Proceedings of the ASME 2021
International Design Engineering Technical Conferences and
Computers and Information in Engineering Conference
IDETC/CIE2021
August 17-20, 2021, Virtual, Online

DETC2021-69337

AN APPLICATION OF MACHINE LEARNING TO PREDICT STIFFNESS DISCRIMINATION THRESHOLDS USING HAPTICS

Ernur Karadoğan
Robotics and Haptics Lab
School of Engineering and Technology
Central Michigan University
Mount Pleasant, MI

ernur.karadogan@cmich.edu

ABSTRACT

The effectiveness of our interaction with the computer-generated environments is subject to our physical limitations in real life such as our ability of discriminating differences in stiffness or roughness. This ability, represented by Weber fractions, is usually quantified by means of psychophysical experimentation. The experimentation process is tedious and repetitive as it requires the same task to be completed by participants until the mastery at a certain stimulus level can be ensured before moving onto the next level. Moreover, these thresholds are dependent on the tested standard stimulus level and, therefore, need to be identified by separate experiments for every possible standard stimulus level. The purpose of the current study is to reduce the amount of experimentation and predict the thresholds for stiffness discrimination of individuals after being tested at a single stimulus level. The prediction models tested provide a moderate level of prediction power, but more features, potentially physical and demographical in nature, are needed to increase their effectiveness. The procedure described herein can be extended to any modality other than stiffness and, therefore, has the potential to predict overall palpation effectiveness of an individual after a feasible amount of data is obtained through experimentation.

Keywords: Haptics, Virtual Environments, Machine Learning, Psychophysics.

1. INTRODUCTION

We use haptics in our daily lives such as in our smart phones and gaming hardware. In addition to these common activities, haptics holds an important place in research, education [1-8] and design. There are, for instance, simulations that help medical professionals practice diagnosis and treatment using palpation in veterinary [9], allopathic [10, 11], and osteopathic medicine [12-14]. Students nowadays have the opportunity to learn by touch—even by “touching” viruses and protein molecules [15]. Therefore, the sense of touch has been and will be in our lives more and more as the technological advances make haptics interfaces ubiquitous and affordable, especially with a high fidelity.

There are several aspects of touch that we utilize while interacting with an object in real life. We receive both cutaneous and proprioceptive feedback from our skin, limbs, and joints. That information could be acquired from several properties of objects such as its roughness, temperature, or stiffness. Therefore, discriminating subtle differences in these properties potentially makes an individual a better palpator, which is a skill that could be learned and improved [16]. This raises the question of how we can measure those abilities of someone to discriminate temperature or stiffness. The answer lies in psychophysics. Psychophysical experiments can quantify the cumulative performance of an individual’s physical abilities and their performance in interpreting the information they received from a physical stimulus. These experiments are tedious and involve repetitive tasks through a certain number of trials; number of trials may depend on the protocol employed and, in the case of adaptive algorithms, even may vary for each participant depending on performance. Another caveat is that the outcomes of these types of psychophysical experiments are dependent on the "standard" stimulus. That is, the outcome is measured relative to a standard stimulus that is presented throughout the experimentation. Therefore, in theory, infinitely many of these tedious and repetitive experiments on an individual must be run to fully comprehend that individual's sensitivity to discriminating differences in a specific modality. If we were to consider multiple modalities, this should signify the importance of a method that will limit the number of such experiments, yet still gives us a confident prediction of the full picture. Machine learning has potential to tackle these challenges by helping interpolate the outcomes of experiments within a certain range of standard stimulus values.

In this paper, we investigated a method to apply machine learning to predict the ability of individuals to discriminate stiffness differences at four different standard stimuli. The data used in the study was obtained from a psychophysical experiment that was designed to acquire the Weber fractions (the ratio of the minimum discriminated stiffness difference between two surfaces to the standard stimulus value) of medical students for the stiffness discrimination task. In the following sections, the methodology, feature selection, and selected model performances are presented along with the discussion of the results.

2. MATERIALS AND METHODS

The data used for this study were obtained by means of a psychophysics experiment that was designed to measure the stiffness discrimination thresholds of individuals. This section describes that experiment, the data obtained as they were transformed into features for machine learning models, the preprocessing of the data (feature selection and preparation), and the selected models with the utilized performance evaluation metrics.

2.1 Data Collection

This section details the psychophysics experiment that was the source of the data used for prediction. The goal of the experiment was to obtain Weber fractions for each participant at every tested standard stimulus level (0.25, 0.50, 1.00 and 1.25 N/mm).

Experimental Setup. The experiments were run on a 2.8 GHz dual Pentium PC with 1 GB RAM and an NVIDIA Quadro 4XGL video adapter. A Geomagic Touch™ (3D Systems, Inc.) haptic interface displayed the stiffnesses to the subjects (Figure 1). The stylus of the haptic device was modified to allow manipulation with the index finger of the dominant hand. The graphical interface was written using Microsoft Visual C++ and the OpenGL® graphic library. The haptic effects were implemented by using the OpenHaptics Toolkit (3D Systems, Inc.). The calibration curve for the stiffness coefficient set using the API and the force-displacement values as read from the device was used to render the desired stiffness values.

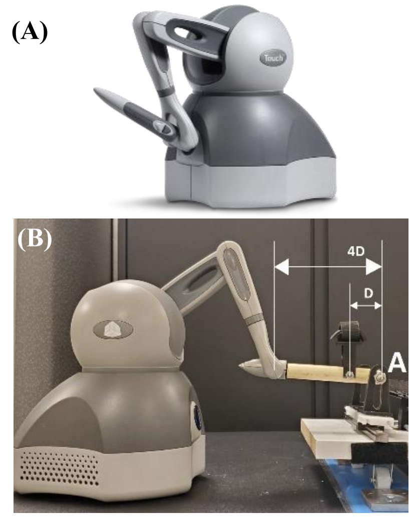

Figure 1. (A) COMMERCIALLY AVAILABLE GEOMAGIC TOUCH™ (3D SYSTEMS, INC.); (B) THE MODIFIED VERSION THAT WAS USED IN THE EXPERIMENT; THE STYLUS CAN ROTATE ABOUT THE AXIS PASSING THROUGH THE POINT A AND SLIDE SIDEWAYS (IN AND OUT OF THE PAGE AS SHOWN)

As shown in Figure 2, the user interface was composed of two virtual cylinders along with a 20-sec timer display, and some other relevant information such as the feedback on the last response. The participants, by using the haptic interface, "felt" the top portions of the virtual cylinders and were asked to identify the stiffer one of these surfaces in 20 seconds for each trial. The amount of stiffness difference between the surfaces were automatically adjusted by the software based on a participant's performance history as described in the Experimental Procedure section.

Figure 2. VISUAL SCENE OF THE EXPERIMENT.

Participants. Thirteen adult subjects (eight female and five male) who were second-year osteopathic medical students participated in the study. None of the participants had any prior known neuromuscular abnormalities.

Experimental Procedure. The stiffness discrimination thresholds for each participant were investigated by using four different standard stimulus levels: 0.25, 0.50, 1.00 and 1.25 N/mm. The participants had two weeks to complete all four sessions, but they were not allowed to perform more than one session at any given day.

A two-alternative forced-choice (2AFC) method was applied by presenting the participants with two surfaces to choose from and enforcing a 20-second time limit per trial. After the expiration of the 20-sec time limit in a trial, the software registered an incorrect answer for that trial. The participants were instructed to select the stiffer of the two surfaces and to keep performing the same task until the software terminated a session. The number of total trials per participant was dependent on the individual performance as, behind the scenes, the software used an adaptive algorithm to adjust the stimulus level, in our case the stiffness difference between the two surfaces. One of the surfaces, called the standard side (STD), had constant stiffness (0.25, 0.50, 1.00 or 1.25 N/mm) at any given session and was randomly displayed on the left or right, whereas the comparison side (CMP) stiffness was adjusted based on the participants' performance, which was always less than the STD stiffness. The Wald decision rule [17] was implemented to decide when to make a change (increase or decrease) in the stimulus level (i.e., the CMP stiffness). Once the decision was made, the amount of change (the step size) was calculated by using Parameter Estimation by Sequential Testing (PEST) [17, 18].

The experimental sessions have ended after seven reversals. A reversal was defined as the change of the CMP stiffness in the opposite direction (e.g. increase of CMP stiffness after a decrease)—the STD stiffness was constant. The average of the difference between STD and CMP stiffness values for the last four reversals were used to calculate the just-noticeable difference (JND). Finally, the Weber fraction at any given standard stimulus level was calculated as [19]:

$$W = \frac{\text{JND}}{\text{standard stiffness value}}$$

A higher Weber fraction for a particular standard stiffness value implied a higher JND, and, therefore, a reduced level of sensitivity to discriminate stiffness differences and vice versa. During the experiments, the amount of force applied by the participants and the fingertip velocities were also recorded.

2.1 Feature Selection and Data Preparation

The data collected from the experiment included the calculated Weber fractions per session per participant (session averages are shown in Figure 3), average palpation force, and fingertip velocity values for every trial as read from the haptic device. As shown in Figure 4, the average force increased monotonically with the standard stiffness level, but the trend was not statistically significant. The average velocity was positively correlated to Weber fraction with $R^2 = 0.65$ (Figure 5).

The palpation force and speed (magnitude of the velocities) were selected as features for the machine learning implementation. The features also included the gender of the participants. Due to the data being aggregated over many trials, two sets of features (14 features in each) were used to be able to potentially capture the individual variety in the amount of force and fingertip velocity. Both sets included one-hot encoded session numbers since the session number represents the standard stimulus level, which affects the Weber fraction directly. Another common future for both feature sets was the gender that was categorized as 0s and 1s, i.e. (0,1) for male and (1,0) for female. Table 1 presents the content details for each feature set.

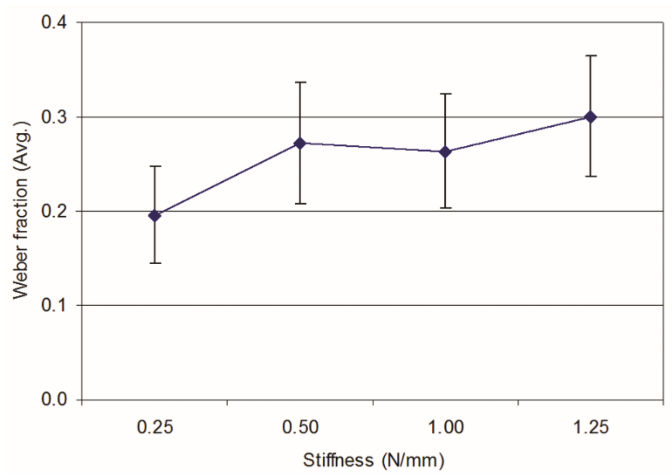

Figure 3. WEBER FRACTION VS. STANDARD STIFFNESS VALUES (MEAN ± STANDARD ERROR)

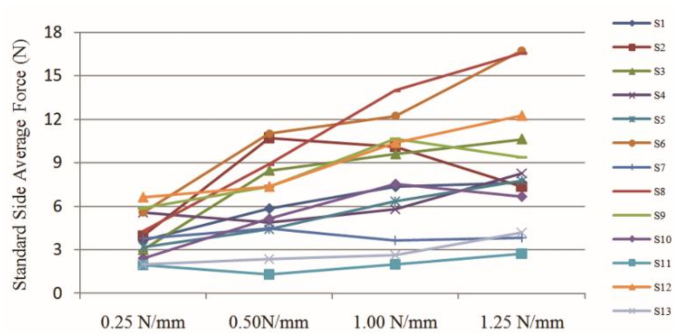

Figure 4. AVERAGE PALPATION FORCE APPLIED BY EACH PARTICIPANT ON THE STANDARD SIDE

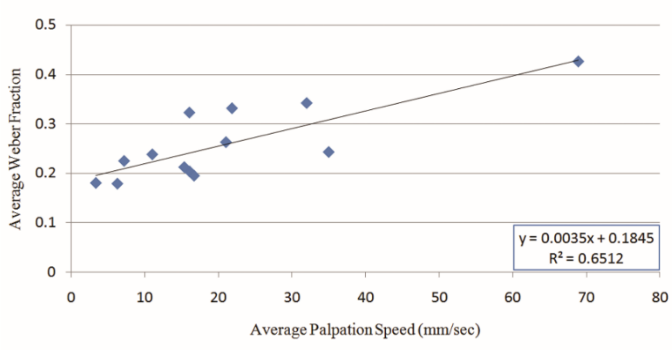

Figure 5. AVERAGE WEBER FRACTION VS. AVERAGE PALPATION SPEED

The main difference between the two feature sets was how the palpation force and fingertip velocities were incorporated. In the first set, the grand average of all trials in a session were included using one-hot encoding. The second feature set included the averages of quartiles for each session. Therefore, in the latter data set all four columns were populated for both force and velocity, whereas in the former set three out of four columns were zeros at any given row—the non-zero column's location was depended on the corresponding session. For instance, for the session with 0.50 N/mm standard stiffness, the second (out of four) column for a participant would be populated with either average fingertip velocity or palpation force. All continuous features were normalized to eliminate any potential scaling effect.

The labeling data were composed of the Weber fractions and they were prepared for both regression and classification. In the case of classification, the classes were defined based on the nature of the Weber fraction as a representation of one's ability to discriminate stiffness differences (higher Weber fraction implies lower sensitivity to differences). The Weber fractions were categorized as 1) a multiclass classification problem (Above Average (AA); Average (A), Below Average (BA)), and 2) a binary classification problem (Above Average (AA); Below Average (BA)). The categorization for the multiclass classification problem was performed by computing the mean (μ) and the standard deviation (σ) of the Weber fractions for each session. In that sense, AA, A, and BA represented individuals with Weber fractions greater than $\mu+\sigma$, between $\mu\pm\sigma$, and less than $\mu-\sigma$, respectively. For binary classification, AA and BA represented greater than and less than μ, respectively. For regression analysis, the Weber fraction values were not modified in any way and continuous.

Table 1. Content of feature sets used for the prediction task.

Feature	Data Type	Feature Set 1	Feature Set 2
Session #1 (0.25 N/mm)	Nominal	☑	☑
Session #2 (0.50 N/mm)	Nominal	☑	☑
Session #3 (1.00 N/mm)	Nominal	☑	☑
Session #4 (1.25 N/mm)	Nominal	☑	☑
Average force for Session #1	Continuous	☑	x
Average force for Session #2	Continuous	☑	x
Average force for Session #3	Continuous	☑	x
Average force for Session #4	Continuous	☑	x
Average force in the first quartile*	Continuous	x	☑
Average force in the second quartile*	Continuous	x	☑
Average force in the third quartile*	Continuous	x	☑
Average force in the fourth quartile*	Continuous	x	☑
Average fingertip velocity for Session #1	Continuous	☑	x
Average fingertip velocity for Session #2	Continuous	☑	x
Average fingertip velocity for Session #3	Continuous	☑	x
Average fingertip velocity for Session #4	Continuous	☑	x
Average fingertip velocity in the first quartile*	Continuous	x	☑
Average fingertip velocity in the second quartile*	Continuous	x	☑
Average fingertip velocity in the third quartile*	Continuous	x	☑
Average fingertip velocity in the fourth quartile*	Continuous	x	☑
Gender (Male)	Nominal	☑	☑
Gender (Female)	Nominal	☑	☑

* The non-zero session number for a given row specifies the corresponding session.

2.2 Selected Machine Learning Models and Performance Evaluation Metrics

Four models were selected to be used during performance evaluations. These models were KNN-Neighbors [20], Random Forest [21], AdaBoost [22], and Logistic Regression [23]. The metrics that were used for evaluation were accuracy, recall, precision, and F1-macro for the binary classification, and accuracy and F1-macro for the multiclass classification. The R^2 metric was used for regression. Each model's evaluation metric was computed using the average of 10-fold cross-validation.

3. PREDICTION RESULTS

This section presents the performance results of the binary and multiclass classification tasks to predict the Weber fraction. The regression results were extremely poor when the R^2 metric was considered and, therefore, they are not included in the results. The 10-fold cross-validation performance of the four models for binary and multiclass classification are presented in Tables 2 and 3.

4. DISCUSSION

In this study, we investigated the feasibility of using machine learning to predict individuals' ability to discriminate stiffness differences by quantifying that ability as Weber fractions. A lower Weber fraction implied higher sensitivity to discriminate these differences in the standard stimulus (stiffness difference in our case). The motivation behind this study was the fact that finding out the Weber fractions by means of only experimentation is tedious and costly (participant incentives, staff time etc.). Additionally, the Weber fraction is dependent on the standard stimulus level tested and, therefore, changes with that standard stimulus level. This implies that, for a meaningful outcome, every individual would need to go through countless experiments to have a confident idea in terms of their abilities to discriminate differences in each modality. We can reduce the number of such experiments with the help of machine learning.

The selected machine learning models were evaluated using 10-fold cross-validation. Several metrics including accuracy, recall, precision, and F1-macro were calculated to have a better view of the performance. As a result, we see that the Feature Set 2 yielded better results for both binary and multiclass classification. Since the number of features are the same (14) in both sets, the difference cannot be simply attributed to the dimensionality (curse of dimensionality [24] etc.). The main difference between the two feature sets we considered was the resolution of the forces and fingertip velocities into quartiles. The Feature Set 2 included the averages of the quartiles, i.e. first quartile, second quartile etc. Even though there was no statistically significant difference between these quartiles, they may have provided more insight into the session data for the models to interpret.

There was not an obvious difference between the binary and multiclass classification. Multiclass classification enabled finer specification—three classes as compared to two in binary classification—of individuals' ability, but it suffered from imbalanced classes. Both Above Average (AA) and Below Average (BA) classes were ~33% of the Average (A) class. The Synthetic Minority Oversampling Technique, or SMOTE [25], and random oversampling were applied to the data, but that revealed inferior results.

One of the limitations of this study was the sample size. Due the aforementioned challenges with experimenting, this was an expected shortcoming. We are, however, still working on adding more data points to improve the outcome of the models. Another limitation was the fact that the data was drawn from second year osteopathic medical students. At this stage, these students are still working on developing their palpation skills via practice in labs and internships. This may have caused some variation in the data that may not have been present if their skills were fully developed. Therefore, increasing the participant numbers and drawing participants from different populations may yield better prediction and generalization of the results.

As future work, we are working on adding more relevant features (mostly anatomical in nature) such as age, anthropometric data (length of the finger digits, forearm, and elbow-to-shoulder distance etc.), and other performance parameters (proprioceptive acuity, reaction time, fingertip sensitivity etc.). We also plan to expand the set of appropriate machine learning models tested to investigate the effect of model selection.

Table 2. Binary classification performance of the tested models (10-fold cross-validation).

Model	Feature Set 1				Feature Set 2			
	Accuracy	Recall	Precision	F1-macro	Accuracy	Recall	Precision	F1-macro
KNN Neighbors	0.60	0.63	0.67	0.58	0.71	0.77	0.74	0.68
Random Forest	0.70	0.80	0.69	0.65	0.71	0.83	0.72	0.65
AdaBoost	0.59	0.67	0.58	0.50	0.72	0.73	0.74	0.68
Logistic Regression	0.66	0.83	0.68	0.60	0.68	0.87	0.72	0.58

Table 3. Multiclass classification performance of the tested models (10-fold cross-validation).

Model	Feature Set 1		Feature Set 2	
	Accuracy	F1 macro	Accuracy	F1 macro
KNN neighbors	0.60	0.29	0.71	0.44
Random Forest	0.64	0.36	0.64	0.38
AdaBoost	0.64	0.30	0.65	0.32
Logistic Regression	0.66	0.31	0.66	0.36

5. CONCLUSION

We investigated a methodology to overcome some of the challenges of psychophysical experimentation by using machine learning. This methodology can be applied to any modality, but the feature selection, as shown from the results, plays an important role as anticipated. The expansion of features that will improve the performance metrics of the models is important. The potential limitations of the study were the limited sample size, associated imbalanced classes, and specific population from which the data were drawn.

REFERENCES

1. Han and Black: Incorporating haptic feedback in simulation for learning physics. (2011).
2. Skulmowski et al.: Embodied Learning using a tangible user interface: the effects of haptic perception and selective pointing on a spatial learning task. (2015).
3. Hightower et al.: Haptic explorers: Supporting science journaling through mobile haptic feedback displays. (2018).
4. Williams et al. Haptics-Augmented Simple-Machine Educational Tools. (2003).
5. Karadogan E., Williams II R.L., Moore D., Luo T.: Haptics-Augmented Training Software for Undergraduate Mechanics, ASME International Design Technical Conferences, 36th Mechanisms and Robotics Conference (Novel Mechanisms, Robots and Applications), Paper # DETC2012-70250, Chicago, IL, August 12-15. (2012).
6. Karadogan E., Williams II R.L., Haptic Modules for Training in Undergraduate Mechanics, American Society for Engineering Education Gulf Southwest Section, San Antonio, TX, March 25-27. (2015).
7. Karadogan E., Karadogan F.: Simulation-based Learning Modules for Undergraduate Engineering Dynamics”, Computer Applications in Engineering Education, 27(4):846-862. (2019).
8. Bateman et al.: A user-centered design and analysis of an electrostatic haptic touchscreen system for students with visual impairments. (2018).
9. Parkes et al.: A Mixed Reality Simulator for Feline Abdominal Palpation Training in Veterinary Medicine. (2009).
10. Escobar-Castillejos.: A review of simulators with haptic devices for medical training. (2016).
11. McWilliams and Malecha.: Comparing Intravenous Insertion Instructional Methods with Haptic Simulators. (2017).
12. Karadogan E., Williams II R.L.: Haptic Modules for Training in Palpatory Diagnosis, IEEE Haptics Symposium, Vancouver, BC, Canada, March 4-7, 223-230. (2012).
13. Karadogan E., Williams II R.L.: Haptic Modules for Palpatory Diagnosis Training of Medical Students, Virtual Reality, 17(1):45-58. (2013).
14. Karadogan E., Williams II R.L., Karadogan F.: Evaluation of Haptic Modules for Training in Undergraduate Mechanics, ASME International Design Technical Conferences, 39th Mechanisms and Robotics Conference, Paper # DETC2015-46060, Boston, MA, August 2-5. (2015).
15. Schonborn, K. J., Bivall, P., & Tibell, L. A.: Exploring relationships between students’ interaction and learning with a haptic virtual biomolecular model. (2011).
16. Williams RL II, Howell JN, Conatser RR Jr, Burns JM, Eland DC.: Palpatory Training on the Virtual Haptic Back Improves Detection of Compliance Difference. Eurohaptics, Paris, France; (2006).
17. Macmillan NA, Creelman CD. Detection Theory: A User’s Guide. Cambridge: Cambridge University Press; (1991).
18. Karadogan E., Williams II R.L., Howell J.N., Conatser Jr. R.R. , 2010, “A Stiffness Discrimination Experiment including Analysis of Palpation Forces and Velocities”, Simulation in Healthcare, 5(5):279-288.
19. Gescheider GA. Psychophysics: the Fundamentals. 3rd ed. Lawrence Erlbaum Associates, Mahwah, NJ; (1997).
20. Bentley, J.L: Multidimensional binary search trees used for associative searching, Communications of the ACM (1975).
21. Breiman: “Random Forests”, Machine Learning, 45(1), 5-32, (2001).
22. Y. Freund: Boosting a weak learning algortihm by majority, Inf. Comput., vol. 121, no. 2, pp. 256–285, (1995).
23. J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd Edition, Morgan Kaufmann, (2011).
24. Murphy, K. Machine learning: a probabilistic approach. MIT Press, 1163, 1169-1175. (2013).
25. Wang, J., Xu, M., Wang, H., & Zhang, J. Classification of imbalanced data by using the SMOTE algorithm and locally linear embedding. In 2006 8th international Conference on Signal Processing (Vol. 3). IEEE. (2006, November)