Users' Responses to a Machine-Learning Decision Support Model: A Randomized Controlled Trial for Prostate-Specific Antigen Screening

Yi-Ting Lin

Department of Urology, St. Joseph Hospital

Han-Sun Chiang

Department of Urology, Fu Jen Catholic University Hospital

Chih-Kuang Liu

Department of Urology, Taipei City Hospital

Yen-Chun Huang

Fu Jen Catholic University

Mingchih Chen (✉ 081438@mail.fju.edu.tw)

Fu Jen Catholic University <https://orcid.org/0000-0002-8278-0033>

Research article

Keywords: machine learning, shared decision-making, prostate-specific antigen, random forest, support vector machine, deep neural network, multilayer perceptron neural network, extreme gradient boosting, logistic regression

Posted Date: July 29th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-38379/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Although a shared decision-making (SDM) process integrates patient values and evidence-based medicine, patients' anxiety and decision conflicts remain. Thus, we propose a new decision-making model integrating a machine-learning algorithm to investigate its feasibility for reducing anxiety, decision conflicts, and increasing satisfaction after making a decision.

Methods: We enrolled participants willing to undergo the SDM process for a prostate-specific antigen (PSA) blood test and obtained data including age, PSA knowledge, if they have a friend with prostate cancer, perceptive risk of prostate cancer, International Prostate Symptom Score and Importance for Physiological and Psychological Impact in PSA Testing scores, personal values, and their final decisions, including "Accept" PSA blood test or "Not now," to build the dataset for training the following machine-learning models: multilayer perceptron neural network, random forest (RF), extreme gradient boosting, support vector machine, and deep learning neural network. Uniform parameter tuning and model comparison were implemented. The best model was used for a randomized controlled trial (RCT), in which we measured the effects of personalized suggestions generated by the machine-learning model on anxiety, decision satisfaction, and decision conflicts.

Results: RF was the best algorithm for building models with our dataset from 507 subjects (mean AUC: 0.8801, mean ACC: 0.8313, Max ACC: 0.8933). Therefore, we used the RF model for RCT with 185 and 182 subjects in the machine-learning suggestion group (MLSG) and control group (CG), respectively. The MLSG patients were calmer, more content, and less worrisome than those in the CG. They also experienced higher decision satisfaction and less decision conflict, including more decision support, advice, assurance of decision, ease of decision-making, and adherence to decision. Moreover, participants who were suggested "Accept" by the model were more likely to make "Accept" their final decision than the CG participants (50.75% vs 24.18%, $\chi^2 = 16.07$, $p < 0.000$). The "Not now" suggestion followed a similar trend.

Conclusions: A highly accurate machine-learning model was constructed using our methods. Personalized suggestions generated from this model yielded increased satisfaction and reduced anxiety and decision conflict. Patients tended to take machine-learning suggestions as their final decision.

Trial name: Shared Decision Making: Decision Tree and Artificial Neural Network Assisted Decision Aid for PSA Screening

Trial registration: ChiCTR, ChiCTR2000034126. Registered 25 June 2020 –

Retrospectively registered, <http://www.chictr.org.cn/ChiCTR2000034126>

Background

The first case of shared decision-making (SDM) was observed in a patient-centered care project performed by Charles in the United States in 1982 [1]. It was considered the evolution of evidence-based medicine. The three Es, evidence, experience, and expectation, of evidence-based medicine is a conceptual framework consisting of the elements involved in a medical decision [2]. SDM has been widely used in numerous developed countries as a routine part of treatment for more than 20 years in numerous medical fields [3]. However, several emerging problems of SDM have also been reported. Decision conflicts, lack of confidence in decision, uncertainty, emotional distress, and complexity of the decision context disturb the standard process of SDM [4]. Various studies have demonstrated that patients would be doubtful about the decision they made even after completing the standard course of SDM [5–9]. To some extent, a decision conflict would exist despite having the entire knowledge of diseases and treatment options. Moreover, some decision-aiding materials would conversely increase the conflict between decisions available, even those made following the standard protocol [10–13]. Allen et al. demonstrated that an increasing number of patients were undecided after undergoing the SDM process [14]. Wakefield et al. reported less value congruence with chosen options in the decision aid group than those in the usual care group [15]. Therefore, a new type of decision aid might be needed.

By contrast, emotion is doubtless involved in the entire SDM process. Emotional distress which results from the disease itself, uncertainty, and medical or surgical interventions during screening or treatment influence the decision-making process [16, 17].

The emotional stress would sometimes compromise logical thinking during the SDM process, possibly leading to an unwise choice [18]. Furthermore, the complexity of considerations and numerous priorities during the decision-making process for certain diseases might become a major challenge for a patient to optimize their choice. In certain circumstances, such as a patient with newly diagnosed cancer, the patients would encounter a very complex decision context about treatment options. With no prior experience, a decision support is in high demand [19]. Similarly, Brenda et al. proposed the following two major barriers to implementing SDM: too many problems to make a decision and patients' lack of trust in their physicians [20]. Therefore, a novel method is required to be devised that can offer suggestions to patients and select the best choice based completely on experiences of peers who confronted similar complex situations and difficult decisions.

Additional challenges were also encountered during SDM in PSA screening although several studies have described the success of SDM in prostate cancer screening for prostate-specific antigen (PSA) [11, 21–24]. First, the usefulness of PSA screening is quite diverse among trials. The Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial have demonstrated that prostate cancer screening did not lower cancer-specific mortality [25, 26]. In this study, 76,693 men aged 55–74 years were randomly assigned into the screening or usual care groups [25]. No significant differences were observed in the primary outcome of prostate cancer mortality (relative risk [RR]: 1.04, confidence interval [CI]: 0.87–1.24) or overall mortality (RR: 0.98, CI: 0.95–1.00) [25]. However, a subsequent modeling study focusing on differences in the screening intensity between PLCO and ERSPC (European Randomized Study of Screening for Prostate Cancer) showed that the intensity of screening has affected the absolute reduced prostate cancer mortality [27]. In 2013, a Cochrane meta-analysis revealed that although the estimated prostate cancer-specific mortality did not reduce among men of the screening group (RR: 0.95, 95% CI: 0.86–1.07), cancer was diagnosed more often in this group (RR: 1.30, 95% CI: 1.02–1.65). The meta-analysis gathered data from five RCTs with 341,342 participants [28]. Because of these controversies, participants have to make their own decision about whether or not to undergo screening according to their personal preferences and values. Furthermore, the decision is not simple. A qualitative study demonstrated that participants need to figure out the priority by considering the following: benefit of early cancer detection, harm from biopsy, false-negative result, false-positive result, overdiagnosis, overtreatment, and emotional impact of all subsequent intervention [29]. These factors make the decision context quite complex. Phyllis Ahiagba et al. demonstrated the conflict between the selected decision and emotional distress among patients who have completed the entire standard course of SDM for PSA blood testing [16]. Therefore, decision support generated by algorithms might have a role in helping patients to make a confident decision consistent with their personal values.

Researchers have constructed machine-learning models that effectively help clinical decision-making among health care professionals [30–34], including diagnosis of diabetes mellitus and glaucoma, predicting the prognosis of patients with colorectal cancer, diagnosis of fatty liver and optimization of liver segmentation, clinical decisions for breast cancer treatment, etc. However, its availability and accessibility for non-health care professionals and evidences of the effect, safety, and consequences of applying machine-learning decisions, which are essential for promoting artificial intelligence in clinical practice, are not yet well established. Few papers have focused on machine-learning suggestions or advice for patients, who are usually not health care professionals [35], and a few others have described the effects and consequences of using decisions generated by the machine-learning algorithm [36]. Therefore, researches surveying non-health care professional users and their responses are urgently needed to aggressively promote machine-learning methods in clinical practice.

Several characteristics can potentially convert the SDM of the PSA screening blood test into very good material for developing a machine-learning mediated decision aid. First, the decision depends heavily on personal values, but not firm evidence. As mentioned earlier, the effect of PSA screening is controversial. If a firm suggestion of whether to undergo screening has already been documented in some clinical practice guidelines, personal values would have no role in making the decision, for example, a Pap smear for cervical cancer [37]. Second, PSA screening is complex in its decision context. Machine-learning models and logistic regression are quite good at dealing with cases with a large amount of variables and have been implemented successfully in several medical fields [38]. Third, the choice is binary (“Accept” to receive it or “Not now”). It is quite difficult to build a high-accuracy machine-learning model based on a dataset with more than two categories. The nature of binary classification in PSA screening is conducive to establishing a machine-learning model with higher accuracy. Fourth, enrollment of participants for PSA screening is easy. A sufficient number of cases is essential for modeling machine-learning algorithms. A large target population makes enrollment of participants easier. Consequently, we hypothesized that a decision suggestion

generated by a machine-learning algorithm is implementable and it is a solution to deal with these complex concerns in SDM of PSA screening. In this study, we used SDM of PSA blood testing as a material. We attempted to construct a highly reliable and valid tool to detect personal preferences: We gathered data for establishing machine-learning models with adequately high accuracy. To observe users' responses, we recorded how participants reacted to suggestions yielded by a machine-learning model, and whether these suggestions positively impacted emotion, satisfaction, and decision conflict. We also wanted to know how the suggestions from the machine-learning model influenced participants' final decisions.

Methods

Establishment of a preference detection tool and validation of questionnaires

We designed a questionnaire, "The Importance for Physiological and Psychological Impact" (IPPI) to understand the value and preference of participants, regarding their concerns about PSA screening based on studies by Eila et al., Lucie Rychetnik et al., and Ferrante et al. [24, 29, 39]. This 10-item questionnaire (Supplement 1; annotations A-J of Table 1) consists of two dimensions with four items evaluating the physiological impact and six items evaluating the psychological impact. It has been validated by 10 experts who are urological specialists in Taiwan with more than 7 years of experience in urological clinical practice. We demonstrated a strong reliability, with Cronbach's alpha of 0.838 and 0.9000, respectively, for the physiological and psychological impact items of the initial 300 participants. The factor analysis of IPPI was performed using the maximum likelihood method, and then the VariMax rotation was applied to extract the following two factors: personal values focusing on physiological impacts (items 1–4) personal values focusing on psychological impacts (items 5–10). We also validated the Chinese version of the Decision Conflict Scale (DCS) [40] and International Prostate Symptom Score (IPSS) [41]. Moreover, the short version of Spielberger State-Trait Anxiety Inventory (SSTI) [42] and Decision Satisfactory Questionnaire [43] were translated by two people who were fluent in both Chinese and English. The research tools are given in Supplements 1, 1a, and 2.

Participant enrollment

This study was received a full ethical review and approved by the ethics Institutional Review Board (IRB) of Fu Jen Catholic University (C105142). The Committee for Medical Research Ethics approved the study, and the requirement to obtain informed consent was waived. We recruited men who visited St. Joseph Hospital randomly from October 2017 to April 2018 by YT Lin. The participants had to meet the inclusion criteria of age ≥ 50 years and interest in information on SDM in PSA testing. The exclusion criteria were any psychological disorder, previous PSA testing, history of prostate surgery or preexisting prostate cancer. We planned to recruit 900 subjects totally. We built machine-learning model repeatedly every 20 new subjects enrolled. When we gathered a study sample of 520 participants, we found the performance of machine learning models reached the plateau. The rest 380 subjects were assigned to RCT. Thirteen subjects in 520 were excluded from the dataset because of poor data quality (i.e., 10 consecutive items with the same score or have rhythm among the scores). Consequently, we used the data of 507 participants (model establishing group, MEG) to train and verify models. We further recruited 380 participants and randomly assigned them into the machine-learning suggestion group (MLSG) and control group (CG) to evaluate the effect of suggestions from a machine-learning model (Fig. 1).

Procedure of SDM and data collection

Steps 1–7 were performed for the patients in the MEG, whereas steps 1–8 were performed for the patients in the RCT (MLSG and CG). These steps are outlined as follows.

Step 1. At the beginning of every interview, the participant was asked to provide data on age, marriage status, education level, and their knowledge of PSA, whether they themselves are or have a friend diagnosed as having prostate cancer, and perceptions of risk of prostate cancer, which were recognized as factors that influenced the willingness to undergo PSA testing in a previous study [24].

Step 2. International Prostate Symptom Score (IPSS) was then filled out by participants.

Step3. Participants then watched a video mediated decision aids detailing the impacts of PSA screening on their body and mind. The tutors offered additional explanations to answer questions from the participants, and discussed with participants about which impact of the screening was important according to participants' personal values.

Step 4. A three-item brief test was performed to evaluate if participants could retain the core information about the decision aid.

Step 5. Participants filled out the IPPI questionnaire and ranked the impacts by importance according to their values.

Step 6. We keyed in the scores of all previous questionnaires to the machine-learning interface based on the best model established with 507 participants in the MEG during the RCT. The interface would then return the predictive choice of each participant. Returns of "control group, no suggestions offered" and "Accept and do it immediately" or "under consideration or refuse" meant that the participant was assigned to the CG and MLSG, respectively. The assignment was generated by R platform based on random number table with the ratio of allocation to MLSG and CG as 1:1.

Step 7. Let participants make their own decisions.

Step 8. Participants filled out the self-report questionnaires, including SSTI, Decision Satisfaction Questionnaire, and DCS in the RCT without any interference.

It took 20–25 minutes to complete the entire procedure for one participant.

The datasets generated during the current study are available in the supplement files with the file name "raw data of MEG" and "raw data of MLSG vs CG"

Method of establishing a machine-learning model

Variables for modeling and data preprocessing

The models were developed using features of subjects, including age; education; marriage status; knowledge about PSA; whether or not they themselves have or have a friend with prostate cancer; IPSS evaluation; IPPI and the first, second, and third concerns of IPPI items; and their decision about PSA testing as dependent variables. Logistic regression was used for exploring the significance of every variable. Then, we used five machine-learning methods to build the models, including the multilayer perceptron neural network (MLP), random forest (RF), extreme gradient boosting (XGB), support vector machine (SVM), and deep learning neural network (DNN). We used the MEG dataset to train and verify the models. For MLP, DNN, and SVM, we standardized the continuous variables and created dummy variables for categorical variables; these steps were not necessary for RF and XGB.

Data splitting and parameter tuning

The sample size of 507 was relatively small in our model establishing dataset. Furthermore, we used the bootstrap method to eliminate the bias during data splitting [44, 45]. For each machine-learning model, 507 bootstrapping iterations would be performed to form 507 pairs of training and test sets. We then calculated the accuracy and area under the ROC curve (AUC) for every pair of training and test sets (Supplement 3a).

We used the mean AUC of the 507 pairs of training sets as an indicator for parameter tuning. Antlion optimizer (ALO) was used to find the best parameter with the maximal mean AUC for every model. ALO has been used in parameter tuning in artificial neural networks and SVMs in previous studies, (46) and has been proved to have better performance than particle swarm optimizer and ant colony optimizer (47). The entire algorithm was well described in Seyedali's work. (48) We also used the design of experiment method to define and narrow the range of searching space for ALO (Supplement 3b). The parameters yielding the best mean AUC are listed in Supplement 4.

Comparison among models and building a website-based user interface for RCT

After finding the best parameters for these five machine-learning algorithms, we compared the ACCs and AUCs of 507 models based on 507 bootstrapped training sets among these five algorithms. We used the nonparametric Kruskal–Wallis test for pairwise comparisons among algorithms. The entire process of model comparison was documented in the study of Hui et al. [45]. The algorithms with the best performance were used to establish a website-based interface. We trained and verified 2000 models using 2000 pairs of bootstrapped training and test sets, applying the best algorithm with the best parameters. The one with the best AUC among the 2000 models was used as the classifier and uploaded onto the Shiny server. The user interface was also constructed and uploaded onto the Shiny server to establish a decision-predictive website (<http://psachoice.shinyapps.io/psapsa/>) [49].

Randomized control trial

In total, 380 participants were randomized into the MLSG and CG. The MLSG received the decision suggestion generated by the decision-predictive website. The tutor would explain the meanings of the prediction/suggestions to the participant. Then, the participant could take the suggestions into consideration to make their own decision about undergoing the PSA blood test. The CG received no prediction/suggestion. After they made their final decision, participants of both groups were asked to fill the self-report questionnaires, including SSTI, Decision Satisfactory Questionnaire, and DCS, without any interference.

Software

R studio (version 3.4.2) was used as the platform to implement the machine learning and ALO. Several R packages were used and are listed in Supplement 5. SPSS (version 21) was used for the nonparametric statistics and chi-square test.

Results

We compared the features of participants who chose “Accept” and “Not now” as their final decision in MEG. We found that participants who chose “Accept” had higher scores than those who chose “Not now” for KnowPSA (2.23 ± 0.81 vs 2.07 ± 0.69 , $p = 0.002$), IPSS1 (1.31 ± 1.48 vs 0.70 ± 1.06 , $p < 0.000$), IPSS2 (1.33 ± 1.42 vs 0.74 ± 1.09 , $p < 0.000$), IPSS3 (1.20 ± 1.53 vs 0.64 ± 1.10 , $p = 0.008$), IPSS4 (1.15 ± 1.46 vs 0.66 ± 1.08 , $p = 0.025$), IPSS5 (0.87 ± 1.26 vs 0.42 ± 0.89 , $p = 0.003$), and IPSS7 (2.02 ± 1.29 vs 1.60 ± 1.10 , $p = 0.045$). This finding suggests that participants with higher IPSS score tended to choose “Accept” for PSA screening as their decision. Regarding IPPI items, participants who chose “Accept” scored high in A (3.93 ± 1.07 vs 3.20 ± 1.29 , $p < 0.000$), B (3.60 ± 1.26 vs 3.26 ± 1.37 , $p < 0.000$), C (3.88 ± 1.15 vs 3.60 ± 1.17 , $p = 0.006$), D (3.42 ± 1.35 vs 3.05 ± 1.59 , $p = 0.001$), E (4.43 ± 0.85 vs 3.31 ± 1.26 , $p < 0.000$), F (4.16 ± 1.00 vs 2.96 ± 1.49 , $p < 0.000$), H (2.94 ± 1.26 vs 2.84 ± 1.67 , $p = 0.004$), I (3.71 ± 1.25 vs 3.47 ± 1.44 , $p = 0.005$), J (3.43 ± 1.29 vs 3.01 ± 1.63 , $p < 0.000$). This notably demonstrated that participants who care about physiological and psychological impacts more were more likely to choose “Accept”. By contrast, subjects who did not care about the positive and negative impact of PSA screening were less likely to receive PSA screening. Age, RiskUThink, IPSS6, and item G of IPPI were similar for both groups (Table 1a).

Participants who chose “Not now” were more likely to be widowed or have a low education level. Important priorities were significantly different between participants who chose “Accept” and those who chose “Not now” (Table 1b). Logistic regression showed that KnowPSA; IPSS7; and IPPI A, D, F, G, J, first and second concerns were statistically significant predictors of the final decision. KnowPSA (odds ratio [OR]: 0.539, CI: 0.336–0.865, $p = 0.010$), IPSS7 (OR: 0.612, CI: 0.432–0.868, $p = 0.006$), A (OR: 0.663, CI: 0.474–0.928, $p = 0.017$), D (OR: 0.623, CI: 0.434–0.895, $p = 0.010$), F (OR: 0.532, CI: 0.350–0.807, $p = 0.003$), J (OR: 0.686, CI: 0.473–0.996, $p = 0.048$) were negative predictors for the “Not now” decision, whereas G (OR: 1.452, CI: 1.028–2.049, $p = 0.034$) was the positive predictor for the same. Regarding the first concern, the answers “A”, “B”, “D”, and “J” were positively associated with “Not now” with “I” as the reference category. For the second concern, the answer “A” was positively associated with “Not now” with “C” as the reference category (Table 1a, 1b).

Table 1

a. Comparison of features of participants in the MEG with their final decisions of "Accept" and "Not now"; continuous and ordinal variables

Features		Model Establish Group (N = 507)							
		Univariate analysis				Multivariate analysis			
		"Accept" (N = 130)	"Not Now" (N = 377)	All (N = 507)		Logistic regression ("Accept":0;"Not now":1)			
		Mean $\pm$ SD	Mean $\pm$ SD	Mean $\pm$ SD	P value	Coefficient	SE	Odds Ratio (95% CI)	P value
Age(y/o)		63.63 $\pm$ 9.41	62.63 $\pm$ 9.80	62.89 $\pm$ 9.70	0.335	0.018	0.021	1.018 (0.977, 1.061)	0.397
KnowPSA ¹		2.23 $\pm$ 0.81	2.07 $\pm$ 0.69	2.11 $\pm$ 0.73	0.002*	-0.619	0.241	0.539 (0.336, 0.865)	0.010*
RiskUThink ²		2.61 $\pm$ 1.12	2.47 $\pm$ 1.02	2.50 $\pm$ 1.05	0.897	-0.149	0.170	0.861(0.617,1.202)	0.380
IPSS	IPSS 1	1.31 $\pm$ 1.48	0.70 $\pm$ 1.06	0.86 $\pm$ 1.21	0.000*	-0.117	0.193	0.890(0.610,1.299)	0.545
	IPSS 2	1.33 $\pm$ 1.42	0.74 $\pm$ 1.09	0.89 $\pm$ 1.21	0.000*	0.003	0.203	1.003(0.674,1.495)	0.986
	IPSS 3	1.20 $\pm$ 1.53	0.64 $\pm$ 1.10	0.78 $\pm$ 1.25	0.008*	-0.258	0.214	0.772(0.507,1.175)	0.228
	IPSS 4	1.15 $\pm$ 1.46	0.66 $\pm$ 1.08	0.78 $\pm$ 1.21	0.025*	-0.156	0.224	0.856(0.551,1.328)	0.487
	IPSS 5	0.87 $\pm$ 1.26	0.42 $\pm$ 0.89	0.53 $\pm$ 1.02	0.003*	-0.232	0.237	0.793(0.499,1.262)	0.328
	IPSS 6	0.81 $\pm$ 1.28	0.58 $\pm$ 1.02	0.64 $\pm$ 1.10	0.518	0.372	0.210	1.451(0.961,2.190)	0.077
	IPSS 7	2.02 $\pm$ 1.29	1.60 $\pm$ 1.10	1.71 $\pm$ 1.16	0.045*	-0.490	0.178	0.612(0.432,0.868)	0.006*
	IPSS Q ³	4.85 $\pm$ 1.45	5.21 $\pm$ 1.28	5.12 $\pm$ 1.33	0.169	-0.088	0.220	0.916(0.595,1.409)	0.690
IPPI	A	3.93 $\pm$ 1.07	3.20 $\pm$ 1.29	3.39 $\pm$ 1.27	0.000*	-0.411	0.172	0.663(0.474,0.928)	0.017*
	B	3.60 $\pm$ 1.26	3.26 $\pm$ 1.37	3.35 $\pm$ 1.35	0.000*	-0.105	0.180	0.901(0.633,1.282)	0.562
	C	3.88 $\pm$ 1.15	3.60 $\pm$ 1.17	3.67 $\pm$ 1.17	0.006*	0.339	0.218	1.403(0.915,2.153)	0.121
	D	3.42 $\pm$ 1.35	3.05 $\pm$ 1.59	3.14 $\pm$ 1.54	0.001*	-0.473	0.185	0.623(0.434,0.895)	0.010*
	E	4.43 $\pm$ 0.85	3.31 $\pm$ 1.26	3.60 $\pm$ 1.26	0.000*	-0.441	0.246	0.644(0.397,1.043)	0.073
	F	4.16 $\pm$ 1.00	2.96 $\pm$ 1.49	3.27 $\pm$ 1.48	0.000*	-0.632	0.213	0.532(0.3500,0.807)	0.003*
	G	2.29 $\pm$ 1.37	2.25 $\pm$ 1.62	2.26 $\pm$ 1.56	0.341	0.373	0.176	1.452(1.0282,0.049)	0.034*
	H	2.94 $\pm$ 1.26	2.84 $\pm$ 1.67	2.86 $\pm$ 1.57	0.004*	0.312	0.207	1.366(0.9112,0.048)	0.132

Features	Model Establish Group (N = 507)							
	Univariate analysis				Multivariate analysis			
	"Accept" (N = 130)	"Not Now" (N = 377)	All (N = 507)		Logistic regression ("Accept":0;"Not now":1)			
	Mean ± SD	Mean ± SD	Mean ± SD	P value	Coefficient	SE	Odds Ratio (95% CI)	P value
I	3.71 ± 1.25	3.47 ± 1.44	3.53 ± 1.40	0.005*	0.075	0.190	1.078(0.7431.564)	0.692
J	3.43 ± 1.29	3.01 ± 1.63	3.12 ± 1.56	0.000*	-0.377	0.190	0.686(0.4730.996)	0.048*

Table 1

b. Comparison of features in MEG subjects between "Accept" and "Not now"; categorical variables

Features (reference)	Model Establish Group (N = 507)							
	Univariate analysis				Multivariate analysis			
	Final decision		χ^2	P value	Logistic regression ("Accept":0,"Not now":1)			
	"Accept" (N = 130)"	Not Now" (N = 377)			Coefficient	SE	Odds Ratio (95% CI)	P value
Marriage ⁴ (Married)				0.000				0.145
	Divorce	5(3.8)	0(0)		-24.02	15358	0.00 (0.00)	0.999
	Single	9(6.9)	9(2.4)		-1.508	0.747	0.221 (0.051,0.957)	0.043*
	Widow	6(4.6)	31(8.2)		0.834	0.698	2.304 (0.587,9.045)	0.232
Education ⁵ (> 12 years)				0.079				0.674
	<=9 years	61(46.9)	195(51.7)		0.130	0.423	1.139 (0.497,2.608)	0.759
	> 9, <=12 years	34(26.2)	115(30.5)		0.375	0.433	1.455 (0.623,3.399)	0.387
PcaFriend ⁶ (Yes)	No	34(26.2)	69(18.3)	0.038	-0.005	0.423	0.995 (0.434,2.280)	0.990
The 1st concern(I) ⁷ Omit insignificant	I	25(19.2)	79(21.0)	0.000				0.003*
	A	32(24.6)	56(14.9)		2.133	0.664	8.442 (2.298,31.017)	0.001*
	B	10(7.7)	75(19.9)		1.641	0.716	5.161 (1.267,21.017)	0.022*
	D	3(2.3)	39(10.3)		3.779	1.026	43.765(5.856,327.091)	0.000*
	J	4(3.1)	25(6.6)		2.164	0.918	8.704 (1.440,52.613)	0.018*
The 2nd concern(C) ⁸ Omit insignificant	C	18(13.8)	69(18.3)	0.000				0.005*
	A	5(3.8)	24(6.4)		2.394	0.991	10.958(1.571,76.422)	0.016*
The 3rd concern(C) ⁹ Omit insignificant	C	28(21.5)	84(22.2)	0.000				0.227
	J	9(6.9)	56(14.9)		1.797	0.798	6.029(1.261,28.831)	0.024*
*: statistically significant, Mann–Whitney U Test for continuous and ordinal variables; χ^2 test for categorical variables.								
1. KnowPSA: item that measures previous knowledge about PSA (Score 1: Not heard, Score 2: Little, Score 3: Much).								

2. RiskUTthink: item that measures the degree of perception of risk for prostate cancer. How much risk of prostate cancer do you have? (1: Rare, 2: Little, 3: Same as others, 4: Much, 5: Very likely).
3. IPSSQ: the quality of life item in the IPSS.
4. Marriage status includes D: divorced, M: married, S: single, and W: widow.
5. Education status was divided into J: diploma less than or equal to junior high school graduation, S: diploma above junior high school, but less than or equal to senior high school graduation, U: diploma above senior high school graduation.
6. PcaFriend: Do you have any friend or relative who has been diagnosed with prostate cancer? Y: yes, N: no.

6. 7. First: From items A to J in the IPPI, which one is the most important factor that influences your decision?
7. 8. Second: From items A to J in the IPPI, which one is the second most important factor that influences your decision?
8. 9. Third: From items A to J in IPPI, which one is the third most important factor that influences your decision?

A: Physiological impact: life prolongation resulting from the PSA blood test.

B: Physiological impact: side effects of unnecessary repeated biopsy resulting from false-positive PSA blood tests.

C: Physiological impact: chance of loss of survival time resulting from false-negative blood tests.

D: Physiological impact: side effects of receiving unnecessary definite treatment for insignificant prostate cancer discovered by a PSA blood test.

E: Psychological impact: being satisfied by PSA test in knowing my health conditions.

F: Psychological impact: Feeling easy resulting from normal results of PSA test.

G: Psychological impact: being tense before PSA blood test.

H: Psychological impact: being anxious after getting abnormal PSA test results.

I: Psychological impact: the psychological impact of being misdiagnosed as normal by PSA test.

J: Psychological impact: being severely anxious after knowing the diagnosis of prostate cancer discovered by PSA test.

The accuracy and AUC of models constructed using the MEG dataset were calculated to find the model with the best performance. Initially, we performed a logistic regression using the same unbiased data-splitting method [45]. We obtained the mean accuracy (0.8140), the highest accuracy among LR models (0.8763), the mean AUC (0.7947), and the highest AUC among LR models (0.8939). Obviously, the DoE-ALO parameter tuning method is not suitable in logistic regression. In terms of machine-learning models, we found the best parameters for all five machine-learning algorithms (Supplement 4). We observed the DNN and RF models to have the highest mean accuracy (0.8429, 0.8313) after parameter tuning. The pairwise comparison showed no significant differences in the accuracy between DNN and MLP models. Moreover, RF models have the highest mean AUC (0.8801) after parameter tuning. Because our MEG dataset is relatively imbalanced, the AUC would be better than accuracy as a performance measurement according to a study published by Charles et al. [50]. Accordingly, we chose the model with the best mean AUC, that is the RF model, as the decision-suggesting tool in our study (Fig. 2). The accuracy of the best model with the best parameters among the models constructed using 2000 iterations of bootstrapping is 0.9000. Thus, the RF model was used to build the user interface for the RCT.

We randomized participants into the MLSG and CG. In total, 380 participants accomplished all steps of the experiment. Five of the MLSG and eight of the CG were dropped because of poor answer quality which we mentioned earlier in the Methods section. There was no important harms or unintended effects in each group. The participants of both groups showed similarity in age, KnowPSA, RiskUThink, IPSS1-3, IPS5-7, IPSS Q, and all items of IPPI. The participants in the MLSG had significantly higher IPSS4 scores than those in the CG. They also scored higher than the participants of the CG in A (3.26 ± 1.466 vs 3.04 ± 1.221), H (2.28 ± 1.933 vs 1.92 ± 1.757), I (2.86 ± 1.641 vs 2.58 ± 1.446) although it did not reach the significance level. The participants of MLSG and CG also showed similarity in marriage status, education level, PcaFriend, and priorities of the importance of impact items (Supplement 5).

Regarding SSTI items, we found that participants in the MLSG were calmer (SSTI1: 2.28 ± 1.210 vs 1.98 ± 1.142 , $p = 0.004$), more content (SSTI5: 2.12 ± 1.219 vs 1.87 ± 1.108 , $p = 0.031$), and less worrisome (SSTI6: 2.00 ± 1.022 vs 2.98 ± 1.166 , $p < 0.000$) than those in the in CG. They also experienced higher satisfaction than those in the CG toward the decision-making process, including more adequately informed (Sa1: 1.75 ± 0.928 vs 3.21 ± 1.560 , $p < 0.000$), assurance that the decision is the best one (Sa2: 1.83 ± 0.886 vs 3.29 ± 1.440 , $p < 0.000$), consistency with personal values (Sa3: 1.77 ± 0.894 vs 3.19 ± 1.605 , $p < 0.000$), willing to carry out the decision (Sa4: 1.67 ± 0.824 vs 3.13 ± 1.698 , $p < 0.000$), and satisfaction (Sap: 1.57 ± 0.818 vs 3.12 ± 1.674 , $p < 0.000$). The

DCS is a five-scale questionnaire with a reverse scoring system, that is, strongly agree: score 0 and strongly disagree: score 5. We found that the participants in the MLSG perceived that they had more decision support (DCS7: 1.85 ± 1.052 vs 2.19 ± 1.244 , $p = 0.012$), decision advice (DCS9: 1.73 ± 0.951 vs 2.30 ± 1.175 , $p < 0.000$), assurance of the decision (DCS10: 1.88 ± 0.993 vs 2.26 ± 1.264 , $p = 0.008$; DCS11: 1.81 ± 0.981 vs 2.38 ± 1.228 , $p < 0.000$), ease of decision-making (DCS12: 1.81 ± 0.939 vs 2.43 ± 1.232 , $p < 0.000$), adherence to the decision (DCS15: 1.78 ± 0.872 vs 2.08 ± 1.168 , $p = 0.040$), and more satisfaction (DCS16: 1.60 ± 0.739 vs 1.98 ± 1.092 , $p = 0.004$; Table 2).

Table 2
Comparison of SSTI, satisfaction, and DCS between MLSG and CG

Features		MLSG vs CG (N = 367)						p value
		MLSG		"CG"		All		
		(N = 185)		(N = 182)		(N = 367)		
		Mean	SD	Mean	SD	Mean	SD	
Anxiety	SSTI1	2.28	1.210	1.98	1.142	2.13	1.185	0.004*
	SSTI2	2.22	0.987	2.08	0.901	2.15	0.947	0.146
	SSTI3	2.21	1.090	2.36	1.051	2.29	1.072	0.176
	SSTI4	2.15	1.197	1.97	1.132	2.06	1.167	0.069
	SSTI5	2.12	1.219	1.87	1.108	2.00	1.170	0.031*
	SSTI6	2.00	1.022	2.98	1.166	2.49	1.198	0.000*
Decision	Sa1	1.75	0.928	3.21	1.560	2.48	1.474	0.000*
Satisfactory	Sa2	1.83	0.886	3.29	1.440	2.55	1.399	0.000*
	Sa3	1.77	0.894	3.19	1.605	2.47	1.478	0.000*
	Sa4	1.67	0.824	3.13	1.698	2.39	1.516	0.000*
	Sa5	1.57	0.818	3.12	1.674	2.34	1.524	0.000*
Decision Conflicts	DCS1	2.09	0.965	2.29	1.115	2.19	1.045	0.143
	DCS2	2.23	1.028	2.18	1.157	2.20	1.093	0.321
	DCS3	2.24	1.097	2.19	1.098	2.22	1.096	0.658
	DCS4	2.23	1.095	2.16	1.187	2.20	1.140	0.377
	DCS5	2.17	1.088	2.33	1.244	2.25	1.169	0.329
	DCS6	2.29	1.089	2.42	1.254	2.36	1.174	0.438
	DCS7	1.85	1.052	2.19	1.244	2.02	1.163	0.012*
	DCS8	2.05	1.178	1.91	1.060	1.98	1.122	0.311
	DCS9	1.73	0.951	2.30	1.175	2.01	1.104	0.000*
	DCS10	1.88	0.993	2.26	1.264	2.07	1.150	0.008*
	DCS11	1.81	0.981	2.38	1.228	2.09	1.146	0.000*
	DCS12	1.81	0.939	2.43	1.232	2.12	1.136	0.000*
	DCS13	1.89	1.053	2.12	1.178	2.01	1.121	0.076
	DCS14	1.96	0.952	2.15	1.061	2.06	1.011	0.108
	DCS15	1.78	0.872	2.08	1.168	1.93	1.039	0.040*
	DCS16	1.60	0.739	1.98	1.092	1.79	0.949	0.004*
*: statistically significant, Mann–Whitney U Test was used.								

We also wondered about the influence of machine-learning suggestions on the final decisions of participants. We observed that 24.18% of participants in the CG chose “Accept”, whereas 75.82% chose “Not now” as their final decision. Participants who were

suggested to “Accept” by the machine-learning model had a higher chance of making “Accept” their final decision than those in CG deciding on “Accept” (50.75% vs 24.18%, $\chi^2 = 16.07$, $p < 0.000$). Similarly, participants who were suggested “Not now” tended to be more likely to make it their final decision, even though it was statistically insignificant (75.82% vs 82.20%, $\chi^2 = 1.72$, $p = 0.190$; Table 3).

Table 3
Effects of machine-learning suggestions on final decision

		Final Decision			
		“Accept”(%)	“Not now”(%)	Sum	Chi-square, p-value,
MLSG	Suggest “Accept”	34(50.75)	33(49.25)	67	$\chi^2 = 16.07$, $p < 0.000^*$
CG	No suggestion	44(24.18)	138(75.82)	182	
Sum		78	171	249	

3a. Comparison of final decision between participants who got suggestion of “Accept” and those who got no suggestions in the CG

		Final Decision			
		“Accept”(%)	“Not now”(%)	Sum	Chi-square, p-value,
MLSG	Suggest “Not now”	21(17.80)	97(82.20)	118	$\chi^2 = 16.07$, $p < 0.000^*$
CG	No suggestion	44(24.18)	138(75.82)	182	
Sum		65	235	300	
3b. Comparison of final decision between subjects got suggestion of “Not now” and those who got no suggestions in the CG					

Discussion

Our study demonstrated that scores of IPSS items are significantly different between participants in the MEG with their final decision as “Accept” and “Not now.” This indicates that IPSS items could serve as input data for building machine-learning models. Jeanne et al. performed a study to explore factors that influence men’s decision to undergo a prostate cancer screening [39] and found that men were prompted to undergo PSA screening by urinary symptoms. This is consistent with our findings of the univariate analysis on IPSS items.

IPPI is a questionnaire which we designed for detecting participants’ attitudes toward the psychological and physiological impact after PSA screening. We found the more serious the attitude of participants toward psychological and physiological impact, the more they were likely to accept PSA screening. Another prior study demonstrated that a perception of high risk is an important factor for undergoing PSA screening [39]. Our findings might result from our decision aid elevating the risk perception in some subjects and make them willing to undergo PSA screening. Simultaneously, these participants also tended to take these impacts seriously. In logistic regression, we observed items A, D, F, G, and J to be predictors of final decisions, as well as the first and second concerns. It is difficult to find a quantitative studies which measured the effect of these psychological and physiological impacts on decision-making. The causal relationships between these concerns and the final decision are still ambiguous. Further studies are needed to clarify these interactions. In factor analysis, items of physiological and psychological impact were grouped separately, instead of comparing positive versus negative impacts. It means participants tended to assign similar importance levels for items in either the physiological or psychological dimensions. Therefore, we further compared the scores of these two dimensions in the “Accept” group. Participants of the “Accept” group tended to score the physiological impacts higher than psychological impacts (3.71 ± 1.38 vs 3.49 ± 1.22 , $p = 0.004$). The participants of the “Not now” group also tended to score the physiological impact higher than psychological impact (3.27 ± 1.38 vs 2.97 ± 1.57 , $p = 0.000$). The univariate analysis of our study also observed education level to be a predictor of the final decision, even if the same trend was not showed

in logistic regression. Participants who had heard more about PSA previously were more likely to make “Accept” their final decision in both univariate and multivariate analyses of our study. This is consistent with the findings of Eila et al. In their study, participants who knew someone who had undergone a PSA test or have already discussed with doctors about the PSA test are more intent to undergo a PSA test [24].

Participants with a higher education level were more likely to choose “Accept” as their final decision. Mehdi et al. reported the positive association between presence of prostate cancer early detection behavior and high education level in Iran [51]. However, the study by RE Myers et al. demonstrated contradicting results in African-American men. The higher their education level, the less intention they had to undergo PSA screening [52]. However, Viet-Thi Tran found decision aid changed decisions favoring PSA screening into disfavoring in high education-level subject in France [53]. The association between education level and intention of prostate cancer screening is not consistent among cultures. The test intention we detected in the MEG, MLSG and CG were 25.64% (130/507), 29.73% (55/185), and 24.18% (44/182), respectively. In a study by Robin WM Vernooij, the association between intention and reduction of mortality was explored. In total, 37% and 44% of the participants intended to undergo prostate cancer screening with different levels of reduction in mortality [54]. An early study reported that the intention to undergo annual prostate cancer screening among African-American men in Philadelphia was 68% [55]. DL Frosch et al. compared decision aid methods and found significant differences in the number of men requesting a PSA test, with the highest rate in the usual care group (97.7%), followed by the discussion (82.2%), video (60.0%), and video plus discussion (50.0%) groups [56]. In conclusion, the intention of screening is quite different among researches and depended on cultures and decision aid methods.

With the dawn of the machine-learning era, few researches are focusing on applying machine-learning in the SDM process in the medical field. In the case regarding modeling and predicting participants’ intentions, we found Eila Watson’s research pioneering. Eila Watson et al. modeled features that influenced the intention to be tested and used univariate and multivariate logistic regression to recognize the features, which significantly affected the intention to be tested. According to them, the most important factors were perceived risk, perceived benefits of the PSA test, attitude toward the PSA test, knowledge, and age group (24). However, their models did not focus on prediction and suggestion. In our study, we successfully constructed a highly accurate model to provide suggestions for the subsequent RCT. To ensure the quality of suggestions, we used unbiased data-splitting, model validation, and model performance comparison methods which were extracted from Chen’s work about differentiating lung nodules (performance comparison of artificial neural network and logistic regression model for differentiating lung nodules in computed tomography) [45]. We used the uniform process ALO to tune parameters for every algorithm, instead of using different processes for each distinct algorithm as done conventionally. Because of its highest mean AUC (0.8801) with maximum AUC (0.9329) and high mean ACC(0.8313), we chose the RF model. With logistic regression models, mean ACC, mean AUC, and maximum AUC came to 0.8140, 0.7947, and 0.8939, respectively. Thus, our results clearly demonstrated that machine-learning algorithms outperformed logistic regression, even though a systemic review could not show the superiority of machine learning over logistic regression (38). More recently, Glenn Salkeld et al. applied a personalized decision support tool developed for prostate cancer screening using a software platform known as Annalisa—an interactive decision aid template based on multicriteria decision analysis (57). The personalized suggestions yielded by the decision aid were found to be of a slightly high quality. However, they did not survey the participants’ response comprehensively. To clarify these changes in intention and psychological variables after computer-generated suggestions, we obtained the anxiety score, satisfaction scale, DCS, and post-suggested decision changes.

The task of our study was to discover the effects of machine-learning suggestions on our participants. Without a doubt, most decision aids have positive effects on decision-making process. Nahara et al. performed a meta-analysis and reviewed four RCTs about SDM and prostate cancer screening (58). Their results demonstrated a reduction in decision conflict with the decision aids. Andrew W Stamm et al. found that participants in the decision aid (DA) + SDM arm were significantly more likely to report that they always felt encouraged to discuss all health concerns (78% DA + SDM vs 72% DA $p = 0.0285$) [59]. Heidi et al. reported the percentage of men with “high anxiety” decreased from 12–7% and decision conflict also decreased with the use of decision aids. In total, 85% of men experienced more ease in making decision [60]. Warlick et al. demonstrated participants reported high decision satisfaction and low decisional conflict [61]. In our research, in addition to the basic decision aid with papers and videos and SDM guided by tutors, we gave highly personalized machine-learning suggestions to the participant, which were generated

from their own features and values. Therefore, participants who received machine-learning suggestions were calmer, more content, and less worrisome than those who did not receive suggestions. In fact, the participants in our study showed more satisfied after receiving machine-learning suggestions. In addition, they experienced more decision support, assurance of the decision, ease in decision-making, and adherence to the decision compared to those who did not receive suggestions. Thus, our study proved the positive effects of machine-learning suggestions.

Traditional decision aids typically reduced the likelihood of being screened. Barr et al. reported an example of this. They recruited 1041 predominantly white, well-educated men and recorded their responses to the pre- and post-viewing questionnaires. After viewing, the proportion of patients leaning away from PSA screening increased significantly [62]. However, machine-learning suggestions have an entirely different effect on final decision. In our study, participants tended to follow the machine-learning suggestions. Participants receiving the “Accept” suggestion tended to make “Accept” their final decision. The same trend was observed in the case of those receiving the “Not now” suggestion, even though it was not statistically significant. We tried to explain this new finding using the degree of trust. Zachary Klaassen’s work demonstrated that the degree of trust an individual had in his physician for cancer information was strongly associated with the likelihood of him undergoing a PSA screening (63). It implies that the more a participant trusts his/her doctor, the more the participant’s willingness to follow the doctor’s orders. Machine learning gained recognition during the 2010s. Since then, machine-learning methods have created benefits in many aspects of our daily life. Therefore, our participants had a very positive attitude toward machine learning and trusted the suggestions received. The other explanation is the principle of authority proposed by Robert Cialdini (64). When participants regarded the machine-learning model as an authority, they would tend to follow the suggestion, even if the machine-learning suggestions were opposed to their own decision. Therefore, we still do not know whether this phenomenon is useful or harmful. Further studies aimed at elucidating the physiological and psychological safety of this phenomenon should be conducted before clinical use.

In this study, we provided a method of machine-learning mediated shared decision making in PSA blood test. It is possible to generalize the methods to other field of shared medical decision making after the safety being examined. This study has some limitations. First, our study population was relatively small for modeling. Moreover, we used simple binary classification to improve the performance of model. Second, this study was not double-blinded. A double-blinded randomized trial with the permission of the IRB is needed for eliminating tutor bias, as well as, observing the effect of reverse suggestions. Third, how machine-learning suggestions impact the psychological status of a participant remains unknown. Future qualitative or quantitative researches should be designed to focus on exploring the psychological impact of machine-learning suggestions and investigating the underlying cause of participants’ tendency to follow these machine-learning suggestions. The psychological impact of suggestions should be examined after these suggestions are executed by participants.

Conclusions

We proved that a highly accurate machine-learning model could be constructed successfully. Personalized suggestions generated from this model would additionally yield more positive effects on increasing satisfaction and decreasing anxiety and decision conflict compared with the traditional decision aid only. Our participants tended to take machine-learning suggestions as their final decision. The influence and safety of this phenomenon deserves further investigation.

Abbreviations

SDM

shared decision making

PSA

prostate specific antigen

IPPI

The Importance for Physiological and Psychological Impact in PSA Testing

SSTI

Short-form of Spielberger State-Trait Anxiety Inventory

DCS
Decision Conflict Scale
AUC
Area under curve
ACC
accuracy
RF
random forest
SVM
support vector machine
DNN
deep neural network
MLP
multilayer perceptron neural network
XGB
extreme Gradient boosting
LR
Logistic regression
ALO
antlion optimizer
DoE
Design of Experiment
MLSG
machine learning suggestions group
CG
control group

Declarations

Ethics approval and consent to participate

This study was received a full ethical review and approved by the ethics Institutional Review Board of Fu Jen Catholic University (IRB-FJU) with the project number C105142. Every subject was informed with the full information about the study and signed consent (written) which was approved by IRB-FJU.

Consent to publish

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing Interests

We have no any financial and non-financial competing interests that could directly undermine, or be perceived to undermine the objectivity, integrity and value of a publication, through a potential influence on the judgements and actions of authors with regard to objective data presentation, analysis and interpretation.

Funding

The research is supported by Funding from 2016 research funds of St. Joseph Hospital with the project number 10503.

Authors' contributions

First author: YTL

Study design, enrolling participants, machine learning model establishment, implement SDN and RCT, data collection and processing, writing paper

HSC

Study design, consultation for IPPI questionnaire, writing paper

CKL

Study design, development for IPPI questionnaire, machine learning model establishment, implement RCT, data collection and processing,

YCH

machine learning model establishment, build user interface of machine learning model, data collection and processing

Corresponding author: MC

Study design, development for IPPI questionnaire, machine learning model establishment, implement RCT, data collection and processing,

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us. We understand that the Corresponding Author is the sole contact for the Editorial process. He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs.

Acknowledgements

This work is supported by National Natural Science Foundation of China (NO.

71801126), Natural Science Foundation of Jiangsu Province (NO. BK20180412),

Aeronautical Science Foundation of China (NO. 2018ZG52080), the MOE Funding Direction Regarding the Development Plan for Universities and Colleges under the project number A0108152 and Fundamental Research Funds for the Central Universities (NO. NR2018003).

CONSORT

We confirm that this study adheres to CONSORT guidelines and have provided checklist as an additional file.

References

1. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). *Soc Sci Med.* 1997;44(5):681–92.
2. Masic I, Miokovic M, Muhamedagic B. Evidence based medicine - new approaches and challenges. *Acta Inform Med.* 2008;16(4):219–25.
3. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. *J Gen Intern Med.* 2012;27(10):1361–7.
4. Stacey D, Legare F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev.* 2017;4:CD001431.

5. Berry DL, Halpenny B, Hong F, Wolpin S, Lober WB, Russell KJ, et al. The Personal Patient Profile-Prostate decision support for men with localized prostate cancer: a multi-center randomized trial. *Urol Oncol*. 2013;31(7):1012–21.
6. Fagerlin A, Dillard AJ, Smith DM, Zikmund-Fisher BJ, Pitsch R, McClure JB, et al. Women's interest in taking tamoxifen and raloxifene for breast cancer prevention: response to a tailored decision aid. *Breast Cancer Res Treat*. 2011;127(3):681–8.
7. Sawka AM, Straus S, Gafni A, Brierley JD, Tsang RW, Rotstein L, et al. How can we meet the information needs of patients with early stage papillary thyroid cancer considering radioactive iodine remnant ablation? *Clin Endocrinol (Oxf)*. 2011;74(4):419–23.
8. Weymiller AJ, Montori VM, Jones LA, Gafni A, Guyatt GH, Bryant SC, et al. Helping patients with type 2 diabetes mellitus make treatment decisions: statin choice randomized trial. *Arch Intern Med*. 2007;167(10):1076–82.
9. Goel V, Sawka CA, Thiel EC, Gort EH, O'Connor AM. Randomized trial of a patient decision aid for choice of surgical treatment for breast cancer. *Med Decis Making*. 2001;21(1):1–6.
10. Morgan MW, Deber RB, Llewellyn-Thomas HA, Gladstone P, Cusimano RJ, O'Rourke K, et al. Randomized, controlled trial of an interactive videodisc decision aid for patients with ischemic heart disease. *J Gen Intern Med*. 2000;15(10):685–93.
11. Gattellari M, Ward JE. A community-based randomised controlled trial of three different educational resources for men about prostate cancer screening. *Patient Educ Couns*. 2005;57(2):168–82.
12. McAlister FA, Man-Son-Hing M, Straus SE, Ghali WA, Anderson D, Majumdar SR, et al. Impact of a patient decision aid on care among patients with nonvalvular atrial fibrillation: a cluster randomized trial. *CMAJ*. 2005;173(5):496–501.
13. Nagle C, Gunn J, Bell R, Lewis S, Meiser B, Metcalfe S, et al. Use of a decision aid for prenatal testing of fetal abnormalities to improve women's informed decision making: a cluster randomised controlled trial [ISRCTN22532458]. *BJOG*. 2008;115(3):339–47.
14. Allen JD, Othus MK, Hart A Jr, Tom L, Li Y, Berry D, et al. A randomized trial of a computer-tailored decision aid to improve prostate cancer screening decisions: results from the Take the Wheel trial. *Cancer Epidemiol Biomarkers Prev*. 2010;19(9):2172–86.
15. Wakefield CE, Meiser B, Homewood J, Ward R, O'Donnell S, Kirk J, et al. Randomized trial of a decision aid for individuals considering genetic testing for hereditary nonpolyposis colorectal cancer risk. *Cancer*. 2008;113(5):956–65.
16. Ahiagba P, Alexis O, Worsley AJ. Factors influencing black men and their partners' knowledge of prostate cancer screening: a literature review. *Br J Nurs*. 2017;26(18):14–21.
17. Chad-Friedman E, Coleman S, Traeger LN, Pirl WF, Goldman R, Atlas SJ, et al. Psychological distress associated with cancer screening: A systematic review. *Cancer*. 2017;123(20):3882–94.
18. Legare F, Thompson-Leduc P. Twelve myths about shared decision making. *Patient Educ Couns*. 2014;96(3):281–6.
19. Katz SJ, Belkora J, Elwyn G. Shared decision making for treatment of cancer: challenges and opportunities. *J Oncol Pract*. 2014;10(3):206–8.
20. Hernandez BY, Wilkens LR, Thompson PJ, Shvetsov YB, Goodman MT, Ning L, et al. Acceptability of prophylactic human papillomavirus vaccination among adult men. *Hum Vaccin*. 2010;6(6):467–75.
21. Blackwelder R, Chessman A. Prostate Cancer Screening: Shared Decision-Making for Screening and Treatment. *Prim Care*. 2019;46(1):149–55.
22. Fedewa SA, Gansler T, Smith R, Sauer AG, Wender R, Brawley OW, et al. Recent Patterns in Shared Decision Making for Prostate-Specific Antigen Testing in the United States. *Ann Fam Med*. 2018;16(2):139–44.
23. Frosch DL, Bhatnagar V, Tally S, Hamori CJ, Kaplan RM. Internet patient decision support: a randomized controlled trial comparing alternative approaches for men considering prostate cancer screening. *Arch Intern Med*. 2008;168(4):363–9.
24. Watson E, Hewitson P, Brett J, Bukach C, Evans R, Edwards A, et al. Informed decision making and prostate specific antigen (PSA) testing for prostate cancer: a randomised controlled trial exploring the impact of a brief patient decision aid on men's knowledge, attitudes and intention to be tested. *Patient Educ Couns*. 2006;63(3):367–79.
25. Pinsky PF, Prorok PC, Yu K, Kramer BS, Black A, Gohagan JK, et al. Extended mortality results for prostate cancer screening in the PLCO trial with median follow-up of 15 years. *Cancer*. 2017;123(4):592–9.

26. Andriole GL, Crawford ED, Grubb RL 3rd, Buys SS, Chia D, Church TR, et al. Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med*. 2009;360(13):1310–9.
27. de Koning HJ, Gulati R, Moss SM, Hugosson J, Pinsky PF, Berg CD, et al. The efficacy of prostate-specific antigen screening: Impact of key components in the ERSPC and PLCO trials. *Cancer*. 2018;124(6):1197–206.
28. Ilic D, Neuberger MM, Djulbegovic M, Dahm P. Screening for prostate cancer. *Cochrane Database Syst Rev*. 2013(1):CD004720.
29. Rychetnik L, Doust J, Thomas R, Gardiner R, Mackenzie G, Glasziou P. A Community Jury on PSA screening: what do well-informed men want the government to do about prostate cancer screening—a qualitative analysis. *BMJ Open*. 2014;4(4):e004682.
30. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2015;13:8–17.
31. Francis NK, Luther A, Salib E, Allanby L, Messenger D, Allison AS, et al. The use of artificial neural networks to predict delayed discharge and readmission in enhanced recovery following laparoscopic colorectal cancer surgery. *Tech Coloproctol*. 2015;19(7):419–28.
32. Birjandi M, Ayatollahi SM, Pourahmad S, Safarpour AR. Prediction and Diagnosis of Non-Alcoholic Fatty Liver Disease (NAFLD) and Identification of Its Associated Factors Using the Classification Tree Method. *Iran Red Crescent Med J*. 2016;18(11):e32858.
33. Ramezankhani A, Pournik O, Shahrabi J, Khalili D, Azizi F, Hadaegh F. Applying decision tree for identification of a low risk population for type 2 diabetes. *Tehran Lipid and Glucose Study. Diabetes Res Clin Pract*. 2014;105(3):391–8.
34. Lamy JB, Sekar B, Guezennec G, Bouaud J, Seroussi B. Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach. *Artif Intell Med*. 2019;94:42–53.
35. Koren G, Souroujon D, Shaul R, Bloch A, Leventhal A, Lockett J, et al. "A patient like me" - An algorithm-based program to inform patients on the likely conditions people with symptoms like theirs have. *Medicine*. 2019;98(42):e17596.
36. Cho I, Jin I. Responses of Staff Nurses to an EMR-Based Clinical Decision Support Service for Predicting Inpatient Fall Risk. *Stud Health Technol Inform*. 2019;264:1650–1.
37. Sawaya GF, Smith-McCune K, Kuppermann M. Cervical Cancer Screening: More Choices in 2019. *JAMA*. 2019;321(20):2018–9.
38. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12–22.
39. Ferrante JM, Shaw EK, Scott JG. Factors influencing men's decisions regarding prostate cancer screening: a qualitative study. *J Community Health*. 2011;36(5):839–44.
40. Lam WW, Kwok M, Liao Q, Chan M, Or A, Kwong A, et al. Psychometric assessment of the Chinese version of the decisional conflict scale in Chinese women making decision for breast cancer surgery. *Health Expect*. 2015;18(2):210–20.
41. Szeto PS. Application of the Chinese version of the International Prostate Symptom Score for the management of lower urinary tract symptoms in a primary health care setting. *Hong Kong Med J*. 2008;14(6):458–64.
42. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *Br J Clin Psychol*. 1992;31(3):301–6.
43. Holmes-Rovner M, Kroll J, Schmitt N, Rovner DR, Breer ML, Rothert ML, et al. Patient satisfaction with health care decisions: the satisfaction with decision scale. *Med Decis Making*. 1996;16(1):58–64.
44. Bradley E, Gail G. A leisurely look at the Bootstrap, the jackknife, and cross-validation. *The American Statistician*. 1983;37(1):36–48.
45. Hui C, Jing Z, Yan X, Budong C, Kuan Z. Performance comparison of artificial neural network and logistic regression model for differentiating lung nodules on CT scans. *Expert Syst Appl*. 2012;39(13):11503–9.
46. Shijie Z, Leifu G, Dongmei Y, et al. Ant Lion Optimizer with Chaotic Investigation Mechanism for Optimizing SVM Parameters[J]. *Journal of Frontiers of Computer Science Technology*. 2016;10(5):722–31.

47. Gupta E, Saxena A. Performance evaluation of an optimization based regulator in automatic generation control of interconnected power system *Journal of Engineering* 2016;4570617. doi.org/10.1155/2016/4570617.
48. Seyedali M. The Ant Lion Optimizer. *Adv Eng Softw.* 2015;83:80–98.
49. Yiting L. User interface of decision making machine. <http://psachoice.shinyapps.io/psapsa/>.
50. Ling CX, Huang J, Zhang H. AUC, a statistically consistent and more discriminating measure than accuracy. *IJCAI*. 2003: 519–524.
51. Mirzaei-Alavijeh M, Ahmadi-Jouybari T, Vaezi M, Jalilian F. Prevalence, Cognitive and Socio-Demographic Determinants of Prostate Cancer Screening. *Asian Pac J Cancer Prev.* 2018;19(4):1041–6.
52. Myers RE, Hyslop T, Jennings-Dozier K, Wolf TA, Burgh DY, Diehl JA, et al. Intention to be tested for prostate cancer risk among African-American men. *Cancer Epidemiol Biomarkers Prev.* 2000;9(12):1323–8.
53. Tran VT, Kisseleva-Romanova E, Rigal L, Falcoff H. Impact of a printed decision aid on patients' intention to undergo prostate cancer screening: a multicentre, pragmatic randomised controlled trial in primary care. *Br J Gen Pract.* 2015;65(634):e295–304.
54. Vernooij RWM, Lytvyn L, Pardo-Hernandez H, Albarqouni L, Canelo-Aybar C, Campbell K, et al. Values and preferences of men for undergoing prostate-specific antigen screening for prostate cancer: a systematic review. *BMJ Open.* 2018;8(9):e025470.
55. Myers RE, Wolf TA, McKee L, McGrory G, Burgh DY, Nelson G, et al. Factors associated with intention to undergo annual prostate cancer screening among African American men in Philadelphia. *Cancer.* 1996;78(3):471–9.
56. Frosch DL, Kaplan RM, Felitti V. The evaluation of two methods to facilitate shared decision making for men considering the prostate-specific antigen test. *J Gen Intern Med.* 2001;16(6):391–8.
57. Salkeld G, Cunich M, Dowie J, Howard K, Patel MI, Mann G, et al. The Role of Personalised Choice in Decision Support: A Randomized Controlled Trial of an Online Decision Aid for Prostate Cancer Screening. *PLoS One.* 2016;11(4):e0152999.
58. Martinez-Gonzalez NA, Neuner-Jehle S, Plate A, Rosemann T, Senn O. The effects of shared decision-making compared to usual care for prostate cancer screening decisions: a systematic review and meta-analysis. *BMC Cancer.* 2018;18(1):1015.
59. Stamm AW, Banerji JS, Wolff EM, Slee A, Akapame S, Dahl K, et al. A decision aid versus shared decision making for prostate cancer screening: results of a randomized, controlled trial. *Can J Urol.* 2017;24(4):8910–7.
60. van Vugt HA, Roobol MJ, Venderbos LD, Joosten-van Zwanenburg E, Essink-Bot ML, Steyerberg EW, et al. Informed decision making on PSA testing for the detection of prostate cancer: an evaluation of a leaflet with risk indicator. *Eur J Cancer.* 2010;46(3):669–77.
61. Warlick CA, Berge JM, Ho YY, Yeazel M. Impact of a Prostate Specific Antigen Screening Decision Aid on Clinic Function. *Urol Pract.* 2017;4(6):448–53.
62. Barry MJ, Wexler RM, Brackett CD, Sepucha KR, Simmons LH, Gerstein BS, et al. Responses to a Decision Aid on Prostate Cancer Screening in Primary Care Practices. *Am J Prev Med.* 2015;49(4):520–5.
63. Klaassen Z, Wallis CJD, Goldberg H, Chandrasekar T, Fleshner NE, Finelli A, et al. The association between physician trust and prostate-specific antigen screening: Implications for shared decision-making. *Can Urol Assoc J.* 2018.
64. Cialdini RB. (1984). *Influence: The Psychology of Persuasion*. New York, United States: Harper Business; 2006 (ISBN 006124189X).

Figures

Flowchart of Modeling and Randomized Control Trial

Figure 1

Flowchart of Modeling and Randomized Control Trial

Pairwise Comparison of AUC and ACC among Algorithms

Non-parametric Kruskal- Wallis test was used to exam the difference on AUC and ACC among groups.

- - - - No significant difference between two groups, p value >0.1
- Borderline difference between two groups, $0.05 \leq p$ value ≤ 0.1
- Statistically significant difference between two groups, p value <0.05
- Model with the best performance

Figure 2

Pairwise Comparison of AUC and ACC among Algorithms

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [rawdataofMEG.csv](#)
- [rawdataofMLSGvsCG.csv](#)
- [Supplement6ComparisonofbasicfeaturesinRCTsubjects.docx](#)
- [Supplement5Rpackagesused.docx](#)
- [Supplement4ParameterswhichyieldbestAUC.docx](#)
- [Supplement3bprocedureofmodeleestablishment.jpg](#)
- [Supplement3aparametertuningfunction.jpg](#)
- [Supplement2expertvalidation.pdf](#)
- [Supplement1IPSSIPISSTIsatisfactoryDCS.pdf](#)

- [Supplement1aenglishtranslationofstudytool.pdf](#)
- [CONSORT2010Checklist.doc](#)