Polykernel	57.14	16	0.04	0.42	0.65	86.63	131.12	0.17	0.12	0.5	0.17	0.250
SVM	Puk	57.14	16	0	0.43	0.65	86.63	131.12	0	0		0	
Regression		57.14	16	0	0.5	0.5	101.07	100.15	0	0		0	
KNN	K=7	53.57	15	0.04	0.51	0.52	103.11	105.32	0.42	0.37	0.45	0.42	0.435
Bagging		53.57	15	-0.04	0.49	0.5	100.75	101.02	0.08	0.12	0.33	0.08	0.133
KNN	K=3	50	14	0.03	0.53	0.59	108.1	118.29	0.67	0.62	0.44	0.67	0.533
Naive Bayes		50	14	-0.02	0.48	0.67	98.62	134.46	0.42	0.44	0.42	0.42	0.417
PART		50	14	-0.02	0.52	0.67	106.87	134.52	0.42	0.44	0.42	0.42	0.417
SVM	PolyKernel	46.42	13	-0.1	0.53	0.73	108.28	146.6	0.33	0.44	0.36	0.33	0.348
Decision Tree J48		46.42	13	-0.1	0.56	0.7	113.24	139.66	0.33	0.44	0.36	0.33	0.348
KNN	K=1	42.85	12	-0.14	0.56	0.73	114.46	145.92	0.42	0.56	0.36	0.42	0.385
Logistic		42.85	12	-0.16	0.57	0.75	116.44	150.95	0.33	0.5	0.33	0.33	0.333
AdaBoost		42.85	12	-0.16	0.54	0.68	109.92	137.19	0.33	0.5	0.33	0.33	0.333
KNN	K=5	35.71	10	-0.28	0.55	0.57	111.02	115.24	0.33	0.62	0.29	0.33	0.308
LWL		35.71	10	-0.28	0.57	0.66	115.87	132.79	0.33	0.62	0.29	0.33	0.308
Jrip		28.57	8	-0.45	0.64	0.71	128.72	142.79	0.16	0.62	0.16	0.16	0.167
	Max Accuracy	**57.14**											**0.533**

Table A.16. Classification Results of Scenario-3 – BH using Selected Features by ReliefF Method

Algorithm	**Option**	**Accuracy**	**Correctly Classified**	**Kappa Statistics**	**MAE**	**RMSE**	**RAE(%)**	**RRSE(%)**	**TPR**	**FPR**	**Precision**	**Recall**	**F-Measure**
KNN	K=3	71.42	20	0.4	0.35	0.47	70.59	94.68	0.58	0.18	0.7	0.58	0.636
KNN	K=5	67.85	19	0.32	0.39	0.48	78.33	96.11	0.5	0.19	0.67	0.5	0.571
KNN	K=7	67.85	19	0.32	0.46	0.5	92.91	101.33	0.5	0.19	0.67	0.5	0.571
Naive Bayes		67.85	19	0.3	0.35	0.56	69.93	112.09	0.42	0.12	0.71	0.42	0.526
SVM	Norm. Polykernel	67.85	19	0.29	0.32	0.57	64.97	113.55	0.33	0.06	0.8	0.33	0.471
KNN	K=1	64.28	18	0.27	0.37	0.57	74.33	115.45	0.58	0.31	0.58	0.58	0.583
Random Forest		64.28	18	0.27	0.44	0.48	88.68	97.65	0.58	0.31	0.58	0.58	0.583
KNN	K=9	64.28	18	0.23	0.46	0.49	93.12	98.41	0.42	0.19	0.62	0.42	0.500
SVM	Puk	64.28	18	0.2	0.36	0.59	72.19	119.7	0.25	0.06	0.75	0.25	0.375
Simple Logistic		60.71	17	0.15	0.47	0.57	94.75	113.56	0.33	0.18	0.57	0.33	0.421
SVM	PolyKernel	60.71	17	0.15	0.39	0.63	79.41	125.54	0.33	0.18	0.57	0.33	0.421
AdaBoost		53.57	15	0.02	0.45	0.61	91.64	122.22	0.33	0.31	0.44	0.33	0.381
Regression		50	14	-0.02	0.52	0.57	105.26	114.62	0.42	0.44	0.42	0.42	0.417
Logistic		46.42	13	-0.06	0.55	0.73	110.38	146.82	0.5	0.56	0.4	0.5	0.444
Bagging		46.42	13	-0.06	0.49	0.52	99.79	104.29	0.5	0.56	0.4	0.5	0.444
PART		42.85	12	-0.16	0.58	0.72	117.3	143.55	0.33	0.5	0.33	0.33	0.333
Decision Tree J48		42.85	12	-0.16	0.59	0.72	120.35	144.11	0.33	0.5	0.33	0.33	0.333
Jrip		39.28	11	-0.25	0.57	0.66	114.69	131.41	0.25	0.5	0.27	0.25	0.261
LWL		35.71	10	-0.31	0.56	0.66	113.14	131.73	0.25	0.56	0.25	0.25	0.250
	Max Accuracy	**71.42**	**20**										**0.636**

Table A.17. Classification Results of Scenario-4 – All Hand Conditions using All Features

		DH			NH			BH		
Algorithm	**Option**	**Accuracy**	**Correctly Classified**	**F-Measure**	**Accuracy**	**Correctly Classified**	**F-Measure**	**Accuracy**	**Correctly Classified**	**F-Measure**
Jrip		**78.57**	**22**	**0.700**	46.42	13	0.348	53.57	15	0.316
Decision Tree J48		71.42	20	0.600	46.42	13	0.286	57.14	16	0.538
PART		67.85	19	0.571	42.85	12	0.273	57.14	16	0.538
AdaBoost		64.28	18	0.444	42.85	12	0.385	46.42	13	0.286
SVM	PolyKernel	60.71	17	0.476	64.28	18	0.583	64.28	18	0.615
Simple Logistic		57.14	16	0.400	46.42	13	0.444	60.71	17	0.560
SVM	Puk	57.14	16	?	57.14	16	?	57.14	16	?
KNN	K=7	57.14	16	0.250	53.57	15	0.581	60.71	17	0.560
LWL		57.14	16	0.143	28.57	8	0.091	35.71	10	0.308
Regression		57.14	16	?	57.14	16	?	57.14	16	?
Random Forest		57.14	16	0.400	**64.28**	**18**	**0.500**	57.14	16	0.400
Naive Bayes		53.57	15	0.552	50	14	0.563	50	14	0.500
Logistic		53.57	15	0.519	57.14	16	0.538	57.14	16	0.571
Bagging		53.57	15	0.133	53.57	15	0.133	53.57	15	0.133
SVM	Normalized Polykernel	50	14	0.000	57.14	16	?	53.57	15	0.235
KNN	K=1	50	14	0.500	46.42	13	0.483	60.71	17	0.621
KNN	K=9	50	14	0.125	53.57	15	0.480	60.71	17	0.522
KNN	K=3	39.28	11	0.320	50	14	0.533	**71.42**	**20**	**0.733**
KNN	K=5	39.28	11	0.190	42.85	12	0.500	67.85	19	0.667
	Max Accuracy	**78.57**	**22**	**0.700**	**64.28**	**18**	**0.583**	**71.42**	**20**	**0.733**

Table A.18. Classification Results of Scenario-4 – All Hand Conditions using Selected Features by ReliefF Method

		DH			NH			BH		
Algorithm	**Option**	**Accuracy**	**Correctly Classified**	**F-Measure**	**Accuracy**	**Correctly Classified**	**F-Measure**	**Accuracy**	**Correctly Classified**	**F-Measure**
SVM	Normalized Polykernel	**82.14**	**23**	**0.783**	46.42	13	0.000	67.85	19	0.571
Jrip		78.57	22	0.700	57.14	16	0.571	42.85	12	0.429
Random Forest		78.57	22	0.700	67.85	19	0.640	64.28	18	0.500
Logistic		75	21	0.720	67.85	19	0.640	71.42	20	0.714
KNN	K=1	71.42	20	0.692	75	21	0.720	53.57	15	0.519
PART		71.42	20	0.636	57.14	16	0.538	57.14	16	0.538
Decision Tree J48		71.42	20	0.636	53.57	15	0.435	53.57	15	0.519
Bagging		67.85	19	0.571	60.71	17	0.560	50	14	0.417
Simple Logistic		64.28	18	0.615	60.71	17	0.522	71.42	20	0.600
SVM	PolyKernel	64.28	18	0.583	71.42	20	0.667	78.57	22	0.769
Regression		64.28	18	0.500	57.14	16	0.538	50	14	0.364
AdaBoost		64.28	18	0.444	60.71	17	0.560	60.71	17	0.560
LWL		64.28	18	0.375	60.71	17	0.522	39.28	11	0.320
KNN	K=3	60.71	17	0.421	67.85	19	0.690	71.42	20	0.692
KNN	K=5	60.71	17	0.421	67.85	19	0.690	71.42	20	0.692
SVM	Puk	60.71	17	0.353	75	21	0.632	60.71	17	0.267
KNN	K=9	57.14	16	0.143	53.57	15	0.581	78.57	22	0.769
Naive Bayes		53.57	15	0.552	71.42	20	0.692	64.28	18	0.545
KNN	K=7	53.57	15	0.133	64.28	18	0.687	67.85	19	0.667
	Max Accuracy	**82.14**	**23**	**0.783**	**75**	**21**	**0.720**	78.57	**22**	**0.769**

APPENDIX B. SELECTED PERFORMANCE FEATURES WITH RELIEFF METHOD

Table B.1. Selected Performance Features Set for Scenario-1

Hand Condition	Accuracy (%)	Selected Feature Set
BH	71.42	Time_Task2, Time_Task3, Time_Task4, Time_Task6, Time_Task10, Time_mean, Time_median, Time_sum, Dist_Task1, Dist_Task3, Dist_Task4, camDist_Task5, camDist_Task6, camDist_Task9, camDist_Task10, camDist_mean, camDist_median, camDist_min, camDist_sum
NH	67.85	Handedness, Time_ Task2, Time_Task4, Time_ Task10, Time_std, Time_max, Time_var, Dist_Task1, Dist_std, Dist_var
DH	60.71	Gender, Handedness, Dist_Task1, Dist_Task6, Time_std, Time_var

Table B.2. Selected Performance Features Set for Scenario-2

Hand Condition	Accuracy (%)	Selected Feature Set
NH	82.14	Time_Task2, Time_Task4, Time_Task6, Time_max, Time_mean, Time_std, Time_sum, Time_var, Dist_Task1, Dist_Task3, Dist_Task4, Dist_Task5, Dist_Task7, Dist_Task8, Dist_Task9, Dist_Task10, Dist_max, Dist_mean, Dist_median, Dist_std, Dist_sum, Dist_var
DH	64.28	Handedness, Time_Task1, Time_Task3, Time_Task5, Time_Task9, Time_Task10, Time_mean, Time_median, Time_sum, Dist_Task2, Dist_Task5, Dist_Task6, Dist_Task10, Dist_mean, Dist_median, Dist_sum, Succ_Task4, Succ_Task6
BH	57.14	Time_Task1, Time_Task4, Time_Task10, Time_max, Time_median, Dist_Task1, Dist_Task3, Dist_Task5, Dist_Task9, Dist_max, Succ_Task6

Table B.3. Selected Performance Features Set for Scenario-3

Hand Condition	Accuracy (%)	Selected Feature Set
BH	78.57	Time_Task2, Time_Task4, Time_max, Time_mean, Time_std, Time_sum, Dist_Task2, Dist_Task3, Dist_min, Catch_Task2, Catch_Task4, Catch_Task9, Succ_Task1, Succ_Task9
NH	75	Handedness, Time_Task3, Time_Task4, Time_Task5, Time_Task8, Dist_Task5, Dist_Task7, Dist_Task8, Dist_median, Dist_min, Catch_Task2, Catch_Task3, Catch_Task4, Catch_Task5, Catch_Task10, Catch_median, Catch_min, Succ_Task1, Succ_Task3, Succ_Task7, Succ_Task8, Succ_mean, Succ_min, Succ_std, Succ_sum, Succ_var
DH	75	Handedness, Time_Task4, Time_Task5, Time_Task6, Time_Task8, Time_mean, Time_median, Time_sum, Dist_Task7, Dist_Task9, Dist_max, Dist_min, Dist_std, Dist_var, camDist_Task4, camDist_Task9, camDist_Task10, camDist_mean, camDist_median, camDist_min, camDist_sum, Catch_Task6, Catch_Task9, Catch_median, Succ_Task6, Succ_mean, Succ_median, Succ_sum

Table B.4. Selected Performance Features Set for Scenario-4

Hand Condition	Accuracy (%)	Selected Feature Set
NH	85.71	Time_Task2, Time_Task10, Time_Task12, Time_Task13, Time_median, errTime_Task5, errTime_Task10, errTime_Task12, errTime_Task13, errTime_median, Dist_Task1, Dist_Task2, Dist_Task3, Dist_Task12, Dist_Task13, errDist_Task1, errDist_Task2, errDist_Task10, errDist_Task12, errDist_Task13, errDist_max, errDist_median, errDist_min, errDist_Taskstd, errDist_var, Devia_Task4, Devia_Task5, Devia_Task12, Devia_Task13, Devia_median
DH	75	Handedness, Time_Task1, Time_Task4, Time_Task5, Time_Task6, Time_Task7, Time_Task9, Time_Task10, Time_Task12, Time_Task15, Time_mean, Time_median, Time_min, Time_sum, errTime_Task1, errTime_Task4, errTime_Task5, errTime_Task6, errTime_Task7, errTime_Task9, errTime_Task10, errTime_Task15, errTime_mean, errTime_median, errTime_min, errTime_sum, Dist_Task7, Dist_Task15, Dist_min, errDist_Task1, errDist_Task4, errDist_Task7, errDist_Task15, errDist_median, errDist_min, camDist_Task3, camDist_Task4, camDist_Task5, Devia_Task2, Devia_Task4, Devia_Task5, Devia_Task6, Devia_Task7, Devia_Task8, Devia_Task10, Devia_Task12, Devia_Task13, Devia_Task15, Devia_mean, Devia_min, Devia_sum, Succ_Task3, Succ_Task10, Succ_Task14
BH	71.42	Handedness, Time_Task1, Time_Task3, Time_Task4, Time_Task5, Time_Task13, Time_Task14, Time_Taskmin, errTime_Task1, errTime_Task4, errTime_Task5, errTime_Task8, errTime_Task13, errTime_Task14, Dist_Task1, Dist_Task2, Dist_Task5, Dist_Task8, Dist_Task10, Dist_Task12, errDist_Task1, errDist_Task2, errDist_Task5, Devia_Task7, Devia_Task8, Devia_Task12, Devia_Task13, Devia_Task15, Devia_max, Devia_std, Devia_var

APPENDIX C. QUESTIONNAIRE FOR OBSERVERS (ORIGINAL)

Değerli Katılımcı,

Bu anket Atılım Üniversitesi Yazılım Mühendisliği Bölümü'nde yapmakta olduğum doktora tezi kapsamında "Cerrahi Eğitim Süreçlerinde Cerrahların Beceri Seviyelerinin 3B Simulasyon Ortamında Tahmini" konulu çalışma ile ilgili bilgi toplamak amacı ile hazırlanmıştır.

Ankette, sizden kimlik belirleyici hiçbir bilgi istenmemektedir. Cevaplarınız tamamıyla gizli tutulacak ve sadece araştırmacılar tarafından değerlendirilecektir; elde edilecek bilgiler bilimsel yayımlarda kullanılacaktır. Çalışma iki bölümden oluşmakta olup, en fazla 30 dakikanızı alacaktır. Çalışma hakkında daha fazla bilgi almak için araştırma görevlisi Damla Topallı (damla.topalli@atilim.edu.tr) ile iletişim kurabilirsiniz.

Çalışmamıza destek verdiğiniz için teşekkür ederiz.

Çalışma Prosedürü

Bu çalışma kapsamında Hacettepe Üniversitesi Beyin Cerrahisi ve KBB Bölümlerinde Uzman ya da Stajer olarak görev yapan 15 katılımcının 3B ortamda geliştirilmiş bir senaryodaki performansları simule edilerek, el ve göz hareketleri ekran üzerinde gösterilecektir. Bu senaryoda amaç anatomik bir burun modeli üzerinde bulunan objeleri cerrahi alet ve kamera (endoskop) görevindeki iki dokunsal cihazı kullanarak toplamaktır. Bu kapsamda katılımcıların el hareketleri dokunsal cihaz kullanılarak, göz hareketleri de göz izleme cihazı kullanılarak kaydedilmiştir. Ekran üzerinde göreceğiniz mavi nokta cerrahi alet görevindeki dokunsal cihazın ekran üzerindeki yerini gösterirken, "göz" simgesi ise katılımcının o anda ekranda baktığı noktayı göstermektedir. İzlenen her performans için, o katılımcıya ait aşağıda verilen yargılardan kendinize uygun olanı seçerek kişinin performansını değerlendirmeniz beklenmektedir. Yapacağınız değerlendirmelerde katılımcının görevleri gerçekleştirirken dokuya teması, hareket tasarrufu (economy of movement), operasyonel akış, aletin 3B ortamdaki konumsal durumu (spatial orientation), derinlik algısı, sağ- sol el kullanım yeteneği, el-göz koordinasyonu ve 3B algının arttırılması gibi parametreler yer alırken, kişinin beceri seviyesini (Uzman ya da Acemi) tahmin etmeniz beklenmektedir.

Değerlendirilen Katılımcı No.

Değerlendirici No:

Bölüm 1. Demografik Bilgiler

Bu kısımda değerlendiriciye ait demografik bilgiler sorulmaktadır.

Üniversite / Bölüm: ....

Eğitim Durumu: ☐ Lisans ☐ Yüksek Lisans ☐ Doktora

Cinsiyet: ☐ Kadın ☐ Erkek

Yaş: ....

Bölüm 2. Katılımcıların Senaryodaki Performanslarının Değerlendirilmesi

Aşağıda verilen yargılar kapsamında katılımcının performansını 1'den 5'e kadar puanlandırarak değerlendiriniz. *(1: En düşük, 5: En yüksek)*

	1	2	3	4	5
Katılımcının 3B ortamdaki derinlik algısı gelişmişti.					
Katılımcının sağ-sol el kullanma yeteneği gelişmişti.					
Katılımcının cerrahi bir alet ile birlikte endoskobu kullanabilme yetisi gelişmişti.					
Katılımcının el- göz koordinasyonu gelişmişti.					
Katılımcının göz hareketlerini incelediğimde cerrahi aletin anlık konumunu takip ediyordu.					
Katılımcının göz hareketlerini incelediğimde operasyon alanının genelini gözlemliyordu.					
Katılımcı görevleri yerine getirirken kendine güvenen ve sakin bir tutum içerisindeydi.					
Katılımcı görevleri yerine getirirken heyecanlı ve tedirgindi.					

Bölüm 3. Katılımcıların Beceri Seviyesi Değerlendirme Ölçeği

3.1 İzlemiş olduğunuz performans ile ilgili katılımcıyı aşağıda verilen yargılardan size uygun olanı seçerek değerlendiriniz.

Not: Bu bölüm Adrales ve ark. (2003) çalışmasından uyarlanmıştır.

Beceri-1	**0**	**1**	**2**	**3**	**4**
Dokuya Saygı (Dokuya Temas)	Gereksiz kuvvet kullanarak sürekli temas eder.	Dokuya sıklıkla temas eder.	Dikkatlidir ancak zaman zaman temas eder.	Oldukça dikkatlidir, dokuya temas çok azdır.	Sürekli olarak dokuya temas etmeden görevleri tamamlar.
Beceri-2	**0**	**1**	**2**	**3**	**4**
Yetenek (hareket tasarrufu)	Çok fazla gereksiz hareket	Fazla gereksiz hareket	Uygun hareketler ancak bazıları hala gereksiz	Uygun, tasarruflu hareketler	Hassas, maksimum hareket tasarrufu
Beceri-3	**0**	**1**	**2**	**3**	**4**
Seri / Eşzamanlı Karmaşıklık-Operasyonel Akış	Bir sonraki hareketten emin değil	Bazı Görevleri gerçekleştirirken seri, yine de hareketleri belirsiz	Seri görevlerde iyi, makul ilerleme	Seri görevlerde çok iyi, iyi düzeyde ilerleme	Seri görevlerde mükemmel, hatasız ilerleme
Beceri-4	**0**	**1**	**2**	**3**	**4**
Konumsal Durum	Çok düşük, sürekli ortamda kayboluyor.	Ortalama, bir kaç denemeden sonra tekrar konumunu bulabiliyor.	İyi, amacına yönelik hareketler ile konumunu belirliyor.	Çok iyi, sürekli olarak amaca yönelik hareketler	Mükemmel, ortamda hiç kaybolmuyor, konumu net.

3.2 Performansını izlediğim katılımcının beceri seviyesinin .... olduğunu tahmin ediyorum.

- □ Acemi
- □ Uzman

APPENDIX D. QUESTIONNAIRE FOR OBSERVERS

Dear Participant,

This questionnaire was prepared with the aim of gathering information about the study on "Estimation of Surgeons Skill Levels in Surgical Training Processes in 3D Simulation Environment" within the scope of my doctoral thesis in Atilim University Software Engineering Department.

In the questionnaire, no identifiable/personal information is required. Your answers will be kept completely confidential and will only be evaluated by researchers; the information to be obtained will be used in scientific publications. The study consists of two parts and will take you up to 30 minutes. You can contact the researcher Damla Topalli (damla.topalli@atilim.edu.tr) for further information about the study.

Thank you for supporting this work.

Procedure

In this study, hand and eye movements will be shown on the screen by simulating the performances of 15 participants in Hacettepe University Neurosurgery and ENT Departments who are surgeons or interns in a scenario developed in 3D environment. In this scenario, the aim is to collect the objects on an anatomical nasal model using two haptic devices, the surgical instrument and the camera (as endoscope). In this context, hand movements of participants were recorded using tactile device and eye movements were recorded using eye tracking device. The blue pointer you will see on the screen shows the location of the tactile device on the screen, while the "eye" symbol indicates where the participant currently looking at the screen. For each participant, you are expected to evaluate the performance of the person by choosing the appropriate alternative for the given statements about their performance. The statements include parameters thah should be considered during the operation such as tissue contact, economy of movement, operational flow, spatial orientation in the 3D environment of the tool, depth perception, ability to use the non-dominant hand (camera control), hand-eye coordination and improving 3D depth perception. At the end, you are expected to predict the skill level of the participant, Intermediate or Novice.

Observer ID:

Assessed Participant ID:

PART- I. Demographics

In this section, the demographic information of the observer is asked.

University /Department: ....

Educational Status: ☐ Bachelor ☐ Masters ☐ Doctorate

Gender: ☐ Female ☐ Male

Age: ....

PART- II. Evaluating Participants' Performances in the Scenario

Assess the participants' performance that you currently observed by choosing the appropriate statements in one of the five alternatives (1: Strongly Disagree, 5: Strongly agree).

	1	2	3	4	5
Participant shows developed depth perception skills in a 3D environment					
Participant shows developed skills to use left-right hand in coordination.					
Participant shows developed skills to use endoscope with a surgical instrument in coordination.					
Participant shows developed skills on hand-eye coordination					
When I examined the participants' eye movements, they were monitoring the instant position of the tool.					
When I examined the participants' eye movements, they observed the general area of operation.					
The participant was confident and calm while performing the surgical tasks.					
The participant was excited and nervous while performing the surgical tasks.					

PART- III. Participants' Skill Level Assessment

3.1 Assess the participant based on his/ her performance that you currently observed. About the following expressions given, choose one of the five alternatives (from 0 to 4).

Note: This part is adapted from the rating scale in Adrales et al. (2003)'s study.

Skill-1	**0**	**1**	**2**	**3**	**4**
Clinical Judgement- Respect for Tissue	Uses unnecessary force	Shows an attempt to respect tissue	Careful but occasionally disrespectful	Very Careful in handling Tissue	Consistently handles tissue with care
Skill-2	**0**	**1**	**2**	**3**	**4**
Dexterity (economy of movement)	Many gross-unnecessary movements	Some gross-unnecessary movements	Appropriate movements but some are still unnecessary	Fine and economical movements	Precise and maximum economy of movement
Skill-3	**0**	**1**	**2**	**3**	**4**
Serial /Simultaneous Complexity (Flow of operation)	Unsure of next move	Some knowledge of serial tasks but still uncertain	Good knowledge of serial tasks; reasonable progression	Very good knowledge of serial tasks; good progression	Excellent knowledge of serial tasks; effortless progression
Skill-4	**0**	**1**	**2**	**3**	**4**
Spatial Orientation	Poor spatial orientation, consistently lost in space.	Moderate spatial orientation; able to adjust after several attempts.	Good Spatial orientation; purposeful movement in space	Very Good Spatial orientation; consistently purposeful movement	Excellent Spatial orientation; precise and purposeful

3.2 I guess that the participants' skill level, which I have currently observed, is ....

□ Novice

□ Intermediate

APPENDIX E. CLASSIFICATION RESULTS ON HAND MEASURES

Table E.1. Classification Results of Scenario-1 – BH Condition using All Features

Algorithm	Option	Accuracy	Correctly Classified	TPR	FPR	Precision	Recall	F-Measure
AdaBoost		**78.57**	22	0.75	0.18	0.75	0.75	0.750
Simple Logistic		64.28	18	0.58	0.31	0.58	0.58	0.583
Decision Tree J48		60.71	17	0.58	0.37	0.54	0.58	0.560
Random Forest		60.71	17	0.33	0.18	0.57	0.33	0.421
PART		60.71	17	0.5	0.31	0.54	0.500	0.522
KNN	K=1	57.14	16	0.58	0.43	0.5	0.58	0.538
SVM	Puk	57.14	16	0	0	?	0	?
Regression		57.14	16	0	0	?	0	?
SVM	Normalized Polykernel	57.14	16	0	0	?	0	?
Bagging		57.14	16	0	0	?	0	?
LWL		57.14	16	0.66	0.5	0.5	0.66	0.571
KNN	K=3	53.57	15	0.58	0.5	0.46	0.58	0.519
KNN	K=7	53.57	15	0.66	0.56	0.47	0.66	0.552
Jrip		53.57	15	0.58	0.5	0.47	0.58	0.519
Naive Bayes		53.57	15	0.58	0.5	0.46	0.58	0.519
KNN	K=5	53.57	15	0.66	0.56	0.47	0.66	0.552
Logistic		53.57	15	0.58	0.5	0.46	0.58	0.519
KNN	K=9	50	14	0.75	0.69	0.45	0.75	0.563
SVM	PolyKernel	50	14	0.58	0.56	0.44	0.58	0.500
	Max Accuracy	**78.57**	**22**	**0.75**	**0.69**	**0.75**	**0.75**	**0.75**

Table E.2. Classification Results of Scenario-1 – BH Condition using Selected Features

				Selected Features- ReliefF (34)				
Algorithm	**Option**	**Accuracy**	**Correctly Classified**	**TPR**	**FPR**	**Precision**	**Recall**	**F-Measure**
KNN	K=1	**85.71**	24	0.91	0.18	0.78	0.91	0.846
Simple Logistic		75	21	0.75	0.25	0.69	0.75	0.720
SVM	Puk	75	21	0.5	0.06	0.85	0.5	0.632
AdaBoost		75	21	0.75	0.25	0.7	0.75	0.72
KNN	K=3	71.42	20	0.83	0.37	0.62	0.83	0.714
KNN	K=7	71.42	20	0.83	0.37	0.62	0.83	0.714
KNN	K=9	71.42	20	0.83	0.37	0.62	0.83	0.714
Regression		67.85	19	0.83	0.43	0.58	0.83	0.690
Jrip		67.85	19	0.5	0.19	0.66	0.5	0.571
Naive Bayes		64.28	18	0.58	0.31	0.58	0.58	0.583
SVM	PolyKernel	64.28	18	0.67	0.37	0.57	0.67	0.615
SVM	Normalized Polykernel	64.28	18	0.67	0.37	0.57	0.67	0.615
KNN	K=5	64.28	18	0.75	0.44	0.56	0.75	0.643
Bagging		64.28	18	0.41	0.18	0.62	0.41	0.500
Decision Tree J48		64.28	18	0.58	0.31	0.58	0.58	0.583
Random Forest		64.28	18	0.5	0.25	0.6	0.5	0.545
Logistic		60.71	17	0.5	0.31	0.54	0.5	0.522
PART		57.14	16	0.58	0.43	0.5	0.58	0.538
LWL		53.57	15	0.58	0.5	0.47	0.58	0.519
	Max Accuracy	**85.71**	**24**	**0.91**	**0.5**	**0.85**	**0.91**	**0.846**

Table E.3. Classification Results of Scenario-3 – BH Condition using All Features

		ALL Features (101)						
Algorithm	**Option**	**Accuracy**	**Correctly Classified**	**TPR**	**FPR**	**Precision**	**Recall**	**F-Measure**
KNN	K=3	**76.92**	20	0.75	0.21	0.75	0.75	0.750
SVM	Normalized Polykernel	50	13	0.08	0.14	0.33	0.08	0.133
KNN	K=5	76.92	20	0.66	0.14	0.8	0.66	0.727
KNN	K=7	61.53	16	0.33	0.14	0.66	0.33	0.444
Naive Bayes		38.46	10	0.08	0.35	0.17	0.08	0.111
KNN	K=1	53.84	14	0.5	0.42	0.5	0.5	0.500
KNN	K=9	53.84	14	0.16	0.14	0.5	0.16	0.250
Random Forest		53.84	14	0.33	0.28	0.5	0.33	0.400
SVM	PolyKernel	50	13	0.41	0.43	0.45	0.42	0.435
SVM	Puk	53.84	14	0	0	?	0	?
Regression		53.84	14	0	0	?	0	?
PART		42.31	11	0.41	0.57	0.38	0.41	0.400
Decision Tree J48		42.31	11	0.41	0.57	0.38	0.41	0.400
Simple Logistic		34.61	9	0.25	0.57	0.27	0.25	0.261
AdaBoost		38.46	10	0.25	0.5	0.3	0.25	0.273
Logistic		46.15	12	0.33	0.43	0.4	0.33	0.364
LWL		38.46	10	0.33	0.57	0.33	0.33	0.333
Bagging		50	13	0.08	0.14	0.33	0.08	0.133
Jrip		42.31	11	0.25	0.42	0.33	0.25	0.286
	Max Accuracy	**76.92**	**20**	**0.75**	**0.57**	**0.8**	**0.75**	**0.667**

Table E.4. Classification Results of Scenario-3 – BH Condition using Selected Features

		Selected Features- ReliefF (42)						
Algorithm	**Option**	**Accuracy**	**Correctly Classified**	**TPR**	**FPR**	**Precision**	**Recall**	**F-Measure**
KNN	K=3	**80.76**	21	0.81	0.22	0.85	0.81	0.797
SVM	Normalized Polykernel	76.92	20	0.6	0.14	0.8	0.66	0.727
KNN	K=5	73.07	19	0.5	0.07	0.857	0.5	0.632
KNN	K=7	73.07	19	0.5	0.07	0.857	0.5	0.632
Naive Bayes		69.23	18	0.58	0.21	0.7	0.58	0.636
KNN	K=1	69.23	18	0.58	0.21	0.7	0.58	0.636
KNN	K=9	69.23	18	0.5	0.14	0.75	0.5	0.600
Random Forest		69.23	18	0.67	0.28	0.66	0.66	0.667
SVM	PolyKernel	57.69	15	0.5	0.35	0.54	0.5	0.522
SVM	Puk	57.69	15	0.25	0.14	0.6	0.25	0.353
Regression		57.69	15	0.66	0.5	0.53	0.66	0.593
PART		57.69	15	0.33	0.21	0.57	0.33	0.421
Decision Tree J48		53.84	14	0.33	0.28	0.5	0.33	0.400
Simple Logistic		50	13	0.41	0.43	0.45	0.41	0.435
AdaBoost		50	13	0.58	0.57	0.46	0.58	0.519
Logistic		46.15	12	0.41	0.5	0.41	0.41	0.417
LWL		46.15	12	0.41	0.5	0.41	0.41	0.417
Bagging		46.15	12	0.41	0.5	0.41	0.41	0.417
Jrip		38.46	10	0.16	0.42	0.25	0.16	0.200
	Max Accuracy	**80.76**	**21**	**0.81**	**0.57**	**0.857**	**0.81**	**0.797**