USING FUZZY SET THEORY TO OBJECTIVELY EVALUATE PERFORMANCE ON MINIMALLY INVASIVE SURGICAL SIMULATORS

by

Ima Hajshirmohammadi
B.Sc., Isfahan University of Technology, 2001

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

In the
School of Engineering Science

SIMON FRASER UNIVERSITY

Spring 2006

APPROVAL

Name: Ima Hajshirmohammadi

Degree: Master of Applied Science

Title of Thesis: Using Fuzzy Set Theory to Objectively Evaluate Performance on Minimally Invasive Surgical Simulators

Examining Committee:

Chair: **Dr. Stephen Hardy**
Professor of Engineering Science

____

Dr. Shahram Payandeh
Senior Supervisor
Professor of Engineering Science

____

Dr. Stephen N. Robinovitch
Supervisor
Associate Professor of Kinesiology

____

Dr. John D. Jones
Internal Examiner
Associate Professor of Engineering Science

Date Defended: Jan. 16/06.

STATEMENT OF ETHICS APPROVAL

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

(a) Human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

(b) Advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

(c) as a co-investigator, in a research project approved in advance,

or

(d) as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

ABSTRACT

Objective surgical performance evaluation is a non-linear and ambiguous problem and hard to model with classic mathematical methods. This thesis explores employing fuzzy set theory as a novel approach to this problem, since the main strength of fuzzy logic is its ability to handle the vagueness and non-linearity of the everyday experiences.

Using a commercial surgical simulator, data were collected from subjects who participated in user study of two surgical procedures. Half of these data were used to design four fuzzy models for surgical skills classification. The remaining data were used to test the constructed models and to investigate the effects of various fuzzy inference properties on their performances.

Our results indicate satisfactory correlation between the surgical skill levels predicted by the fuzzy models and the actual skill levels of the user. Thus, fuzzy classifiers can be considered as effective tools to handle the fuzziness of objective performance evaluation.

Keywords:

Surgical performance evaluation, Objective performance assessment, Minimally invasive surgical simulators, Surgical skill level, Fuzzy classifiers.

DEDICATION

To my caring husband and my supportive parents

ACKNOWLEDGEMENTS

I wish to thank my senior supervisor, Professor Shahram Payandeh, for his guidance and support throughout the course of this work. Professor Payandeh directed my attention and interest towards the subject of this thesis, and his continuous attention and valuable discussions made the completion of this work possible. His suggestions, advice, and support are greatly appreciated.

Many thanks to the staff of the School of Engineering Science for their prompt and helpful responses to all my queries.

I would like to thank my friends Veis, Vamegh, and Shaghayegh, and the Khayambashi, Dokhani, and Sokhansanj families. Their genuine friendship and support made my stay in Vancouver much more pleasant. My very special thanks to my dear friend Somayeh, who has always been there for me, even when we were miles apart.

My most sincere gratitude to my immediate and extended family for their help and support in every aspect of my life, and for their particular interest in my academic career. Countless thanks to my dear husband, Saba, for all his love and care, for being my inspiration and source of strength, and for his constant support ever since I got to know him. Special thanks go to my loving and supportive parents; as my first and best teachers they taught me how to love, care, and live with pride. Many thanks to my mentor and best friend, my sister, Atousa. Words cannot express my appreciation of her help and support during these years, and especially during the last days of this work.

My deepest thanks to Persi, whose companionship brightened my loneliest days. The happiness and joy she brought me in her short life is immeasurable. She will always be loved and remembered.

Above all, I am thankful to God for giving me the opportunity for a fulfilling life, and the strength, dedication and ability to pursue and achieve my goals so far.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Minimally Invasive Surgery

Contrary to the traditional surgical approaches which have used incisions designed to provide maximum exposure of the operative site, Minimally Invasive Surgery (MIS) uses small incisions through which cameras and instruments are passed to accomplish the operation from within a body cavity.

MIS was introduced to the world of medicine first by demonstrating a cystoscopy in the early 1800's in France, and was then occasionally used for procedures such as gastroscopy, endoscopic investigation of the abdominal cavity, appendectomy, etc. until the late 1900's (Wayand, 2004). The MIS technology was developed during these years, but it was in 1989 that the video demonstration of a laparoscopic cholecystectomy (surgical removal of the gallbladder) marked the starting signal for MIS throughout Europe. In the course of the following two years, practically all European countries, and then other parts of the world started to operate laparoscopically (Wayand, 2004).

The MIS techniques of surgical access and exposure have a great advantage over traditional incisions by significantly reducing trauma to the body, which results in better cosmetic outcome, reduction of post-operative pain, earlier patient recovery time, and therefore faster patient discharge from the hospital (NYU medical, 2005).

With the introduction of minimally invasive surgery however, surgeons must learn radically new and complex skills and procedures. Traditional methods of training that were adequate for conventional surgery may not be as effective in demonstrating and teaching these substantially new skills. A new standardized structured program for minimally invasive surgery training has been of interest to many researchers over the past decade. Recent efforts to develop such programs have generally involved the use of training boxes or computer-based virtual reality simulations (Rosen, Hannaford, Richards, & Sinanan, 2001; Chen, Yeasin, & Sharma, 2003; Stylopoulos et al., 2004).

1.2 Training

Studies have shown that minimally invasive procedures such as laparoscopic cholecystectomy may have a higher rate of operative complications than the traditional open surgery. For instance, *Regan, Yuan, and McAfee in 1999 investigated the feasibility and safety of minimally invasive spine surgery compared with that of open procedures. Their study concluded that the complications were higher for the minimally invasive approach (4.9%) compared to the traditional approach (4.2%). (Regan, Yuan, & McAfee, 1999). Another study has shown that* the odds ratio for intra-operative injury in laparoscopic cholecystectomy compared with open cholecystectomy was 1.79, *in Western Australia* in the period 1988 to 1994 (Fletcher et al., 1999*).*

The higher rate of complications in MIS could be due to the following limitations of these procedures: First, a laparoscopic surgeon has only a two-dimensional view, with a restricted field of vision. Second, physical feedback is limited, and thus visual cues become more critical in identifying anatomy. Additionally, the increasing reliance of the surgeon on technology adds an intangible layer of separation between the doctor and patient.

Although these limitations are real, research shows that laparoscopic procedures are associated with a learning curve, which once mastered, are effective and safe when compared with traditional techniques of surgery (Regan, Yuan, & McAfee, 1999).

Surgical skills are principally obtained under the long-established apprenticeship model in animal and human subjects. Residents learn by observation followed by participation, taking more active roles in the operation as their experience increases (Moorthy, Munz, Sarker, & Darzi, 2003).

However, methods of training that were suitable for the traditional surgical scheme may not be as efficient in training significantly new MIS procedures. Performing MIS involves a multi-dimensional series of tasks requiring a combination of visual information and the kinematics and dynamics of the surgical tools (Rosen et al., 2001).

Obtaining these skills under the traditional apprenticeship model requires extensive number of hours of training in the operating rooms. This is in face of the increasing limitations on the time and resources needed to train surgeons, which have reduced the opportunity to acquire surgical skills in the operating theatre (Moorthy et al., 2003). This has led to a further emphasis on finding innovative ways to teach or enhance the MIS skills outside the operating room (Tendick et al., 2000).

Virtual reality training has been a long-term goal of numerous investigators and has been proposed as a method to both instruct surgical students and objectively evaluate their skills in performing surgical operations. (Rosen et al., 2001; Chen et al., 2003; Stylopoulos et al., 2004). Researchers from different backgrounds such as engineers, scientists, computer programmers, anatomists, and experienced surgeons are making effort to improve these systems by creating more realistic surgical scenarios and practice sessions that help residents master their skills during their course of training.

Computer simulators are used for training in other industries such as military, nuclear, maritime, transportation and most prolifically, aviation. Flight simulators have been of interest since the early 1900s and have been used to familiarize flight crews in normal and emergency operating procedures. In such training programs the trainee does not advance to operating the real system (aircraft, tank, automobile, ship, etc.) until a set of pre-defined criteria as determined by testing "experts" with their performance level as the gold standard, has been met on the simulator (Simulators for Training, 2003). Similar approach could be taken in the surgical field; the resident does not operate upon a patient until the satisfactory performance level has been achieved on the simulator.

Using simulators as training tools has several advantages for both the trainees and the educational system. A less stressful training environment, lower training costs, and a lower risk to the trainees are to name a few. For example in the aviation industry, because powered flight is hazardous to attempt untrained, flight simulators are used to enable new pilots to get the feel of the controls without actually being airborne. In addition, pilots are able to train for situations that they would be unable to do safely in an actual aircraft, such as complete power loss, engine fires, electrical faults, storms, slippery runways, navigational system failures and countless other problems which the crew need to be familiar with and act upon (Wikipedia, Modern simulators, ¶ 4).

Translated into the surgical environment, computer simulators provide a less stressful learning environment for the surgical students by eliminating the risks of operating on real patients. In addition, they allow the trainees to deal with emergency operating procedures as well as rare but critical surgical scenarios that the students may not necessarily face in the operating theatre during their course of training. To master their skills, the students also have the option of practicing their weaknesses on the simulators as many times as they want, without having to face the limitations of real surgical operations. Other advantages of using simulators as surgical training tools include a lower risk to the patient, a lower cost, and a standardized method of skills evaluation (Richards, Rosen, Hannaford, Pellegrini, & Sinanan, 2000).

1.2.1 Performance Evaluation

In recent years, there has been a growing interest in competence assessment in medical practice and especially so in surgery (Darzi, 2001). Until very recently, the only method of evaluating the level of competence in minimally invasive surgery has been based on examinations, log books, opinion of experienced surgeons observing the operation on the viewing monitors, and the overall outcome of the procedure (Cuschieri, 2001; Moody, Baber, Arvanitis, & Elliott, 2003). All these methods are largely *subjective* and lack validity and reliability. As mentioned by Moorthy, Munz, Sarker, and Darzi (Moorthy, Munz, Sarker, & Darzi, 2003), deficiencies in training and performance are difficult to correct without an *objective* feedback, making it essential to have a standardized objective method for assessment of surgical competency. He addresses a series of objective surgical assessment methods such as checklists, global rating scales, and dexterity analysis systems, which have been developed in recent years and are being used in various surgical training programs (Moorthy et al., 2005).

Checklists are one of the commonly used methods of performance evaluation. The residents are required to perform a series of procedures while being observed by an expert. There is a different checklist for each procedure, specifying the steps that need to be taken during the operation. The supervisor verifies whether the student has performed or missed any particular step (The royal Australian, Surgical Skills Competence, ¶ 1; A. Nagy, personal communication, December 4, 2003).

The Objective Structured Assessment of Technical Skills (OSATS) is an example of a global rating scales system which was developed in Toronto, Canada (Martin et al., 1997). The system consists of six stations for residents and trainees to perform procedures in a fixed time period on live animals or bench models. An expert observes and evaluates the students during the performance of tasks at each station by using checklists specific to the operation or task and a global rating scale (Moorthy et al., 2005).

It has been said that using checklists removes the subjectivity of the evaluation process by turning the examiners into observers, rather than interpreters of behavior (Regehr, MacRae, Reznick, & Szalay, 1998). However, using checklists and global rating scales could be time consuming, as it requires the presence of multiple expert observers at the examination scene or extensive video watching (Moorthy et al., 1997).

One of the important aspects of technical skill, especially in MIS, is the psychomotor skills or the dexterity required to perform the planned procedure (Darzi, Smith, & Taffinder, 1999; Darzi, 2001). This is more than being able to perform a procedure with quick and fluent movements that may look impressive to an observer. As Darzi (2001) points out, "it includes, for example, being able to suture tissue accurately and tie knots that are functional and prevent fluid leaking, but are not so tight as to cause tissue damage" (Darzi, 2001).

Dexterity analysis systems such as the Imperial College Surgical Assessment Device (ICSAD) or the Advanced Dundee Endoscopic Psychomotor Trainer (ADEPT) are developed to objectively evaluate the dexterity of surgical residents and trainees (Moorthy et al., 1997). ICSAD is a commercially available electromagnetic tracking system (Isotrak II, Polhemus, United States) and consists of an electromagnetic field generator and two sensors that are attached to the surgeon's hands at standardized positions. The positional data generated by the sensors is converted to dexterity measures such as the number and speed of hand movements, the distance travelled by the hands and the time taken for the task.

Many studies have established the validity of dexterity analysis systems for surgical performance evaluation (Datta, Mackay, Mandalia, & Darzi, 2001; Taffinder, Smith, Mair, Russell, & Darzi, 1999; Smith, Torkington, Brown, Taffinder, & Darzi, 2002). One drawback of this approach however, is that it is impossible to ensure standardization, since the students operate on real patients and all patients are different. In addition, other factors such as patients' safety can affect the performance of the trainees (Moorthy et al., 1997).

After the important role of computer-based simulators in MIS training was realized, attention was turned towards development of objective technical skill evaluation methods on the simulators. Computerized systems have enabled the recording of quantitative dexterity parameters including time to perform the surgical tasks, economy of hand movements, smoothness of instrument motion and applied forces, which cannot be measured easily with conventional instruments alone. New technologies have been developed in the past decade that address the issues of objective assessment of surgical dexterity to some extent.

1.3 Related Work

Chen et al. (2003) specify three requirements for a system to accurately measure the technical competence of the surgical performance: (1) the system must have adequate sensing techniques to monitor the user's operation; (2) the system must extract appropriate features from the sensing data; and (3) the system needs a good computational model to generate a "score", representing the skill demonstrated in the operation based on the relevant sensing data (Chen et al., 2003).

Computer based surgical simulators satisfy the first requirement by having the ability to monitor and record the performance of operating users. The other two requirements however, have been the main challenge in defining applicable methods of performance evaluation. Several researchers over the past decades have focused on identifying reliable metrics that are representative of users' dexterity in a procedure, and a system to translate those metrics into values that correlate with the user's real level of expertise.

Operative speed is known as an important factor in objective measurement of technical skill. Van-Rij et al. (1995) have used time to quantify skill in junior surgeons (Van-Rij et al., 1995), and Hanna, Shimi, and Cuschieri (1998) have utilized time as a means of skill evaluation in experienced surgeons (Hanna, Shimi, & Cuschieri, 1998). However, evaluating competence simply by setting time targets for certain procedures is crude and unacceptable. As Darzi et al. (1997) point out, "a fast surgeon is not necessarily a good surgeon". (Darzi et al., 1997).

Other studies have used electromagnetic, mechanical, and optical motion tracking systems to analyze the hand and tool movements during surgical operations. Software is used to convert the positional data generated by the sensors to dexterity measures (Darzi, 2001).

The Imperial College Surgical Assessment Device (ICSAD), is an example of the electromagnetic motion tracking systems that has been used in several studies to determine a number of dexterity metrics such as the number and speed of hand movements, the distance traveled by the hands and the time taken for the task at hand. (Taffinder, Smith, Huber, & Russell, 1999; Datta et al., 2002).

The Blue DRAGON (Brown, Rosen, Chang, Sinanan, & Hannaford, 2004) is another system for acquiring the kinematics and the dynamics of the endoscopic tools during an operation. It includes two four-bar mechanisms equipped with position and force/torque (F/T) sensors for measuring the positions and orientations of two endoscopic tools along with the forces and torques applied by the surgeon's hands. In addition, the synchronized view of the surgical scene is incorporated into a graphical user interface displaying the data in real-time. For each surgical task different types of the tool-tip/tissue interaction are decomposed into discrete tool manoeuvres (states), each with a unique F/T pattern using a fully connected, finite-states Markov model. Their study showed that major differences between residents at different skill levels were: the types of tool/tissue interactions being used, the transitions between tool/tissue interactions being applied by each hand, time spent while performing each tool/tissue interaction, the overall completion time, and the variable F/T magnitudes being applied by the subjects through the endoscopic tools.

Chen et al. (2003) also used Hidden Markov Models (HMMs) to model and evaluate hand movements in a typical surgical exercise such as surgical knot-tying (Chen et al., 2003)[1]. They used a video-based technique for tracking the hand movements during a surgical knot-tying task. Their method for surgical skill assessment is based on the log probability of an observation sequence for a specific skill model. The probability measures the stochastic similarity between the performance of the observation sequence and the performance represented by the model – the higher the probability, the closer is the observation sequence to the model. Although they only consider the hand movements in their analysis of performance evaluation, they suggest that the same HMM-based approach could be taken to model other components of human skills, such as the forces applied by hands, the orientation of the hands, and the hand-eye coordination.

Payandeh, Lomax, Dill, MacKenzie, and Cao in their studies in 2002 video taped a series of surgical tasks performed in an animal lab and conducted time-line studies of tool movements during the operations. By evaluating the video taped training sessions they showed that surgical tasks could be decomposed into a series of subtasks. They identified and analyzed five basic motions that the surgeon/tool performed during various procedures: reach and orient, grasp and hold/cut, push, pull, and release. The study concluded that the length of time taken to complete these subtasks could be a measure of performance differences between novices and experts (Payandeh, Lomax, Dill, MacKenzie, & Cao, 2002).

A different approach was taken by Cotin et al. (2002) (Cotin et al., 2002). Their system, the Computer-Enhanced Laparoscopic Training System or CELTS (Stylopoulos et al.,

---

[1] Hidden Markov Models are capable of characterizing two embedded stochastic processes with one underlying process that is not observable, but can only be observed through another process that produces observation sequences. In the case of surgical knot tying for example, the skill of tying surgical knots is the hidden stochastic process, and the other process is a video sequence of continuous hand movements during the operation of tying a surgical knot.

2004), uses a five degree of freedom device capable of tracking the motion of two laparoscopic instruments. Using the kinematics analysis method, software converts the raw data into the following parameters: the spatial distribution of the tip of the instrument, smoothness of motion, depth perception, response orientation, and ambidexterity. Also included in the measurements is the time to perform the task and outcome of the task as two other aspects of success of a procedure. At the end of each procedure, the software calculates a standardized z-score for each of the performance parameters by computing their distances from the results set by a group of experts. A final score is then determined for each instrument by calculating a weighted average of all the z-scores, providing instant feedback for the user.

Commercial systems such as The Minimally Invasive Surgical Trainer-Virtual Reality, MIST-VR (Mentice, Gothenburg, Sweden) have also developed methods of surgical skills assessment. At the end of each procedure, the MIST-VR provides a "score" for the user by calculating a weighted average of the user's performance metrics. Several studies to date have confirmed that the MIST-VR has validity as an assessment tool (Torkington, Smith, Rees, & Darzi, 2001; Jordan, Gallagher, McGuigan, & McClure, 2001).

Recently, a non-traditional mathematical approach was proposed, using fuzzy logic to evaluate surgical performance as judged by expert surgeons (Hajshirmohammadi & Payandeh, 2005; Huang, Payandeh, Doris, & Hajshirmohammadi, 2005).

1.4 Motivation

Despite the recent efforts to develop standardized structured performance evaluation systems in minimally invasive surgery, none of these models has been widely accepted and officially integrated into a surgical training program or any other official training course. The main challenge in designing a standard method of evaluation is to design a scoring scheme that correlates with the subjective opinion of experienced surgeons, or in other words, to formulate the expert's judgment. Formulating a subjective opinion is not a simple task, as human beings consider an enormous number of factors in their decisions, and there are times that even themselves cannot specify reasons behind their decisions.

During interviews with experienced laparoscopic surgeons (A. Nagy, personal communication, December 4, 2003), it was found to be impossible to get a structured answer to the question "What is called a satisfactory performance in MIS?". As they described, "surgery is a combination of art and science… we cannot look at two procedures and call one of them superior to the other, as long as they both have satisfactory outcomes".

Experts' opinions seem to be too ambiguous and fuzzy to be formulated with conventional mathematical techniques. That triggered the idea of using Fuzzy Set Theory for this purpose, as the main strength of fuzzy logic is its ability to deal with the vagueness and ambiguity in human's natural language (Cox, 1999; Fuzzy Logic, 1997).

Over the past few decades, fuzzy logic has been introduced to successfully solve the problem of imprecision or fuzziness common in various fields of study such as sociology, physics, biology, finance, marketing, engineering, psychology, health management, and computer programming. Fuzzy logic has also been effectively employed in computer simulated systems. Examples include development of a fuzzy logic performance control system to reduce variable error and overshoots in a reconfigurable general aviation simulator (Beringer, 2002). Ota, Loftin, Saito, Lea, and Keller (1995) also outline how the task of dissecting a blood vessel in a virtual environment can be evaluated by fuzzy logic (Ota, Saito, Lea, & Keller, 1995).

This work explores the possibility of employing fuzzy set theory to evaluate the performance of the users of minimally invasive surgical simulators. We proposed a simple method to develop a fuzzy-based if-then rule system, or more specifically a fuzzy classifier based on performance data collected in a virtual reality trainer.

Research shows that surgical procedures can be divided into a number of tasks and subtasks, and overall surgical competence can be evaluated based on performance in each task (Payandeh et al., 2002). Thus, in this project, the focus was to design fuzzy systems that can objectively evaluate a trainee's skill level on two important tasks in MIS: laparoscopic suturing and laparoscopic knot-tying.

The long-term goal of this research is to create an automatic skill evaluation scheme to be incorporated in computer-assisted surgical training systems. With proper assessment and validation, such systems can provide both initial and ongoing assessment of operator skill throughout one's career, while enhancing patient safety through reduced risk of intraoperative error. Additionally, a computerized trainer can provide either terminal (post-task completion) or concurrent (real time) feedback during the training episodes, enhancing skills acquisition.

1.5 Contribution

This project explores the feasibility of fuzzy logic-based performance classifiers as a novel solution to the subjective nature of the traditional methods of surgical skills evaluation. The goal was to develop a method that is capable of *objectively* predicting a surgeon's level of expertise, based on his/her performance metrics during a procedure on a surgical simulator. To achieve this goal, a user study was conducted to collect performance data from 26 subjects in three different surgical skill levels: novice, intermediate, and expert. Each user performed two surgical tasks on a virtual reality surgical simulator, namely laparoscopic suturing and laparoscopic knot-tying.

The collected data were divided into two halves: the training dataset which was used to train the fuzzy models, and testing dataset to test the constructed systems. Two different approaches were taken to separate the data into the testing and training datasets, as explained in section 4.

A new algorithm was developed to design four fuzzy classifiers, two for each surgical task. Various fuzzy inference properties were then applied to the constructed classifiers, and the optimal combination of fuzzy inference properties were determined for the models. The performances of the classifiers were then tested with the testing datasets, and the reliability of each model was determined by calculating the number of systems' correct answers and the amount of error in the systems' results.

A simple statistical method was also used to analyse the user study data and its results were compared with the fuzzy models' outcomes.

The results were promising. It was concluded that fuzzy classifiers may have the potential to effectively handle the complexity and fuzziness of objective surgical performance evaluation.

1.6 Thesis Organization

The organization of this thesis is as follows:

Section 1 introduces the minimally invasive surgery and the current issues in surgical performance evaluation. Research works related to this topic are also discussed in this section.

Section 2 describes the process of collecting users' performance data on a surgical simulator. Some analyses of the raw data collected in the user study are also discussed.

Section 3 provides a brief background information about fuzzy logic and fuzzy set theory. The history and applications of fuzzy logic are discussed. Basic principles of fuzzy set theory and the design process of fuzzy inference systems are also explained.

Section 4 is a step-by-step explanation of the design process of fuzzy classifiers for surgical performance evaluation. The primary analysis of performance of classifiers are also demonstrated.

Effect of various fuzzy inference properties on performance of the classifiers are explored in section 5, and the optimal combination of these properties for the designed fuzzy models are identified. The reliability of the classifiers are also verified.

The results are summarized in section 6 , and compared with results of a simple statistical approach.

Finally section 7 concludes this thesis.

2 DATA COLLECTION

2.1 Method

We conducted a user study to test surgical performance of subjects of different MIS skill levels and collected their performance data to design and test a fuzzy classifier for each of the surgical tasks in the study. The Minimally Invasive Surgical Trainer-Virtual Reality (MIST-VR) (Mentice Corp., 2004) was used to conduct the user study. Subjects were selected and categorized into three different skill levels based on their MIS experience, and were to complete two surgical tasks available in MIST-VR.

This experiment, which was a follow up on our preliminary user study (Huang et al., 2005), provided us with promising results and a probable new way of categorizing surgical performance in computer-based simulators (For a summary of our preliminary study, please refer to Appendix A: Preliminary User Study).

2.2 User Study

2.2.1 Experimental Set Up

The user study was conducted in an isolated room in Surrey Memorial Hospital, BC, Canada.

Our set up (Figure 1) included:

• A two-handed laparoscopic device with needle-driver handles (Virtual Laparoscopic Interface, by Immersion Inc.)

• A dual Intel Xeon 2.8Ghz computer

• A 19” eye-level LCD monitor

Figure 1: Experimental Set up

The simulator that we used for the study was the MIST-VR, which is a fully validated (MIST User Manual, 2002), commercially available laparoscopic simulator. It has been shown that training on MIST-VR has lead to faster adaptation to the novel psychomotor restrictions encountered by laparoscopic surgeons (Torkington et al. 2001; Jordan, Gallagher, McGuigan, & McClure, 2001).

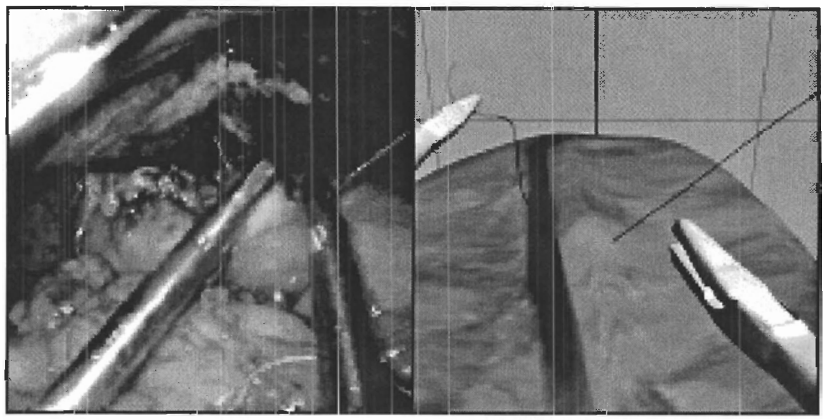
MIST-VR has tutorial, examination, analysis, and configuration modes. To prepare the surgery residents for the operating rooms, the system offers a series of tasks, from basic laparoscopic concepts such as target manipulation and placement, transferring objects between instruments, and diathermy, to more complicated surgical tasks such as laparoscopic suturing and knot-tying. The user's performance metrics are measured and recorded in a database for later reference and performance assessment.

For this study, the plan was to employ complicated surgical tasks to increase the chances of finding a noticeable pattern in performance data of users in different levels of expertise. We selected the **Stitch** and the **Half-Square Knot** tasks defined in MIST for this experiment, as laparoscopic stitch and knot tying skills are considered to be of the most complex tasks in MIS (Tendick et al., 2000).

Stitch Task

Figure 2: The Stitch Task

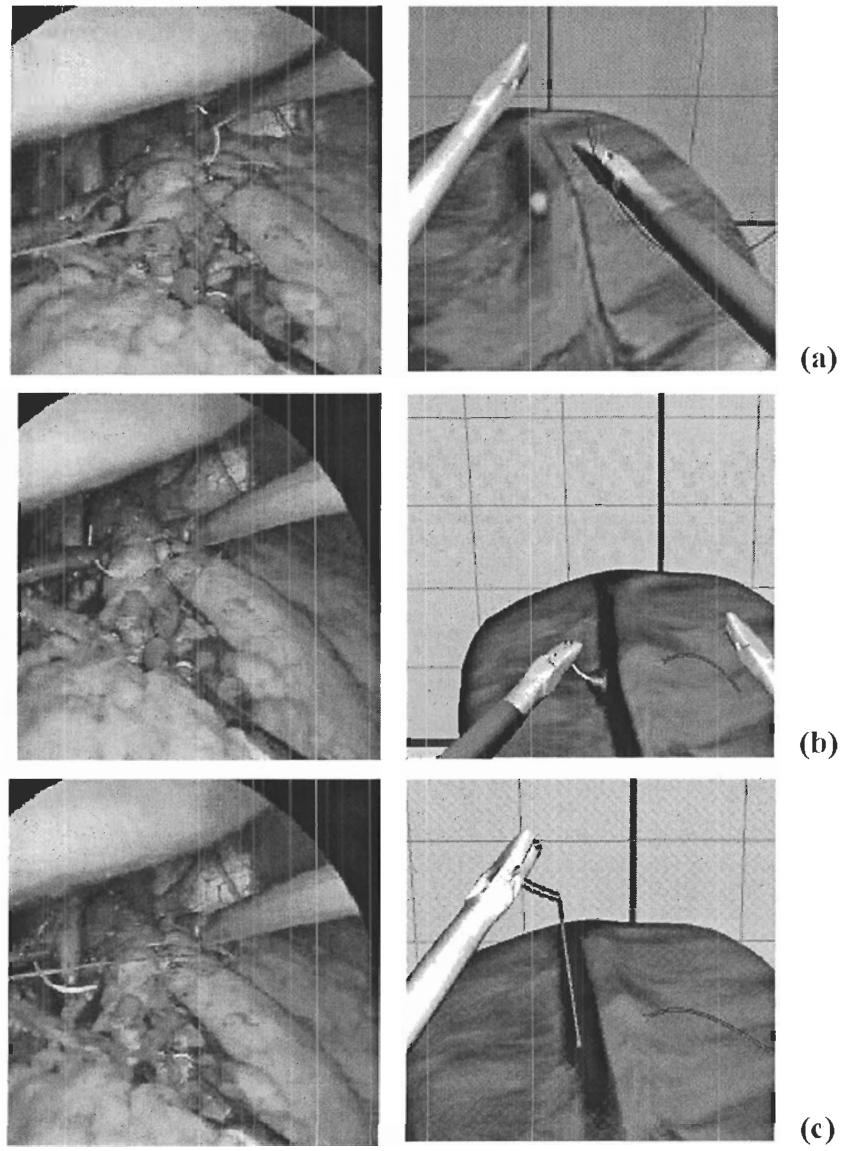
Left: Stitching in real operation (Adopted from: MIST Suturing Module, 2003, by permission) , Right: MIST-VR stitch task (Source: MIST Suturing Module, 2003, by permission)



This task is a test of the user's ability to pierce the tissue with the needle and then pull the needle through the tissue to make a complete stitch (Figure 2 - Right). The objective is to fulfil a complete stitch with maximal accuracy and minimal tension on the tissue. This skill involves accurate 3-D placement, penetrating force with minimal strain to the tissue and curved movement of the needle.

Visual clues guide the user throughout the whole task. For example, a green band around one of the two grippers indicates which one to use at the moment, or target spheres on the tissue mark the correct location to penetrate or pull out the needle (Figure 3 - a).

Figure 3: **Snap shots of MIS stitching in real operation (Left column) and in MIST-VR (Right column) for a right-handed.**
(a) Penetrating the needle into the tissue with the right tool. (b) Grabbing the needle with the left tool to exit the tissue. (c) Pulling out the needle to complete the stitch user (Adopted from: MIST Suturing Module, 2003, by permission)



The task starts with the tool corresponding to the user's dominant hand being active[2] (e.g. the right tool becomes active for a right-handed user to start the task). To make a complete stitch, the subject should use the active gripper to grab the needle, then move the needle towards the tissue and penetrate the tissue with the needle at the entry target

[2] MIST-VR gives us the option of specifying and setting up the scene for left or right-handed users.

sphere. The entry sphere will change color to green when the tissue is penetrated at the correct position (Figure 3 - a).

The subject should then push the needle out through the tissue at the exit target sphere (which will change color to green when the needle exits the tissue at the correct position), and use the other gripper to grab the needle at the tip (Figure 3 - b) and pull it through in a curved motion, following a path defined by the needle's radius (Figure 3 - c).

In order to evaluate the user's dexterity, the system accounts for two categories of performance metrics; Dynamic Evaluation Measures and Errors.

Dynamic Evaluation Measures for the Stitch task include:

- Time: Total time spent to complete each trial (i.e. each stitch).
- Entry/Exit Hit-Target Distance: Distance between the point that the needle enters or exits the tissue and the target points marked by target spheres (as shown in Figure 3).
- Maximum Entry/Exit Deformation: Maximal tissue deformation of entry/exit stitch from entry/exit hit point which is calculated through a model of the tissue's force-deflective (stiffness) behaviour. These values are associated with the amount of penetrating force on the needle and its path of motion while penetrating/exiting the tissue (because of the arched shape of the needle, to minimize the strain on the tissue, the needle has to be pulled through in a curved motion, following a path defined by its radius).

Errors mainly take account of number of inappropriate collisions between the tools or between tools and the tissue within the operating space. Table 1 demonstrates a complete list of performance metrics for the Stitch task.

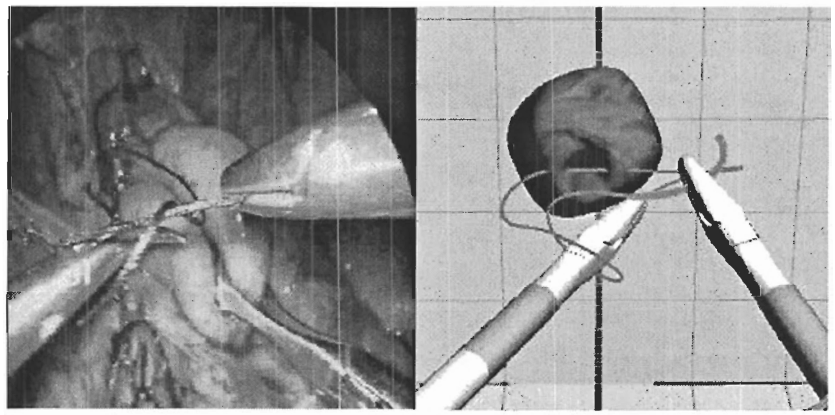
Table 1: **Performance metrics for the Stitch task**

	Metrics	Description
Dynamic Evaluation Measures	Time	Total time spent to complete the task
	Entry Hit-Target Distance (Hit Ent)	Distance between entry hit point and entry target point
	Exit Hit-Target Distance (Hit Ex)	Distance between exit hit point and exit target point
	Max Entry Deformation (Def Ent)	Maximal tissue deformation of entry stitch from entry hit point
	Max Exit Deformation (Def Ex)	Maximal tissue deformation of exit stitch from exit hit point.
Errors	Tool-Tool Collision (To-To)	When any segment of the left tool touches any segment of the right tool
	Hit Outside Entry Target (Out Ent)	When needle hits the surface outside the entry target area
	Entry Overstretch (Over Ent)	When the entry stitch is deforming the surface more than a given limit
	Hit Outside Exit Target (Out Ex)	When needle hits the surface outside the exit target area
	Exit Overstretch Error (Over Ex)	When the exit stitch is deforming the surface more than a given limit
	Incomplete Entry Stitch (Inc Ent)	When the needle is pulled out again through the entry point
	Incomplete Exit Stitch (Inc Ex)	When the needle is pulled out again through the exit point
	Closed Needle Entry (Cl Ent)	When trying to acquire the needle with the tool, but the grips are closed
	Tip Removed (Top Rem)	If an active tool entered the target object with open grips but was subsequently withdrawn without closing the grips (which is a failure to acquire the object)
	Wrong Section Grip (Wr Sect)	When the tool grabs the needle outside the target section
	Unexpected Tool (Unex To)	If the wrong tool grabs the needle
	Unexpected Stitch (Unex St)	The needle hits the tissue without the target area being defined by a target sphere (i.e. a red sphere appears when the stitch is made)
	Needle Dropped (Ne Dr)	The grips of the first tool open up after acquiring the needle
	Needle Pushed Out of Reach (Tar Out)	The tool has pushed the needle out of reach

Half-Square Knot Task

Figure 4: The Half-Square Knot Task

Left: knotting in real operation (Adopted from: MIST Suturing Module, 2003, by permission), Right: MIST HS Knot task (Adopted from: MIST Suturing Module, 2003, by permission)

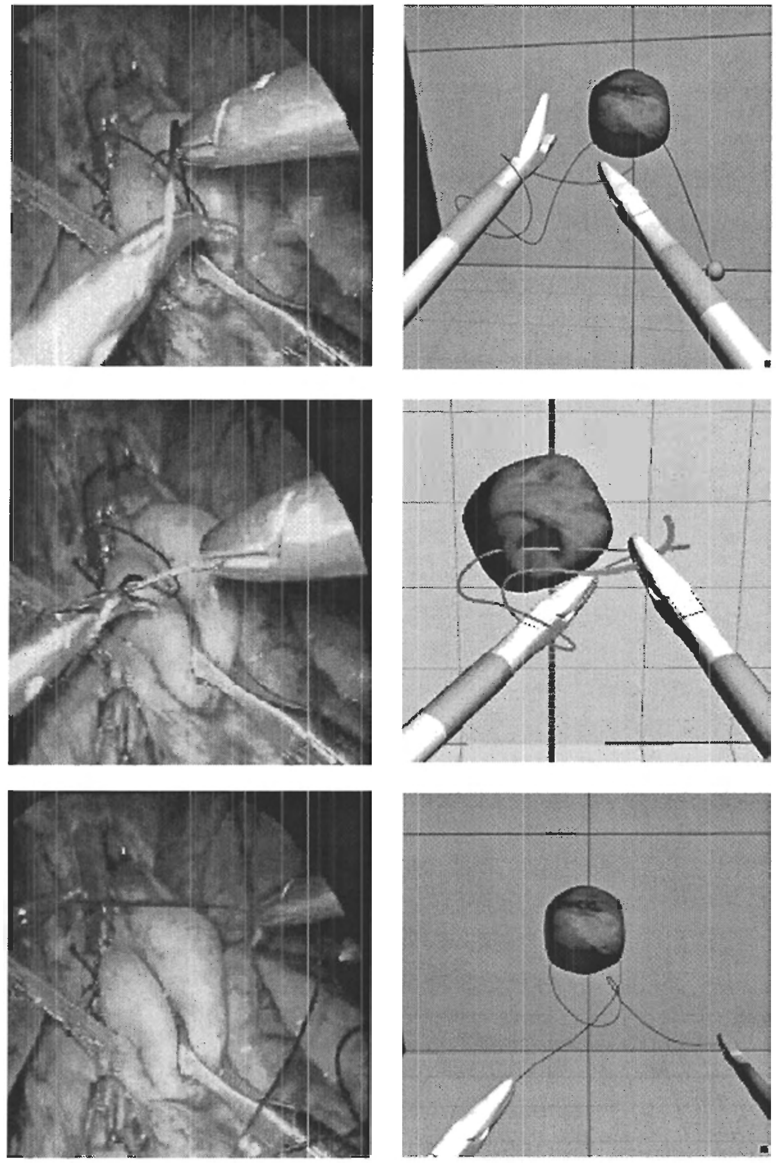


The purpose of the Half-Square Knot (HS Knot) task is to train the correct movements in the first half of a square knot. This includes the winding of the thread around one tool, and then to tightening the knot (Figure 4).

Similar to the Stitch task, visual clues guide the user throughout the procedure. Examples of these clues include the green bands around the tool(s), marking the active gripper at the time, and target spheres spotting the part of the thread that needs to be grabbed in the next move (Figure 5 - a). The thread itself also changes color into red, if it is overstretched, and into green, when the knot is tightened enough at the end of the procedure.

To start the task, the user must acquire the needle with the active tool (which is the grasper corresponding to the user's dominant hand), and wind the thread around the other tool while holding the needle with the first grasper (Figure 5 - a). Once a complete 360° loop around the second tool is acquired, a target sphere appears at the free end of the thread (Figure 5 - a), which is a clue for the user to grab the end of the thread with the second tool (with thread loops intact as shown in Figure 5 - b) and move the left and right instruments in opposite directions to form a knot (Figure 5 - c).

Figure 5: **Snap shots of MIS Knotting in real operation (left column) and in MIST-VR (right column) for a right-handed user**
(a) Winding the Thread, (b) Forming the knot, (c) Tightening the knot (Adopted from: MIST Suturing Module, 2003, by permission)



(a)

(b)

(c)

Similar to the Stitch task, performance metrics that MIST-VR collects for the HS Knot task fall into two categories; Dynamic Evaluation Measures and Errors, which are summarized in Table 2.

Table 2: **Performance metrics for the HS Knot task**

	Metrics	Description
Dynamic Evaluation Measures	Time	Total time to complete the task
	Max Winding Overstretch	Maximal thread overstretch during winding (Instead of pulling the thread through the stitch)
	Max Tightening Overstretch	Maximal thread overstretch during tightening
Errors	Tool-Tool Collision	When any segment of the left tool touches any segment of the right tool
	Closed Needle Entry	When the active tool enters the needle with a closed grip
	Tip Removed	If an active tool entered the target object with open grips but was subsequently withdrawn without closing the grips (which is a failure to acquire the object)
	Wrong Section Grip	When the tool grabs the needle outside the target section
	Dropped Thread	When the free thread end is dropped after first been grabbed correctly
	Needle Dropped	The grips of the first tool open up or are too loose after acquiring the needle
	Needle Pushed Out Of Reach	The tool has pushed the needle out of reach

Dynamic Evaluation Measures for the HS Knot task include:

- *Time:* Total time spent to complete each trial (i.e. each knot).
- *Maximum Winding Overstretch:* The thread can be overstretched during winding. The maximal thread overstretch, which is calculated through a model of the thread's force-deflective behaviour, will therefore represent the likelihood of pulling the thread through the stitch before a knot is formed, and is measured as one of the performance metrics.
- *Maximum Tightening Overstretch:* Maximal thread overstretch during tightening the knot is also measured through a model of the thread's force-

deflective behaviour as another metric representing the user's performance..

2.2.2 Experimental Design

The total of 26 subjects (8 Experts, 8 Intermediates, and 10 Novices) participated in the user study. Subjects were selected based on their experience in MIS surgery; senior surgeons who had performed more than 50 operations were considered to be experts, surgical assistants who were surgeons mainly responsible for controlling the camera or holding forceps in the operating room, and had not performed more than 20 MIS (or any other type of surgery) were in the intermediate level, and OR (Operating Room) nurses, who were familiar with laparoscopic surgery by observing surgeries, but had no MIS experience themselves, consisted our group of novices[3].

Participants were from different age groups (all over 29 years of age), and were consisted of 16 females and 10 males. They were asked to fill out a questionnaire (Appendix B: User Study Questionnaire) to give us information about their previous surgical experience and their prior MIS training. Most of the participants in the expert group had at least 20 hours of previous MIS training under the supervision of an experienced surgeon in the operating room, and in some cases for a few hours in animal labs. In the intermediate group, except for two of the participants, no one had significant hours of prior MIS training. As for surgical simulators, only one of the experts and one of the intermediates had training experience with physical simulators for more than 2 hours, but no one was previously trained with virtual simulators. In the novice group, none of the participants had any type of MIS training before.

---

3 Participants were selected and categorized based on experienced surgeons' suggestions at Surrey Memorial Hospital, BC, Canada.

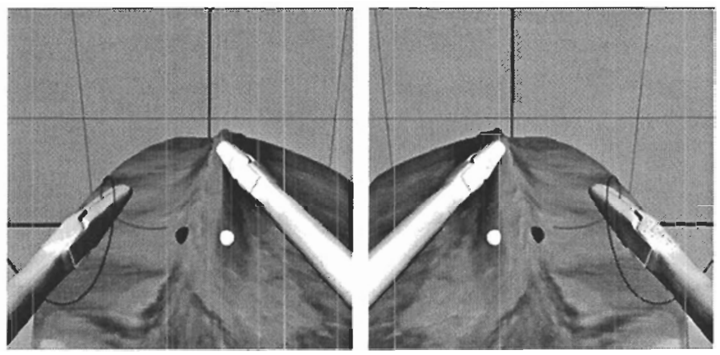
Table 3: Experiment steps and the approximate timing

	Experiment steps	Approximate time (min.)
Before starting the experiment	Sign the consent form	1
	Fill out the questionnaire	2
Stitch Task	Demo of the Stitch task	0.5
	Errors explained	1
	1 practice trial (not counted in the results)	1.5
	The actual test - 4 trials (counted in the results)	6
HS Knot task	Demo of the HS Knot task	0.5
	Errors explained	1
	1 practice trial (not counted in the results)	1.5
	The actual test - 4 trials (counted in the results)	6
		Total: 21

The experiment took less than 25 minutes for each subject. Participants were required to complete the Stitch task first, and the HS Knot task second. For each task, the presenter introduced the subject to the task, by demonstrating an error-free trial of the task and explaining what constituted as errors. Each subject was then allowed to practice the task once, before performing the four trials, which were counted in the results. Table 3 recaps the steps in the experiment and the approximate timings.

For the two tasks employed in this experiment, MIST gives us the option of specifying and setting up the operating space for a left or a right-handed user. The operating scenario for a left-handed user is symmetrical to the one for a right-handed user (e.g. in the Stitch task, a left-handed user starts the task by grabbing and inserting the needle into the tissue with the left tool, and continues by pulling out the needle from the other side with the right gripper, as opposed to a right-handed user who starts the task with the right tool and completes the stitch with the left tool - Figure 6). To eliminate the effect of handdominancy in our results, we set up the tasks for each user based on their dominant-hands, and only considered data values corresponding to the dominant hand in our analysis.

Figure 6: Symmetrical scenes for right and left-handed users. Left: MIST Stitch task for a left-handed user, Right: MIST Stitch task for a right-handed user



2.3 Data Analysis

As shown in Table 1 and Table 2, MIST VR collects 16 different performance metrics for the Stitch task and 10 for the HS Knot task. Studies to date have proven that MIST performance metrics have validity in terms of surgical skills assessment (Darzi, 2001; Gallagher, McClure, McGuigan, Crothers, & Browning, 1999; Wilson, Middlebrook, Sutton, Stone, & McCloy, 1997), and therefore are appropriate to be used as inputs to our fuzzy classifiers. However, in designing a fuzzy system without the help of automated fuzzy rule generating software, having a large number of input parameters can lead to an unmanageably large number of possible combinations and therefore expert rules (Fayek, & Sun, 2001). Thus, to yield a more manageable model we reduced the number of parameters by combining each group of parameters of the same nature into a new data value, and ignoring factors that were believed to be less effective on the results. The changes include:

- Compiling all the different errors that MIST accounts for, into a new parameter called *Number of Errors*, for both Stitch and HS Knot tasks. (Number of Errors = Sum of all Errors)

- Compiling *Maximum Entry Tissue Deformation* and *Maximum Exit Tissue Deformation* in the Stitch task, into a new parameter called *Maximum Tissue Deformation* (Max Tissue Deformation = Maximum Entry Tissue Deformation + Maximum Exit Tissue Deformation)

- Compiling *Maximum Winding Overstretch* and *Maximum Tightening Overstretch* in the HS Knot task, into a new parameter called *Maximum Thread Overstretch* (Max Thread Overstretch = Maximum Winding Overstretch + Maximum Tightening Overstretch)

- Ignoring the *Entry/Exit Hit-Target Distance* values in the Stitch task.

Table 4 and Table 5 demonstrate the resulting new parameters that were employed in designing the fuzzy classifier.

Table 4: Performance metrics used for Stitch task in data analysis

Metrics	Description
Time	Total time spent to complete the task
Max Tissue Deformation	Sum of Max Entry and Exit Tissue Deformation (representing max tissue deformation throughout the whole task of stitching)
Number of Errors	Sum of all the errors

Table 5: Performance metrics used for HS Knot task in data analysis

Metrics	Description
Time	Total time spent to complete the task
Max Thread Overstretch	Sum of Max Winding Overstretch and Max Tightening Overstretch (representing max thread overstretch throughout the whole task of knotting)
Number of Errors	Sum of all the errors

Also, to have analogues ranges of data values for different parameters, we normalized the data collected by MIST, by dividing all the performance measurements for each parameter to the maximum value among them. From now on, we will only work with normalized data (ranging between 0 and 1) rather than the raw data collected in the user study.

An example of all the modifications applied to the raw data (collected by MIST-VR) to transform it into appropriate data for our analysis is shown in Figure 7. Data values shown in the example are an expert user's performance metrics for one trial of the Stitch task.

Figure 7: Example of modifications applied to the raw data collected by MIST-VR

Performance metrics collected by MIST-VR (raw data)																		
Dynamic Evaluation Metrics					Errors													
Time	Hit Ent	Def Ent	Hit Ex	Def Ex	Tar Out	Inc Ent	Inc Ex	Ne Dr	Out Ent	Out Ex	Over Ent	Over Ex	Unex To	To To	Cl Ent	Tip Rem	Wr Sect	Unex St
26.5	1.6	4.8	2.9	5.7	0	0	0	0	1	0	1	1	0	0	0	0	0	0

+

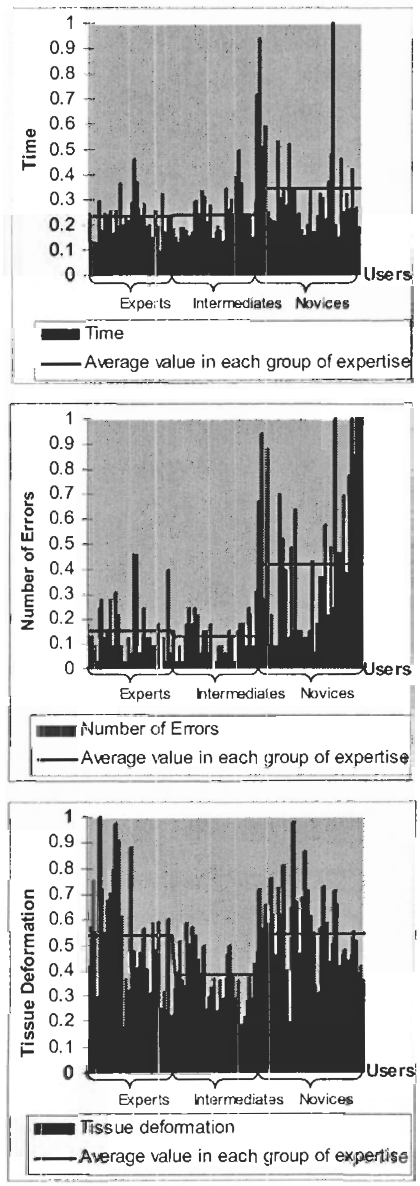
+

Reduced Number of Parameters		
Time	Max Tissue Deformation	Number of Errors
26.5	10.5	3.0

Normalization

Performance metrics used in data analysis		
Time	Max Tissue Deformation	Number of Errors
0.15	0.59	0.09

Figure 8: **Stitch Task - User's individual and group average performance metrics (a) Time, (b) Num. of Errors, (c) Max. Tissue Deformation (x-axis: Users, y-axis: Normalized performance metrics)**

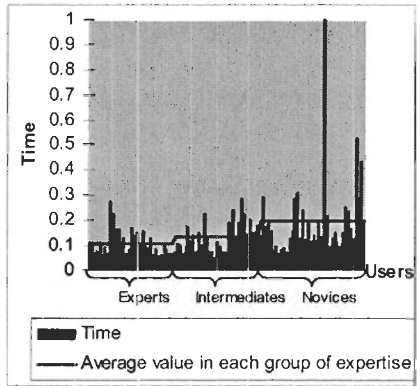


(a)

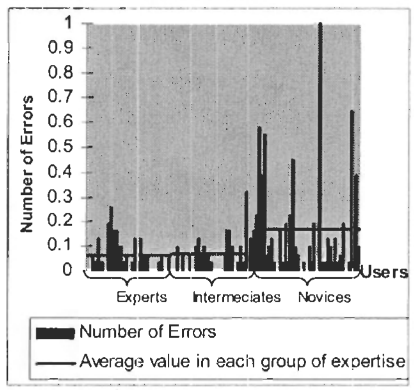
(b)

(c)

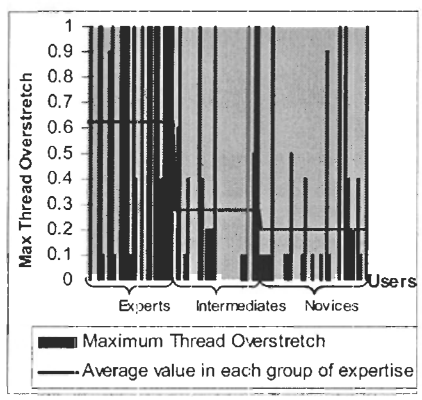
Figure 9: HS Knot Task - User's individual and group average performance metrics (a) Time, (b) Num. of Errors, (c) Max. Thread Overstretch (x-axis: Users, y-axis: Normalized performance metrics)



(a)



(b)



(c)

Figure 8 and Figure 9 represent normalized users' performance metrics, with columns representing the individual values and horizontal lines showing the average values for each group of users. Table 6 compares average values of performance metrics between the three groups of users for both Stitch and HS Knot tasks. Please note that in this study lower values for individual metrics and overall scores mean a better performance (e.g. the lower the time value is, the faster the user has performed the procedure). By looking at the average values, it seems that Experts and Intermediates have generally performed faster, with less constraint on the tissue or the thread, and with fewer numbers of errors. However, looking at individual test results, it seems impossible to find a well-defined pattern to categorize the subjects (e.g. as shown in Figure 8, in the Stitch task, the big range of "Time", "Number of Errors" and "Maximum Tissue Deformation" values in the Novice group does not follow any specific pattern).

Table 6: Average values of performance metrics in each group of expertise for Stitch and HS Knot tasks

	Stitch Task			HS Knot Task		
Input	Time	Max Tissue Deformation	Number of Errors	Time	Max Thread Overstretch	Number of Errors
Experts	0.231	0.534	0.154	0.108	0.625	0.0635
Intermediates	0.238	0.381	0.131	0.136	0.275	0.066
Novices	0.342	0.540	0.414	0.196	0.202	0.169

In the following section simple statistical methods are used to analyse the user study data and to find the relationship between the Stitch and the HS Knot task performance metrics (represented in Table 4 and Table 5) and the users' surgical skill levels.

2.3.1 Correlation and Regression Analysis

Correlation analysis

Correlation analysis is the statistical tool that can be used to describe the degree to which one variable is linearly related to another. In other words, correlation analysis is used to measure the degree of association between two variables.

The strength of the linear relationship between two variables x and y (for a sample of n measurements on x and y) is measured by the *coefficient of correlation, r*, as follows:

$$r = \frac{ss_{xy}}{\sqrt{ss_{xx}ss_{yy}}}$$

Equation 1

Where

$$ss_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}), \qquad ss_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2, \qquad ss_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

The coefficient of correlation has two important characteristics:

1. The magnitude of the correlation coefficient is independent of the scales of measurement for variables. This means that the correlation coefficient can compare the relationship between variables regardless of what is being represented by them.

2. The value of a correlation coefficient is between +1.0 and -1.0. A value of either +1.0 or -1.0 shows a 100% correlation between the variables, meaning that the movement of the two variables is in an absolute similar or complementary direction.

All variables used in correlation analysis must have numerical values. In this problem, the surgical skill level takes three non-numerical values: Expert, Intermediate, and Novice. To be able to perform the correlation analysis, we assigned a number to each of these three values as shown in Table 7. These numbers are selected by dividing the interval from 0 to 1 into three equal regions, and assigning the centre of each region to one surgical skill level. It should be noted that the magnitudes of these values do not affect the results of correlation analysis.

Table 7: Numerical values assigned to surgical skill levels

Skill Level	Numerical Value
Expert	0.167
Intermediate	0.5
Novice	0.833

Figure 10: Assigning numerical values to surgical skill levels

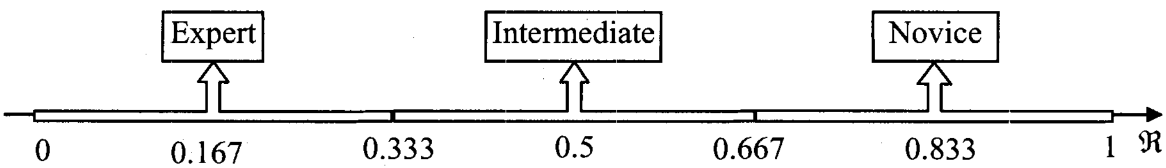


Table 8 and Table 9 show the correlation coefficient, calculated by MATLAB (The MathWorks, Inc, 1994-2006), for each two pairs of performance metrics and the users' surgical skill levels, for the Stitch and the HS Knot tasks.

Table 8: Coefficient of correlation for Stitch task performance metrics

Stitch Task Correlation Analysis Results				
	Time	**Number of Errors**	**Max Tissue Deformation**	**Skill Level**
Time	1	0.464631	0.419861	0.391149
Number of Errors	0.464631	1	0.464631	0.606139
Tissue Deformation	0.419861	0.464631	1	0.644996
Skill Level	0.391149	0.606139	0.644996	1

Table 9: **Coefficient of correlation for Stitch task performance metrics**

HS Knot Task Correlation Analysis Results				
	Time	**Number of Errors**	**Max Thread Overstretch**	**Skill Level**
Time	1	0.56043	-0.24502	0.453937
Number of Errors	0.56043	1	-0.14152	0.408959
Thread Overstretch	-0.24502	-0.14152	1	-0.45765
Skill Level	0.453937	0.408959	-0.45765	1

As shown in Table 8, in the Stitch task, surgical skill level has a fairly high correlation with "Number of Errors" and "Tissue Deformation", but not as high with "Time". For the HS Knot task (Table 9), correlation between the skill level and "Time", "Number of Errors", and the "Thread Overstretch" are almost equal, but less than 50%.

This may indicate that the performance metrics selected for the Stitch and the HS Knot tasks are not entirely related to surgical skill level, or it could be due to the fact that our sample data is not representative of the population. However, as we will see later in section 4.3, the poorly correlated data may result in contradictory fuzzy rules, which consequently help eliminate the effect of such data in making the final decision.

Linear Regression Analysis

Linear regression analysis is another statistical method which involves finding the best straight-line relationship to explain how the variation in an outcome (or dependent) variable, *Y*, depends on the variation in a predictor (or independent) variable, *X*. When *Y* is a function of more than one independent variable, *Multiple Regression* is used to estimate the relationship between *Y* and the independent variables. This estimate can be used to build a regression equation of the form:

$$Y = c + a1X1 + a2X2 + \dots + anXn$$

Equation 2

Where X_i's are the independent variables, a_i's are the regression coefficients representing the amount *Y* changes when the corresponding x_i changes 1 unit, and *c* is the constant representing the amount of *Y* when all the independent variables are 0. Once the regression equation is built, it can be used to predict the value of Y based on a set of measured X_i's.

We used MATLAB to find the regression equation for the Stitch and the HS Knot tasks. For each task, we split the dataset into two equal halves using two different data separation methods, the DAU and the DHU methods (which will be explained later in section 4). Half of the data, which we call the training dataset, were used to build the regression equation, and the other half, or the testing dataset, were used to test the ability of the regression equations in predicting the value of Y.

For the Stitch task, the independent variables are the "Time", the "Number of Errors", and the "Maximum Tissue Deformation" performance metrics. For the HS Knot task, the independent variables are the "Time", the "Number of Errors", and the "Maximum Thread Overstretch" performance metrics, and the dependent variable for both tasks is the "Skill Level". Similar to the correlation analysis, we assigned numerical values shown in Table 7 to the three surgical skill levels. The resulting regression equations are as follows:

Stitch task (DAU):

$$\textit{Skill Level} = -0.0663 - 0.127 \times \textit{Time} + 0.4292 \times \textit{Num. Of Errors} - 1.0668 \times \textit{Max Tissue Deformation}$$

Equation 3

Stitch task (DHU):

$$\textit{Skill Level} = -0.1427 - 0.0586 \times \textit{Time} + 0.5636 \times \textit{Num. Of Errors} - 0.4255 \times \textit{Max Tissue Deformation}$$

Equation 4

HS Knot task (DAU):

$$\textit{Skill Level} = 0.5215 + 0.2719 \times \textit{Time} + 0.3598 \times \textit{Num. Of Errors} - 0.2163 \times \textit{Max Thread Overstretch}$$

Equation 5

HS Knot task (DHU):

$$\textit{Skill Level} = 0.04112 + 1.0366 \times \textit{Time} + 0.4619 \times \textit{Num. Of Errors} - 0.2573 \times \textit{Max Thread Overstretch}$$

Equation 6

Performance metrics in the corresponding testing datasets were then substituted in Equation 3 to Equation 6 to predict surgical skill levels. The results are represented later in section 6.

Even though the results of statistical analysis on our user study data do not suggest a high correlation between the Stitch and HS Knot performance metrics and surgical skill levels, a particular pattern of a complex combination of all the parameters, which is not recognizable with simple statistical methods, may exist and could be used for performance categorization. If such a pattern exists, we hypothesise that a fuzzy classifier will be an appropriate means to recognize and model this pattern, as one of the important applications of fuzzy logic is in pattern recognition.

3 FUZZY LOGIC: BASIC PRINCIPLES AND APPLICATIONS

3.1 History and Applications

Most of the phenomena that we encounter every day carry a certain degree of ambiguity and fuzziness in the description of their nature. "The weather is *hot* today" is a typical example of a fuzzy expression. What temperature is considered hot? How much does it need to be decreased to be considered warm, and not hot? If the weather is hot for me, is it hot for my neighbour as well? This kind of imprecision or fuzziness associated with continuous phenomena is common in almost any field of study: sociology, physics, biology, finance, marketing, engineering, psychology, health management, etc.

Before the introduction of fuzzy theory, conventional mathematical methods were the only means of modelling natural processes. The underlying logic of these methods is the precise Boolean logic, which is based on the law of *Excluded Middle*. This logic has only two states, "0" and "1" or "True" and "False". In other words, every proposition must either be true or false; no intermediate values are allowed. Conventional mathematical methods however, require detailed and precise information to operate. They can not handle the uncertainty of the natural phenomena and the human natural language. Albert Einstein faced the same dilemma:

> *"So far as the laws of mathematics refer to reality, they are not certain. And so far as they are certain, they do not refer to reality."*

Fuzzy logic was introduced by Lotfi Zadeh in 1965 as a means to model the uncertainty of natural language (Lotfi Zadeh, 1965). It could be considered as a superset of conventional (Boolean) logic that handles the concept of partial truth or truth-values between "completely true" and "completely false". Despite the conventional logic systems that focus on the quantitative aspects of objects, fuzzy logic describes their qualitative nature, which in many ways are related to the rules of grammar that focus on descriptive adjectives and adverbs.

Even though the fuzzy logic theory was largely ignored in the western world, it attracted the attention of industrial designers and inventors in Asian countries such as Japan and China almost as soon as it was proposed by Dr. Lotfi Zadeh. Over the past few years, fuzzy modeling and identification methodologies have been successfully used in a number of real-world applications and for various aims such as analysis, design, medical instrumentation, monitoring, decision making, patter recognition, and industrial process control.

Examples include application of fuzzy logic in products ranging from large-scale electro-mechanical processes, like subway systems and elevators, to mass-market consumer applications such as cameras, camcorders, washing machines, and microwave ovens (Fuzzy logic tool box, what is fuzzy logic, ¶1, 2004; World Technology Evaluation, Fuzzy Logic research and LIFE, ¶4, 2005). Fuzzy expert systems have also been used for engineering design performance evaluation (Vanegas & Labib, 2005), or project performance prediction and evaluation (Fayek & Sun, 2001). They have also served as monitoring systems for intrusion detection in networked computers (Gómez & Dasgupta, 2002), or as decision support systems to assist operators (Hartog et al., 1997), and to support decisions in medical domains (Gorzałczany & Grądzki, 1999).

One of the important applications of fuzzy logic is in the area of pattern recognition. A common thrust of this problem area is the search for structures in data, where the issue is to compare, in terms of relevant features, the categories identified in data with given perfect categories (Klir, St.Clair., & Yuan, 1997).

The utility of fuzzy logic is also well established in the design of automatic controllers. Especially in the aviation industry, because of the high degrees of nonlinearity, uncertainty, and complexity of the aerospace systems and the involvement of human beings, fuzzy logic-based methodologies have been widely used in the design of flight control systems (Dote & Ovaska, 2001). For instance as intelligent helicopter navigation systems (Rahbari, Leach, Dillon, & DaSilva, 2002), as complex aircraft controllers (Mengali G., 2000), or as flight control systems in aviation simulators (Beringer, 2002).

The linguistic interpretability of fuzzy systems make them suitable for another important application; modelling of human decisions or experience. Sundaram , Naidu, and Das (2004) used fuzzy multi attribute decision making approach to evaluate the quality of food products as judged by human senses such as vision, taste, smell, and touch (Sundaram, Kalpana Naidu, & Das, 2004). Kumar, Stoll and Stoll (2003) used fuzzy expert systems to approximate patients' physical fitness based on real world physiological parameter measurements. Without the use of expert systems, the only solution to this problem was the advice of an expert (Kumar, Stoll, & Stoll, 2003).

The successful applications of fuzzy logic theory and the rapid growth of research involving fuzzy logic suggest that the impact of this revolutionary approach to computing will be felt more strongly in the coming years. Fuzzy logic is likely to play an important role in science and engineering, but eventually its influence may extend much farther (The Berkeley Initiative, A glimpse into the future, ¶1, 2005).

3.2 Fuzzy Set Theory

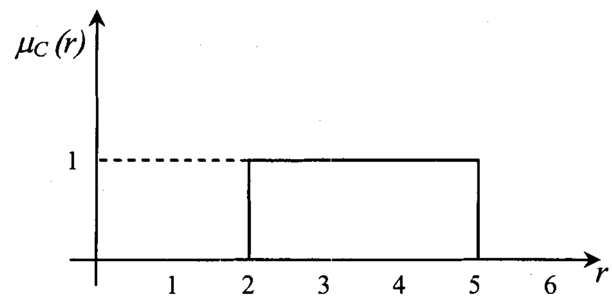
Fuzzy logic is almost synonymous with the theory of fuzzy sets; a theory which relates to classes of objects with unsharp boundaries in which membership is a matter of degree. Conventional (i.e., crisp) sets contain objects that satisfy precise properties required for membership. The set C of real numbers from 2 to 5 is crisp; we write $C = \{r \in \Re \mid 2 \leq r \leq 5\}$, where $\Re$ is the set of real numbers. Equivalently, C is described by its *membership function* (MF), $\varphi_C : \Re \rightarrow \{0,1\}$, defined as:

$$\varphi_c(r) = \begin{cases} 1 & 2 \leq r \leq 5 \\ 0 & Otherwise \end{cases}$$

Equation 7

In logic, values of φ_C are called truth-values with reference to the question, "Is r in C?" The answer is yes if and only if $\varphi_C(r) = 1$, and no, otherwise. As shown in Figure 11, in a conventional set there is a clear-cut differentiation between the elements that belong to the set and those that do not.

Figure 11: Membership function for the conventional set C



Defining the real numbers between 6 and 8 is a problem that is intrinsically crisp and would not require the use of fuzzy sets. A situation closer to what we encounter in everyday life however, is for example deciding if the weather is hot or not in a particular day. Reasoning according to the conventional logic, we would need to define a temperature threshold that divides hot from not-hot weather. If the temperature is higher than the threshold (even by 0.001 of a degree) then the weather is hot, otherwise, not hot. This is obviously far from the way human beings make their judgments. Our perception of the weather temperature is better described as a sort of soft switching rather than a threshold mechanism. This is also why we often add a modifier to the word "hot" (i.e., not, not very, somewhat, very, etc.) in order to express "degrees of hotness" rather than absolute true or false answers.

A fuzzy set could well accommodate the way human beings make their decisions. In the fuzzy set "hot weather temperatures" (which will be defined as an example in the following section) a degree of hotness is defined, thus providing a *continuum* rather than an abrupt transition from true to false.

3.2.1 Elements of Fuzzy Set Theory

Fuzzy sets, Membership functions, and Universe of discourse

As mentioned before, a fuzzy set is a class of objects with unsharp boundaries. In other words, elements of a fuzzy set may belong to it to *partial degrees*, from the full belongingness to the full non-belongingness through all intermediate values. Hence the membership function of a fuzzy set is allowed to have values *between* 0 and 1 that denote the *degree of membership* of an object in the given set.

Consider X as a space of objects and let x be a generic element of X. A fuzzy set F in X is defined as a set of ordered pairs

$$F = \{(x, \mu_f(x)\}$$

Equation 8

Where $\mu_f : X \rightarrow [0,1]$ is called the *membership function* (MF) for the fuzzy set F, and maps each element of X to a membership degree $\mu_f(x) \in [0,1]$.

X is often referred to as the *universe of discourse* (universe, universal set, referential, reference set, etc.) and contains all elements relevant for the particular concept. It may consist of discrete (ordered or non-ordered) objects or it can be a continuous space.

The construction of a fuzzy set depends on two things: the identification of a suitable universe of discourse and the specification of an appropriate membership function. There are two possible approaches in defining the MFs; the most straightforward approach is to ask the experts to draw the MFs. The functions could be either defined by one expert, or as the average of the membership functions defined by several experts, so that:

$$\underset{x \in X}{\forall} \mu(x) = \frac{1}{n} \sum_{i=1}^{n} \mu_i(x)$$

Equation 9

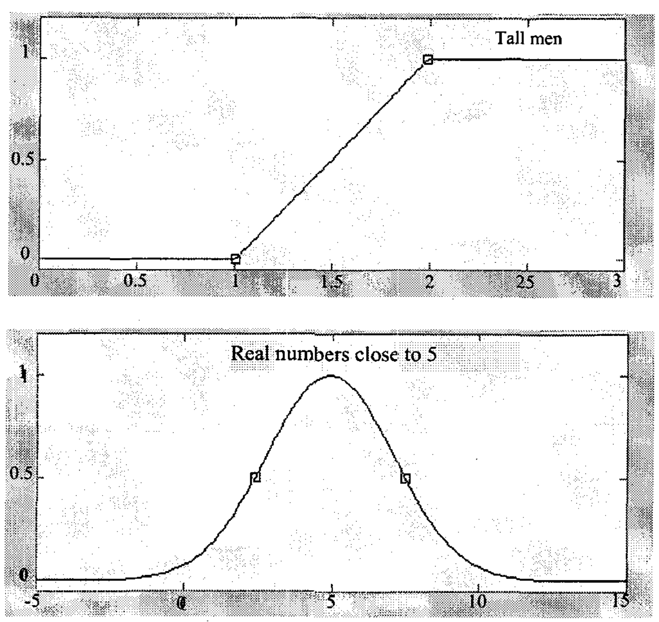
Where $\mu_i(x)$ is the MF defined by expert i. While averaging the MFs defined by different experts reduces the subjectivity, the resulting MFs will probably have a rough shape that is not consistent with the way of human thinking. Therefore, usually an approximation by a standard shape is used to smooth the membership function (Szczepaniak, Lisboa, & Kacprzyk, 2000, page 34).

Another possible approach is to define the membership functions on the basis of numerical data. In this case, a standard shape is usually selected for the MFs, and the sampled data is used to define the functions parameters.

Even though there is no restriction in the shapes chosen for membership functions, there are some standard functions that have been used more in literature. Common choices for membership functions are Gaussian or S-shaped, Triangular/Trapezoidal and Bell shaped functions (Szczepaniak, et al., 2000, page 34). Examples of Triangular and Gaussian membership functions are shown in Figure 12. Because of their simplicity, Triangular and Trapezoidal functions are the most popular choices among the standard MF shapes at present and most authors have found them efficient enough to use in their systems (Fayek & Sun, 2001; Dadone, 2002).

Figure 12: Examples of membership functions

(a): Triangular membership function representing the fuzzy of "tall men" (x-axis: Height (m), y-axis: Truth-values) (b):Gaussian membership function representing the fuzzy set of "real numbers close to 5" (x-axis: Height (m), y-axis: Truth-values)



(a)

(b)

The following example can further clarify the concept of fuzzy sets.

Example: The "hot weather temperature" fuzzy set

Let's go back to the question "what weather temperatures are hot?". As mentioned earlier, the conventional set theory fails to characterize the "hot weather temperatures" as judged by human beings. In this example we employ the fuzzy logic approach to deal with the problem.

Let's define a fuzzy set named "hot weather temperatures". The first step in specifying the fuzzy set is to define its universe of discourse. We define $T \subset \Re$ so that:

$$T = \{t \mid -100 \leq t \leq +100\}$$

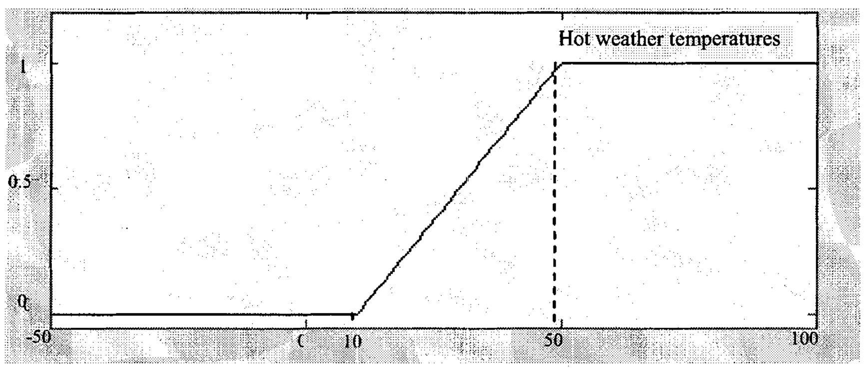
Equation 10

T covers all the possible weather temperatures (in degrees Celsius). If the fuzzy set representing the "hot weather temperatures" is H, the membership function for H is defined as: μ_H: $T \rightarrow [0,1]$, such that $\mu_H(t) \in [0,1]$ is the degree to which an element $t \in T$ belongs to the fuzzy set H.

The next step is defining the membership functions for our fuzzy set. We do so by choosing a standard shape like the Trapezoidal function for the MF, and identifying the critical points in the universe of discourse and assigning them the appropriate truth-values. For example we know that almost anyone considers a temperature of 45°C a hot weather temperature. Therefore we can say that 45°C belong to the fuzzy set H, so that $\mu_H(45)=1$. If 45°C is considered hot, then any temperature higher than 45°C would also be considered hot, therefore $\mu_H(t \geq 45)=1$. On the other hand, weather temperatures of 15°C and lower are most certainly not judged as hot temperatures and therefore do not belong to H. In other words, $\mu_H(t \leq 15)=0$. For weather temperature between 15°C and 45°C however, we can not confidently say whether they belong to H or not. 30°C is a very hot temperature for people living in Alaska, but could be a normal temperature for those living in Africa. From an intuitive point of view however, we can say that the degree of belongingness to H increases from 0 to 1, as the temperature goes up from 15 to 45°C.

The above information is enough for us to draw the Trapezoidal membership function, as shown in Figure 13. The horizontal axis represents the weather temperatures, and the vertical axis shows the truth-value for each temperature.

Figure 13: **Fuzzy set "Hot weather temperatures"**
(x-axis: weather temperatures in degrees Celsius, y-axis: truth-values)



Other definitions and terms in fuzzy set theory

This section briefly presents basic definitions and properties related to fuzzy sets, concentrating more on those that are relevant to this work.

Empty sets: A fuzzy set A in X is said to be empty, written A=Ø, if and only if:

$$\forall x \in X, \mu_A(x) = 0$$

Equation 11

Equal sets: Two fuzzy sets A and B defined in the same universe of discourse X are said to be equal, written A=B, if and only if

$$\forall x \in X, \mu_A(x) = \mu_B(x)$$

Equation 12

Subset: A fuzzy set A defined in X is said to be a subset of a fuzzy set B in X, written $A \subseteq B$, if and only if

$$\forall x \in X, \mu_A(x) \le \mu_B(x)$$

Equation 13

Normal and *subnormal sets*: A fuzzy set A defined in X is said to be normal if and only if

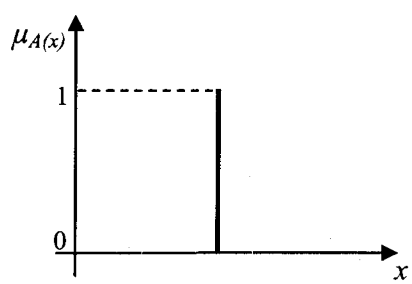
$$\max_{x \in X} \mu_A(x) = 1$$

Equation 14

and it is said to be subnormal otherwise.

Fuzzy singleton: A fuzzy set A defined in X is called a fuzzy singleton, if its support is a single point in X with $\mu_A(x) = 1$.

Figure 14: Graphical representation of a singleton fuzzy set



Support: The support of a fuzzy set A in X is the set of all points with nonzero membership degree in A

$$Supp(A) = \{x \in X \mid \mu_A(x) > 0\}$$

Equation 15

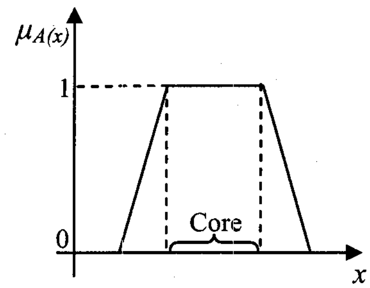
Figure 15: Graphical representation of support of a fuzzy set

Core: The core of a fuzzy set A is the set of all points with unit membership degree in A

$$Core(A) = \{x \in X \mid \mu_A(x) = 1\}$$

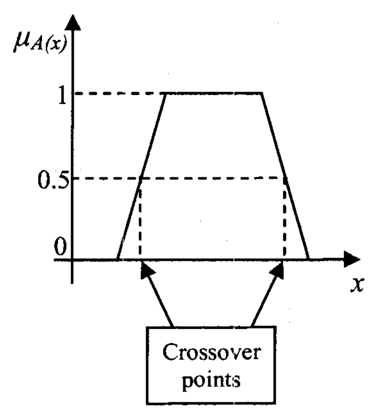
Equation 16

Figure 16: Graphical representation of the core of a fuzzy set



Crossover points: A point $x \in X$ at which $\mu_A(x) = 0.5$ is called the crossover point of a fuzzy set A in X.

Figure 17: Graphical representation of the crossover points



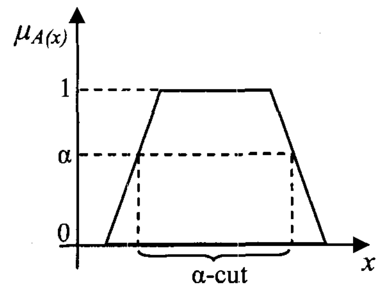
α-cut, strong α-cut: The α-cut or α-level set of a fuzzy set A is a crisp set, written A_α, and defined as the following set

$$A_\alpha = \{x \in X \mid \mu_A(X) \geq \alpha\}$$

Equation 17

If we replace the "≥" in Equation 17 with ">", then we have the strong α-cut, or strong α-level set of the fuzzy set A.

Figure 18: Graphical representation of the fuzzy α-cut



3.2.2 Basic Operations on Fuzzy Sets

Similar to the conventional set theory, the basic operations in fuzzy set theory are the complement, intersection, and union.

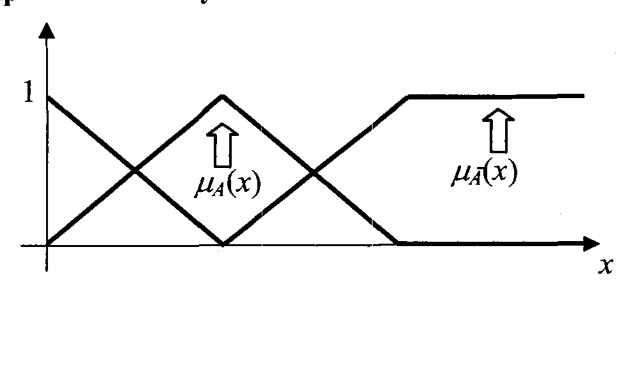
Complement or (*negation*): The complement of a fuzzy set A in X, denoted by $\bar{A}$, corresponds to the negation "not", and is defined as

$$\forall x \in X, \mu_{\bar{A}}(x) = 1 - \mu_A(x)$$

Equation 18

The complement can be represented as in Figure 19 where $\mu_{\bar{A}}(x)$ is shown in heavy lines.

Figure 19: The complement of a fuzzy set



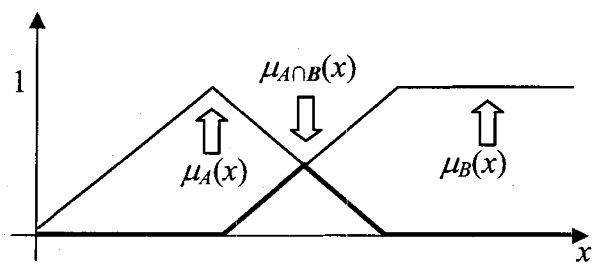
Intersection: The intersection of two fuzzy sets A and B in X, written as $A \cap B$, is defined as

$$\forall x \in X, \mu_{A \cap B} = \min\{\mu_A(x), \mu_B(x)\}$$

Equation 19

The intersection can be illustrated as in Figure 20 Where $\mu_{A \cap B}$ is shown in heavy lines.

Figure 20: Graphical representation of the intersection of two fuzzy sets



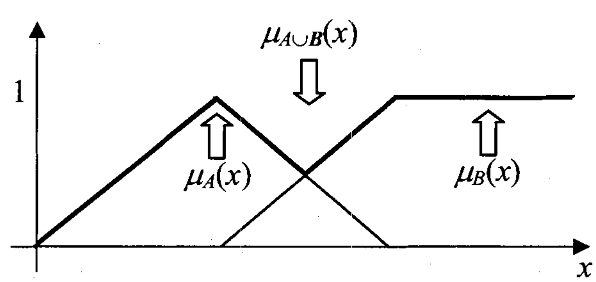
Union: The union of two fuzzy sets A and B in X, written as $A \cup B$, is defined as

$$\forall x \in X, \mu_{A \cup B}(x) = \max\{\mu_A(x), \mu_B(x)\}$$

Equation 20

The union can be demonstrated as in Figure 21 Where $\mu_{A \cup B}$ is shown in heavy lines.

Figure 21: Graphical representation of the union of two fuzzy sets



The above definitions of fuzzy intersection and union are well established and widely used. However, similar to the traditional set theory, these operations could also be defined through the general *t-norm* and *s-norm* (or *t-conorm*) operators (Szczepaniak et al. page 34; Fayek & Sun, 2001; Dadone, 2001).

A *t-norm*, t: $[0,1] \times [0,1] \rightarrow [0,1]$, is defined such that for each $x, y, z \in [0,1]$:

1. Its unit element is 1: $t\,(x\,,\,1) = x$
2. It is monotone: $x \leq y \Rightarrow t\,(x\,,\,z) \leq\; t\,(y\,,\,z)$
3. It is commutative: $t\,(x\,,\,y) = t\,(y\,,\,\mathrm{x})$
4. It is associative: $t\,[x,\,t\,(y\,,\,z)] = t\,[t\,(x\,,\,y),\,z]$

The minimum operator is the most widely used t-norm operator in the fuzzy set theory. Some other examples of t-norm operators are the *algebraic product* ($t\,(x, y) = x.y$) and the *Lukasiewicz* t-norm ($t\,(x, y) = \max\,(0, x + y\,-1)$).

An *s-norm*, s: $[0,1] \times [0,1] \rightarrow [0,1]$, is defined such that for each $x, y, z \in [0,1]$:

1. Its unit element is 0: $s\,(x\,,\,0) = x$
2. It is monotone: $x \leq y \Rightarrow s\,(x\,,\,z) \leq\; s\,(y\,,\,z)$
3. It is commutative: $s\,(x\,,\,y) = s\,(y\,,\,\mathrm{x})$
4. It is associative: $s\,[x,\,s\,(y\,,\,z)] = s\,[s\,(x\,,\,y),\,z]$

The most popular s-norm operator in fuzzy set theory is the maximum operator. Other examples are the *probabilistic product* ($s\,(x, y) = x + y - xy$), and the *Lukasiewicz s-norm* ($s\,(x, y) = \min\,(x + y\,,\,1)$).

3.3 Fuzzy Inference Systems

Fuzzy inference is using fuzzy logic to formulate the mapping of a given input to an outpu. Fuzzy inference systems (FISs) are rule based systems in which the relationship between the inputs and outputs of the system are retrieved in the form of if-then rules. In the process of inference, the inputs are first *fuzzified*, (i.e. converted from crisp numbers to fuzzy sets). After going through the fuzzy rules contained in a rule-base, the output for each set of inputs is computed in the form of a fuzzy set. The output fuzzy sets are then composed and *defuzzified* (i.e., converted from a fuzzy set to a crisp number), since the desired output is usually a crisp number rather than a fuzzy set.

There are two basic types of fuzzy inference systems: Mamdani-Assilian (or Mamdani), 1975 by Ebrahim Mamdani (Mamdani & Assilian, 1975) and Takagi-Sugeno-Kang (or Sugeno), introduced in 1985 (Takagi & Sugeno 1985). These two types of inference systems vary somewhat in the way outputs are determined. In Mamdani systems, which is the most common methodology (Fuzzy logic toolbox, fuzzy inference systems, ¶ 3), both the input and output are represented with linguistic terms (such as "tall", "short", "hot", "cold"). The antecedent and consequent of an if-then rule are typically Bollean expression of simple clauses. A simple form of the Mamdani system is of the form:

If x is A, then y is B.

In which A and B are linguistic terms defined by fuzzy sets on the ranges (universes of discourse) X and Y respectivelt.

In Sugeno systems, the antecedant is a Bollean expression of simple clauses, but the consequent is a function of the input (usually a polynomial). This can be represented in the form:

If x is A, then y is f(x).

In which A is a linguistic term defined by a fuzzy set on the universe of discourse X and $f(x)$ is a function of the input x.

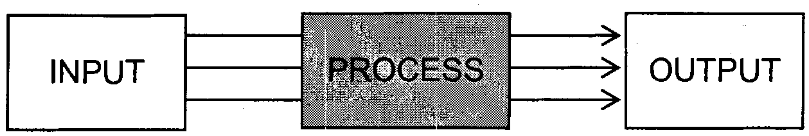
The Sugeno fuzzy inference systems are faster and work well with linear techniques. The Mamdani systems however, are intuitive and suitable for human inputs. Therefore, we employed the Mamdani FISs to solve the problem of surgical performance evaluation.

We will briefly overview the process of fuzzy inference, and explain the design process of Mamdani-type FISs in the following sections. For more information about Sugeno systems, please refer to (Takagi & Sugeno, 1985).

3.3.1 Overview of Fuzzy Inference Process

Figure 22 shows the three major elements of a fuzzy inference system. The information flows from left to right, or from the inputs to the outputs of the system. The purpose is to map an input space to an output space, and the key mechanism for doing this is a list of if-then statements called *fuzzy if-then rules*. All rules are evaluated in parallel, and the order of the rules is unimportant.

Figure 22: Major elements of a fuzzy inference system



Fuzzy if-then rules

A single fuzzy if-then rule is in the form:

if x is A then y is B

Where A and B are linguistic values defined by fuzzy sets on the universes of discourse X and Y, respectively. The if-part of the rule, "x is A", is called the antecedent or premise, while the then-part of the rule, "y is B", is called the *consequent* or conclusion. An example of such a rule might be

If you are late for your meeting, *then* you should walk fast.

Note that "late" is represented by a fuzzy set, and so the antecedent could be interpreted by a single number between 0 and 1, depending on the "degree of lateness". On the other hand, the consequent is the "fast" fuzzy set, which should later be *defuzzified* to assign a single numerical value to the output.

Both the antecedent and the consequent of a rule can have multiple parts. In that case all parts of the antecedent are calculated concurrently to determine a single number, using the logical operators described in the previous section. On the other hand, all consequents are affected equally by the result of the antecedent. An example of a fuzzy rule with two antecedent and two consequent parts could be:

If you are late for your meeting *and* the meeting is important,

then you should walk fast or catch a taxi.

In the case of traditional or binary logic, interpreting the if-then rules does not present much difficulty. If the premise is true, then the conclusion is true. When dealing with a fuzzy if-then rule however, the premise or the antecedent could be only partially true, which will affect the consequent of the rule.

The following steps are to be taken when interpreting a fuzzy if-then rule:

1. Fuzzifying the inputs: Means resolving all parts of the antecedent to a degree of membership between 0 and 1. In other words, it means calculating the truth-value for all fuzzy expressions in the antecedent.

2. Applying fuzzy operators: When the antecedent has multiple parts, fuzzy operators (t-norm or s-norm) need to be applied to resolve the antecedent (by combining the truth values of all fuzzy expressions) to a single number between 0 and 1, called the *degree of support* for the rule.

3. Applying *implecation* methods: means using the degree of support for the rule to shape the output fuzzy set. If the antecedent is only partially true, (i.e., is assigned a value less than 1), then the output fuzzy set is truncated according to the implication method. The most common implication methods are the minimum (which removes the α-cut for α = "degree of support" from the output fuzzy set), and the product (which multiplies the output fuzzy set by the degree of support).

To make it more clear, lets consider the following example:

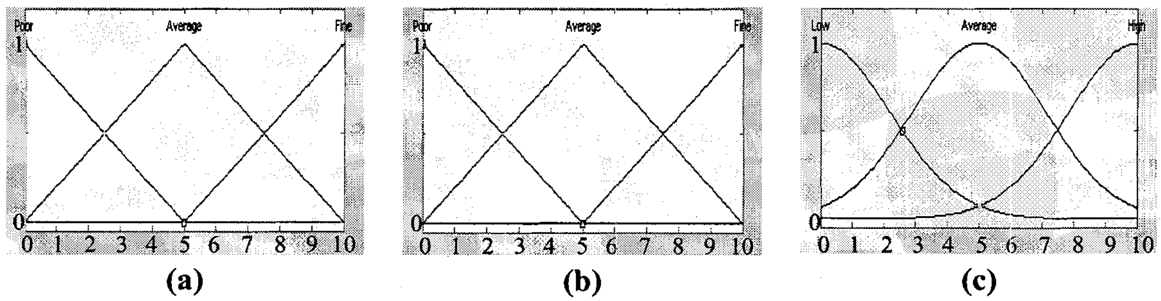
The restaurant rating problem: Given two sets of numbers between 0 and 10 (where 10 is excellent) that respectively represent the quality of the service and the quality of the food at a restaurant, in a scale of 1 to 5 (5 being excelent) what should the rating of the restaurant be?

In this problem, we have two inputs, "the quality of service" and "the quality of food", and one output, "the restaurant rating". A series of rules (in linguistic terms) can be written to connect these inputs to the output. Suppose our set of rules is consist of the following three rules:

1. If "the quality of service" is *fine* and "the quality of food" is *fine*, then the "rating of the restaurant" is *high*.
2. If "the quality of service" is *average* and "the quality of food" is *fine*, then the "rating of the restaurant" is *medium*.
3. If "the quality of service" is *poor* or "the quality of food" is *poor*, then the "rating of the restaurant" is *low*.

Three different linguistic terms are used in our set of rules that describe the inputs (fine, average, and poor) and the output (high, medium, and low). Figure 23 demonstrates the fuzzy sets that represent these linguistic terms. For each variable, the three corresponding fuzzy sets are shown in one graph (e.g. "poor", "average", and "fine" quality of service fuzzy sets in Figure 23 (a)).

Figure 23: Input and output fuzzy sets for the "restaurant rating" problem (a): fuzzy sets describing the "quality of service", (b): fuzzy sets describing the "quality of food", (c): fuzzy sets describing the "restaurant rating" (x-axis: variable's universe of discourse, y-axis: truth-values).



As represented in Figure 23, fuzzy sets have divided the universe of discourse of the inputs and the output of the system into regions, demonstrating the linguistic terms that describe each variable. For example in Figure 23 (a), the "poor" fuzzy set is representative of the "poor quality of service" values. Therefore, the low values assigned to the "quality of service" input, which mean a lower level of service in a restaurant, have a high truth-value in the "poor" fuzzy set.

Suppose restaurant "X" whose "quality of service" and "quality of food" are characterized by numbers 6 and 8, respectively, is to be rated with our fuzzy model. To do so, we first need to interpret the model's three rules. Let's start with a step-by-step interpretation of rule 1:

If "the quality of service" *is good and* "the quality of food" *is good,*

then the "rating of the restaurant" *is* high

1. Fuzzifying the inputs: the numbers 6 and 8, representing the service and food quality in restaurant "X" should be fuzzified. In other words, their truth-values in the corresponding fuzzy sets (i.e. "fine" fuzzy set in "quality of service" and "fine" fuzzy set in "quality of food") need to be determined. As shown in Figure 23 (a), The truth-value of number 6 is 0.2 in the "fine quality of service" fuzzy set. Also, the truth-value of number 8 is 0.6 in the "fine quality of food" fuzzy set.

2. Applying fuzzy operators: The degree of support for rule 1 needs to be calculated by resolving the antecedent into a single number, using fuzzy t-norm and s-norm operators. In this example, we will apply the minimum and maximum operators as the fuzzy t-norm and s-norm operators respectively. The word "AND" that connects the two parts of the antecedent tells us that we need to use a t-norm (the minimum) operator. Therefore the degree of support for rule 1 will be calculated as:

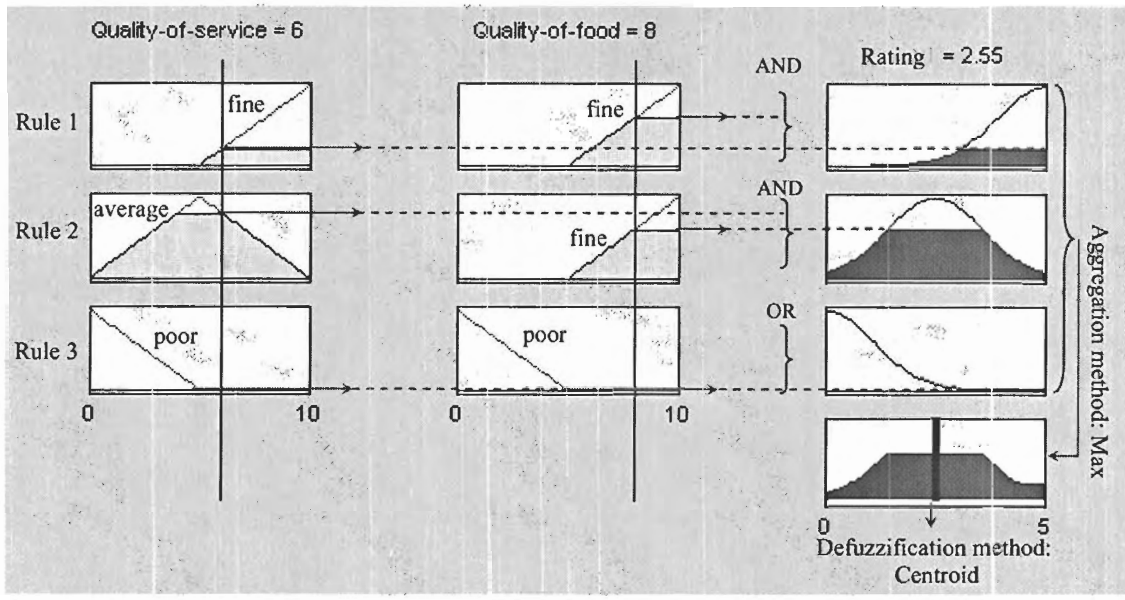
Rule 1's degree of support = $min\{0.2, 0.6\} = 0.2$

3. Applying the implication method: We apply the minimum implication method to truncate the "high" fuzzy set in the output, by the rule's degree of support (i.e. 0.2)

The other two rules of the restaurant rating problem are interpreted in a similar way. Each row in Figure 24 demonstrates the process of interpreting one of the three rules. For example in row 1, the truth-value for number 6 in the "fine quality of service" fuzzy set, and number 8 in the "fine quality of food" fuzzy set is determined. The minimum of the two truth-values, or the degree of support for rule 1, is carried over and truncated the "high restaurant rating" fuzzy set in the output.

The output fuzzy sets for each rule need to be combined in some manner, so that they all contribute to the final output of the FIS. This is called the *aggregation* process.

Figure 24: Fuzzy inference process for the restaurant rating problem



The aggregation process

The input of the aggregation process is the list of truncated output functions returned by the implication process for each rule. The output of the aggregation process is one fuzzy set for each output variable. The aggregation method is commutative, therefore the order in which the rules are executed is unimportant. Two of the common methods of aggregation are the *maximum* (maximum of all rules output sets), and *sum* (the sum of all rules output sets). The last row in Figure 24 shows an example of the aggregation process (maximum) in the restaurant rating problem.

The result of the aggregation process is a fuzzy set that needs to be resolved to a crisp number, as the final desired output in an FIS is generally a single number. This process is called the *defuzzification* process.

The defuzzification process

Defuzzification is the final stage in the fuzzy inference process. The input for the defuzzification process is a fuzzy set (the output of the aggregation process), and the output is a single number. Some of the defuzzification methods are the centroid method (which calculates the center of mass under the fuzzy set), the bisector (which returns bisector of area under the curve), middle of maximum (the average of the maximum value of the output set), largest of maximum (largest of the maximum values of the output set), and smallest of maximum.

Figure 25: Different Defuzzification methods applied to an example fuzzy output curve. The vertical line shows location of the numerical fuzzy output over the output curve. Methods applied: (a): Centroid, (b): Bisector, (c): Middle of maximums, (d): Largest of maximums, (e): Smallest of maximums

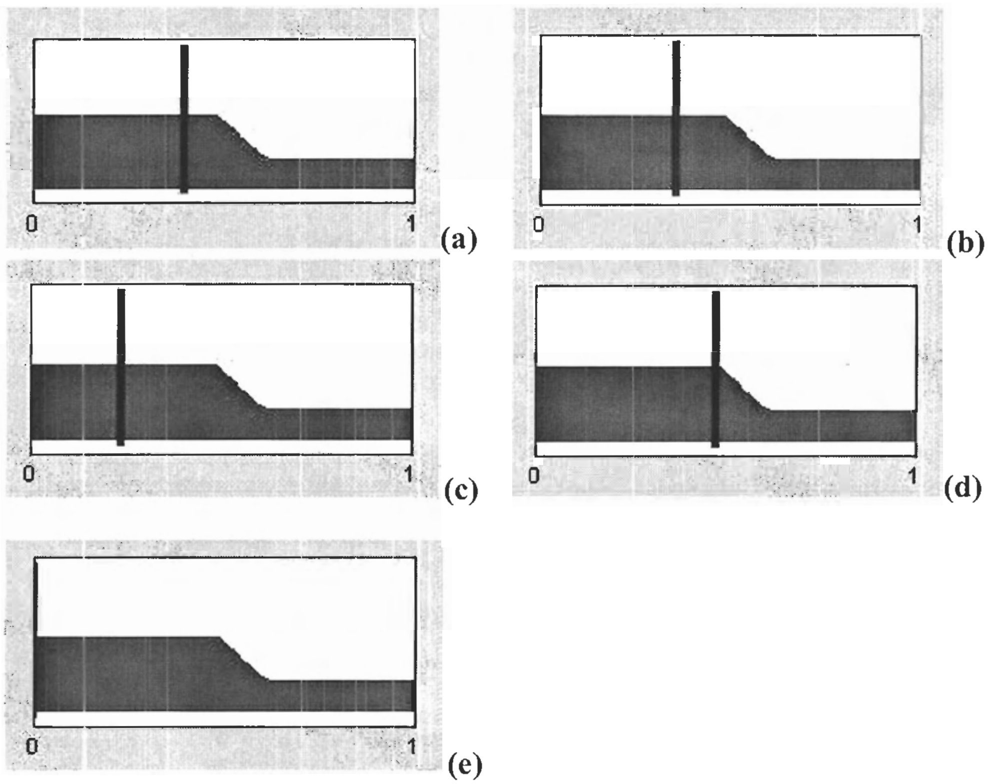


Figure 25 demonstrates examples of these methods. The vertical line in each figure shows the location of the defuzzified output value over the output fuzzy set.

We used the centroid method to determine the final output for the restaurant rating problem. As shown in Figure 24, the final result or "the rating of the restaurant" is calculated to be 2.55 out of 5.

3.3.2 Design Process of Fuzzy Inference Systems

This section is a step-by-step explanation of the process of designing a Mamdani-type fuzzy inference system. Essentially, there are three fundamental stages in the construction of a fuzzy model:

- Selecting the input and output (control) variables
- Defining the fuzzy sets
- Constructing the relationship between input and output spaces (rules)
- Selecting the fuzzy inference properties

Selecting the input and output variables

The first step in designing a FIS is to define the system's two major elements: the information (data points) that flows into the system, and the data elements that are eventually the outputs of the system. This involves identifying the inputs and outputs of the system based on the initial information and the goal of the problem, and specifying the universe of discourse for each of the input and output variables.

Defining the fuzzy sets

Fuzzy sets need to be defined to classify the input and output variables into categories (or classes of data) that represent possible states of that variable. Linguistic terms are usually used to identify these data categories. For example in the restaurant rating problem introduced in section 3.3.1, fuzzy sets *fine*, *average*, and *poor* categorize the "quality of service" input into three states.

Membership functions are curves that specify how each point in a fuzzy set's space maps to a membership value between 0 and 1. They could be drawn by intuition, which is derived from the intelligence and understanding of human beings and involves contextual and semantic knowledge about an issue (similar to the approach taken in the restaurant rating problem in section 3.3.1), or defined on the basis of numerical data. In the latter case, groups of data that produce a concise representation of the variables behaviour characterize the fuzzy sets and membership functions. *Clustering* of numerical data is a way of identifying the natural groupings of data from a large dataset. *Fuzzy c-means* (FCM) is an example of a clustering technique in which each data point belongs to a cluster that is defined by a membership degree. The algorithm starts with an initial guess for the cluster centres (meant to mark the mean location of each cluster) and a membership grade for each cluster assigned to every data point. The cluster centres move to the right location by iteratively minimizing a function that represents the distance from any given data point to a cluster centre weighted by that data point's membership grade.

Figure 23 in section 3.3.1 demonstrates how fuzzy sets define the different linguistic term for the input/output variables in the restaurant rating problem.

Constructing the relationship between input and output spaces (rules)

Fuzzy *conditional statements*, or simply fuzzy if-then rules, describe the relationship between the input and output variables in an FIS. Several methods have been proposed for generating fuzzy rules. Many of these methods, similar to the approach taken in this project, are based on clustering techniques (Yager & Filev, 1994; Hong & Lee 1996; Hong & Chen 1999). These methods can be categorized into two phases:

1. Partitioning (or clustering) the variable spaces into classes of data
2. Identifying fuzzy rules for each class of data

The process of designing fuzzy if-then rules in this project is described in more details in section 4.3.

Selecting the Fuzzy Inference Properties

Properties of the FIS, such as shapes of the membership functions, type of the t-norm and t-conorm processors, the aggregation method, and the defuzzification method need to be defined during the design process of the system. These properties may be selected intuitively (based on the nature of the problem and judgment of the system's expert) or derived from the numerical data.

3.3.3 Fuzzy Logic Toolbox in MATLAB

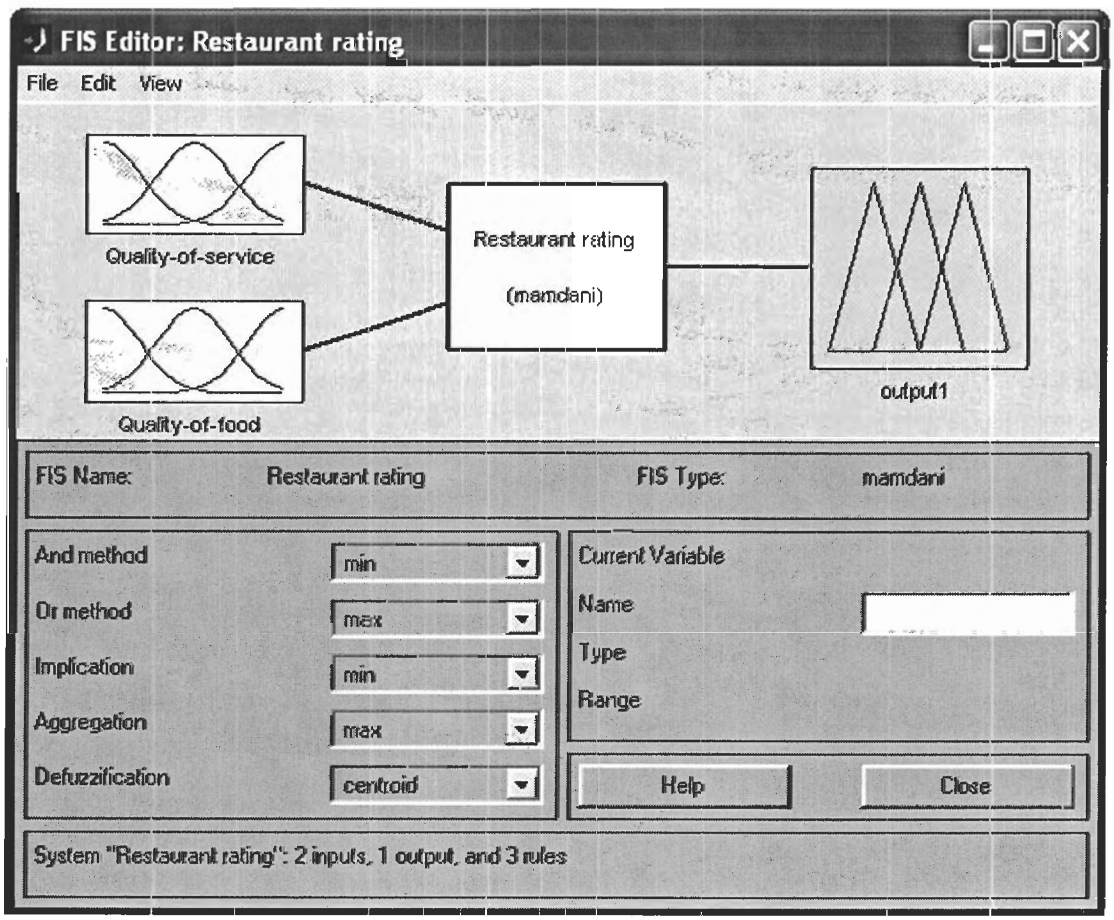
In this project, the fuzzy logic toolbox in MATLAB (Fuzzy logic toolbox, 2004) was used for modeling the fuzzy systems. The fuzzy logic toolbox is a collection of functions built on the MATLAB numeric environment. It relies on the graphical user interface (GUI) tools that provide an environment for fuzzy inference system design, analysis and implementation. Five primary GUI tools help building, editing, and observing fuzzy inference systems: The Fuzzy Inference or FIS editor, the membership function editor, the rule editor, the rule viewer, and the surface viewer. These tools are dynamically linked, thus changes made to the FIS using one of them affects the other four. In addition, the toolbox includes the Adaptive Neuro-Fuzzy Inference System (ANFIS) editor, which is used for building and analyzing Sugeno-type FISs.

The fuzzy inference system editor

The FIS editor handles the high level issues for the system such as the number of input and output variables and their names, types of the "And" and "Or" operators, and the aggregation and defuzzification methods.

Figure 26 shows the FIS editor for the restaurant rating problem as an example. It displays general information about the system. The diagram at the top of the window shows the names of the input and output variables. Popup menus on the bottom left allow the user to modify the FIS properties and the fields on the bottom right display the name, the membership type and the range for each of the input or output variables.

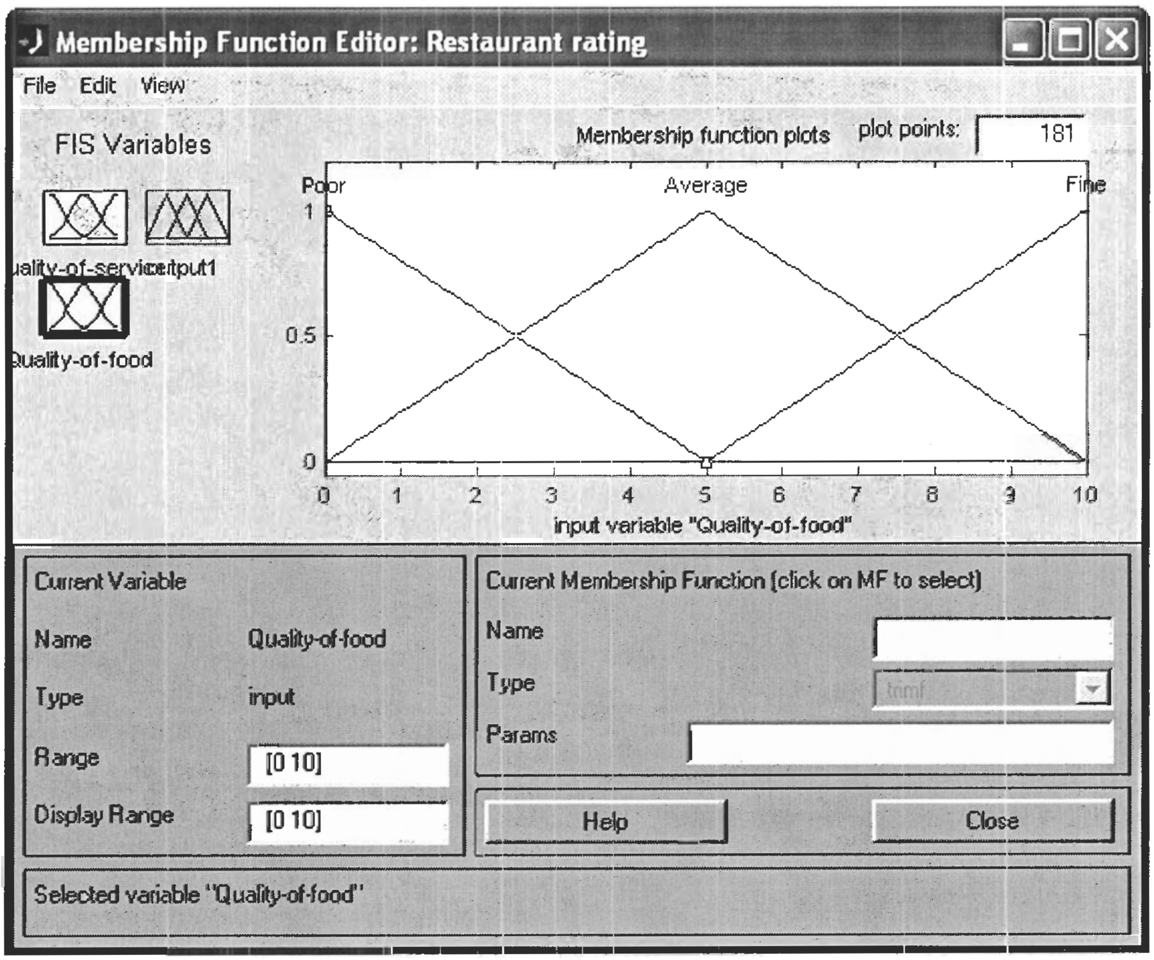
Figure 26: Example of the "FIS editor" window



The membership function editor

The membership function editor is used to define the properties of the membership functions for the system's variables.

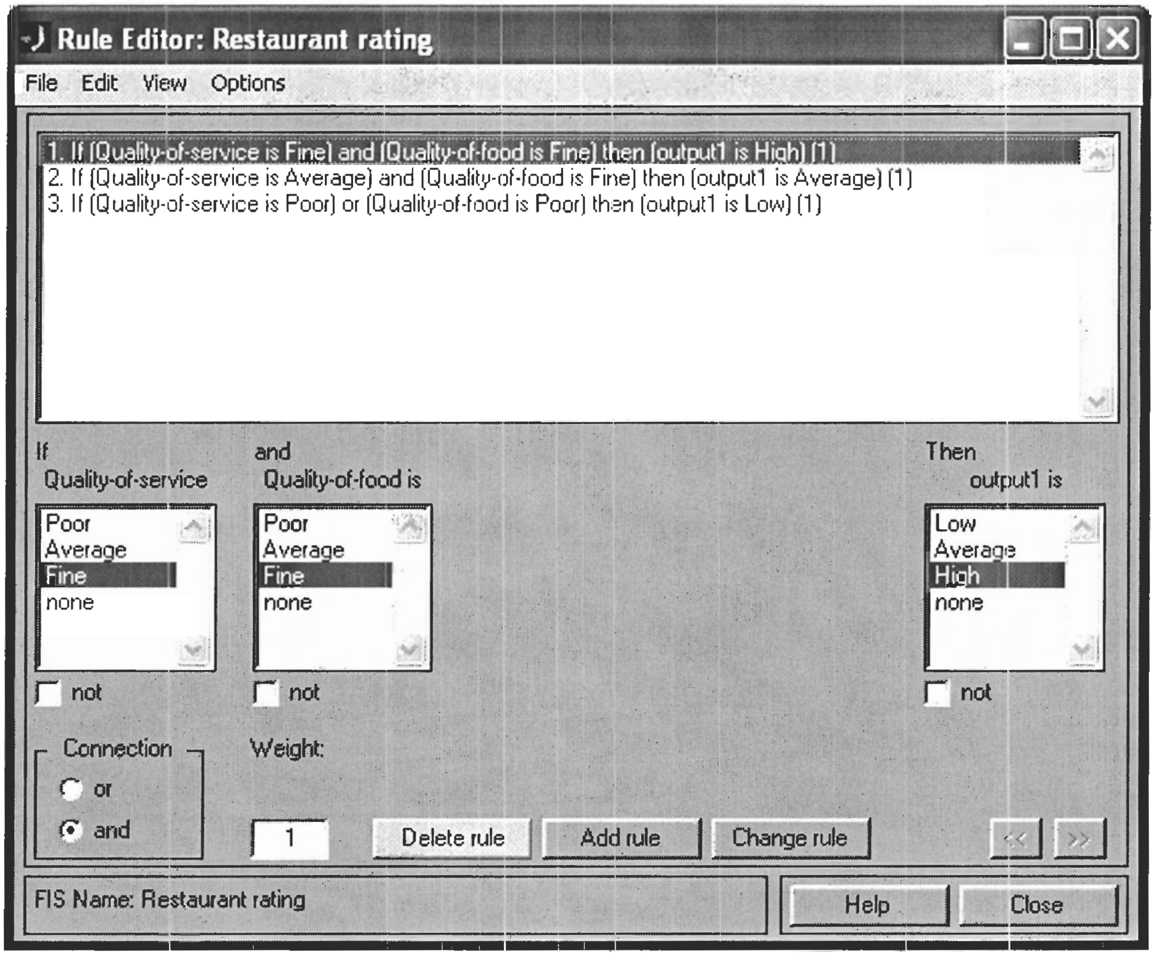
Figure 27: Example of the "membership function editor" window



The rule editor

The rule editor enables the user to define and edit the list of rules that describe the behaviour of the system.

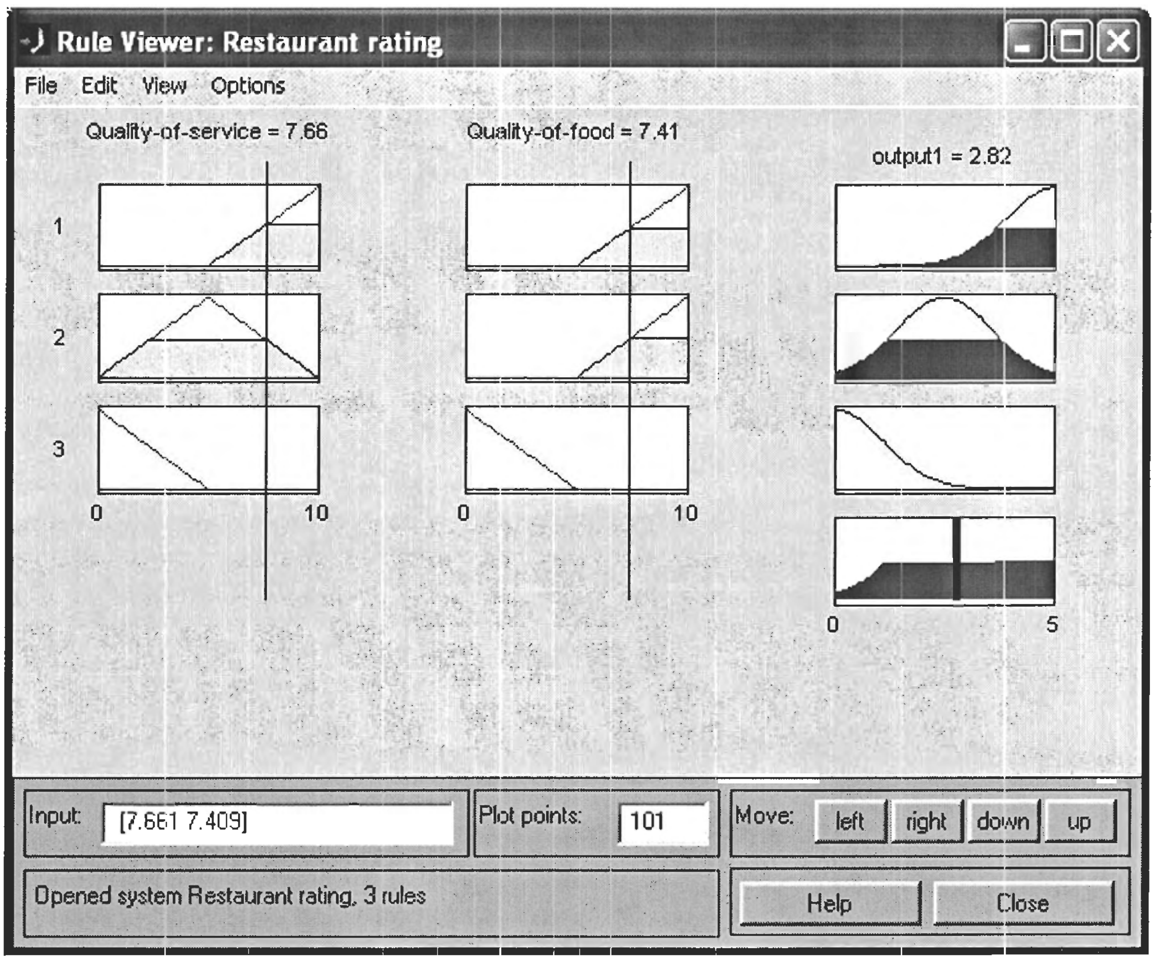
Figure 28: Example of the "rule editor" window



The rule viewer

The rule viewer is a read only tool that displays the whole fuzzy inference diagram. An example of the rule viewer window is demonstrated in Figure 29. Each column illustrates one variable and each row of small plots represent the antecedents and the consequents of one rule in the FIS. The rule viewer shows for a given set of inputs, which rules are active, what is the output, and how individual membership function are affecting the output.

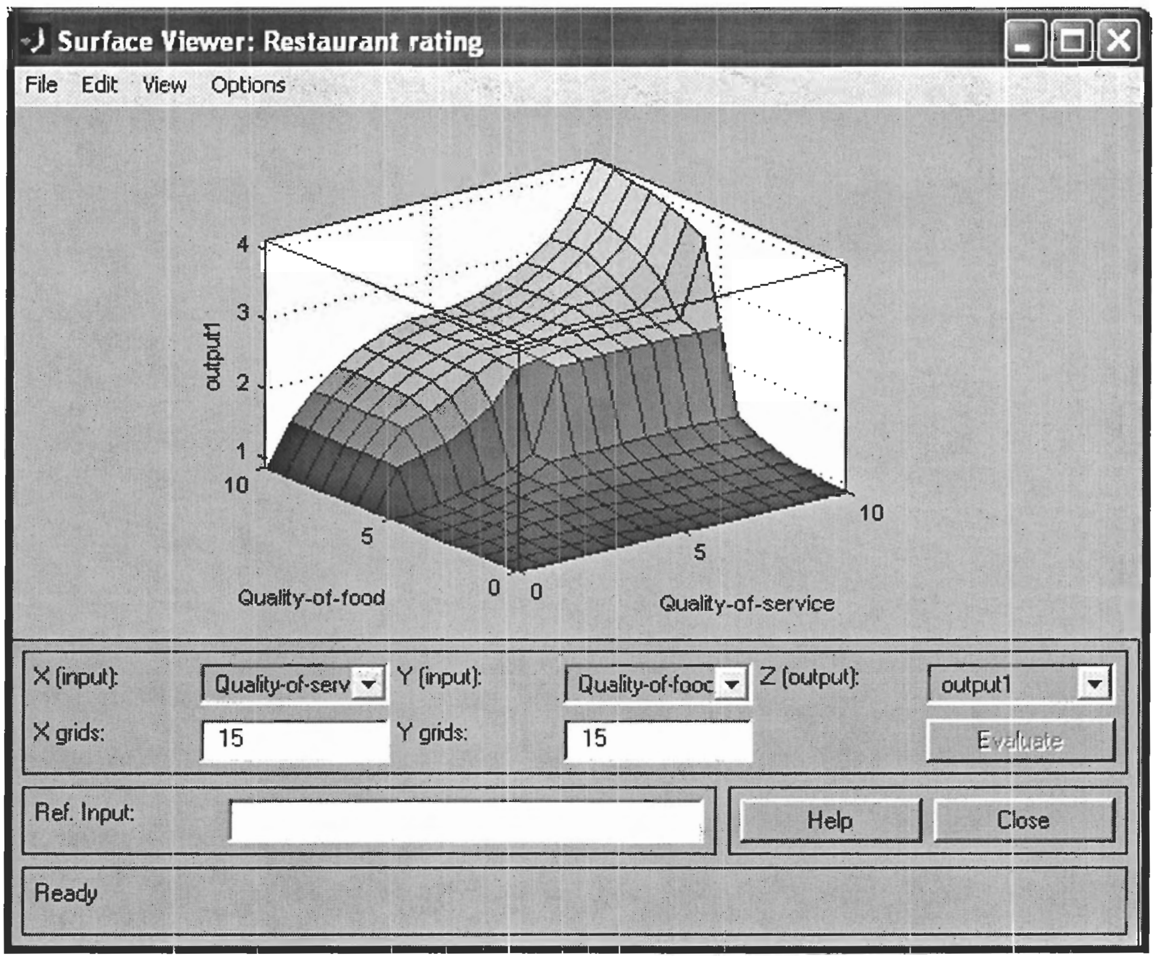
Figure 29: Example of the "rule viewer" window



The surface viewer

The surface viewer is also a read-only tool. It is used to display how an output is dependant on any one or two of the inputs. For instance Figure 30 shows the relationship between the output (z-axis) and the two inputs (x and y axes) in the restaurant rating problem. The popup menus at the bottom of the window allow the user to choose which variables to be plotted.

Figure 30: Example of the "surface viewer" window



4 DESIGN PROCESS OF FUZZY CLASSIFIERS FOR SURGICAL PERFORMANCE EVALUATION

Having 26 subjects participating in the experiment provided us with 26 full sets of data for each of the Stitch and HS Knot tasks. The data for each task was organized into 104 vectors in total (26 subjects who did 4 trials each). Each vector contained the 3 performance metrics for the trial. For instance the Stitch task data vectors were in the form of:

V (Time, Max Tissue Deformation, Number of Errors)

We then split each full set of data into two equal halves; 52 data vectors were used as Training Data Set (to design and train the classifier), and the other 52 as Testing Data Set (to test the constructed model). There were two possible approaches in getting separate training and testing data sets:

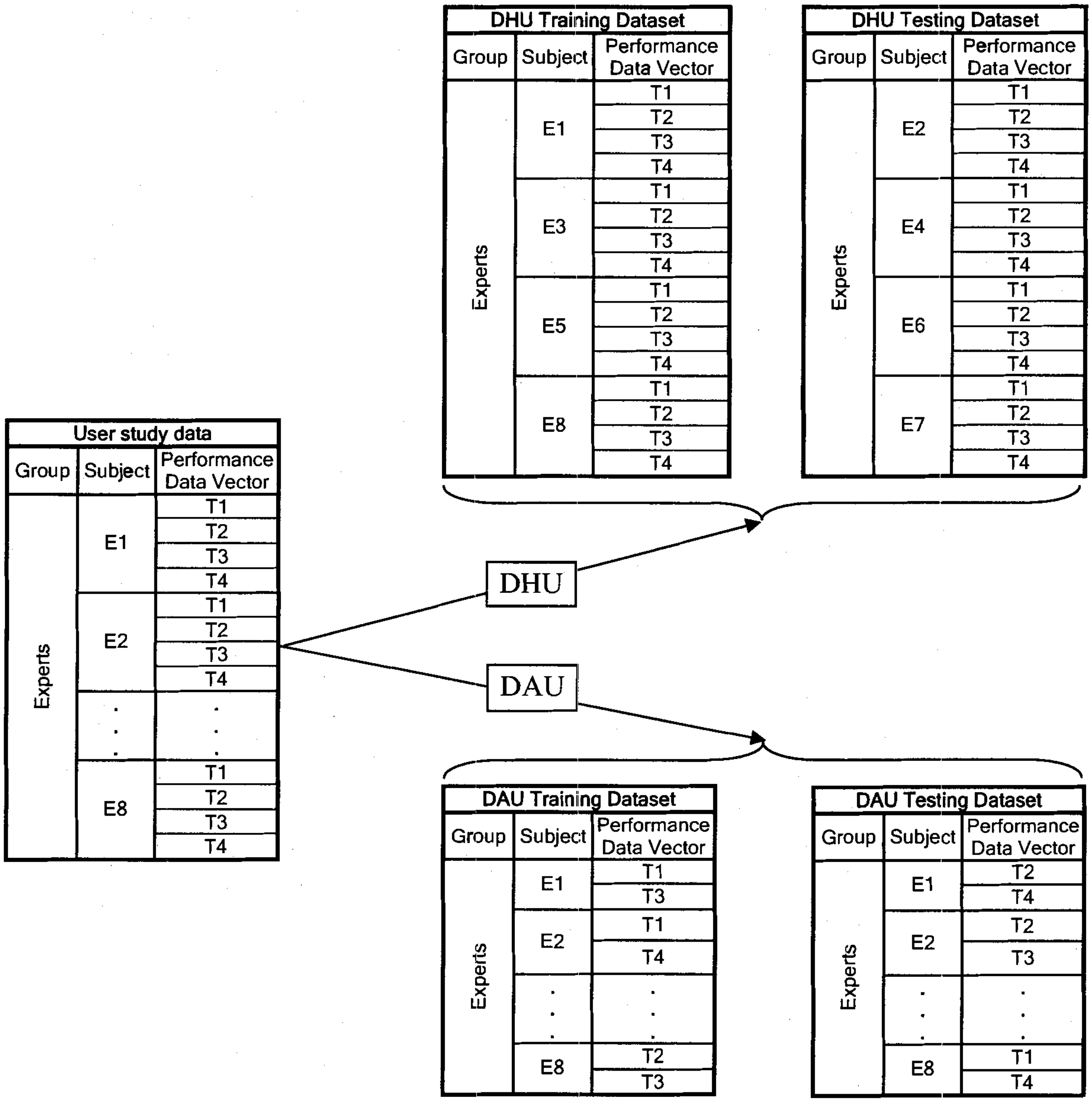
1. Using data from all participants for both sets, separating each subject's data into two groups randomly (we call this approach the "Data from All Users", or DAU method).

Or,

2. Dividing the subjects in each level of expertise into two groups randomly. Then using the data from one group to train the system and from the other group to evaluate the system (e.g. having 8 expert subjects, randomly select data from 4 of them to be in one group and from the other 4 to be in the second group). We call this approach the "Data from Half of the Users" or DHU method.

Figure 7 shows the DHU and DAU methods applied to data from the experts group as an example. Data from the Intermediate and Novice groups were split in the same way.

Figure 31: Example of the DHU and DAU data splitting methods



Either approach has costs and benefits; the limit of the DAU approach is that if our sample of surgeons is not representative of the overall population, we can train our system, and successfully evaluate it. But when we use trainees from outside our training group, the system doesn't work. This would be because our system had trained itself on features that are present in our group but not in the population at large.

The DHU approach suffers from the complementary risk: If our evaluation set does not do very well, is that because the group we trained it on is abnormal or the evaluation group is non-standard? Hard to say! Therefore each method was employed once to create a fuzzy classifier for each of the Stitch and HS Knot tasks, providing us with four fuzzy classifiers in total.

The training and testing datasets for the fuzzy models are demonstrated in Table 10 to Table 13.

Table 10: Stitch task (DAU) - Left: Training dataset, Right: Testing dataset

Data vector	Time	Max Tissue Deformation	Number of Errors	Level of Expertise
1	0.244507	0.413408	0.129032	E
2	0.128451	0.413408	0.090909	E
3	0.291831	0.413408	0.242424	E
4	0.178028	0.413408	0.272727	E
5	0.251831	0.413408	0.272727	E
6	0.167887	0.413408	0.090909	E
7	0.362817	0.413408	0.090909	E
8	0.201127	0.413408	0.030303	E
9	0.287887	0.413408	0.060606	E
10	0.264225	0.413408	0.060606	E
11	0.231549	0.413408	0.121212	E
12	0.28	0.413408	0.242424	E
13	0.252394	0.413408	0.090909	E
14	0.19493	0.413408	0.181818	E
15	0.322254	0.413408	0.121212	E
16	0.170704	0.413408	0.030303	E
17	0.180845	0.413408	0.151515	I
18	0.151549	0.413408	0.030303	I
19	0.155493	0.413408	0.242424	I
20	0.169577	0.413408	0.181818	I
21	0.294648	0.413408	0.212121	I
22	0.329577	0.413408	0.151515	I
23	0.310423	0.413408	0.151515	I
24	0.141972	0.413408	0	I

Data vector	Time	Max Tissue Deformation	Number of Errors	Level of Expertise
1	0.131831	0.575419	0.129032	E
2	0.127887	0.296089	0.030303	E
3	0.242817	0.547486	0.121212	E
4	0.23831	0.670391	0.151515	E
5	0.256901	0.972067	0.30303	E
6	0.190986	0.905028	0.212121	E
7	0.224789	0.368715	0.030303	E
8	0.216338	0.324022	0.121212	E
9	0.458028	0.47486	0.454545	E
10	0.371268	0.407821	0.454545	E
11	0.194366	0.402235	0.121212	E
12	0.2	0.312849	0.121212	E
13	0.149296	0.586592	0.090909	E
14	0.094648	0.251397	0	E
15	0.222535	0.603352	0.393939	E
16	0.23831	0.223464	0.060606	E
17	0.126761	0.513966	0.090909	I
18	0.190423	0.363128	0.030303	I
19	0.190423	0.340782	0.121212	I
20	0.179155	0.586592	0.181818	I
21	0.286761	0.536313	0.242424	I
22	0.240563	0.385475	0.060606	I
23	0.211831	0.296089	0.060606	I
24	0.27662	0.335196	0.181818	I

Data vector	Time	Max Tissue Deformation	Number of Errors	Level of Expertise
26	0.130704	0.413408	0.060606	I
27	0.342535	0.463687	0.060606	I
28	0.232676	0.290503	0.030303	I
29	0.491268	0.184358	0.181818	I
30	0.225352	0.27933	0.090909	I
31	0.230986	0.379888	0.242424	I
32	0.286197	0.469274	0.30303	I
33	0.72	0.715084	0.666667	N
34	0.593239	0.581006	0.878788	N
35	0.221408	0.620112	0.212121	N
36	0.216338	0.458101	0.090909	N
37	0.532394	0.620112	0.69697	N
38	0.303662	0.810056	0.515152	N
39	0.517183	0.642458	0.484848	N
40	0.237746	0.463687	0.151515	N
41	0.157183	0.709497	0.121212	N
42	0.19662	0.614525	0.151515	N
43	0.332394	0.547486	0.424242	N
44	0.234366	0.569832	0.363636	N
45	0.322254	0.726257	0.363636	N
46	0.301408	0.586592	0.575758	N
47	0.47493	0.709497	0.242424	N
48	1	0.653631	1	N
49	0.459155	0.480447	0.692308	N
50	0.260845	0.553073	1	N
51	0.417465	0.513966	0.846154	N
52	0.185352	0.363128	1	N

Data vector	Time	Max Tissue Deformation	Number of Errors	Level of Expertise
26	0.175775	0.290503	0.090909	I
27	0.259718	0.497207	0.151515	I
28	0.296901	0.374302	0.121212	I
29	0.388732	0.363128	0.151515	I
30	0.365634	0.217877	0.181818	I
31	0.239437	0.290503	0.212121	I
32	0.148732	0.424581	0.090909	I
33	0.940282	0.564246	0.939394	N
34	0.508732	0.653631	0.272727	N
35	0.251268	0.759777	0.090909	N
36	0.210704	0.72067	0.090909	N
37	0.342535	0.396648	0.393939	N
38	0.276056	0.195531	0.121212	N
39	0.341408	0.977654	0.151515	N
40	0.330141	0.664804	0.636364	N
41	0.242817	0.681564	0.151515	N
42	0.176338	0.865922	0.121212	N
43	0.170141	0.340782	0.060606	N
44	0.170141	0.312849	0.181818	N
45	0.218592	0.430168	0.212121	N
46	0.369014	0.502793	0.484848	N
47	0.198873	0.413408	0.461538	N
48	0.146479	0.458101	0.461538	N
49	0.247887	0.435754	0.384615	N
50	0.317746	0.458101	0.769231	N
51	0.258028	0.385475	1	N
52	0.265352	0.413408	1	N

Table 11: Stitch task (DHU) - Left: Training dataset, Right: Testing dataset

Data vector	Time	Max Tissue Deformation	Number of Errors	Level of Expertise
1	0.244507	0.413408	0.129032	E
2	0.131831	0.575419	0.129032	E
3	0.128451	0.75419	0.090909	E
4	0.127887	0.296089	0.030303	E
5	0.287887	0.877095	0.060606	E
6	0.458028	0.47486	0.454545	E
7	0.371268	0.407821	0.454545	E
8	0.264225	0.469274	0.060606	E
9	0.231549	0.458101	0.121212	E
10	0.28	0.564246	0.242424	E
11	0.194366	0.402235	0.121212	E
12	0.2	0.312849	0.121212	E
13	0.149296	0.586592	0.090909	E
14	0.252394	0.480447	0.090909	E
15	0.19493	0.592179	0.181818	E
16	0.094648	0.251397	0	E
17	0.190423	0.340782	0.121212	I
18	0.179155	0.586592	0.181818	I
19	0.155493	0.502793	0.242424	I
20	0.169577	0.569832	0.181818	I
21	0.286761	0.536313	0.242424	I
22	0.294648	0.441341	0.212121	I
23	0.240563	0.385475	0.060606	I
24	0.329577	0.497207	0.151515	I
25	0.342535	0.463687	0.060606	I
26	0.259718	0.497207	0.151515	I

Data vector	Time	Max Tissue Deformation	Number of Errors	Level of Expertise
1	0.291831	1	0.242424	E
2	0.178028	0.715084	0.272727	E
3	0.242817	0.547486	0.121212	E
4	0.23831	0.670391	0.151515	E
5	0.251831	0.698324	0.272727	E
6	0.167887	0.793296	0.090909	E
7	0.256901	0.972067	0.30303	E
8	0.190986	0.905028	0.212121	E
9	0.362817	0.614525	0.090909	E
10	0.201127	0.173184	0.030303	E
11	0.224789	0.368715	0.030303	E
12	0.216338	0.324022	0.121212	E
13	0.322254	0.318436	0.121212	E
14	0.170704	0.24581	0.030303	E
15	0.222535	0.603352	0.393939	E
16	0.23831	0.223464	0.060606	E
17	0.180845	0.329609	0.151515	I
18	0.151549	0.391061	0.030303	I
19	0.126761	0.513966	0.090909	I
20	0.190423	0.363128	0.030303	I
21	0.310423	0.24581	0.151515	I
22	0.211831	0.296089	0.060606	I
23	0.27662	0.335196	0.181818	I
24	0.141972	0.368715	0	I
25	0.172394	0.240223	0.030303	I
26	0.194366	0.363128	0.090909	I

Data vector	Time	Max Tissue Deformation	Number of Errors	Level of Expertise
28	0.232676	0.290503	0.030303	I
29	0.230986	0.379888	0.242424	I
30	0.239437	0.290503	0.212121	I
31	0.148732	0.424581	0.090909	I
32	0.286197	0.469274	0.30303	I
33	0.72	0.715084	0.666667	N
34	0.940282	0.564246	0.939394	N
35	0.508732	0.653631	0.272727	N
36	0.593239	0.581006	0.878788	N
37	0.242817	0.681564	0.151515	N
38	0.176338	0.865922	0.121212	N
39	0.157183	0.709497	0.121212	N
40	0.19662	0.614525	0.151515	N
41	0.47493	0.709497	0.242424	N
42	1	0.653631	1	N
43	0.198873	0.413408	0.461538	N
44	0.146479	0.458101	0.461538	N
45	0.459155	0.480447	0.692308	N
46	0.247887	0.435754	0.384615	N
47	0.317746	0.458101	0.769231	N
48	0.260845	0.553073	1	N
49	0.417465	0.513966	0.846154	N
50	0.258028	0.385475	1	N
51	0.265352	0.413408	1	N
52	0.185352	0.363128	1	N

Data vector	Time	Max Tissue Deformation	Number of Errors	Level of Expertise
28	0.130704	0.368715	0.060606	I
29	0.388732	0.363128	0.151515	I
30	0.491268	0.184358	0.181818	I
31	0.365634	0.217877	0.181818	I
32	0.225352	0.27933	0.090909	I
33	0.251268	0.759777	0.090909	N
34	0.221408	0.620112	0.212121	N
35	0.216338	0.458101	0.090909	N
36	0.210704	0.72067	0.090909	N
37	0.532394	0.620112	0.69697	N
38	0.303662	0.810056	0.515152	N
39	0.342535	0.396648	0.393939	N
40	0.276056	0.195531	0.121212	N
41	0.517183	0.642458	0.484848	N
42	0.341408	0.977654	0.151515	N
43	0.330141	0.664804	0.636364	N
44	0.237746	0.463687	0.151515	N
45	0.332394	0.547486	0.424242	N
46	0.170141	0.340782	0.060606	N
47	0.170141	0.312849	0.181818	N
48	0.234366	0.569832	0.363636	N
49	0.322254	0.726257	0.363636	N
50	0.301408	0.586592	0.575758	N
51	0.218592	0.430168	0.212121	N
52	0.369014	0.502793	0.484848	N

Table 12: HS Knot task (DAU) - Left: Training dataset, Right: Testing dataset

Data vector	Time	Max Thread Overstretch	Number of Errors	Level of Expertise
1	0.100583	1	0	E
2	0.073615	0	0.032258	E
3	0.102041	1	0.129032	E
4	0.057337	1	0.032258	E
5	0.276482	0.9	0.193548	E
6	0.163022	0	0.16129	E
7	0.095481	1	0.096774	E
8	0.128037	1	0.096774	E
9	0.165452	0.1	0	E
10	0.106414	0	0.032258	E
11	0.154762	1	0.129032	E
12	0.103984	0	0.064516	E
13	0.05345	1	0	E
14	0.061467	0.4	0.032258	E
15	0.084548	1	0	E
16	0.067541	1	0.064516	E
17	0.108844	0.6	0	I
18	0.099611	1	0.096774	I
19	0.171283	0.4	0	I
20	0.140428	0	0.064516	I
21	0.149417	1	0.096774	I
22	0.225705	0	0.064516	I
23	0.074344	0.2	0.064516	I
24	0.070943	0.2	0.064516	I
25	0.093294	0	0	I
26	0.071672	0	0	I

Data vector	Time	Max Thread Overstretch	Number of Errors	Level of Expertise
1	0.095967	0	0	E
2	0.067784	0	0.064516	E
3	0.088921	0.1	0.032258	E
4	0.065112	0	0	E
5	0.223518	1	0.258065	E
6	0.163508	0.1	0.16129	E
7	0.068513	1	0.032258	E
8	0.082604	1	0.064516	E
9	0.139456	1	0.032258	E
10	0.152089	0.4	0.129032	E
11	0.087949	1	0.064516	E
12	0.129495	1	0.064516	E
13	0.061224	0	0	E
14	0.082604	1	0	E
15	0.101312	1	0.064516	E
16	0.066569	1	0	E
17	0.069971	0	0	I
18	0.065355	0	0	I
19	0.094752	0.1	0.064516	I
20	0.117833	0	0	I
21	0.074587	0	0	I
22	0.098154	0.4	0.129032	I
23	0.12415	0.2	0.096774	I
24	0.050292	1	0.032258	I
25	0.110787	0	0	I
26	0.072157	0	0	I

Data vector	Time	Max Thread Overstretch	Number of Errors	Level of Expertise
28	0.127551	0	0.096774	I
29	0.285957	0.1	0.032258	I
30	0.225705	0.1	0.096774	I
31	0.20068	0	0.322581	I
32	0.179057	1	0.16129	I
33	0.293246	0.1	0.580645	N
34	0.195335	0.1	0.548387	N
35	0.079446	0	0.096774	N
36	0.065355	0	0.032258	N
37	0.089893	0.1	0.16129	N
38	0.126093	0.5	0.193548	N
39	0.285957	0	0.225806	N
40	0.126093	0.1	0.096774	N
41	0.183916	0	0.032258	N
42	0.119776	0.1	0	N
43	0.119776	0	0.032258	N
44	0.184159	0.1	0.193548	N
45	1	0.9	1	N
46	0.126336	0	0.129032	N
47	0.112245	1	0.032258	N
48	0.149417	1	0.032258	N
49	0.249271	0.4	0.064516	N
50	0.233479	0.2	0.193548	N
51	0.431001	0	0.387097	N
52	0.184888	1	0.096774	N

Data vector	Time	Max Thread Overstretch	Number of Errors	Level of Expertise
28	0.238095	0	0.16129	I
29	0.18829	0	0	I
30	0.146501	1	0.032258	I
31	0.152089	0.5	0.064516	I
32	0.152332	1	0.129032	I
33	0.137512	0.1	0.225806	N
34	0.172012	0.1	0.387097	N
35	0.155491	1	0.064516	N
36	0.092566	0	0.129032	N
37	0.084548	0	0	N
38	0.071672	0.1	0.032258	N
39	0.157434	0	0.032258	N
40	0.304665	0	0.451613	N
41	0.241011	0.4	0.064516	N
42	0.123664	0	0	N
43	0.137998	0	0.096774	N
44	0.133625	0	0	N
45	0.216472	0.1	0.032258	N
46	0.096453	0	0.032258	N
47	0.152818	0	0.096774	N
48	0.094266	0	0.129032	N
49	0.165695	0.2	0	N
50	0.130466	0	0.032258	N
51	0.527697	0.4	0.645161	N
52	0.283285	0.1	0.193548	N

Table 13: HS Knot task (DHU) - Left: Training dataset, Right: Testing dataset

Data vector	Time	Max Thread Overstretch	Number of Errors	Level of Expertise
1	0.095967	0	0	E
2	0.100583	1	0	E
3	0.067784	0	0.064516	E
4	0.073615	0	0.032258	E
5	0.102041	1	0.129032	E
6	0.057337	1	0.032258	E
7	0.088921	0.1	0.032258	E
8	0.065112	0	0	E
9	0.095481	1	0.096774	E
10	0.128037	1	0.096774	E
11	0.068513	1	0.032258	E
12	0.082604	1	0.064516	E
13	0.165452	0.1	0	E
14	0.139456	1	0.032258	E
15	0.152089	0.4	0.129032	E
16	0.106414	0	0.032258	E
17	0.094752	0.1	0.064516	I
18	0.171283	0.4	0	I
19	0.117833	0	0	I
20	0.140428	0	0.064516	I
21	0.12415	0.2	0.096774	I
22	0.074344	0.2	0.064516	I
23	0.070943	0.2	0.064516	I
24	0.050292	1	0.032258	I
25	0.18829	0	0	I
26	0.285957	0.1	0.032258	I

Data vector	Time	Max Thread Overstretch	Number of Errors	Level of Expertise
1	0.276482	0.9	0.193548	E
2	0.223518	1	0.258065	E
3	0.163508	0.1	0.16129	E
4	0.163022	0	0.16129	E
5	0.154762	1	0.129032	E
6	0.103984	0	0.064516	E
7	0.087949	1	0.064516	E
8	0.129495	1	0.064516	E
9	0.061224	0	0	E
10	0.082604	1	0	E
11	0.05345	1	0	E
12	0.061467	0.4	0.032258	E
13	0.101312	1	0.064516	E
14	0.084548	1	0	E
15	0.066569	1	0	E
16	0.067541	1	0.064516	E
17	0.069971	0	0	I
18	0.108844	0.6	0	I
19	0.099611	1	0.096774	I
20	0.065355	0	0	I
21	0.074587	0	0	I
22	0.149417	1	0.096774	I
23	0.098154	0.4	0.129032	I
24	0.225705	0	0.064516	I
25	0.110787	0	0	I
26	0.093294	0	0	I

Data vector	Time	Max Thread Overstretch	Number of Errors	Level of Expertise
28	0.146501	1	0.032258	I
29	0.20068	0	0.322581	I
30	0.152089	0.5	0.064516	I
31	0.152332	1	0.129032	I
32	0.179057	1	0.16129	I
33	0.137512	0.1	0.225806	N
34	0.293246	0.1	0.580645	N
35	0.172012	0.1	0.387097	N
36	0.195335	0.1	0.548387	N
37	0.084548	0	0	N
38	0.089893	0.1	0.16129	N
39	0.071672	0.1	0.032258	N
40	0.126093	0.5	0.193548	N
41	0.157434	0	0.032258	N
42	0.285957	0	0.225806	N
43	0.304665	0	0.451613	N
44	0.126093	0.1	0.096774	N
45	0.137998	0	0.096774	N
46	0.119776	0	0.032258	N
47	0.184159	0.1	0.193548	N
48	0.133625	0	0	N
49	0.249271	0.4	0.064516	N
50	0.233479	0.2	0.193548	N
51	0.165695	0.2	0	N
52	0.130466	0	0.032258	N

Data vector	Time	Max Thread Overstretch	Number of Errors	Level of Expertise
28	0.071672	0	0	I
29	0.191448	0	0.064516	I
30	0.18829	0	0.16129	I
31	0.238095	0	0.16129	I
32	0.127551	0	0.096774	I
33	0.155491	1	0.064516	N
34	0.079446	0	0.096774	N
35	0.092566	0	0.129032	N
36	0.065355	0	0.032258	N
37	0.241011	0.4	0.064516	N
38	0.123664	0	0	N
39	0.183916	0	0.032258	N
40	0.119776	0.1	0	N
41	1	0.9	1	N
42	0.216472	0.1	0.032258	N
43	0.096453	0	0.032258	N
44	0.126336	0	0.129032	N
45	0.152818	0	0.096774	N
46	0.112245	1	0.032258	N
47	0.094266	0	0.129032	N
48	0.149417	1	0.032258	N
49	0.527697	0.4	0.645161	N
50	0.283285	0.1	0.193548	N
51	0.431001	0	0.387097	N
52	0.184888	1	0.096774	N

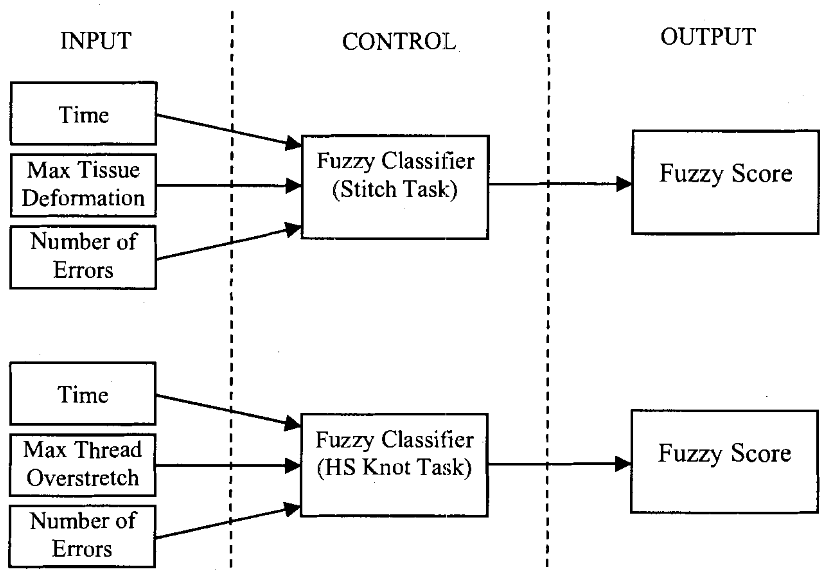
We designed our classifiers based on Mamdani's Fuzzy Inference Method (explained in section 3.3.2), through the following four steps:

- Selecting the input and output (control) variables
- Defining the fuzzy sets
- Constructing the relationship between input and output spaces (rules)
- Selecting the fuzzy inference properties

4.1 Selecting the Input and Output Variables

The nature of our classifiers suggests the inputs to the fuzzy systems to be user's performance metrics, and the outputs to be a fuzzy score representing user's surgical skill level. Thus, as shown in Figure 32, each system has three inputs and one output; "Time", "Maximum Tissue Deformation", and "Number of Errors" are inputs to the Stitch task classifiers, and "Time", "Maximum Thread Overstretch", and "Number of Errors" are inputs to the HS Knot task classifiers. Since we used normalized data values to design the classifiers, the universe of discourse for each of our input and output variables is between zero and one.

Figure 32: Control Variables in Stitch and HS Knot Fuzzy Systems

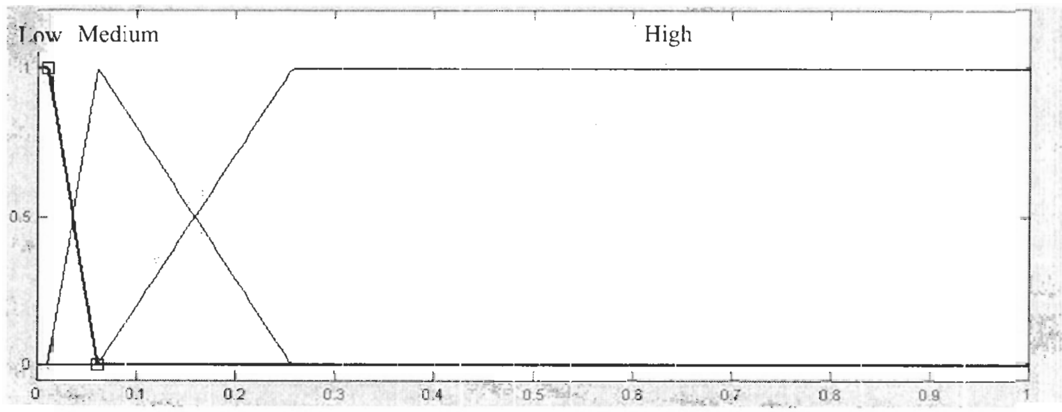


4.2 Defining the Fuzzy Sets

Knowing the input and output variables of the systems, fuzzy sets were required to be defined that classify these variables into different categories, so that each category represents a possible state of the corresponding variable. We defined three linguistic descriptors; Low, Medium, and High, each of which can be represented by a fuzzy membership function (or a fuzzy set). Low, Medium, and High membership functions for an input correspond to relatively low, medium or high values of the input, respectively. Figure 33 shows an example of the three membership functions for the "Number of Errors" input in the HS Knot task.

Figure 33: **Example of input membership functions**

"Number of Errors" input, HS Knot Task (x-axis: "Number of Errors" values, y-axis: truth-values)



We used the training datasets to determine the boundaries of the low, medium and high categories for each input variable, as follows:

The data values corresponding to each input variable were first sorted in the ascending order. We then used the Fuzzy Logic Toolbox in MATLAB (introduced in Section 3.3.3) to find three clusters of data among data values corresponding to each input variable, based on the *Fuzzy c-means* (FCM) clustering method (fuzzy logic toolbox, Fuzzy C-Means Clustering section, ¶ 1), explained in section 3.3.3. These data clusters represent the low, medium, and high categories for each input variable.

Example: Defining the Low, Medium, and High categories for the "Number of Errors" input variable in the HS Knot task

The left-side table in Figure 34 shows the data values in the training dataset corresponding to the "Number of Errors" variable in the HS Knot task. To determine the three categories mentioned earlier, we first arranged the data in the ascending order (Figure 34, Right-side table). MATLAB was then used to find three clusters of data values within the sorted data. These three groups represent the Low, Medium, and High values for the "Number of Errors" variable in the HS Knot task.

Figure 34: Example of clustering the data into Low, Medium, and High categories for the "Number of Errors variable in the HS Knot task

Training dataset
Number of Errors
0
0
0.064516
0.032258
0.129032
0.032258
0.032258
0
0.096774
0.096774
0.032258
0.064516
0
0.032258
0.129032
0.032258
0.064516
0
0
0.064516
0.096774
0.064516
0.064516
0.032258
0
0.032258
0.096774
0.032258
0.322581
0.064516
0.129032
0.16129
0.225806
0.580645
0.387097
0.548387
0
0.16129
0.032258
0.193548
0.032258
0.225806
0.451613
0.096774
0.096774
0.032258
0.193548
0
0.064516
0.193548
0
0.032258

Sort in ascending order

Sorted data	
Data Index	**Number of Errors**
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0.032258
12	0.032258
13	0.032258
14	0.032258
15	0.032258
16	0.032258
17	0.032258
18	0.032258
19	0.032258
20	0.032258
21	0.032258
22	0.032258
23	0.032258
24	0.064516
25	0.064516
26	0.064516
27	0.064516
28	0.064516
29	0.064516
30	0.064516
31	0.064516
32	0.096774
33	0.096774
34	0.096774
35	0.096774
36	0.096774
37	0.096774
38	0.129032
39	0.129032
40	0.129032
41	0.16129
42	0.16129
43	0.193548
44	0.193548
45	0.193548
46	0.225806
47	0.225806
48	0.322581
49	0.387097
50	0.451613
51	0.548387
52	0.580645

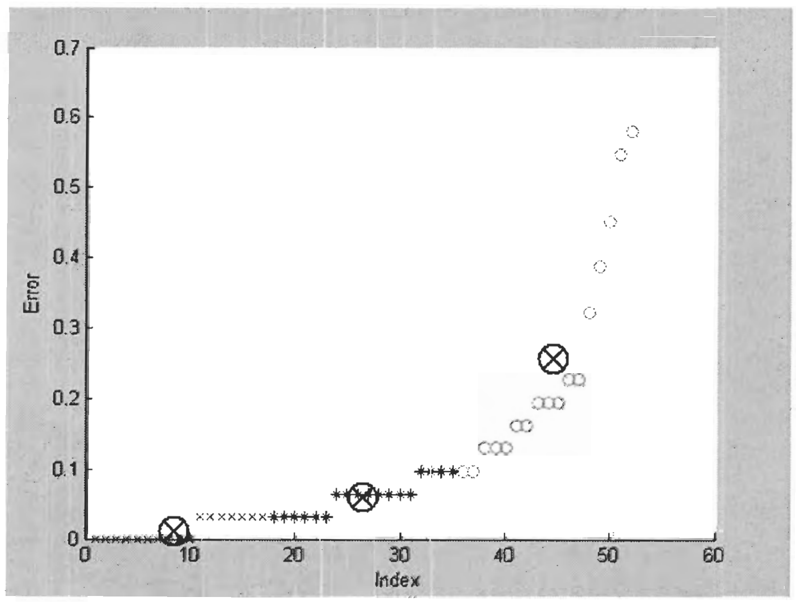
Low category

Medium category

High category

An index number was assigned to each data value to plot the 1-dimensional in 2-D, as shown in Figure 35. The three categories of Low, Medium, and High are shown with x's, *'s, and o's, respectively. The ⊗'s mark the centre values for each category or class of data (0.011, 0.061, and 0.267 are the centre values for Low, Medium, and High classes of data in "Number of Errors" in the HS Knot task).

Figure 35: Example of clusters of data in the "Number of Errors" variable in the HS Knot Task (x-axis: data index, y-axis: Number of Errors)



The commonly used triangular and trapezoidal shapes were selected to represent the membership functions (or fuzzy sets), and the following rules were followed to generate these functions:

1. In each class of data, the centre of the class has the maximum membership value (i.e. 1) in the corresponding membership function.

2. Membership functions representing the smallest or the largest linguistic term (i.e. Low and High categories) are trapezoidal in shape, since all values below the biggest, or above the smallest value with the highest membership, respectively, are considered to have the same maximum

membership value. The Medium linguistic term, which is in between the two end-of-the-range categories, is represented with triangular membership functions.

3. For the Low and High membership functions, all the values below or above the centre value, respectively, have the maximum truth-value (i.e. 1).

4. The overlap between each input's membership functions are chosen so that the sum of the truth-values of all points through the overlapping fuzzy sets is equal to one. This is called the "S*um-to-one (or less)*" rule (Cox, 1999).

Example: Defining the membership functions for the "Number of Errors" variable in the HS Knot task

Figure 36 shows the membership functions for the "Number of Errors" variable in the HS Knot task. The following steps were taken in defining these functions:

1. 0.011, 0.061, and 0.267, which are the centres of the Low, Medium, and High categories, respectively, were given truth-value of 1 in the corresponding membership functions (Figure 36).

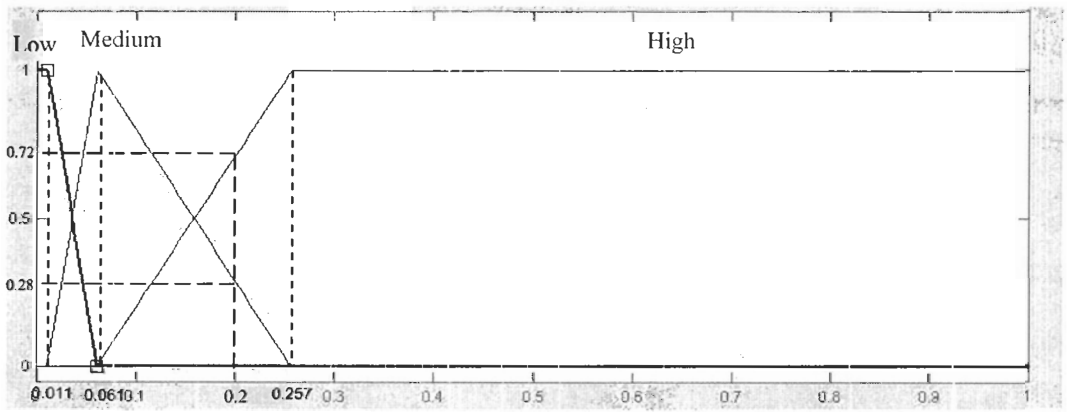
2. Trapezoidal functions were used to represent the Low and High fuzzy sets, and triangular function for the Medium fuzzy set.

3. In the Low fuzzy set, all the values below the centre of the low category (i.e. 0.011), were assigned the truth-value of 1. Also in the High membership function, all the values above 0.267, which is the centre of the High category were given the highest truth-value (Figure 36).

4. The "Sum-to-one" rule was used to determine the overlaps between the neighbouring membership functions. It means that for any value of "Number of Errors" input, the sum of the truth-values in all fuzzy sets is equal to one. For example, as shown in Figure 36, the truth-value for a

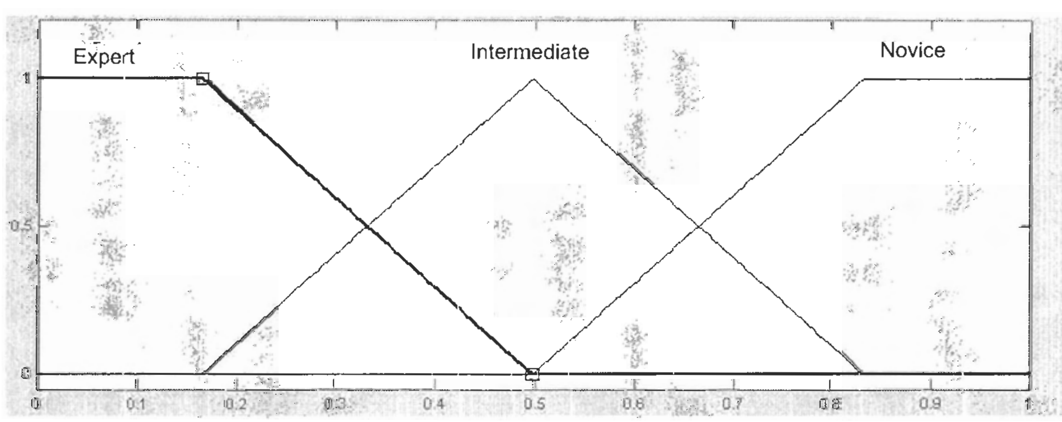
"Number of Errors" value of 0.2 is 0 in the Low fuzzy set, 0.28 in the Medium fuzzy set, and 0.72 in the High fuzzy set (0 + 0.28 + 0.72 = 1).

Figure 36: Membership functions for the "Number of Errors" input in HS Knot Task (x-axis: Input value, y-axis: Input's truth-value)



To categorize outputs of the systems, the fuzzy score (or **score)**, the output space was divided into three equal regions. As shown in Figure 37, the membership function corresponding to the lowest score values was named Experts, as those scores were expected to be achieved by the expert users. Same is for the Intermediate and Novice membership functions in the system's output. After testing the classifiers the output fuzzy sets could be adjusted iteratively to improve the performance of the system.

Figure 37: Output of the fuzzy classifier- (x-axis: output (score) value, y-axis: output's truth-value)



4.3 Rules

Data from the Training Datasets was used to construct the rules for the four fuzzy models: the Stitch task DAU, Stitch task DHU, HS Knot task DAU, and HS knot task DHU classifiers. For each user, performance metrics were first categorized into the class of data (Low, Medium, or High) that would best characterize their value, or in other words, the class of data that had the highest truth-value in the corresponding fuzzy set. We then constructed one rule per user, based on the classified user performance metrics and their level of expertise. The process of constructing the rules is explained further through the following example.

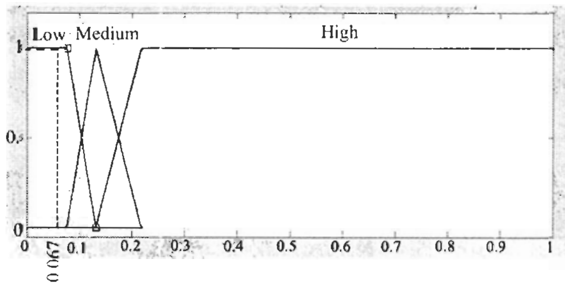
Example: Constructing a fuzzy rule, based on an expert user's performance metrics

Table 14 shows performance metrics of an expert user, for one trial of the HS Knot task.

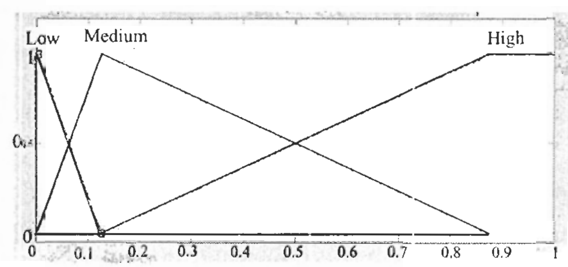
Table 14: Example of categorizing a user's performance metrics into one of the three classes of data (HS Knot task)

	Time	Max Thread Overstretch	Number of Errors
Performance metrics	0.067784	0	0.064516
Class of data	Low	Low	Medium

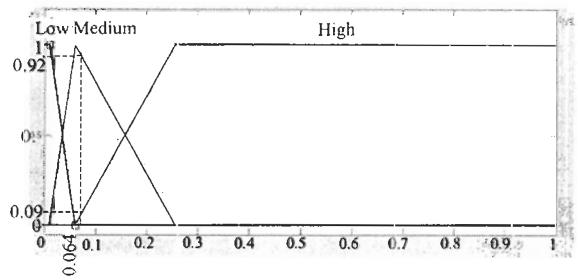
Figure 38: Membership functions for the HS Knot task
(a) Time, (b) Maximum Thread Overstretch, (c) Number of Errors - (x-axis: Input value, y-axis: Input's truth-value)



(a) Time



(b) Max Thread Overstretch



(c) Number of Errors

To classify the user's performance metrics, we look at the membership functions for the HS Knot task input variables, shown in Figure 38, to see which membership function has the highest truth-value for each of the user's performance metrics. For example, the user's "Time" value is approximately "0.067". Looking at Figure 38 (a), membership functions for the "Time" input variable, we will see that "0.067" has the truth-value of "1" in the Low fuzzy set, and "0" in the Medium and High fuzzy sets. Therefore, the Low fuzzy set is the best representative for the user's "Time" value, or in other words, *the user's "**Time**" is **Low***. Similarly, we can see that "0" is a **Low** value for "“**Maximum Thread Overstretch**”", and "0.064" is a **Medium** value for the "**Number of Errors**" input (in Figure 38 (c), the truth-value for "0.064" is "0.92" in the Medium fuzzy set, which is higher than its truth-value in the Low fuzzy set, "0", and the Medium fuzzy set, "0.09"). These classifications are shown in Table 14.

Knowing that the performance metrics shown in Table 14 are those of an expert user, if we feed them to the fuzzy classifier as inputs, the resulting fuzzy score is expected to be a rather low score, best represented by the **Expert** membership function in the output fuzzy set (Figure 37). To guarantee this outcome, we generate the following rule for our fuzzy system:

If "**Time**" is **Low** and "**Maximum Thread Overstretch**" is **Low** and "**Number of Errors**" is **Medium**,

then

the **Score** is (that of an) **Expert**

We can now be sure that if we test our system with a set of performance metrics that are Low, Low, and Medium in "Time", "Maximum Thread Overstretch", and "Number of Errors" inputs, respectively, the system will predict the user with those metrics to be an expert, unless the effect of this rule is influenced by contradictory rules as explained later.

For each fuzzy model, the same method was used to construct 52 rules, one for each data vector in the training dataset. Table 15 shows the categorized values of the HS Knot task (DHU) training dataset as an example.

Table 15: Categorized performance metrics in the training dataset, HS Knot task-DHU (L: Low, M: Medium, H: High, E: Expert, I: Intermediate, N: Novice)

Data vector	Time		Max Thread Overstretch		Number of Errors		Level of Expertise
	Value	Category	Value	Category	Value	Category	
1	0.095967	L	0	L	0	L	E
2	0.100583	L	1	H	0	L	E
3	0.067784	L	0	L	0.064516	M	E
4	0.073615	L	0	L	0.032258	L	E
5	0.102041	L	1	H	0.129032	M	E
6	0.057337	L	1	H	0.032258	L	E
7	0.088921	L	0.1	M	0.032258	L	E
8	0.065112	L	0	L	0	L	E
9	0.095481	L	1	H	0.096774	M	E
10	0.128037	M	1	H	0.096774	M	E
11	0.068513	L	1	H	0.032258	L	E
12	0.082604	L	1	H	0.064516	M	E
13	0.165452	M	0.1	M	0	L	E

Data vector	Time		Max Thread Overstretch		Number of Errors		Level of Expertise
	Value	Category	Value	Category	Value	Category	
14	0.139456	M	1	H	0.032258	M	E
15	0.152089	M	0.4	M	0.129032	M	E
16	0.106414	M	0	L	0.032258	L	E
17	0.094752	L	0.1	M	0.064516	M	I
18	0.171283	M	0.4	M	0	L	I
19	0.117833	M	0	L	0	L	I
20	0.140428	M	0	L	0.064516	M	I
21	0.12415	M	0.2	M	0.096774	M	I
22	0.074344	L	0.2	M	0.064516	M	I
23	0.070943	L	0.2	M	0.064516	M	I
24	0.050292	L	1	H	0.032258	L	I
25	0.18829	H	0	L	0	L	I
26	0.285957	H	0.1	M	0.032258	L	I
27	0.225705	H	0.1	M	0.096774	M	I
28	0.146501	M	1	H	0.032258	L	I
29	0.20068	H	0	L	0.322581	H	I
30	0.152089	M	0.5	M	0.064516	M	I
31	0.152332	M	1	H	0.129032	M	I
32	0.179057	H	1	H	0.16129	H	I
33	0.137512	M	0.1	M	0.225806	H	N
34	0.293246	H	0.1	M	0.580645	H	N
35	0.172012	M	0.1	M	0.387097	H	N
36	0.195335	H	0.1	M	0.548387	H	N
37	0.084548	L	0	L	0	L	N
38	0.089893	L	0.1	M	0.16129	H	N
39	0.071672	L	0.1	M	0.032258	M	N
40	0.126093	M	0.5	M	0.193548	H	N
41	0.157434	M	0	L	0.032258	L	N
42	0.285957	H	0	L	0.225806	H	N
43	0.304665	H	0	L	0.451613	H	N
44	0.126093	M	0.1	M	0.096774	M	N
45	0.137998	M	0	L	0.096774	M	N
46	0.119776	M	0	L	0.032258	L	N
47	0.184159	H	0.1	M	0.193548	H	N
48	0.133625	M	0	L	0	L	N
49	0.249271	H	0.4	M	0.064516	M	N
50	0.233479	H	0.2	M	0.193548	H	N
51	0.165695	M	0.2	M	0	L	N
52	0.130466	M	0	L	0.032258	L	N

As demonstrated in Table 15, 52 rules can be generated from the HS Knot task training dataset. Some of these rules however, are repeated a few times . For instance in Table 15, data vectors 5, 9, and 12, generate the same rule;

> *if* "Time" *is* Low, *and* "Maximum Thread Overstretch" *is* High, *and* "Number of Errors" *is* Medium, *the user is an* Expert.

Instead of repeating the recurring rules, we assigned each rule a weighting proportional to the frequency of its appearance. For example the weighting for the above-mentioned rule, which is repeated three times, is three times the weighting for a rule that is extracted only once.

In the case of having a contradiction (meaning that two rules with the same premise have different conclusions), we included both rules in the rule set. For example in Table 15, data vectors 1 and 4 suggest the same rule;

> *if* "Time" *is* Low, *and* "Maximum Thread Overstretch" *is* Low, *and* "Number of Errors" *is* Low, *the user is an* Expert.

Data vector 37 however, generates a second rule with the same premise, but with a different conclusion;

> *if* "Time" *is* Low, *and* "Maximum Thread Overstretch" *is* Low, *and* "Number of Errors" *is* Low, *the user is a* Novice.

These contradictory rules in the rule set eliminate each other's effects during the fuzzy inference process, and therefore do not affect the final decision.

The resulting rules along with their corresponding weightings formed our set of rules for each fuzzy model.

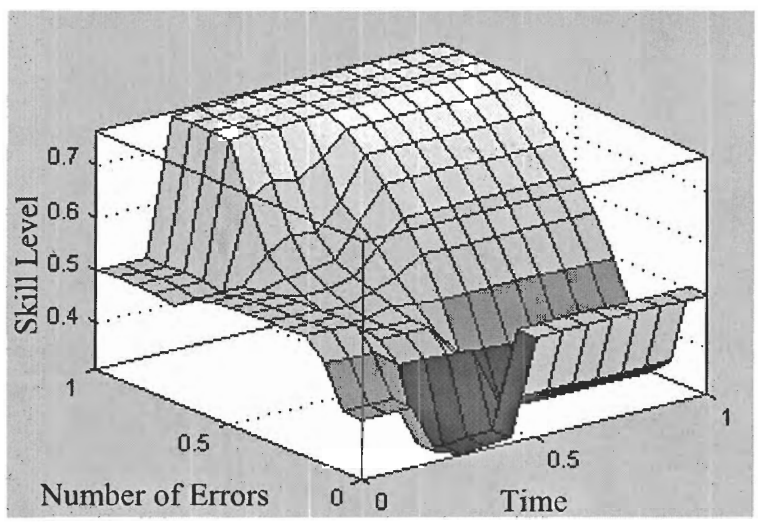
4.4 Selecting the Fuzzy Inference Properties

Properties of the fuzzy systems such as shapes of the membership functions, the t-norm and t-conorm operators, and the aggregation and defuzzification methods need to be specified (refer to section 3 for more information). We selected the most commonly used fuzzy inference properties to design our systems. They include the triangular/trapezoidal membership functions, the maximum and minimum t-norm and t-conorm operators, respectively, the maximum aggregation method, and the centroid defuzzification method. Particulars of these methods are explained in section 3. The effect of other fuzzy inference properties on the classifiers are studied in section 5.

4.5 Results and Analysis

Each of the DAU and DHU data separation methods were employed once to create a fuzzy classifier for each of the Stitch and HS Knot tasks, which provided us with four fuzzy classifiers in total.

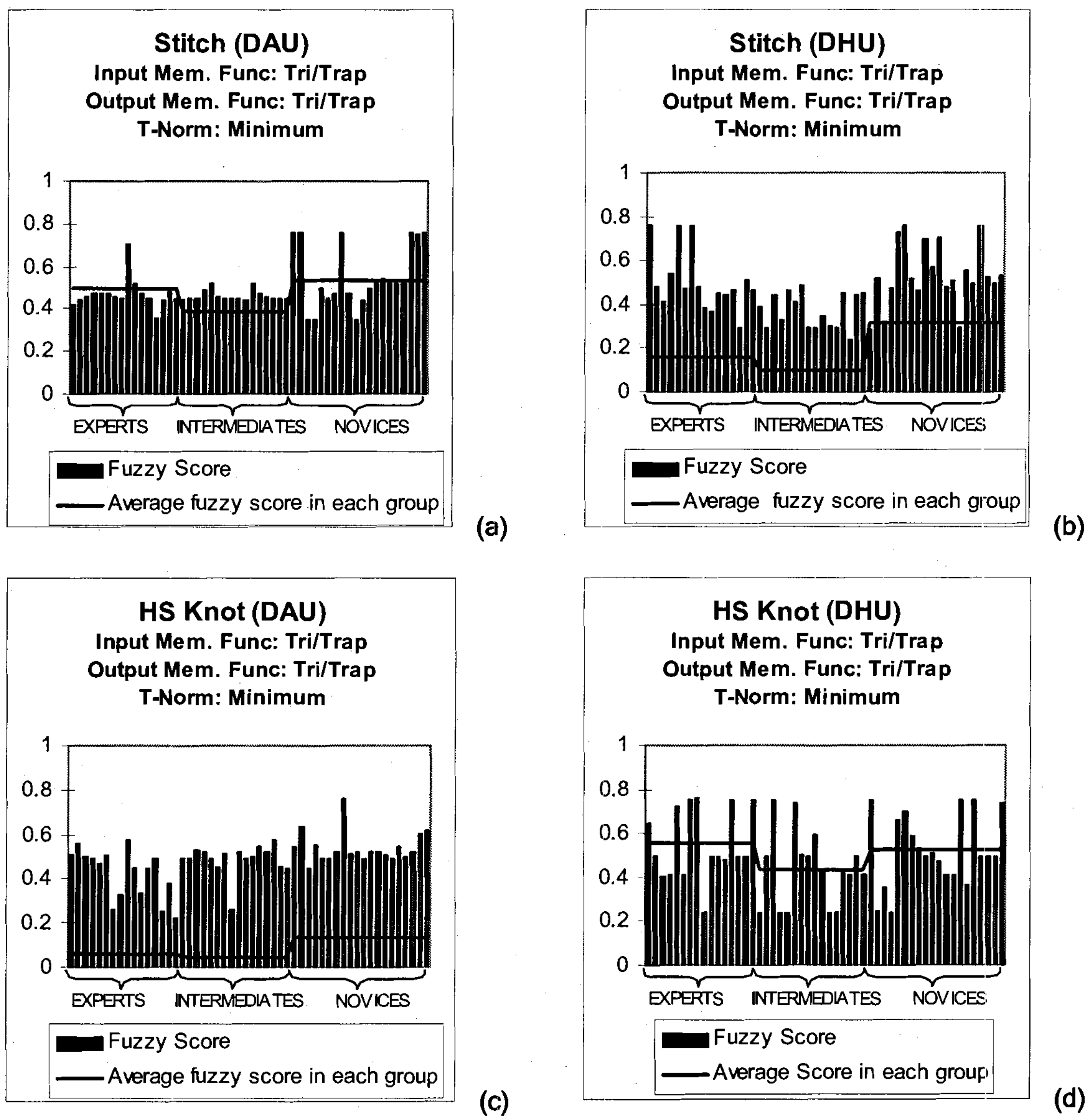
Figure 39: The effect of Time and Number of Errors on the Stitch task's fuzzy score (x-axis: Number of Errors, y-axis: Time, z-axis: fuzzy score)



The constructed models demonstrated a highly non-linear and non-monotonic relationship between the inputs and the output of the systems. Figure 39 shows the effect of "**Time**" and "**Number of Errors**" inputs on the "**Score**" for the Stitch task model as an example. As we can see in the figure, for higher values of "Time" and "Number of Errors" ("Time"> 0.5, "Number of Errors"> 0.3), there is a monotonic relationship between the output (score) and the two inputs. For lower values of "Time" (< 0.5), however, the effect of the "Number of Errors" input on the Score does not follow any specific pattern. Similarly, for "Number of Errors" values below 0.3, the effect of "Time" on the output is irregular and serrated.

We used the testing datasets (each containing 52 data vectors) to test our classifiers. The results are shown in Figure 40.

Figure 40: Fuzzy Scores for the Stitch and HS Knot tasks
(a): DAU Stitch testing dataset, (b): DHU Stitch testing dataset, (c): DAU HS Knot testing dataset, (d): DHU HS Knot testing dataset - (x-axis: data vectors, y-axis: fuzzy score).



Based on the numerical fuzzy scores, the users' levels of expertise can be predicted by the fuzzy classifiers. Each user belongs to one of the three categories (expert, intermediate, or novice) that has the highest truth-value for his/her fuzzy score in the classifier's output fuzzy sets.

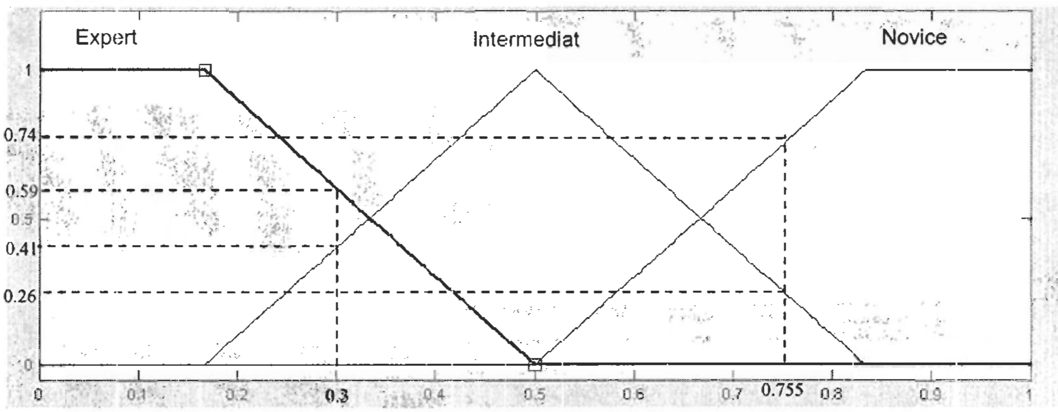
Example: predicting the users' levels of expertise based on their fuzzy scores

The fuzzy scores for 5 users are shown in Table 16 as an example. Each subject has a real level of expertise (depending on their experience in MIS as explained in section 2.2.2), and a predicted level of expertise, which is determined based on the score they achieved in the user study. For example, subject 1 in Table 16, who was in the expert group, has achieved a fuzzy score of 0.3. In the classifiers' output fuzzy sets, 0.3 has the truth-value of 0.59 in the Expert fuzzy set, 0.41 in the Intermediate fuzzy set and 0 in the Novice fuzzy set. Therefore, because user 1's score has the highest truth-value in the Expert fuzzy set, his predicted level of expertise would also be expert. For user 3 however, the predicted level of expertise is novice (his fuzzy score is 0.755, which has the highest truth-value in the Novice fuzzy set), although his real level of expertise is intermediate. Thus, we have a match between the real and predicted levels of expertise for user 1, but not for user 3.

Table 16: Example of users' predicted and real levels of expertise. Table shows the fuzzy score, real levels of expertise, predicted levels of expertise, and whether there is a match between the real and predicted levels of expertise for 5 participants.

User	Real Level of Expertise	Fuzzy Score	Predicted Level of Expertise	Match?
1	Expert	0.3	Expert	Yes
2	Expert	0.708	Novice	No
3	Intermediate	0.755	Novice	No
4	Intermediate	0.494	Intermediate	Yes
5	Novice	0.757	Novice	Yes

Figure 41: Predicting users' levels of expertise based on the Output fuzzy sets. (x-axis: fuzzy score (output), y-axis: output's truth-value)



Performance of the classifiers can be evaluated according to the number of matches between the users' real and predicted levels of expertise. Table 17 shows the percentage of matches between the real and predicted levels of expertise, after all the users have been categorized.

Table 17: Percentage of correct results in all groups of expertise and in total for each classifier

	Stitch Task (DAU)	Stitch Task (DHU)	HS Knot Task (DAU)	HS Knot Task (DHU)
Experts	0%	6.25%	25%	43.75%
Intermediates	100%	56.25%	93.75%	81.25%
Novices	30%	25%	5%	15%
Total	42.31%	28.85%	38.46%	44.23%

5 EXPLORING THE EFFECT OF VARIOUS FUZZY INFERENCE PROPERTIES ON THE PERFORMANCE OF FUZZY CLASSIFIERS

In the design of our fuzzy classifiers we employed the simplest, or the most commonly used Fuzzy Inference System (FIS) properties. For example, we used the simple triangular/trapezoidal shapes for the membership functions, and used the common minimum t-norm operator and the centroid defuzzification method in our systems. In this section we explore the effect of some other FIS properties on our fuzzy classifiers.

It is shown that amongst the different FIS properties, the shapes of the *membership functions*, the t-*norm operators* (the AND method), and the types of the *defuzzifiers* are the most significant factors in the fuzzy inference process (Dadone, 2001). Therefore, we tested and compared the performance of our classifiers with different combinations of these three factors.

To do so, we first examined the effect of various membership functions and t-norm operators on our classifiers to find the system with the optimal combination of these two factors. We then applied different defuzzification methods to that system and identified the most functional grouping of FIS properties for our fuzzy classifiers.

To be able to compare performance of the different classifiers, and in order to find the one with the optimal combination of FIS properties, we took two different approaches to compare the functionalities of the systems after they were tested with the testing dataset.

One approach is based on the number of correct classifications for each system, or the number of matches between users' real and predicted levels of expertise. Since the objective of the test is to classify users correctly, this method sounds like a reasonable approach.

However, it should also be considered that there are different levels of *fallacy* among the systems' incorrect answers. For instance a classifier that predicts a novice user to be an expert is making a major error, which may cause an inexperienced surgeon to advance to the operating rooms. In comparison, another system that classifies the same user also incorrectly, but as an intermediate, is less likely to cause such a fatal error and therefore is more reliable. Considering only the number of correct classifications for the systems does not take this problem into account, and therefore the other approach that we used to compare the functionalities of our classifiers is based on the total error in the predicted levels of expertise for the user in the testing dataset, and is explained in the following section.

5.1 The "Root Mean-Squared Error"

The Root Mean-Squared Error

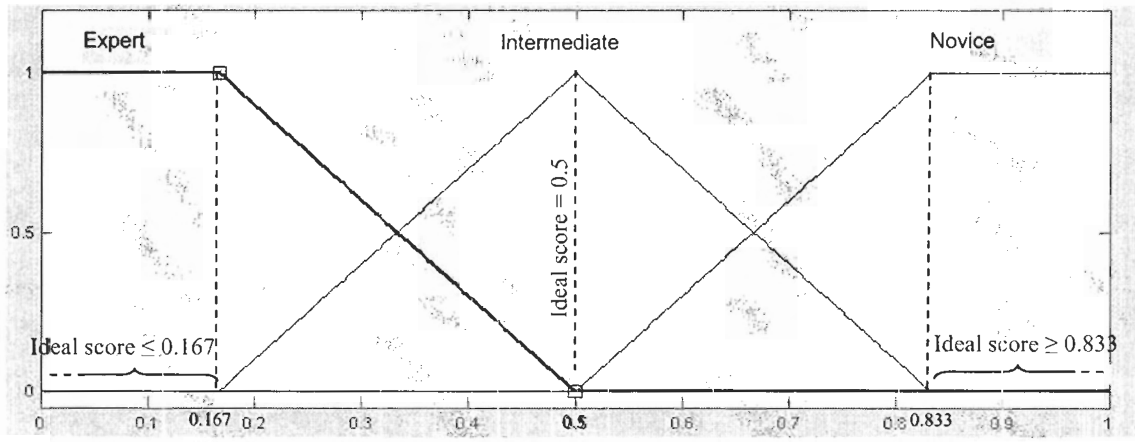
Root Mean-squared error (RMS error), determined by calculating the deviations of points from their true position, summing up the squares of the measurements, then taking the square root of the sum, and then dividing the result by the number of points, is the most commonly used measure of success of numeric prediction (GRB Research, GRP Tool Shed, ¶ 3.2.4, 2005). For example if a_1, a_2, …, a_n, are the system's predicted values, with corresponding true values of c_1, c_2, …, c_n, the RMS error for a_1 to a_n is calculated as:

$$\sqrt{\frac{(a_1 - c_1)^2 + (a_2 - c_2)^2 + \ldots + (a_n - c_n)^2}{n}}$$

Equation 21

To determine the RMS error for each classifier's fuzzy scores, the deviations of the predicted scores from their true values need to be calculated. The true value of score for each user is the value that best represents the group of expertise that the user belongs to. Considering the output membership functions for our classifiers (Figure 42), we can see that the true (or ideal) score for each group of expertise is the score that has the highest truth-value in the corresponding membership functions. For example the ideal score for an expert user is the score that has the highest truth-value (i.e. 1) in the Expert fuzzy set in the output membership functions.

Figure 42: Ideal score for each group of expertise in the output fuzzy sets. (x-axis: output (score) value, y-axis: output's truth-value)



As shown in Figure 42, all the values below 0.167 and above 0.833 have the maximum truth-value in the Expert and Novice fuzzy sets, respectively. Therefore the ideal score for the expert group could be any value below 0.167, and above 0.833 for the novice group. Because the Intermediate fuzzy set is triangular in shape, there is only one score with the maximum truth-value for the intermediate group. Therefore the ideal score for an intermediate user will be 0.5. These values are shown in Table 18.

Table 18: Ideal score for users in each group of expertise

User	Ideal (true) score
Expert	≤ 0.167
Intermediate	0.5
Novice	≥ 0.833

The deviations of the fuzzy scores from their true values can now be calculated through the following equations:

$$
\textit{Experts' Scores Deviation} = \begin{cases} 0 & \text{Fuzzy Score} \leq 0.167 \\ \text{Fuzzy Score} - 0.167 & \text{Fuzzy Score} > 0.167 \end{cases}
$$

Equation 22

$$
\textit{Intermediates' Scores Deviation} = \text{Fuzzy Score} - 0.5
$$

Equation 23

$$
\textit{Novices' Scores Deviation} = \begin{cases} 0 & \text{Fuzzy Score} \geq 0.833 \\ \text{Fuzzy Score} - 0.833 & \text{Fuzzy Score} < 0.833 \end{cases}
$$

Equation 24

To determine the RMS Error for each classifier after being tested with the 52 data vectors in the testing dataset, user's fuzzy and ideal scores were substituted in:

$$
RMS \quad Error = \sqrt{\frac{\sum_{i=1}^{16} \Delta E_i^2 + \sum_{j=1}^{16} \Delta I_j^2 + \sum_{k=1}^{20} \Delta N_k^2}{52}}
$$

Equation 25

In which ΔE_i, ΔI_j, and ΔN_k are the score deviations for the i^{th} expert, j^{th} intermediate, and k^{th} novice data vectors, and are calculated by substituting the testing data set fuzzy scores in Equation 22 to Equation 24.

Relative Root Mean-Squared Error

The *relative RMS error* is defined to be the ratio of the RMS Error to the maximum RMS Error value possible for each classifier, and indicates the relative reliability of the classifiers. The maximum RMS Error is calculated as follows:

$$\max\left(RMS \quad Error\right) = \max\left(\sqrt{\frac{\sum_{i=1}^{16} \Delta E_i^2 + \sum_{j=1}^{16} \Delta I_j^2 + \sum_{k=1}^{20} \Delta N_k^2}{52}}\right)$$

$$= \sqrt{\frac{\sum_{i=1}^{16} \max\left(\Delta E_i^2\right) + \sum_{j=1}^{16} \max\left(\Delta I_j^2\right) + \sum_{k=1}^{20} \max\left(\Delta N_k^2\right)}{52}}$$

Equation 26

Since the fuzzy scores belong to [0,1], from Equations 2 to 4 we have:

$$\max\left(\Delta E_i^2\right) = (1-0.167)^2 = (0.833)^2 = 0.694$$

$$\max\left(\Delta I_j^2\right) = \left\{\begin{matrix} (1-0.5)^2 \\ or \\ (0-0.5)^2 \end{matrix}\right\} = (\pm 0.5)^2 = 0.25$$

$$\max\left(\Delta N_k^2\right) = (0-0.833)^2 = (0.833)^2 = 0.694$$

By substituting these values in Equation 26, we will have:

$$\max\left(RMS \quad Error\right) = \sqrt{\frac{\sum_{i=1}^{16} 0.694 + \sum_{j=1}^{16} 0.25 + \sum_{k=1}^{20} 0.694}{52}}$$

$$= \sqrt{\frac{16 \times 0.694 + 16 \times 0.25 + 20 \times 0.694}{52}} = 0.746$$

Equation 27

The relative RMS Error for each classifier would be its RMS Error value divided by the result of Equation 27.

5.2 Various Fuzzy Inference Properties

5.2.1 Membership Functions

Various functions such as Triangular/Trapezoidal, Gaussian, Sigmoid, Bell Shaped, and Polynomial-based curves are used to represent the membership functions in fuzzy inference systems. We studied the effect of Gaussian and Triangular/Trapezoidal shapes on the input and outputs membership functions in our fuzzy classifiers.

5.2.2 T-norm Operators

It is shown that the optimal behaviour of the fuzzy controllers is achieved by the differentiable t-norm operators such as the Product operator (Dadone, 2001). The minimum t-norm operator is also commonly used in fuzzy systems (Szczepaniak, Lisboa, & Kacprzyk, 2000). Thus, we tested each of our classifiers, employing the minimum and the Product operators.

To find the best functions for the input and output fuzzy sets, and the most effective t-norm operator in our systems, for each of our four classifiers, we created all the possible combinations of the Triangular/Trapezoidal and Gaussian membership functions, and the Product and minimum t-norm operators. The resulting 8 combinations for each classifier (two functions for each of the input and output membership functions and two t-norm operators lead to 2^3 or 8 combinations of FIS properties) were then tested with the testing dataset. Table 19 to Table 22 show the percentage of correct results and the average scores in each group of expertise for each of the classifiers. The results are also represented in Figure 43 to Figure 46 (vertical bars correspond to the individual scores and horizontal lines represent the average score in each group of expertise).

Table 19: Results of various combinations of membership functions and t-norm operators for the Stitch Task (DAU) fuzzy classifier

Stitch Task (DAU) FIS Properties			Percentage of Correct Results				Average Score in each group		
Input Membership Function	Output Membership Function	T-norm Method	Experts	Intermediates	Novices	Total	Experts	Intermediates	Novices
Tri/Trap	Tri/Trap	Min	0	100	30	42.31	0.49	0.39	0.53
Tri/Trap	Tri/Trap	Product	0	100	30	42.31	0.47	0.46	0.55
Tri/Trap	Gaussian	Min	0	100	10	34.62	0.47	0.48	0.53
Tri/Trap	Gaussian	Product	0	100	10	34.62	0.48	0.48	0.53
Gaussian	Tri/Trap	Min	0	100	30	42.31	0.48	0.47	0.57
Gaussian	Tri/Trap	Product	0	100	30	42.31	0.48	0.47	0.57
Gaussian	Gaussian	Min	0	100	15	36.54	0.48	0.48	0.55
Gaussian	Gaussian	Product	0	100	5	32.70	0.48	0.48	0.54

Table 20: Results of various combinations of membership functions and t-norm operators for the Stitch Task (DHU) fuzzy classifier

Stitch Task (DHU) FIS Properties			Percentage of Correct Result				Average Score in each group		
Input Membership Function	Output Membership Function	T-norm Method	Experts	Intermediates	Novices	Total	Experts	Intermediates	Novices
Tri/Trap	Tri/Trap	Min	6.25	56.25	25	28.85	0.50	0.37	0.54
Tri/Trap	Tri/Trap	Product	6.25	56.25	25	28.85	0.50	0.37	0.54
Tri/Trap	Gaussian	Min	6.25	81.25	5	28.85	0.49	0.40	0.52
Tri/Trap	Gaussian	Product	0	100	15	36.54	0.57	0.53	0.61
Gaussian	Tri/Trap	Min	0	93.75	20	36.54	0.50	0.42	0.54
Gaussian	Tri/Trap	Product	0	93.75	20	36.54	0.50	0.42	0.55
Gaussian	Gaussian	Min	0	100	0	30.77	0.49	0.43	0.53
Gaussian	Gaussian	Product	0	100	0	30.77	0.50	0.44	0.53

Table 21: Results of various combinations of membership functions and t-norm operators for the HS -Knot Task (DAU) fuzzy classifier

HS Knot Task (DAU) FIS Properties			Percentage of Correct Result				Average Score in each group		
Input Membership Function	Output Membership Function	T-norm Method	Experts	Intermediates	Novices	Total	Experts	Intermediates	Novices
Tri/Trap	Tri/Trap	Min	25.00	93.75	5.00	38.46	0.42	0.49	0.54
Tri/Trap	Tri/Trap	Product	31.25	93.75	5.00	40.38	0.42	0.49	0.55
Tri/Trap	Gaussian	Min	18.75	93.75	5.00	36.54	0.44	0.49	0.53
Tri/Trap	Gaussian	Product	18.75	93.75	5.00	36.54	0.44	0.49	0.53
Gaussian	Tri/Trap	Min	0.00	100.00	10.00	34.62	0.47	0.50	0.56
Gaussian	Tri/Trap	Product	0.00	100.00	15.00	36.54	0.46	0.50	0.56
Gaussian	Gaussian	Min	0.00	100.00	5.00	32.69	0.47	0.50	0.54
Gaussian	Gaussian	Product	0.00	100.00	5.00	32.69	0.46	0.50	0.53

Table 22: Results of various combinations of membership functions and t-norm operators for the HS -Knot Task (DHU) fuzzy classifier

HS Knot Task (DHU) FIS Properties			Percentage of Correct Result				Average Score in each group		
Input Membership Function	Output Membership Function	T-norm Method	Experts	Intermediates	Novices	Total	Experts	Intermediates	Novices
Tri/Trap	Tri/Trap	Min	43.75	81.25	15.00	44.23	0.39	0.47	0.50
Tri/Trap	Tri/Trap	Product	43.75	81.25	20.00	46.15	0.39	0.47	0.50
Tri/Trap	Gaussian	Min	6.25	93.75	5.00	32.69	0.44	0.42	0.51
Tri/Trap	Gaussian	Product	12.50	100.00	0.00	34.62	0.41	0.46	0.50
Gaussian	Tri/Trap	Min	37.50	93.75	15.00	46.15	0.42	0.51	0.54
Gaussian	Tri/Trap	Product	37.50	93.75	15.00	46.15	0.40	0.49	0.53
Gaussian	Gaussian	Min	6.25	100.00	10.00	36.54	0.42	0.47	0.52
Gaussian	Gaussian	Product	6.25	100.00	10.00	36.54	0.43	0.49	0.53

Figure 43: **Results of various combinations of membership functions and t-norm operators for the Stitch Task (DAU) fuzzy classifier.**
The vertical bars represent the individual scores and horizontal lines show the average score in each group of expertise.

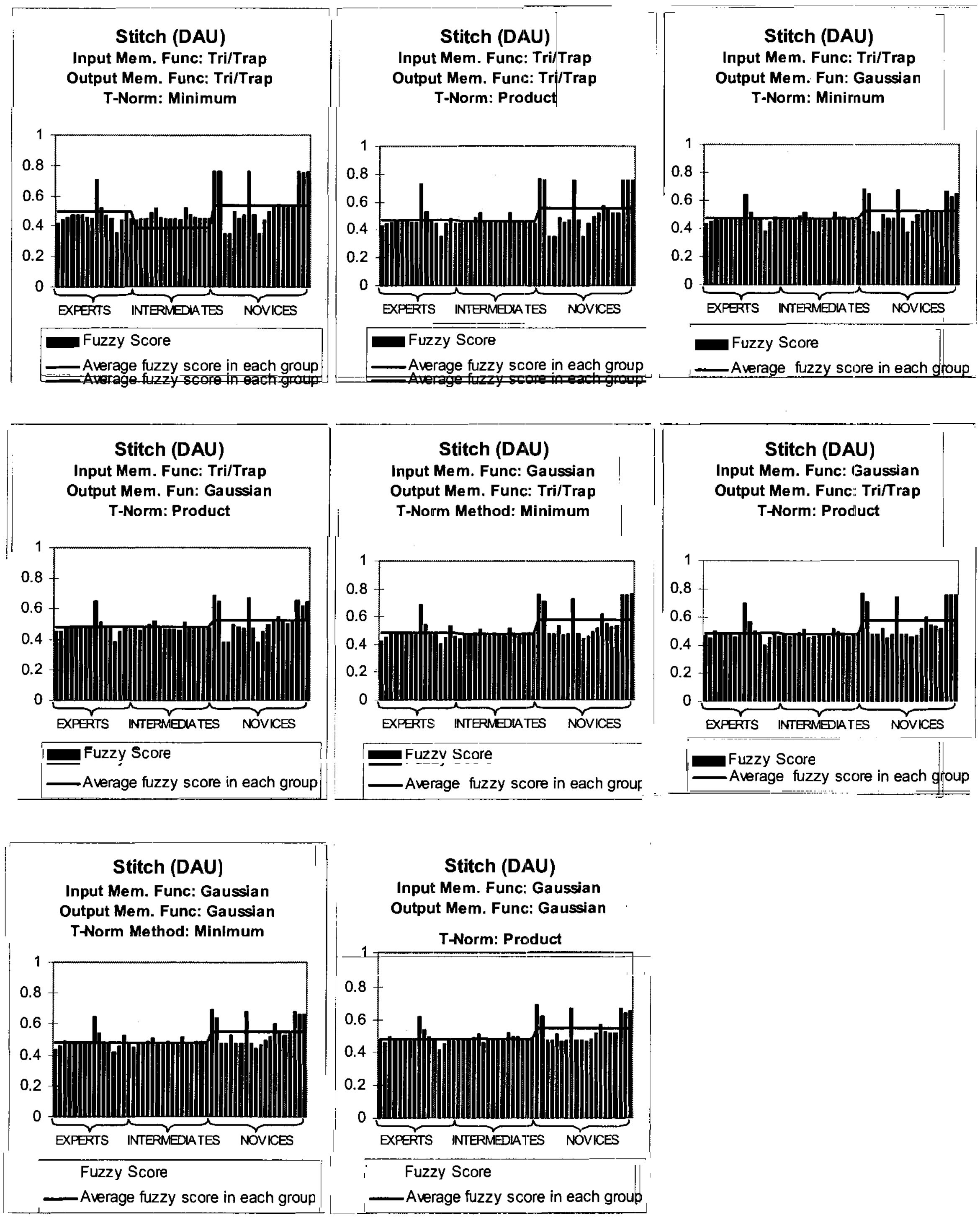


Figure 44: Results of various combinations of membership functions and t-norm operators for the Stitch Task (DHU) fuzzy classifier.

The vertical bars represent the individual scores and horizontal lines show the average score in each group of expertise.

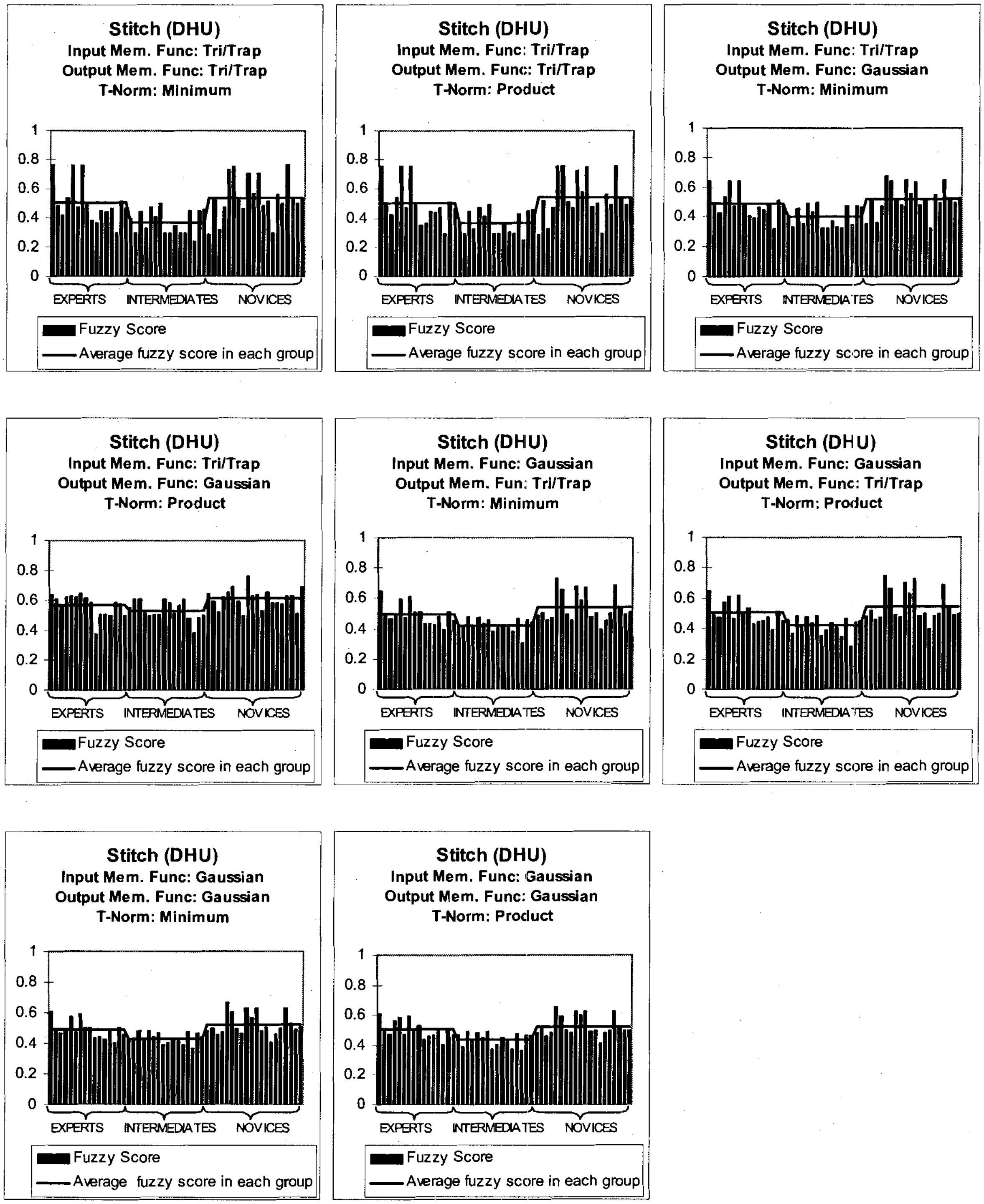


Figure 45: Results of various combinations of membership functions and t-norm operators for the HS Knot Task (DAU) fuzzy classifier.

The vertical bars represent the individual scores and horizontal lines show the average score in each group of expertise.

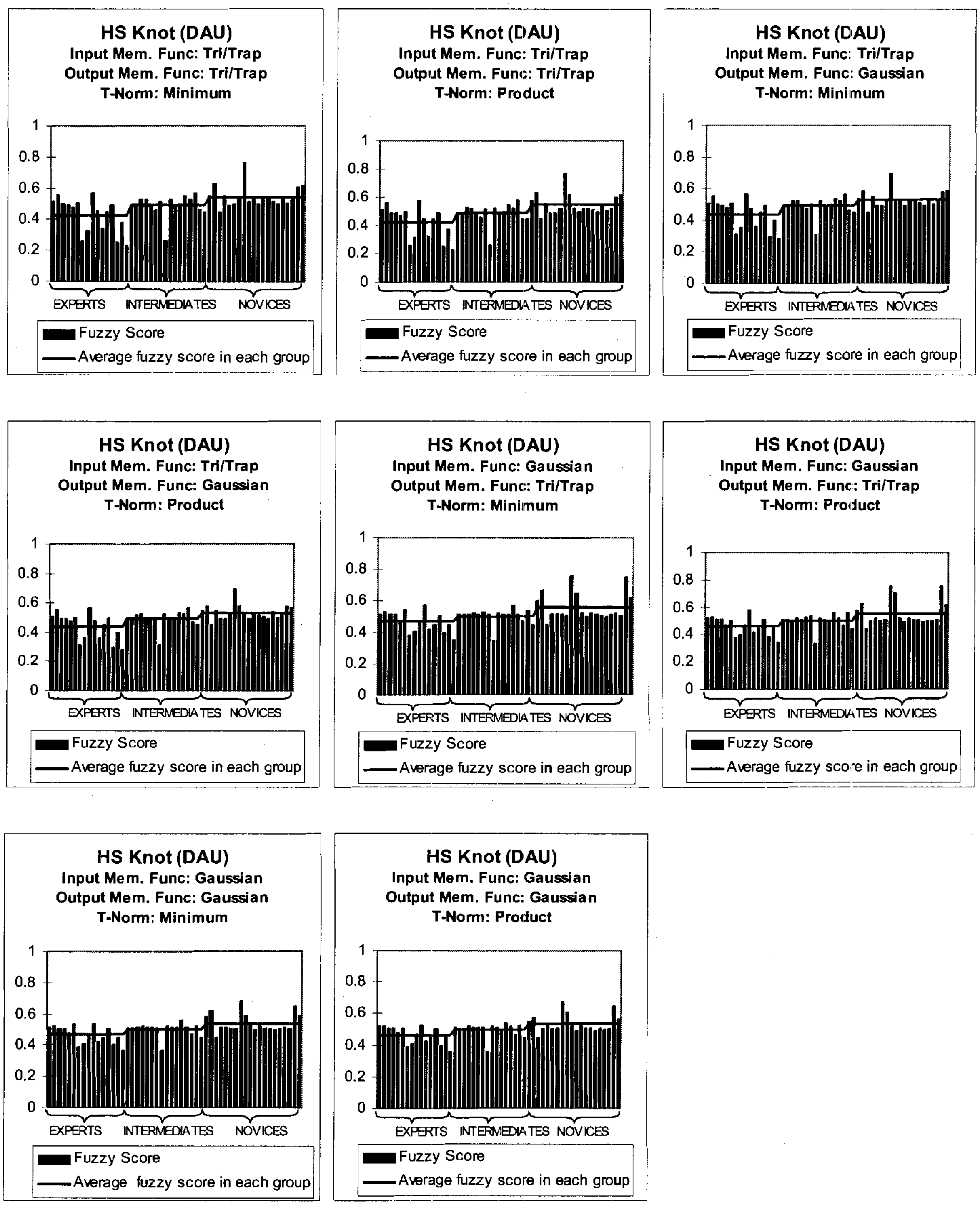


Figure 46: Results of various combinations of membership functions and t-norm operators for the HS Knot Task (DHU) fuzzy classifier.

The vertical bars represent the individual scores and horizontal lines show the average score in each group of expertise.

To compare the performance of these fuzzy classifiers, we tested them with the testing dataset, and calculated the RMS Error value for each system. These values are shown in Table 23 to Table 26. Highlighted rows correspond to the classifier with the lowest RMS error values.

Table 23: Percentage of matches and RMS Error for classifiers with various combinations of FIS properties – Stitch Task (DAU)

Stitch Task (DAU) FIS Properties			Percentage of Matches	RMS Error
Input MF	Output MF	T-norm Method		
Tri/Trap	Tri/Trap	Min	42.31	0.261
Tri/Trap	Tri/Trap	Product	42.31	0.261
Tri/Trap	Gaussian	Min	34.62	0.263
Tri/Trap	Gaussian	Product	34.62	0.264
Gaussian	Tri/Trap	Min	42.31	0.25
Gaussian	Tri/Trap	Product	42.31	0.253
Gaussian	Gaussian	Min	36.54	0.254
Gaussian	Gaussian	Product	32.69	0.256

Table 24: Percentage of matches and RMS Error for classifiers with various combinations of FIS properties – Stitch Task (DHU)

Stitch Task (DHU) FIS Properties			Percentage of Matches	RMS Error
Input MF	Output MF	T-norm Method		
Tri/Trap	Tri/Trap	Min	28.85	0.296
Tri/Trap	Tri/Trap	Product	28.85	0.295
Tri/Trap	Gaussian	Min	28.85	0.281
Tri/Trap	Gaussian	Product	36.54	0.269
Gaussian	Tri/Trap	Min	36.54	0.27
Gaussian	Tri/Trap	Product	36.54	0.273
Gaussian	Gaussian	Min	30.77	0.27
Gaussian	Gaussian	Product	30.77	0.273

Table 25: Percentage of matches and RMS Error for classifiers with various combinations of FIS properties – S Knot Task (DAU)

HS Knot (DHU) FIS Properties			Percentage of Matches	RMS Error
Input MF	Output MF	T-norm Method		
Tri/Trap	Tri/Trap	Min	38.46	0.243
Tri/Trap	Tri/Trap	Product	40.38	0.24
Tri/Trap	Gaussian	Min	36.54	0.247
Tri/Trap	Gaussian	Product	36.54	0.246
Gaussian	Tri/Trap	Min	34.62	0.247
Gaussian	Tri/Trap	Product	36.54	0.247
Gaussian	Gaussian	Min	32.69	0.251
Gaussian	Gaussian	Product	32.69	0.252

Table 26: Percentage of matches and RMS Error for classifiers with various combinations of FIS properties – HS Knot Task (DHU)

HS Knot (DHU) FIS Properties			Percentage of Matches	RMS Error
Input MF	Output MF	T-norm Method		
Tri/Trap	Tri/Trap	Min	44.23	0.275
Tri/Trap	Tri/Trap	Product	46.15	0.275
Tri/Trap	Gaussian	Min	32.69	0.27
Tri/Trap	Gaussian	Product	34.62	0.263
Gaussian	Tri/Trap	Min	46.15	0.251
Gaussian	Tri/Trap	Product	46.15	0.255
Gaussian	Gaussian	Min	36.54	0.259
Gaussian	Gaussian	Product	36.54	0.254

As we can see in Table 23 to Table 26, for all four models, fuzzy classifiers with the lowest amount of RMS Error also have the highest percentage of matches between users' real and predicted levels of expertise (highlighted rows). Thus, it can be concluded that the best combination of the input and output membership functions and t-norm operators for our models are:

- Stitch task (DAU): Gaussian input MF, Triangular/Trapezoidal output MF, and Minimum T-Norm operator

- Stitch task (DHU): Triangular/Trapezoidal input MF, Gaussian output MF, and Product T-Norm operator

- HS Knot task (DAU): Triangular/Trapezoidal input MF, Triangular/Trapezoidal output MF, and Product T-Norm operator

- HS Knot task (DHU): Gaussian input MF, Triangular/Trapezoidal output MF, and Minimum T-Norm operator

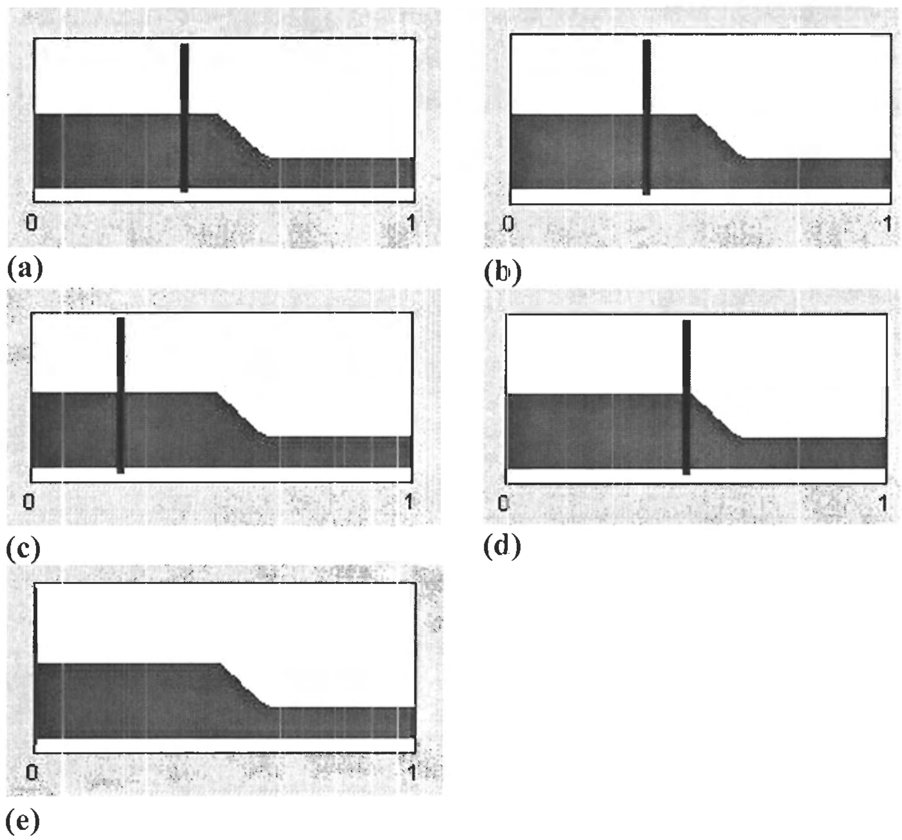
5.2.3 Defuzzification Methods

Our classifiers were initially designed using the centroid defuzzification method. In this section, we explore the effect of some of the other defuzzification methods on our fuzzy systems.

There are five different defuzzification methods supported in MATLAB; centroid (which returns the center of area under the curve), bisector (which returns bisector of area under the curve), middle of maximum or MOM (the average of the maximum value of the output set), largest of maximum or LOM (largest of the maximum values of the output set), and smallest of maximum or SOM. Figure 47 shows examples of these methods. The vertical line in each figure shows the location of the defuzzified output value over the output curve.

We implemented each of these five methods in our system with the optimal combination of membership functions and t-norm operators (determined in the previous section) to find the most appropriate defuzzifier for our systems.

Figure 47: Different Defuzzification methods applied to an example fuzzy output curve. The vertical line shows location of the numerical fuzzy output over the output curve. Methods applied: (a): Centroid, (b): Bisector, (c): Middle of maximums, (d): Largest of maximums, (e): Smallest of maximums



The resulting classifiers were tested against the testing dataset. Table 27 to Table 30, and Figure 48 to Figure 51 represent the results.

Table 27: Results of various defuzzification methods for the Stitch Task (DAU) fuzzy classifier

Stitch Task (DAU) -	Percentage of Correct Results				Average Score in each group		
Defuzzification Method	Experts	Intermediates	Novices	Total	Experts	Intermediates	Novices
Centroid	0	100	30	42.31	0.48	0.47	0.58
Bisector	0	100	30	42.31	0.48	0.47	0.57
MOM	31.25	100	45	57.69	0.47	0.49	0.57
LOM	0	0	75	28.85	0.72	0.79	0.80
SOM	81.25	0	0	25	0.21	0.20	0.30

Table 28: Results of various defuzzification methods for the Stitch Task (DHU) fuzzy classifier

Stitch Task (DHU) -	Percentage of Correct Results				Average Fuzzy Score in each group		
Defuzzification Method	Experts	Intermediates	Novices	Total	Experts	Intermediates	Novices
Centroid	0	100	15	36.54	0.57	0.53	0.61
Bisector	0	87.5	5	28.85	0.53	0.55	0.58
MOM	6.25	56.25	65	44.23	0.54	0.56	0.64
LOM	0	6.25	100	40.38	0.89	0.90	0.96
SOM	68.75	37.5	0	32.69	0.20	0.23	0.32

Table 29: Results of various defuzzification methods for the HS Knot Task (DAU) fuzzy classifier

HS Knot Task (DAU) -	Percentage of Correct Results				Average Fuzzy Score in each group		
Defuzzification Method	Experts	Intermediates	Novices	Total	Experts	Intermediates	Novices
Centroid	31.25	93.75	5.00	40.38	0.42	0.51	0.54
Bisector	31.25	93.75	5.00	40.38	0.40	0.52	0.55
MOM	43.75	68.75	65.00	59.62	0.36	0.49	0.57
LOM	0.00	6.25	100.00	40.38	0.61	0.73	0.82
SOM	87.50	25.00	0.00	34.62	0.12	0.25	0.32

Table 30: **Results of various defuzzification methods for the HS Knot Task (DHU) fuzzy classifier**

HS Knot Task (DHU) -	Percentage of Correct Results				Average Fuzzy Score in each group		
Defuzzification Method	Experts	Intermediates	Novices	Total	Experts	Intermediates	Novices
Centroid	37.50	93.75	15.00	46.15	0.42	0.49	0.55
Bisector	50.00	62.50	20.00	42.31	0.42	0.49	0.56
MOM	62.50	12.50	45.00	40.38	0.40	0.454	0.67
LOM	0.00	43.75	80.00	44.23	0.67	0.83	0.93
SOM	87.50	43.75	0.00	40.38	0.13	0.24	0.40

Figure 48: **Results of various defuzzification methods applied to the Stitch Task (DAU) model with the optimal combination of membership functions and t-norm operator.**

The vertical bars represent the individual scores and horizontal lines show the average score in each group of expertise.

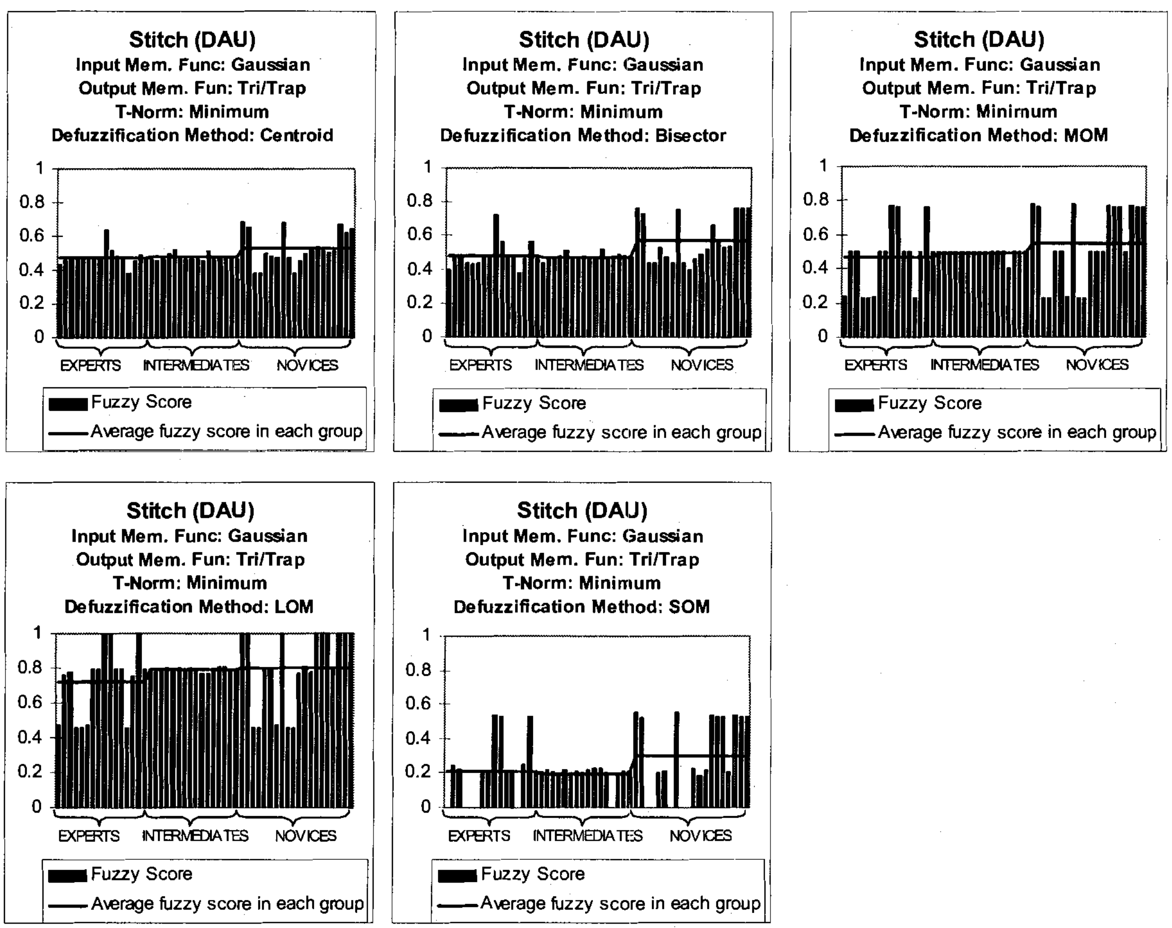
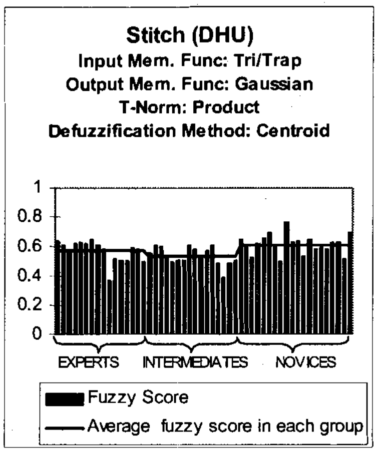
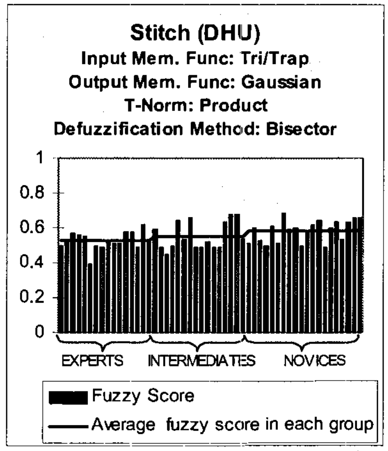
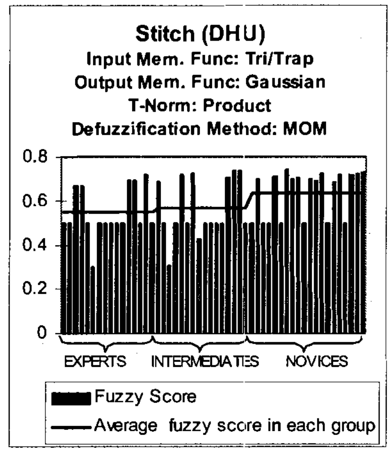


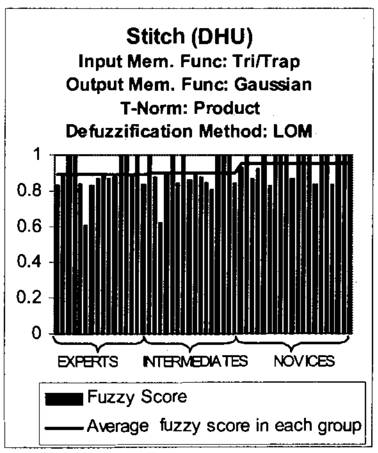
Figure 49: **Results of various defuzzification methods applied to the Stitch Task (DHU) model with the optimal combination of membership functions and t-norm operator.**

The vertical bars represent the individual scores and horizontal lines show the average score in each group of expertise.









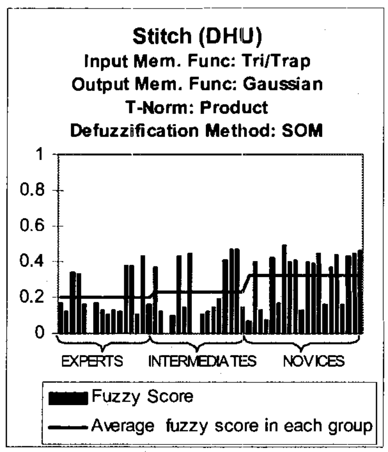


Figure 50: Results of various defuzzification methods applied to the HS Knot Task (DAU) model with the optimal combination of membership functions and t-norm operator.

The vertical bars represent the individual scores and horizontal lines show the average score in each group of expertise.

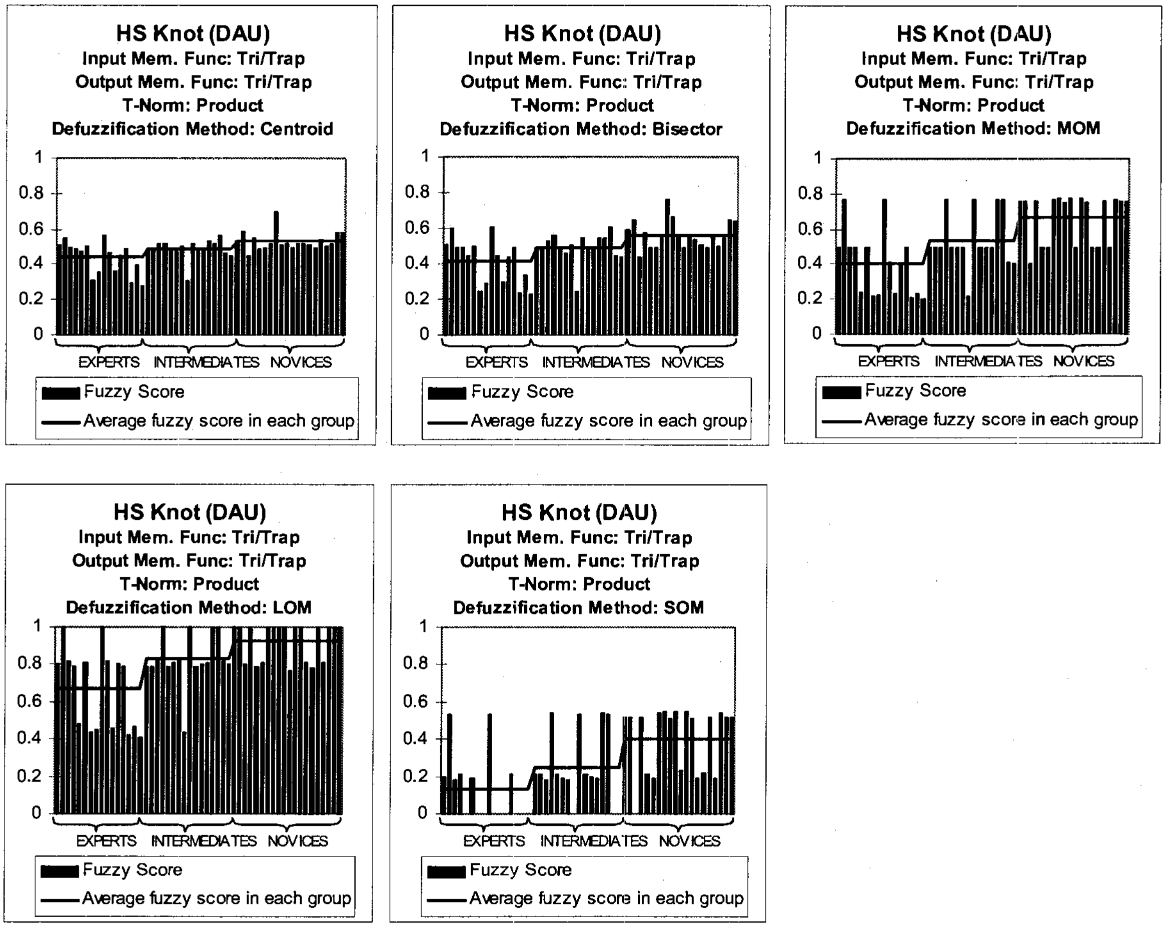
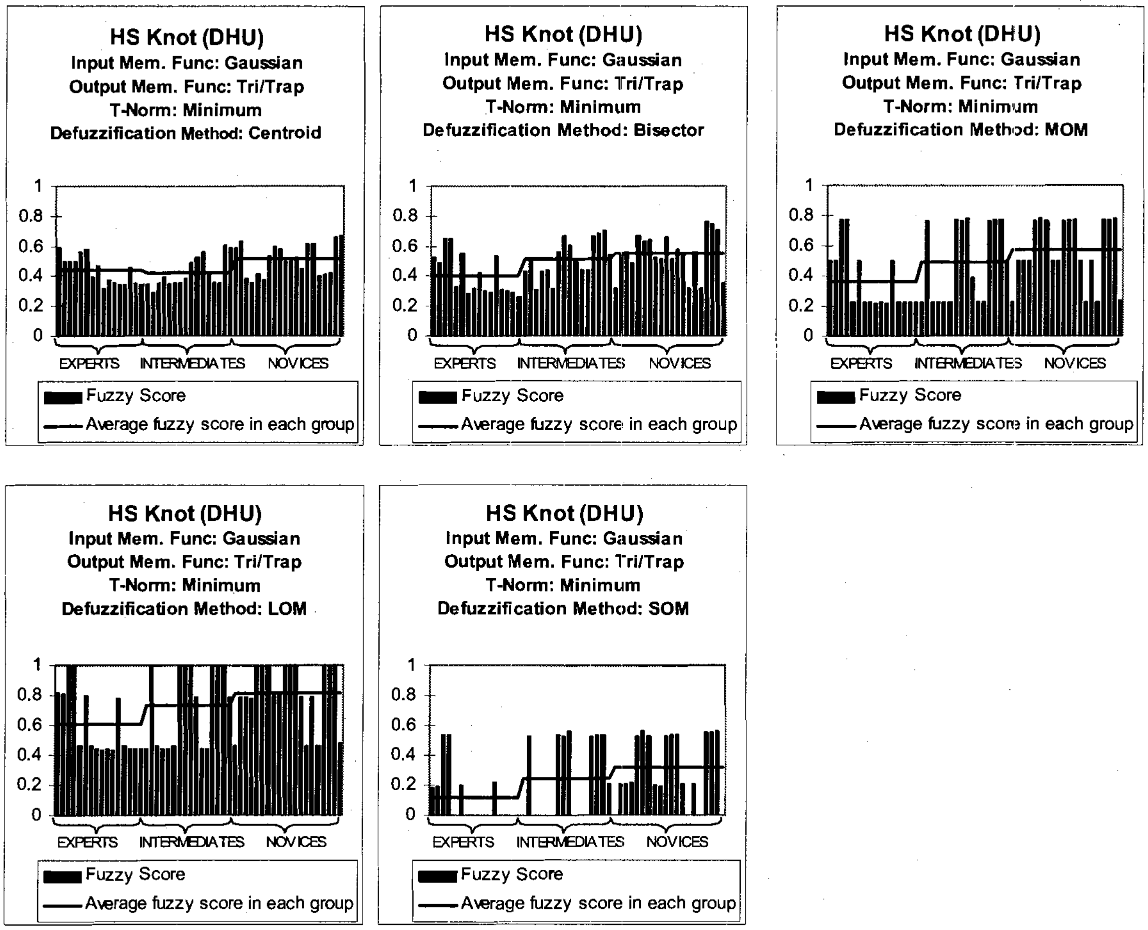


Figure 51: **Results of various defuzzification methods applied to the HS Knot Task (DHU) model with the optimal combination of membership functions and t-norm operator.**

The vertical bars represent the individual scores and horizontal lines show the average score in each group of expertise.



To compare the efficiency of these defuzzifiers, we calculated the RMS Error value for each model. These values are shown in Table 31 (a), (b), (c), and (d) for each of the four models with the five defuzzification methods applied.

Table 31: RMS Errors for various defuzzification methods applied to classifiers with optimal combination of membership functions and t-norm operator (minimum value(s) highlighted)

Stitch Task (DAU)		
Defuzzification Method	**Percentage of Matches**	**RMS Error**
Centroid	42.31	0.250
Bisector	42.31	0.254
Middle of Max	57.69	0.293
Largest of Max	28.85	0.387
Smallest of Max	25	0.403

(a)

Stitch Task (DHU)		
Defuzzification Method	**Percentage of Matches**	**RMS Error**
Centroid	36.54	0.2814
Bisector	28.85	0.260
Middle of Max	44.23	0.268
Largest of Max	40.38	0.464
Smallest of Max	32.69	0.378

(b)

HS Knot Task (DAU)		
Defuzzification Method	**Percentage of Matches**	**RMS Error**
Centroid	40.38	0.24
Bisector	40.38	0.235
Middle of Max	59.62	0.227
Largest of Max	40.38	0.362
Smallest of Max	34.62	0.342

(c)

HS Knot Task (DHU)		
Defuzzification Method	**Percentage of Matches**	**RMS Error**
Centroid	46.15	0.251
Bisector	42.31	0.255
Middle of Max	40.38	0.293
Largest of Max	44.23	0.348
Smallest of Max	40.38	0.406

(d)

As shown in Table 31 (a), (b), (c), and (d), the Middle of Maximum (MOM) defuzzification method has the highest percentage of matches in all models except for the HS Knot task (DHU), in which the highest percentage of matches is achieved by the Centroid method. For the HS Knot (DAU) and HS Knot (DHU) models, the systems with the highest percentage of matches are also the ones with the lowest RMS Error, but this is not the case for the Stitch task models. However, because the differences between the RMS Error values is insignificant, we consider the models with the highest percentage of matches (highlighted rows in Table 31) to have the optimal combination of fuzzy inference properties. Thus, it could be concluded that the most appropriate defuzzification method for the Stitch task (DAU), Stitch task (DHU), and the HS Knot task (DAU) is the Middle of Maximum defuzzification method and Centroid method for the HS Knot task (DHU) model.

6 ANALYSIS OF RESULTS

6.1 Results summary

As explained in the previous section, the optimal combination of fuzzy inference properties for each of the four fuzzy classifiers was identified. Table 32 summarizes these combinations and the percentage of matches and RMS Error values for each of the four models. The Relative RMS Error in Table 32 shows the ratio of the RMS Error for each model to the maximum possible RMS Error for our classifiers, which was calculated in Equation 27 in section 5.1.

Table 32: Summary of results for the fuzzy models with the optimal combination of fuzzy inference properties

	Best fuzzy inference properties combination				Percentage of Matches	RMS Error	Relative RMS Error
	Input MF	Output MF	T-Norm Operator	Defuzzification method			
Stitch (DAU)	Gaussian	Tri/Trap	Minimum	MOM	57.69	0.293	0.393
Stitch (DHU)	Tri/Trap	Gaussian	Product	MOM	44.23	0.268	0.359
HS Knot (DAU)	Tri/Trap	Tri/Trap	Product	MOM	59.62	0.227	0.304
HS Knot (DHU)	Gaussian	Tri/Trap	Minimum	Centroid	46.15	0.251	0.336

Comparing the percentage of matches between the DAU and DHU models for each task shows that models with the DAU data separation have performed better for both the Stitch and the HS Knot tasks. This could be due to the fact that our sample population has not been large enough to represent the overall population, and dividing the users in two groups in the DHU method has resulted in even smaller training and testing datasets that do not share common characteristics. Different results might be obtained in a system with a larger sample size.

6.2 Comparing Fuzzy Classifiers with Conventional Mathematical Methods

As mentioned in section 2.3.1, we applied regression analysis as a linear statistical method to our training datasets. Equation 3 to Equation 6 in section 2.3.1 represent the resulting regression equations for the four models (Stitch (DAU), Stitch (DHU), HS Knot (DAU), and HS Knot (DHU)), which estimate the value of "skill level" based on the Stitch and HS Knot performance metrics. We substitute the performance metrics from the testing datasets in the corresponding regression equations to predict the related surgical skill levels. These values are represented in Table 33 and Table 34.

It was mentioned in section 2.3.1 that we assigned a numerical value to each surgical skill level by dividing the interval from 0 to 1 into three equal regions, and assigning the center of each region to one surgical skill level (Figure 52). After predicting numerical values for the "skill level", we took a reverse action to classify those values into the three categories of Expert, Intermediate, and Novice, as follows:

$$\text{Skill Level Category} = \begin{cases} \text{Expert} & 0 \le \text{Skill Level Value} < 0.333 \\ \text{Intermediate} & 0.333 \le \text{Skill Level Value} < 0.667 \\ \text{Novice} & 0.667 \le \text{Skill Level Value} \le 1 \end{cases}$$

The predicted categories of surgical skill levels are shown in Table 33 and Table 34.

Figure 52: Assigning numerical values to surgical skill levels

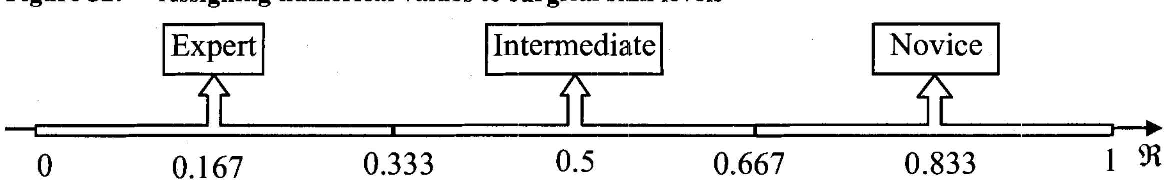


Table 33: **Test results of regression equations with Stitch task (DAU) and Stitch task (DHU) testing datasets**

Data Vector	Stitch Task (DAU)			Stitch Task (DHU)		
		Predicted Skill Level			Predicted Skill Level	
	Real skill level	Value	Category	Real skill level	Value	Category
1	E	0.5862	I	E	0.6877	N
2	E	0.2464	E	E	0.5902	I
3	E	0.539	I	E	0.4297	I
4	E	0.6837	N	E	0.4994	I
5	E	1.0682	N	E	0.5788	I
6	E	0.966	N	E	0.5216	I
7	E	0.3115	E	E	0.712	N
8	E	0.3039	E	E	0.6361	I
9	E	0.5772	I	E	0.4341	I
10	E	0.5167	I	E	0.2217	E
11	E	0.3902	I	E	0.3035	E
12	E	0.2941	E	E	0.3362	I
13	E	0.5796	I	E	0.3276	E
14	E	0.1899	E	E	0.2544	E
15	E	0.7182	N	E	0.6084	I
16	E	0.1679	E	E	0.258	E
17	I	0.5049	I	I	0.3577	I
18	I	0.3099	E	I	0.3173	E
19	I	0.3251	E	I	0.4052	I
20	I	0.6148	I	I	0.3031	E
21	I	0.5735	I	I	0.3145	E
22	I	0.3404	I	I	0.2904	E
23	I	0.2487	E	I	0.3716	I
24	I	0.3342	I	I	0.2913	E
25	I	0.3354	I	I	0.2519	E
26	I	0.2603	E	I	0.337	I
27	I	0.4962	I	I	0.3072	E
28	I	0.3473	I	I	0.3261	E
29	I	0.3368	I	I	0.3598	I
30	I	0.1978	E	I	0.2948	E
31	I	0.3043	E	I	0.3164	E
32	I	0.4068	I	I	0.2996	E

Data Vector	Stitch Task (DAU)			Stitch Task (DHU)		
		Predicted Skill Level			Predicted Skill Level	
	Real skill level	Value	Category	Real skill level	Value	Category
33	N	0.8194	N	N	0.5025	I
34	N	0.6835	N	N	0.5131	I
35	N	0.7514	N	N	0.3762	I
36	N	0.7148	N	N	0.4882	I
37	N	0.4824	I	N	0.7681	N
38	N	0.1593	E	N	0.7599	N
39	N	0.9984	N	N	0.5134	I
40	N	0.8741	N	N	0.278	E
41	N	0.695	N	N	0.659	I
42	N	0.8871	N	N	0.624	I
43	N	0.3017	E	N	0.7648	N
44	N	0.3239	E	N	0.4114	I
45	N	0.4559	I	N	0.5952	I
46	N	0.6313	I	N	0.3119	E
47	N	0.5476	I	N	0.3683	I
48	N	0.6019	I	N	0.5763	I
49	N	0.5322	I	N	0.6378	I
50	N	0.7122	N	N	0.6991	N
51	N	0.7414	N	N	0.4325	I
52	N	0.7703	N	N	0.6082	I

Table 34: **Test results of regression equations with HS Knot task (DAU) and Stitch task (DHU) testing datasets**

Data Vector	Stitch Task (DAU)			Stitch Task (DHU)		
		Predicted Skill Level			Predicted Skill Level	
	Real skill level	Value	Category	Real skill level	Value	Category
1	E	0.5475	I	E	0.5556	I
2	E	0.5631	I	E	0.5048	I
3	E	0.5356	I	E	0.6295	I
4	E	0.5392	I	E	0.6547	I
5	E	0.4587	I	E	0.3739	I
6	E	0.6023	I	E	0.5488	I
7	E	0.3354	I	E	0.2748	E
8	E	0.3508	I	E	0.3179	E
9	E	0.3546	I	E	0.4747	I
10	E	0.5227	I	E	0.2395	E
11	E	0.3522	I	E	0.2093	E
12	E	0.3635	I	E	0.3869	I
13	E	0.5381	I	E	0.2887	E
14	E	0.3276	E	E	0.2415	E
15	E	0.3559	I	E	0.2229	E
16	E	0.3232	E	E	0.2537	E
17	I	0.5405	I	I	0.4837	I
18	I	0.5392	I	I	0.3696	I
19	I	0.5488	I	I	0.3018	E
20	I	0.5535	I	I	0.4789	I
21	I	0.5417	I	I	0.4885	I
22	I	0.508	I	I	0.3535	I
23	I	0.5468	I	I	0.4696	I
24	I	0.3304	I	I	0.675	N
25	I	0.5516	I	I	0.526	I
26	I	0.5411	I	I	0.5079	I
27	I	0.5967	I	I	0.486	I
28	I	0.6442	I	I	0.4855	I
29	I	0.5727	I	I	0.6395	I
30	I	0.3566	I	I	0.6809	N
31	I	0.4779	I	I	0.7325	N
32	I	0.393	I	I	0.5881	I

Data Vector	Stitch Task (DAU)			Stitch Task (DHU)		
		Predicted Skill Level			Predicted Skill Level	
	Real skill level	Value	Category	Real skill level	Value	Category
33	N	0.6185	I	N	0.3449	I
34	N	0.6859	N	N	0.5383	I
35	N	0.3706	I	N	0.5668	I
36	N	0.5931	I	N	0.4938	I
37	N	0.5444	I	N	0.5879	I
38	N	0.5309	I	N	0.5394	I
39	N	0.5759	I	N	0.6168	I
40	N	0.7668	N	N	0.5096	I
41	N	0.5237	I	N	1.6782	N
42	N	0.5551	I	N	0.6248	I
43	N	0.5938	I	N	0.5261	I
44	N	0.5578	I	N	0.6018	I
45	N	0.5703	I	N	0.6143	I
46	N	0.5593	I	N	0.2851	E
47	N	0.5978	I	N	0.5685	I
48	N	0.5935	I	N	0.3237	E
49	N	0.5232	I	N	1.1533	N
50	N	0.5685	I	N	0.7685	N
51	N	0.8105	N	N	1.0368	N
52	N	0.6465	I	N	0.3902	I

In Table 34, percentage of matches and the RMS Error values for the results of regression analysis are compared with those obtained by fuzzy classifiers. Performance of the fuzzy models has been slightly better in three of the four models. Stitch task (DAU), Stitch task (DHU), and HS Knot task (DAU) fuzzy models have a higher percentage of matches with lower amounts of RMS Errors compared to the results of regression analysis. The only fuzzy model that did not perform better than the regression analysis is the HS Knot task (DHU), with equal percentage of matches and a slightly higher amount of RMS Error.

Table 35: Comparing Regression Analysis and Fuzzy Classifier results

	Regression Analysis Results		Fuzzy Classifier Results	
	Percentage of Matches	RMS Error	Percentage of Matches	RMS Error
Stitch (DAU)	51.92	0.309	57.69	0.294
Stitch (DHU)	26.92	0.290	46.15	0.268
HS Knot (DAU)	40.38	0.232	59.62	0.227
HS Knot (DHU)	46.15	0.248	46.15	0.251

7 DISCUSSION AND CONCLUSION

Fuzzy set theory makes it possible to express heuristic claims about complicated facts in mathematical language, and is a powerful tool to handle imprecision or fuzziness associated with continuous phenomena dealt with in a large number of practical problems (Cox, 1999). Examples of such problems can be found in various fields of study such as physics, sociology, biotechnology, ecology, finance, medicine, and especially in engineering.

In most cases, the underlying phenomena in such systems are not clearly understood and the most significant source of information is the knowledge of human experts. This knowledge may be too vague and inexact to be expressed by mathematical functions. It is, however, often possible to describe the performance of systems by means of natural language, in the form of if-then rules.

In addition, the nature of many real-world systems is non-linear and cannot be represented by linear models used in conventional system identification. Artificial neural networks and fuzzy models are two of the most popular model structures used for the identification of non-linear systems from measured data.

Fuzzy modelling and identification methodologies have been successfully used for various aims in a broad range of real-world applications. Applications of fuzzy set theory are considerably more developed in engineering than in other areas of research (Klir et al., 1997). Two of the important applications of fuzzy logic are in the problem areas of decision-making and pattern recognition.

Fuzzy methods have been broadly and successfully developed in virtually all branches of decision-making, including multiobjective, multiperson, and multistage decision-making. These methods are, in general, more realistic than their classical counterparts (Klir et al, 1997).

The utility of fuzzy set theory is also well established in the problem area of pattern recognition. This is quite understandable since most categories we commonly encounter and use do not have precise boundaries (Pal, & Dutta Majumder, 1986).

Fuzzy rule-based systems seem to be appropriate tools to handle the problem of surgical performance evaluation, as the nature of the variables is continuous, the relationship between the inputs and outputs of the system is non-linear and complicated, and the only accepted methods of evaluation are based on the complex, imprecise and subjective opinion of experts. Although an objective method of surgical performance evaluation has been the focus of a number of studies, no formal framework has been set.

In this project we investigated the use of fuzzy classifiers as a new approach in objective surgical skills assessment, based on numerical performance metrics collected by a surgical simulator. The goal was to create a skill evaluation scheme to be incorporated in computer-assisted surgical training systems. With proper assessment and validation, such systems can provide feedback during the training episodes, enhancing skills acquisition.

Twenty six subjects with three different surgical skill levels (novice, intermediate, and expert), completed one suturing and one knot-tying task available in the MIST-VR surgical simulator. The performance data collected in the experiment were divided into two equal datasets: the training dataset, which was used to train the fuzzy classifiers, and the testing dataset, used to evaluate the resulting models. This was achieved with the use of two different data separation methods, the DAU and the DHU.

The initial analysis of the user study data revealed some inconsistencies between the surgical skill levels and performance metrics collected by MIST-VR. In the Stitch task for instance, the Maximum Tissue Deformation values were generally higher in the expert and the novice groups, but low in the intermediate group (Figure 8). Similarly in the HS Knot task, the Maximum Thread Overstretch values had the highest values in the expert group (Figure 9). Basic statistical analysis on the collected data also did not suggest a strong correlation between the performance metrics collected by MIST-VR and the surgical skill levels (Table 8 and Table 9). However, it was hypothesised, and was proved to a great extent later, that fuzzy logic-based classifiers can recognize hidden patterns in even poorly correlated data and conquer the imperfections in the sample data.

The Mamdani's Fuzzy Inference Method (Mamdani, 1977) was used to design four fuzzy classifiers: Stitch task DAU, Stitch task DHU, HS knot task DAU, and HS Knot task DHU. The initial models were designed employing the most popular fuzzy inference properties. The constructed models were then tested with the testing dataset.

The effects of a few different fuzzy inference properties were then explored on the performance of the classifiers. Various combinations of membership functions, t-norm operators, and defuzzification methods were applied to our models and each model was tested against the testing dataset.

Performance of the classifiers with various combinations of fuzzy inference properties were compared based on the amount of "Root-Mean-Squared Error" in each system's results, and the number of matches between each system's predicted surgical skill levels and users' real levels of expertise. The best combination of fuzzy inference properties was identified for models with the highest number of matches, which also had a low amount of RMS Error.

Systems with the DAU data separation method provided more reliable results, which may indicate that our sample population was not large enough to represent the overall population. This was due to the extremely busy schedule of the potential participants of this study, and also time and financial limitations.

Regression analysis was used on the user study data as a simple statistical method to identify the input/output relationship in the problem of surgical performance evaluation. The resulting regression equations were then tested with the testing datasets. The comparison between outcomes of the regression analysis and the fuzzy classifications showed that in general, fuzzy models have performed slightly better than the statistical method (Table 35).

The preliminary verifications of our novel approach confirmed that fuzzy classifiers may have the potential to distinguish between various surgical skill levels, and the results of this research can be used as a basis for further improved models.

Since our fuzzy classifiers were designed based on data values collected in a user study, the logistic difficulties in acquiring participants limited the fuzzy systems' source of data. A larger sample size which represents the population more accurately could generate more efficient classifiers.

Further controlling the experimental conditions may also improve the results of the user study. For instance, the presenter in our study was not blind to the participants' surgical skill levels, which may have negatively affected the results.

Other factors such as age, gender, fatigue, and even previous experience with computer games may also have affected the results of this study. Further considerations when acquiring participants may improve the results.

Another issue with the design of fuzzy classifiers in this study was the low correlation between the performance metrics collected by MIST-VR and the participants' surgical skill levels. Using more relevant metrics may improve performance of the classifiers.

The initial design of the fuzzy classifiers was based on an ad hoc procedure. Investigating the effect of various fuzzy inference properties in section 5 showed that performance of these models could be improved by modifying the fuzzy inference properties. In addition, other characteristics of the systems such as the membership functions' attributes could be further adjusted with the help of fuzzy adaptive learning systems (Abony, Nagy, & Szeifert, 1999; Jang, 1993).

Additionally, using more sophisticated performance metrics such as force and torque measurements in advanced surgical training environments may result in more effective systems for assessment of a trainee's skill level.

Finally, the combined effect of the fuzzy classifiers designed for different surgical manoeuvres (such as the Stitch and HS Knot tasks in this study) could be considered as an alternative way of predicting a trainee's surgical expertise.

BIBLIOGRAPHY

MIST Suturing Module - User Manual(2003). Mentice AB.

Simulators for training: Assessment, validation and acceptance strategies, executive summary. (2003). Medicine meets virtual reality, MIST User Manual(2002). Mentice AB.

Fuzzy logic and fuzzy expert systems tutorial. (1997). Retrieved April 20, 2004 from http://www-2.cs.cmu.edu/Groups/AI/html/faqs/ai/fuzzy/part1/faq.html

The Berkeley initiative in soft computing. Retrieved November 17, 2005 from http://www-bisc.cs.berkeley.edu/BISCProgram/default.htm

Flight simulator from wikipedia, the free encyclopaedia. Retrieved November 6, 2005 from http://en.wikipedia.org/wiki/Flight_simulator

Fuzzy logic toolbox user guide. Retrieved April 10, 2004 from http://www.mathworks.com/access/helpdesk/help/toolbox/fuzzy/index.html

GRB research group at Minnesota state university, Mankato. Retrieved September 29, 2005 from http://grb.mnsu.edu/grbts/doc/manual/ Error_Measurements.html#SECTION00032400000000000000

Mentice corporation., 2004 from www.mentice.com

NYU medical center, benefits of minimally invasive surgery. Retrieved Aug/19, 2005 from http://www.med.nyu.edu/cvsurgery/en/info/cardiac/mininvasive2.html

Personal Communication: Dr. Nagy, A.; Vancouver General Hospital, BC, Canada, 2003, Dr. Doris, P.; Surrey Memorial Hospital, BC, Canada, 2004, Dr. Lomax, A.J.; Simon Fraser University, BC, Canada, 2003

The royal Australian and New Zealand college of obstetricians and gynaecologists. Retrieved November 5, 2005 from (http://www.ranzcog.edu.au/trainees/trainees-surgical-skills-assessment.shtml

World technology evaluation centre (WTEC). Retrieved November 16, 2005 from http://www.wtec.org/loyola/kb/c5_s4.htm

Abonyi, J., Nagy, L., & Szeifert, F. (1999). Adaptive fuzzy inference system and its application in modeling and model based control. Chem. Eng. Res. Des, 77(A), 281-290.

Berg, D., Berkley, J., Weghorst, S., Raugi, G., Turkiyyah, G., & Ganter, M. et al. (2002). Issues in validation of a dermatologic surgery simulator. Medicine meets virtual reality, Newport Beach, CA, 10

Beringer, D. B. (2002). Applying performance-controlled systems, fuzzy logic, and fly-by-wire controls to general aviation (Tech. Rep No. DOT/FAA/AM-02/7). Washington, D.C.: Civil Aerospace Medical Institute PUBLICATIONS.

Brown, J. D., Rosen, J., Chang, L., Sinanan, M. N., & Hannaford, B. (2004). Quantifying surgeon grasping mechanics in laparoscopy using the blue DRAGON system. Studies in health technology and informatics - medicine meets virtual reality, Newport Beach,

Chen, J., Yeasin, M., & Sharma, R. (2003). Visual modelling and evaluation of surgical skill. Pattern Analysis and Applications, 6(1), 1-11.

Cotin, S., Stylopoulos, N., Ottensmeyer, M., Neumann, P., Rattner, D., & Dawson, S. (2002). Metrics for laparoscopic skills trainers: the weakest link!. MICCAI, lecture notes in computer science, 2488 35-43.

Cox, E. (1999). The fuzzy systems Handbook: A practitioner's guide to building, using, and maintaining fuzzy systems (2nd ed.). San Diego, CA; London: AP Professional.

Cuschieri, A. (2001). Human reliability assessment in surgery: A new approach for improving surgical performance and clinical outcome. Annals of the Royal College of Surgeons of England, 135(7), 551-555.

Dadone, P. (2001). Design optimization of fuzzy logic systems.

Darzi, A. (2001). Assessment of surgical competence. Quality in Health Care, 10, ii64-ii69.

Darzi, A., Smith, S., & Taffinder, N. (1999). Assessing operative skill. BJM (British Medical Journal), 318, 887-888.

Datta, V., Mackay, S., Mandalia, M., & Darzi, A. (2001). The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. Journal of American College of Surgeons, 193, 479-485.

Datta, V., Mandalia, M., Mackay, S., Chang, A., Cheshire, N., & Darzi, A. (2002). Relationship between skill and outcome in the laboratory-based model. Surgery, 131(3), 318-323.

Dote, Y., & Ovaska, S. J. (2001). Industrial applications of soft computing: A review. IEEE, 89(9), 1243-1265.

Fayek, A. R., & Sun, Z. (2001). A fuzzy expert system for design performance prediction and evaluation. Canadian Journal of Civil Engineering, 28(1), 1-25.

Gallagher, A. G., McClure, N., McGuigan, J., Crothers, I., & Browning, J. (1999). Virtual reality training in laparoscopic surgery: A preliminary assessment of minimally invasive surgical trainer virtual reality (MIST VR). Endoscopy, 31(4), 310-313.

Gallagher, A. G., Richie, K., McClure, N., & McGuigan, J. (2001). Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality: A preliminary comparison of a USA and British sample. Worl Journal of Surgery, 25(11), 1478-1483.

Gómez, J., & Dasgupta, D. (2002). Evolving fuzzy classifiers for intrusion detection . Third annual IEEE information assurance workshop, New Jersey,

Gorzałczany, M. B., & Grądzki, P. (1999). Computational intelligence in medical decision support - a comparison of two neuro-fuzzy systems. IEEE international symposium on industrial electronics, Bled, Słowenia, 1 408-413.

Hajshirmohammadi, I., & Payandeh, S. (2006). Study of fuzzy set theory to evaluate performance on surgical simulators. Submitted after revision.

Hanna, G. B., Shimi, S. M., & Cuschieri, A. (1998). Randomised study of influence of two-dimensional versus three-dimensional imaging on performance of laparoscopic cholecystectomy. Lancet, 351, 248-251.

Hartog, d., M.H., Babuˇska, R., Deketh, H. J. R., Grima, M. A., Verhoef, P. N. W., & Verbruggen, H. B. (1997). Knowledge-based fuzzy model for performance prediction of a rock cutting trencher. International Journal of Approximate Reasoning, 16(1), 43-66.

Hong, T. P., & Chen, J. B. (1999). Finding relevant attributes and membership functions. Fuzzy Sets and Systems, 103(3), 389-404.

Hong, T. P., & Lee, C. Y. (1996). Induction of fuzzy rules and membership functions from training examples. Fuzzy Sets and Systems, 84(1), 33-47.

Huang, J., Payandeh, S., Doris, P., & Hajshirmohammadi, I. (2005). Fuzzy classification: Towards evaluating performance on a surgical simulator. Medicine meets virtual reality 13, Long Beach, California, 194-200.

Jang, J. S. (1993). ANFIS: Adaptive-network-based fuzzy inference system. IEEE Transactions on Systems, Man, and Cybernetics, 23(3), 665-684.

Jordan, J. A., Gallagher, A. G., McGuigan, J., & McClure, N. (2001). Virtual reality training leads to faster adaptation to the novel psychomotor restrictions encountered by laparoscopic surgeons. Surgical Endoscopy, 15(10), 1080-1084.

Kumar, M., Stoll, R., & Stoll, N. (2003). Regularized adaptation of fuzzy inference systems. modelling the opinion of a medical expert about physical fitness: An application. Fuzzy Optimization and Decision Making, 2(4), 317-336.

Lotfi Zadeh. (1965). Fuzzy sets. Information and Control, 8, 338-353.

Mamdani, E. H. (1977). Application of fuzzy logic to approximate reasoning using linguistic systems. Fuzzy Sets and Systems, 26, 1182-1191.

Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Studies, 7, 1-13.

Martin, J. A., 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.

McNatt, S. S., & Smith, C. D. (2001). A computer-based laparoscopic skills assessment device differentiates experienced from novice laparoscopic surgeons. Surgical Endoscopy, 15(10), 1085-1089.

Mengali, G. (2000). The use of fuzzy logic in adaptive flight control systems. The Aeronautical Journal, 104(1031), 31-37.

Moody, L., Baber, C., Arvanitis, T. N., & Elliott, M. (2003). Objective metrics for the evaluation of simple surgical skills in real and virtual domains. Presence: Teleoperators & Virtual Environments, 12(2), 207-221.

Moorthy, K., Munz, Y., Sarker, S., K., & Darzi, A. (2003). Objective assessment of technical skills in surgery. BMJ (British Medical Journal), 327, 1032-1037.

Ota, D., Loftin, B., Saito, T., Lea, R., & Keller, J. (1995). Virtual reality in surgical education. Computers in Biology and Medicine, 25(2), 125-127.

Payandeh, S., Lomax, A. J., Dill, J., MacKenzie, C. L., & Cao, C. G. L. (2002). On defining metrics for assessing laparoscopic surgical skills in a virtual training environment. Medicine meets virtual reality, Newport Beach, CA, 10

Rahbari, R., Leach, B. W., Dillon, J., & DaSilva, C. W. (2002). Adaptive tuning of a kalman filter using the fuzzy integral for an intelligent navigation system. Vancouver, BC, 252-257.

Regehr, G., MacRae, H., Reznick, R. K., & Szalay, D. (1998). Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Academic Medicine, 73(9), 993-997.

Richards, C., Rosen, J., Hannaford, B., Pellegrini, C., & Sinanan, M. (2000). Skills evaluation in minimally invasive surgery using force/torque signatures. Surgical Endoscopy, 14, 791-798.

Rosen, J., Hannaford, B., Richards, C., & Sinanan, M. (2001). Markov modeling of minimally invasive surgery based on Tool/Tissue interaction and Force/Torque signatures for evaluating surgical skills. IEEE Transactions on Biomedical Engineering, 48(5), 579-591.

Smith, S. G., Torkington, J., Brown, T. J., Taffinder, N. J., & Darzi, A. (2002). Motion analysis. Surgical Endoscopy, 16(4), 640-645.

Stylopoulos, N., Cotin, S., Maithel, S. K., Ottensmeyer, M., Jackson, P. G., & Bardsley, R. S. et al. (2004). Computer-enhanced laparoscopic training system (CELTS): Bridging the gap. Surgical Endoscopy, 18(5), 782-789.

Sundaram, J., Naidu, B. K., & Das, H. (2004). Fuzzy logic a multi attribute decision making approach in product development and sensory evaluation. ASAE Annual Meeting,

Szczepaniak, P. S., Lisboa, P. J. G., & Kacprzyk, J. (2000). Fuzzy systems in medicine (1st ed.). Heidelberg ; New York: Physica Verlag.

Taffinder, N., Smith, S., Mair, J., Russell, R., & Darzi, A. (1999). Can a computer measure surgical precision? reliability, validity and feasibility of the ICSAD. Surgical Endoscopy, 13(1), 81.

Taffinder, N., Smith, S. G., Huber, J., Russell, R. C., & Darzi, A. (1999). The effect of a second-generation 3D endoscope on the laparoscopic precision of novices and experienced surgeons. Surgical Endoscopy, 13(11), 1087-1092.

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man, and Cybernetics, 15, 329-346.

Tendick, F., Downes, M., Goktekin, T., Cavusoglu, M., Feygin, D., & Wu, X. et al. (2000). A virtual environment testbed for training laparoscopic surgical skills. Presence, 9(3), 236-255.

Torkington, J., Smith, S. G., Rees, B. I., & Darzi A. (2001). Skill transfer from virtual reality to a real laparoscopic task. Surgical Endoscopy, 15(10), 1076-1079.

Vanegas, L. V., & Labib, A. W. (2005). Fuzzy approaches to evaluation in engineering design. ASME Journal of Mechanical Design, 127

VanRij, A. M., McDonald, J. R., Pettigrew, R. A., Putterill, M. J., Reddy, C. K., & Wright, J. (1995). Cusum as an aid to early assessment of the surgical trainee. British Journal of Surgery, 82(1500), 1503.

Wayand, W. (2004). The history of minimally invasive surgery. Retrieved Dec./21, 2005 from http://www.touchbriefings.com/pdf/952/Wayand.pdf

Wilson, M. S., Middlebrook, A., Sutton, C., Stone, R., & McCloy, R. F. (1997). MIST VR: A virtual reality trainer for laparoscopic surgery assesses performance. Annals of the Royal College of Surgeons of England, 79(6), 403-404.

Yager, R. R., & Filev, D. (1994). Template-based fuzzy systems modeling. Journal of Intelligence and Fuzzy Systems, 2(1), 39-54.

Zeng, X. J., & Singh, M. G. (1995). Approximation theory of fuzzy systems –MIMO case. IEEE Trans., 3(2), 219-235.

Appendices

Appendix A: Preliminary User Study

A primary user study of 12 subjects in three categories of expertise (novice, intermediate, and expert) was conducted, employing Acquire Place and Transfer Place tasks available in MIST-VR. Both tasks involved the manipulation of a ball with grippers and were among the most basic tasks defined in MIST-VR. Half of the performance metrics were used to design a fuzzy classifier for each task, and the rest to test the models. Our classifiers however, did not offer the optimal results. We believed that the results could be improved by:

- Testing a lager sample size that would be a better representative of the performance variation of our three categories of participants, and
- Employing more complicated surgical tasks, which could increase chances of success of a classifier by providing a larger skill gap between the three categories of expertise.

Our following experiment, which is explained in this thesis, was designed by incorporating the above-mentioned factors.

Appendix B: User Study Questionnaire

Using Fuzzy Set Theory to Evaluate performance on Surgical Simulators

Study Questionnaire Form

Part 1: About the participant

1. Participant's position:

☐ Senior Surgeon ☐ Surgical Resident ☐ Intern

☐ Other (Please specify): ____

2. Participant's Age:

☐ 19-29 ☐ 29-39 ☐ 39-49 ☐ 49-59 ☐ Over 59

3. Participant's dominant hand:

☐ Left ☐ Right

Part 2: Minimally Invasive Surgery training history

1. Have you ever had training in minimally invasive surgery?

☐ No (Go to Part 3)

☐ Yes

2. Have you had any training in computer-based simulators?

☐ No (Go to question 3)

☐ Yes

For how many hours?

☐ Less than 2 ☐ 3-5 ☐ 6-9 ☐ 10-13 ☐ Over 13

3. Have you had any training using physical simulators?

☐ No (Go to question 4)

☐ Yes

For how many hours?

☐ Less than 2 ☐ 3-5 ☐ 6-9 ☐ 10-13 ☐ Over 13

4. Have you had any training in animal labs?

☐ No (Go to question 5)

☐ Yes

For how many hours?

☐ Less than 2 ☐ 3-5 ☐ 6-9 ☐ 10-13 ☐ Over 13

5. Have you had any training in an OR, by observing a surgery?

☐ No (Go to question 6)

☐ Yes

How many surgeries have you been present in? ______________________

6. Have you had any training in an OR, by assisting a surgery?

☐ No (Go to question 7)

☐ Yes

Please specify in how many surgeries and the type of assistance? __________

____

7. Have you had any training in an OR, by performing a surgery under the supervision of an expert surgeon?

☐ No (Go to part 3)

☐ Yes

How many surgeries? ____

Part 3: Previous surgical experience

1. For purposes other than training, have you ever been present in the operating room as a performer/observer of a minimally invasive surgery?

☐ No (Go to Part 4)

☐ Yes

2. How many minimally invasive surgeries have you observed?

☐ 0 ☐ 1-20 ☐ 21-40 ☐ 41-60 ☐ Over 60

3. How many minimally invasive surgeries have you assisted in?

☐ 0 ☐ 1-20 ☐ 21-40 ☐ 41-60 ☐ Over 60

Type of assistance: ______________________________

4. How many minimally invasive surgeries have you performed?

☐ 0 ☐ 1-20 ☐ 21-40 ☐ 41-60 ☐ Over 60

Part 4: Previous experience with surgical simulators

1. For purposes other than training, have you ever used computer-based surgical simulators?

☐ No (Go to question 2)
☐ Yes
For how many hours?
☐ Less than 2 ☐ 2-5 ☐ 6-9 ☐ 10-13 ☐ Over 13

2. For purposes other than training, have you ever used physical surgical simulators?

☐ No (Go to part 5)
☐ Yes
For how many hours?
☐ Less than 2 ☐ 2-5 ■ 6-9 ☐ 10-13 ☐ Over 13

Part 5: Comments

Thank you for participating in this experiment and for filling this questionnaire.