

An application for classifying perceptions on my health bank in Taiwan using convolutional neural networks and web-based computerized adaptive testing

A development and usability study

Chen-Fang Hsu, MD^{a,b,c,d}, Tsair-Wei Chien, MBA^e, Yu-Hua Yan, PhD^{f,g,*}

Abstract

Background: The classification of a respondent's opinions online into positive and negative classes using a minimal number of questions is gradually changing and helps turn techniques into practices. A survey incorporating convolutional neural networks (CNNs) into web-based computerized adaptive testing (CAT) was used to collect perceptions on My Health Bank (MHB) from users in Taiwan. This study designed an online module to accurately and efficiently turn a respondent's perceptions into positive and negative classes using CNNs and web-based CAT.

Methods: In all, 640 patients, family members, and caregivers with ages ranging from 20 to 70 years who were registered MHB users were invited to complete a 3-domain, 26-item, 5-category questionnaire asking about their perceptions on MHB (PMHB26) in 2019. The CNN algorithm and k-means clustering were used for dividing respondents into 2 classes of unsatisfied and satisfied classes and building a PMHB26 predictive model to estimate parameters. Exploratory factor analysis, the Rasch model, and descriptive statistics were used to examine the demographic characteristics and PMHB26 factors that were suitable for use in CNNs and Rasch multidimensional CAT (MCAT). An application was then designed to classify MHB perceptions.

Results: We found that 3 construct factors were extracted from PMHB26. The reliability of PMHB26 for each subscale beyond 0.94 was evident based on internal consistency and stability in the data. We further found the following: the accuracy of PMHB26 with CNN yields a higher accuracy rate (0.98) with an area under the curve of 0.98 (95% confidence interval, 0.97–0.99) based on the 391 returned questionnaires; and for the efficiency, approximately one-third of the items were not necessary to answer in reducing the respondents' burdens using Rasch MCAT.

Conclusions: The PMHB26 CNN model, combined with the Rasch online MCAT, is recommended for improving the accuracy and efficiency of classifying patients' perceptions of MHB utility. An application developed for helping respondents self-assess the MHB cocreation of value can be applied to other surveys in the future.

Abbreviations: AUC = area under the curve, AVE = average variance explained, CAT = computerized adaptive testing, CCC = category characteristic curve, CNN = convolutional neural networks, CR = construct reliability, CV = convergent validity, EFA = exploratory factor analysis, KIDMAP = KID able to read the map, MCAT = multidimensional CAT, MNSQ = mean square error,

Editor: Mohammed Nader Shalaby.

Ethics approval and consent to participate and to publish: This study was approved by the institutional review board of Show Chwan Memorial Hospital (IRB no. 1061107).

Availability of data and materials: All data used in this study are available in Supplemental Digital Content files.

Funding: This study was supported by grand funds (MOST 106-2410-H-217-001-SSS, MOST 107-2410-H-217-001-SSS) from the Taiwan Technology Department in 2017 and 2018 to accomplish the research.

The authors report no conflicts of interest.

Supplemental Digital Content is available for this article.

The datasets generated during and/or analyzed during the current study are publicly available.

^a Department of Pediatrics, Chi Mei Medical Center, Tainan, Taiwan, ^b School of Medicine, College of Medicine, Chung Shan Medical University, Taichung, Taiwan, ^c School of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan, ^d School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan, ^e Department of Medical Research Department, Chi-Mei Medical Center, Tainan, Taiwan, ^f Superintendent Office, Tainan Municipal Hospital (Managed by Show Chwan Medical Care Corporation), Tainan, Taiwan, ^g Department of Hospital and Health Care Administration, Chia Nan University of Pharmacy and Science, Tainan, Taiwan.

* Correspondence: Yu-Hua Yan, Tainan Municipal Hospital (Managed by Show Chwan Medical Care Corporation), No. 670, Chungde Road, East Dist., Tainan 70173, Taiwan (e-mail: 2d0003@mail.tmh.org.tw).

Copyright © 2021 the Author(s). Published by Wolters Kluwer Health, Inc.

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and build upon the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Hsu CF, Chien TW, Yan YH. An application for classifying perceptions on my health bank in Taiwan using convolutional neural networks and web-based computerized adaptive testing: a development and usability study. *Medicine* 2021;100:52(e28457).

Received: 6 July 2021 / Received in final form: 2 December 2021 / Accepted: 9 December 2021

<http://dx.doi.org/10.1097/MD.00000000000028457>

MPRSA = matching personal response scheme to adapt for the correct classification, NHI = National Health Insurance, PMHB26 = perceptions about my health bank, ROC = receiver operating characteristic, SD = standard deviation, SE = standard errors, Sens = sensitivity, SM = self-management, Spec = specificity.

Keywords: computerized adaptive testing, convolutional neural network, my health bank, Rasch model, receiver operating characteristic curve, value cocreation

Highlights:

- We introduced the My Health Bank (MHB) used in Taiwan by observing the perception of users on the matter.
- The CNN model was used to classify clusters derived from a 26-item 5-category questionnaire asking about their perceptions of MHB (PMHB26), which is novel and innovative.
- Rasch multidimensional CAT (MCAT) applied to the CNN model can save respondents' time in answering items in a survey.

1. Introduction

As health information technology advances, numerous innovations have been applied to engage people with their own health data in self-management (SM) behaviors to improve their health outcomes,^[1] including Blue Button adopted in the United States,^[2,3] My Health Record, or Personally Controlled Electronic Health Record in Australia,^[4] and Summary Care Record under the United Kingdom's National Health Service.^[5,6] Such a system is also under development in Taiwan, called My Health Bank (MHB), by its universal, comprehensive, and single-payer program of the National Health Insurance (NHI).^[7]

1.1. Perceptions on MHB need to understand

The MHB is iconographic to cater to users who have become used to pictorial electronic devices. The new MHB version also flashes out individualized notes, reminding the user to make an appointment for the next dental scaling, physical check-up, or refill the prescription for chronic illness.^[7]

The Taiwanese government-run National Health Insurance Administration set up the My Health Bank (MHB) system in September 2014 to narrow the gap between medical service providers and patients with respect to medical information, allowing the insured to retrieve personal health data on the Internet, making medical information more transparent and convenient for both patients and health care providers^[7,8] (Supplemental Digital Content 1, <http://links.lww.com/MD/G568>). By the end of 2018, 1.03 million people registered on this platform.^[9] The basic concept behind this platform is to match the needs of dual parties and multiparty to achieve "interaction" and "exchange" of medical resources by either an actual platform or a virtual platform.^[10–13] Accordingly, the MHB system possessing the potential to become a patron saint for people's health needs a survey to classify MHB satisfaction using an appropriately scientific approach that is the first challenge we encountered in this study.

1.2. Many constructs surveyed in a study

Another major challenge we faced was multidimensional questions in a questionnaire. In the context of survey research, a construct is an abstract idea, underlying theme, or subject matter that one wishes to measure using survey questions.^[14] Some constructs are relatively simple (such as burnout, being bullied, or mental illness at the workplace^[15–17]) and can be measured using only one or a few questions, whereas other constructs are more complex (ie, a satisfaction or perception survey in this study) and may require a whole battery of questions to fully elucidate the construct to meet the research goals.

Complex constructs contain multiple dimensions that are bound together by some commonalities that compose the construct as a whole, such as a questionnaire regarding the perceptions on the MHB (named PMHB26) containing 26 five-point items. It is thus necessary to develop an application that can deal with such questionnaires involving multidimensional scales, particularly using multidimensional computer adaptive testing (MCAT)^[18,19] for efficiency (with less item length to reduce respondent burden) and accuracy (without compromising the precision of assessment).^[20]

1.3. The convolutional neural networks might be helpful

Convolutional neural networks (CNNs) have been successfully used in healthcare settings in several forms,^[14,20–27] with their greatest impact being in the field of health informatics.^[28] The CNN, a famous deep learning method, can improve the prediction accuracy (up to 7.14%).^[29] CNN architecture was described as an interleaved set of feedforward layers implementing convolutional filters followed by reduction, rectification, or pooling layers^[15–17,30,31] and performed in R-language,^[32] Python-based utility software,^[33] MATLAB, Octave, Fortran, C#, Java# to write costs fast, or in C or C++ with Juann-Esteban to write fast codes.^[34]

1.4. CNN and MCAT used in MS Excel

A few have used CNNs on a small number of cases (eg, <1000) for supervised training in MS Excel, imitating 2-dimensional images with pixels to predict classification numbers from 2 to 4 using survey data.^[15–17] Missing data in the questionnaires are a major problem and another challenge that should be overcome. That is, all questions must be answered by all respondents.

Although applications were designed for users to assess their health status, including burnout, being bullied, or mental illness at the workplace,^[15–17] the need to shorten items to reduce respondents' burdens (ie, efficiency) without compromising the assessment purpose (ie, accuracy) exists.^[35] Unfortunately, the reliability of a test or questionnaires would decrease if the number of items in a measurement instrument was shortened according to the Spearman-Brown prophecy formula.^[36–38] Nonetheless, the CAT based on item response theory adapts to an examinee's

ability level^[39–42] and provides the examinee with the next suitable item for answering the question.^[16] Accordingly, each patient under the CAT scenario is able to answer the fewest possible items with less burden, which could reach more or equivalently accurate outcomes.^[35,43] We were motivated to conduct this study incorporating CNNs into CAT to resolve the problems (ie, all questions must be answered) in previous studies.^[15–17] As such, it is thus necessary to develop an MCAT application incorporated with CNN for (1) efficiently collecting data regarding users' perceptions of MHB satisfaction and effectively classifying MHB satisfaction into 2 classes: unsatisfied (–) and satisfied (+).

1.5. An online MHB-APP using CNN and MCAT

As with all forms of web-based technology development, no online MCAT assessment has incorporated a CNN to assess individual perceptions about satisfaction on an issue available until now. After solving the problems of missing responses in MCAT affecting the CNN computation, the next challenge is to develop an APP using CNN and MCAT for collecting the perceptions on MHB from patients' experiences.

1.6. Study aims

Sentiment classifications play an important role in the age of artificial intelligence. Classifying a client's opinions online into positive and negative classes needs to be studied. The aims of this study are to design an online app that could accurately and efficiently classify a respondent's perceptions into positive and negative classes using CNNs and web-based MCAT.

2. Methods

2.1. Data source

Examinees invited to answer the MHB perception questions were selected from the outpatient department in a hospital in southern Taiwan. Before delivering the questionnaire to examinees, they were asked whether they had registered to MHB to meet the preliminary criteria used in this study. If so, the questionnaire could be filled anonymously. In all, 401 questionnaires were delivered, of which 391 were valid after removing those with incomplete or inaccurate data.

The sample size was determined and recommended by Dillman's^[44] instruction based on the 1.03 million registered users in MHB and the 95% confidence intervals required for computation. As a result, the minimal sample size was set at 384, which is less than the final 391 eligible respondents in this study.

This study was approved by the institutional review board of Show Chwan Memorial Hospital (IRB no. 1061107).

2.2. Contents of the questionnaire

The questionnaire contained the following 3 parts: satisfaction of MHB and its platform (12 items), agreement with the MHB functionality of resource exchange and integration (6 items), and satisfaction of the doctor-patient interaction and cocreation values existing in MHB (8 items) (Supplemental Digital Content 2, <http://links.lww.com/MD/G569>).

The 26-item 5-point Likert-type questionnaire was analyzed to examine the construct validity of the instrument. Higher scores mean a higher level of satisfaction with MHB.

2.3. Application displaying the classification on smartphones

A cell phone application was designed to predict the classification of MHB perception satisfaction using the CNN algorithm and model parameters. The resulting classification appears on participants' smartphones. The visual representation with binary (satisfied + and unsatisfied –) category probabilities would be shown on a dashboard using Google Maps. The forest plot was applied to present the comparison of 3 domain scores yielded in the CNN PMHB26 model.

2.4. Exploratory factor analysis of extract factors

Exploratory factor analysis (EFA) was performed to determine the number of factors retained in the survey data. The criterion of factor loadings was set at 0.50. Furthermore, Fornell and Larcker^[45] suggested the use of validity significance (eg, average variance explained [AVE] > 0.5 or named convergent validity [CV]^[46]; Eq. 1) and construct reliability (CR) (> 0.60; Eq. 2)^[47] or Cronbach reliability > 0.70 (Eq. 3).^[48]

$$\frac{(\sum \lambda_i^2)}{(\sum \lambda_i^2) + (\sum \varepsilon_i)}, \quad (1)$$

where λ is the item loading to the construct domain, squared λ indicates the community to the factor, and ε denotes the measurement error.

$$\frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + (\sum \varepsilon_i)}, \quad (2)$$

where λ and ε are defined as in Eq. 1.

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^I \sigma_i^2}{\sigma_x^2} \right), \quad (3)$$

where k is the item length, σ_i denotes the variance for item i , and σ_x represents the variance of examinees on the summation score.

2.5. Rasch analysis

2.5.1. Construct validity. Rasch analysis^[49] was performed using Winsteps software^[50] to transform the ordinal responses into interval logit (ie, log odds) scores.^[51] The construct fitting to the Rasch model's expectation was examined using criteria of mean square errors in Infit and Outfit statistics (ie, mean square error, MNSQ), within an acceptable range of 0.5 to 1.5^[52] compared with the classic test theory with factor loadings in EFA mentioned in the previous sections.

2.5.2. Rasch CAT. Efficient and accurate assessment can be simultaneously achieved through CAT.^[39–43] The online MCAT was designed based on the MHB perception instrument consisting of several constructs mentioned in Section 2.2.

Not all items were answered in CAT (deemed missing responses on CNNs); therefore, we used the Rasch rating scale model^[53] to generate the expected responses and overcame the drawback of not having all the items answered in CAT.^[16,39–43] As such, all the feature variables were applied to the CNN PMHB26 model.

2.5.3. Rasch KIDMAP. Rasch KIDMAP^[54] has been used in educational fields.^[54–56] All responses were compared with those of the Rasch model and examined by misfit statistics (ie, Z score = [observed score – expected value]/standard deviation). Through KIDMAP, any aberrant behaviors on responses to the questionnaire were displayed and interpreted. The overall fit statistics were observed by examining the Outfit MNSQ (= $[\sum Z - score_i]/L$, where L is the item length) for an examinee on the assessment.

2.6. CNN model created in MS Excel

2.6.1. Feature variables. Feature variables were determined by observing the extracted items using the EFA and Rasch analysis mentioned in the previous sections.

2.6.2. Unsupervised and supervised learning. Unsupervised learning^[16,53,54] was performed using k-means clustering,^[55] which is a traditional, simple machine learning algorithm that is trained on a test dataset and then able to classify a new dataset using a prime (ie, k clusters) defined *a priori*. In contrast, supervised learning was preceded using the CNN model with the feature variables to predict the labels that were defined by k-means clustering.

2.6.3. CNN model used in MS Excel. Similar to previous studies, we implemented the CNN PMHB26 model in MS Excel (MP4 video in Supplemental Digital Content 3, <http://links.lww.com/MD/G570>). The feature variables were transformed into the CNN model. Model parameters were estimated using the Solver add-in algorithms in MS Excel. The accurate rate, sensitivity, specificity, and receiver-operating characteristic (ROC) curve (area under the curve, AUC) were then computed.

2.7. Application displaying the classification on smartphones

A cell phone application was designed to predict the classification of MHB perception satisfaction using the CNN algorithm and model parameters. The resulting classification appears on participants’ smartphones. The visual representation with binary (satisfied + and unsatisfied –) category probabilities is shown on a dashboard using Google Maps. The forest plot was applied to present the comparison of 3 domain scores yielded in the CNN PMHB26 model.

2.8. Statistical tools and data analysis

MedCalc 9.5.0.0 for Windows (MedCalc Software) and Statistical Package for the Social Sciences for Windows (version 18.0; IBM Corp., Armonk, NY) were used to calculate the sensitivity (Sens), specificity (Spec), corresponding AUC (Eqs. from 4 to 7), and k-means clustering to yield the class labels (ie, 0 for unsatisfied and 1 for satisfied).

$$AUC = \frac{\text{one triangle area} + \text{one ladder area}}{2} = \frac{(1 - \text{Spec}) \times \text{Sens}}{2} + \frac{(\text{Sens} + 1) \times \text{Spec}}{2}, \quad (4)$$

$$SE \text{ for AUC} = \sqrt{\frac{AUC \times (1 - AUC)}{n}}, \quad (5)$$

$$z^* = \text{NORMSINV}(1 - (1 - 0.95)/2), \quad (6)$$

$$AUC \text{ error bound}(95\%CI) = AUC \pm z^* \times SE \text{ for AUG}, \quad (7)$$

With reference to the frequency and percentage of the variables related to sex, a χ^2 test was performed to examine the difference in count distributions. The significance level was set at 0.05.

Visual representations were used to display the classification effect using 2 curves based on the Rasch category characteristic curve (CCC)^[15–17,53] (ie, the curve from the bottom left corner to the top right corner denotes the success feature [the satisfied +], and the curve from the top left corner to the bottom right side denotes the failure attribute [the unsatisfied –]). The study flowchart is shown in Figure 1.

3. Results

3.1. Demographic data of the 391 cases

Of the 391 participants, 60.3% were women, 87% were over 54 years’ old, and 54% were single (Table 1). The mean annual income of 40.4% of the participants was between US \$930 and \$1560 per month. More than 60% graduated from a 4-year university or above, and 58.8% purchased additional commercial insurance. Regarding occupation, 19.4% worked in the private sector, 19.2% worked in service departments, and 20.2% endorsed others. The χ^2 test indicated that 4 variables were significantly different in sex distribution: age ($P < .001$), marital status ($P < .001$), mean income per month ($P < .001$), and occupation ($P < .001$) (Table 1).

3.2. Analyses of reliability and validity of the subscales

Three construct factors were extracted from PMHB26 (Table 2 and Supplemental Digital Content 2, <http://links.lww.com/MD/G569>). All 26 items with Cronbach $\alpha \geq 0.94$, CR ≥ 0.80 , and AVE ≥ 0.78 for each subscale were observed and showed internal consistency and stability in the study data (Table 2).

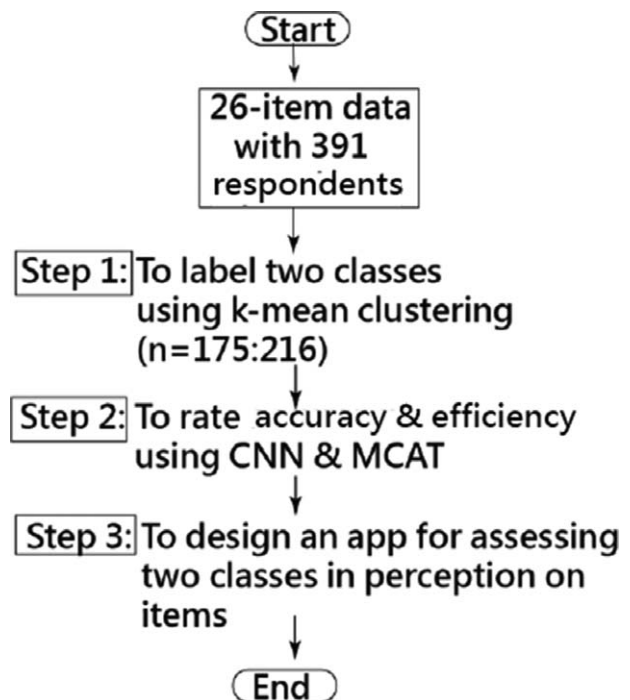


Figure 1. Study flowchart.

Table 1
Descriptive statistics of personal characteristics (N = 391).

Variable	Female	%	Male	%	All	%	χ^2	ρ
Age, y							23.758 (2)	.001
≤54	221	56.5	119	30.4	340	87.0		
55~64	11	2.8	29	7.4	40	10.2		
≥65	4	1.0	7	1.8	11	2.8		
Marital status							11.920 (1)	.001
Single	144	36.8	67	17.1	211	54.0		
Married	92	23.5	88	22.5	180	46.0		
Mean income per month (USD)							18.307 (3)	.001
No income	35	9.0	13	3.3	48	12.3		
≤930	82	21.0	49	12.5	131	33.5		
930~1560	100	25.6	58	14.8	158	40.4		
≥1560	19	4.9	35	9.0	54	13.8		
Educational level							9.493 (4)	.050
Junior high school	11	2.8	9	2.3	20	5.1		
Senior high school	18	4.6	24	6.1	42	10.7		
2-y College	38	9.7	29	7.4	67	17.1		
4-y University	145	37.1	74	18.9	219	56.0		
Graduate School	24	6.1	19	4.9	43	11.0		
Business life insurance coverage							1.497 (1)	.132
Yes	133	34.0	97	24.8	230	58.8		
No	103	26.3	58	14.8	161	41.2		
Occupation							21.381 (5)	.001
Medical sector	48	12.3	15	3.8	63	16.1		
Private sector	32	8.2	44	11.3	76	19.4		
Government employee	21	5.4	16	4.1	37	9.5		
Financial sector	33	8.4	28	7.2	61	15.6		
Service sector	47	12.0	28	7.2	75	19.2		
Others	55	14.1	24	6.1	79	20.2		

Rasch analysis indicated that all items have acceptable fit statistics (ie, MNSQ ≤1.5)^[52]; however, the one (ie, #1 equipment required [eg, card reader]) with an Infit MNSQ of 1.64 and an Outfit MNSQ of 1.62 indicated that the data fit the

Rasch model rather well. All threshold difficulties for each subscale exhibit monotonic increases from lower to higher with a gap (ie, between 2 threshold difficulties in difficulty) >1.5 logits^[61] (Table 2).

Table 2
Validity and average variable extracted.

Construct	Max. Infit	Max. Outfit	Cronbach α	CR	AVE
Platform operation (12 items)	1.64	1.62	0.94	0.92	0.80
Threshold difficulties	-7.80				
	-1.49				
	3.02				
	6.27				
Resource exchange and integration (6 items)	1.38	0.73	0.96	0.80	0.96
Threshold difficulties	-15.05				
	-3.13				
	5.48				
	12.71				
Values co-creation (8 items)	1.20	1.48	0.98	0.96	0.78
Threshold difficulties	-14.65				
	-2.77				
	5.65				
	11.77				

AVE=average variable extracted, CR=construct reliability.

Table 3
CNN applied to prediction of the MHB utility (PMHB26) (n=391).

CNN classifications and ACC	True condition		Statistics	
	Satis (+)	Unsatis (-)	PPV/FOR	FDR/NPV
Positive	214	5	0.98	0.02
Negative	2	170	0.01	0.99
Sensitivity	0.99	—	—	—
FPR	0.01	—	—	—
FNR (Miss rate)	0.03	—	—	—
Specificity	0.97	—	—	—
AUC (95% CI)	0.98 (0.97–0.99)	—	—	—
Accuracy (ACC)	0.98	—	—	—

— = Not applicable, ACC=accuracy, AUC=area under the receiver operating characteristic curve, CNN = convolutional neural networks, FDR=1-PPV, FNR=false-negative rate, FOR=1-NPV, FPR=false-positive rate, NPV=negative predictive value, PPV=positive predictive value.

3.3. Comparison of prediction accuracies in the CNN PMHB26 model

Two classes (ie, n1=175 and n2=216) divided by k-means clustering were labeled 0 (unsatisfied) and 1 (satisfied). The cutoff point was set at 85 points using ROC analysis. The 26-item CNN PMHB26 model yielded higher accuracy rates with an accuracy of 98%, an AUC of 0.99 (0.99–1.00), and much smaller Type I and Type II errors (≤3%) based on the 391 cases (Table 3).

Another visual representation that displays the classification effect is depicted using box plots, AUCs, and bimodal distributions in Figure 2. The AUC and its 95% CI were computed by Eqs. from 4 to 7 for AUC ($=0.98=(1-Spec) \times Sens/2 + (Senc+1) \times Spec/2 = (1-0.97) \times (0.99)/2 + (0.99+1) \times 0.97/2$) and the 95%CI ($0.97-0.99=0.98 \pm 19.6 \times SE$, where $SE=0.0069 = \sqrt{\frac{0.98 \times (1-0.98)}{391}}$), respectively.

Interested readers are encouraged to see the study process in Supplemental Digital Content 3, <http://links.lww.com/MD/G570> using the parameters based on the 391 cases to predict the accuracy of the CNN PMHB26 model.

3.4. Application detecting satisfactory classes on a web-based assessment

A CNN PMHB26 application for examinees predicting satisfactory classes was developed and is demonstrated at the top of Figure 3. One resulting example of the satisfied class is presented at the bottom of Figure 3 on the CCC (ie, category 0 from the top left corner to the bottom right corner, category 1 from the bottom left side to the top right side) based on the Rasch rating scale model,^[53] which is novel and innovative when using a visual display shown on Google Maps. A comparison of 3 domain scores for the respondent was made using the forest plot to display in Figure 4.

Readers are invited to click on the link in reference^[62] to practice the CNN PMHB26 application on their own or see the operational process in Supplemental Digital Content 3, <http://links.lww.com/MD/G570>. Note that all 38 model parameters for classifying individual binary levels are involved in the Rasch online MCAT module.

3.5. Rasch KIDMAP for detecting aberrant responses in an assessment

Once the bubble of satisfied or unsatisfied was clicked (Fig. 5), 3 icons for the KIDMAPs of the 3 constructs appeared.^[63–65] One,

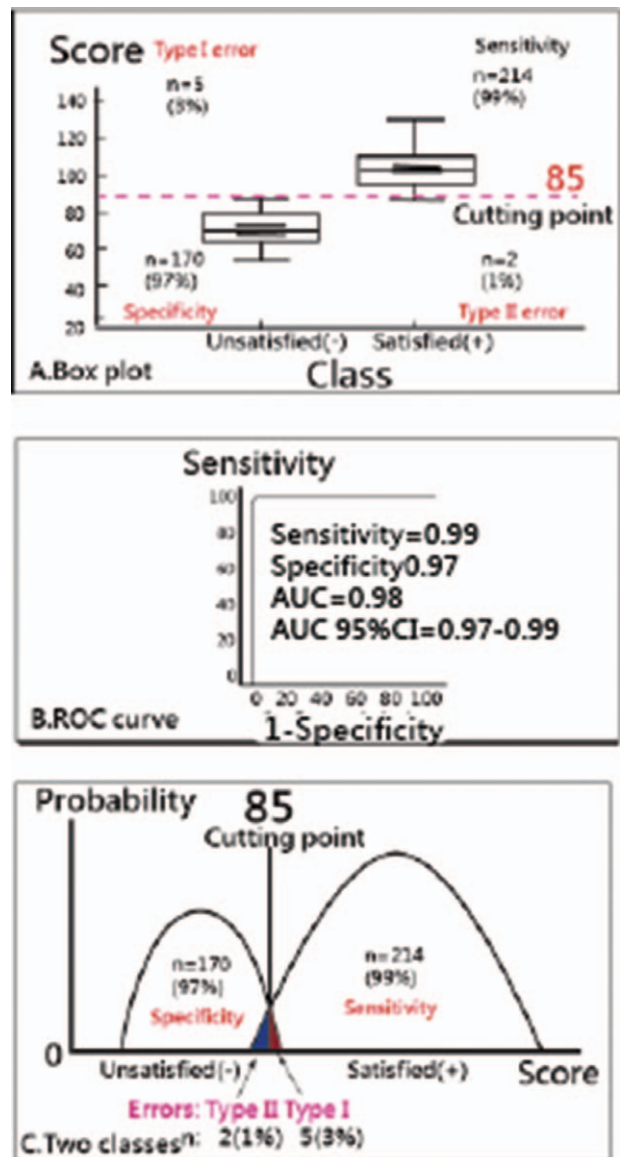
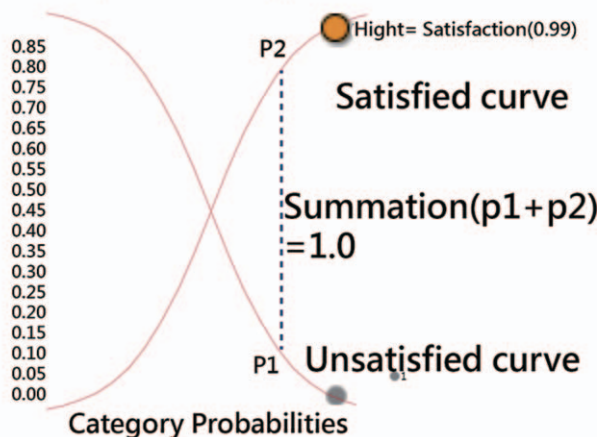


Figure 2. Two classes grouped using the convolutional neural networks approach.



A. Snapshot on App



B. Output of the assessment

Figure 3. Snapshot of the application and the assessment output.

the resulting KIDMAP (ie, the 12 items regarding satisfaction of the MHB operation and the platform), is shown in Figure 5.

Z scores and Rasch logit scores are on the x- and y-axes, respectively. Bubbles are measured using the standard errors (SEs) of the item difficulties. Higher bubbles mean that the items (or the examinee shown in yellow) are more difficult (or higher) on the y-axis. The red color denotes an aberrant response due to a Z score of >2.0, indicating that the response is a misfit to the model's expectation. That is, the response was endorsed as too

high (ie, much more satisfied) based on the person estimated measure (2.55 logits). The overall person Outfit MNSQ (2.4) is greater than the cutoff point at 2.0.^[61]

3.6. Online dashboards shown on Google maps

All dashboards in the figures appear once the QR code is scanned or the links^[62-65] are clicked. Readers are advised to examine the details about the information for each entity.

4. Discussion

4.1. Principal findings

In this study, we observed 2 aspects: the accuracy of the CNN-MCAT PMHB26 model for the 2 categories having a higher accuracy rate (0.98) with an AUC of 0.98 (95% CI, 0.97-0.99) and the efficiency of not having to answer all 26 items without compromising the accuracy of the measurement. All 26 items were successfully used to design an application for the respondents to evaluate MHB utility in Taiwan. As such, this study aims to design an online app that could accurately and efficiently classify a respondent's perceptions into positive and negative classes using CNNs and web-based MCAT.

4.2. Review of research findings

CNN can improve the prediction accuracy (up to 7.14%),^[29] Although >1408 articles were found using "convolutional neural network" searched in the title,^[66] we have not seen any study incorporating CNN with MCAT to solve the problem of missing responses in questionnaires. The perception of MHB^[7,8] (Supplemental Digital Content 1, <http://links.lww.com/MD/G568>) was illustrated in this study to verify that the MCAT combined with CNN could be viable and feasible in the application.

4.2.1. MCAT rule and calculation. We set one of the stop rules for the measurement of SEs smaller than a criterion, that is, <0.51 (SD × SQRT [1 - alpha]=2.09 × SQRT[1 - 0.94]),^[67] where alpha is Cronbach reliability and SD denotes the person standard deviation of an assessment so that the minimum number of questions required for completion could be efficiently satisfied because CAT achieves a similar measurement precision to that with only approximately half the test length compared with traditional paper-based examinations.^[35,67-70]

In addition, we set at least 5 items in CAT to be completed as another stop rule, which might inflate the test length. The test length of CAT might be slightly higher than that of CAT in previous studies.^[68-70]

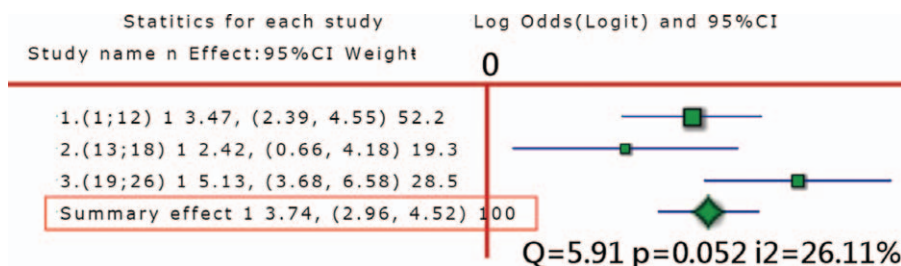


Figure 4. Comparison of 3 domain scores was made using the forest plot to display.

The initial question was randomly selected from a subscale similar to previous studies^[23–27,30,59] using an item selection strategy. The provisional person measure was estimated using the iterative Newton–Raphson procedure^[36,69] after 3 items were answered, avoiding all item answers corresponding to either 0 or 4 as the extreme category in PMHB26, limiting the effective estimation in CAT or MCAT.

The next question selected was the one with the highest information value among the remaining unanswered questions weighted against the provisional person measure.^[67,69] Details of the CAT procedures are shown in Supplemental Digital Contents 3 to 5, <http://links.lww.com/MD/G570>, <http://links.lww.com/MD/G571>, <http://links.lww.com/MD/G572>.

4.2.2. The CNN–MCAT PMHB model created in this study.

We illustrated an example of perceptions about MHB utility in Taiwan^[71] (Supplemental digital Content 1, <http://links.lww.com/MD/G568>). Patients can use the MHB platform to obtain knowledge about their disease and treatment history, which can motivate them to actively participate in their healthcare.^[71] This survey on PMHB26 showed that if patients properly use MHB information, the MHB platform will help them immediately understand their own health and medical conditions. The information can be used as a reference when seeking medical treatment, avoiding unnecessary physical checkups, reducing duplicated prescriptions, and promoting doctor–patient interaction and value cocreation.^[72–74] The CNN–MCAT PMHB model can be used in other survey-based questionnaires,^[14–16] not just limited to perception surveys of MHB.

4.3. Implications and applications in the future

4.3.1. Improvement made in accuracy further in the future.

Although the accuracy of the CNN–MCAT PMHB26 model could reach 98%, improvements can be further made by adjusting a few known responses to avoid letting the CNN fail in the classification, such as a scheme named the matching personal response scheme to adapt for the correct classification in the model (MPRSA)^[15] designed for driving the model's accuracy toward 100%.

Only 7 misclassified cases were found in the CNN PMHB26 model (Table 3 and Fig. 2). All these summation scores were close to the cutoff point of 85. The way in which we applied the MPRSA to achieve the 100% accuracy goal can be designed if the same response string encountered in the model is automatically adjusted to the designated class.^[15]

4.3.2. The calculation of AUC and its 95% CI in Eqs. from 4 to 7.

CNN can improve the prediction accuracy (up to 7.14%).^[29] More than 1408 articles were found using CNN^[66] in PubMed Central. Thus far, we have not seen any study, except for the 3 aforementioned studies,^[15–17] using Microsoft Excel to perform the CNN in the literature, which is novel and useful in studies with small sample sizes or a survey-based questionnaire, as we did in this study. Furthermore, we did not see any previous 1408 articles clearly and completely describing how to compute the AUC and its 95% CI, as we did in Eqs. from 4 to 7.

The curves of category probabilities based on the Rasch rating scale model^[53] are sophisticated in Figure 5. Binary categories (eg, success and failure on an assessment in the psychometric field) have been broadly used in health-related outcomes.^[15–17,19,43,44,75] However, the animation-type dashboard with the KIDMAP (Fig. 5)^[63–65] shown on Google Maps is particularly

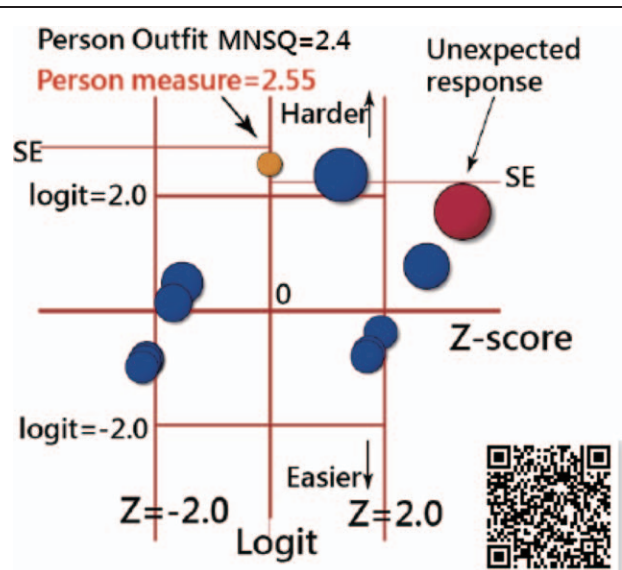


Figure 5. KIDMAP presented for examining the aberrant responses for an examinee.

designed for the CNN–MCAT PMHB26 model used in this study.

4.3.3. Missing responses in MCAT can be solved. Finally, not all questions were answered in the CAT. Different from those using the mean value^[69] over the entire dataset to fill the missing values and those applying CAT^[16,40] instead of MCAT,^[28,29] we used the expected value in the model for each unanswered item to fill the missing data, as done in previous studies.^[40,41] By doing so, the expected responses and CNN parameters can be incorporated to predict the classes of interest.

4.4. Strengths of this study

This study has 4 strengths. First, we demonstrated the CNN module incorporated with the Rasch MCAT classifying opinions online into positive and negative levels that have not been found in the traditional Rasch analysis in the literature. Second, all 26 person responses and 38 model parameters were embedded into the CNN–MCAT MHB26 model that can be used for predicting the underlying classes, as no such CNN was found incorporating MCAT into a prediction model ever before in academics.^[76] Third, Microsoft Excel was used to execute the CNN prediction model, which is familiar and convenient to ordinary researchers who can understand the easy-to-use CNN in applications more than ever before using other Non-MS-Excel-based software with the CNN model. Fourth, an application was designed to display results using the visual dashboard on Google Maps (see the MP4 video in Supplemental Digital Content 3, <http://links.lww.com/MD/G570>).

As with all forms of web-based technology, advances in health communication technology are rapidly emerging.^[77] Mobile online CAT assessment is promising and worth considering in several fields of health assessment. It is recommended that readers click on links at references^[62–65] to practice the application on their own for more details about the online app.

4.5. Limitations and future studies

Although our study aim has been achieved through the CNN–MCAT PMHB model that could accurately (with fewer standard errors) and efficiently (with fewer items) classify respondents' perceptions into positive and negative classes, some limitations exist in this study.

First, although the psychometric properties of PMHB26 have been validated for measuring perceptions on MHB (Table 2), no evidence exists to support the fact that the instrument is suitable for use on MCAT assessment when referred to the item (ie, #1 equipment required [eg, card reader]) with an Infit MNSQ of 1.64 and an Outfit MNSQ of 1.62 greater than the criterion (<1.5).^[52,61] Additional studies are needed.

Second, although the 2 classes were determined to have a high accuracy rate (98%) using k-means clustering and the CNN algorithm (Table 3), we cannot guarantee that this CNN is the only statistical approach to improve the classification accuracy. Future studies are encouraged to seek other prediction methods that can also improve the model prediction, such as logistic regression, naive Bayes, decision trees, random forests, gradient tree boosting,^[78–87] and other artificial neural networks.^[88]

Third, item difficulties were calibrated using Winsteps software^[50] under the Rasch rating scale model.^[53] The ideal way to gain item difficulties is to use Rasch ConQuest software^[89] because the Rasch multidimensional model^[90] was used in this study.

Fourth, all the model parameters (ie, item difficulties, step-threshold difficulties, and the 38 CNN model parameters) were derived from this study. If any environment or occasion/condition is changed (eg, the survey subjects and contents), the result (eg, the model's parameters) will be different from the present study and will need further verifications in the future.

Fifth, missing responses were filled by the expected scores only held under the Rasch model because a one-dimensional subscale was verified (eg, Infit and Outfit MNSQs less than 1.5). Otherwise, the remaining items that were not answered in CAT could not be obtained as the expected responses and used for the CNN algorithm. Future studies should be cautious about this matter.

Sixth, the application accessed via references^[56–65] should be further developed for Android and IOS users. The efficiency with a lower number of questions answered to the questions of PMHB26 must be investigated as to whether a similar measurement (or classification) precision to that with only about half the test length^[36,67–70] exists using a simulation method in the future.

Finally, the study sample was taken from a patient-perception survey. The model parameters estimated for PMHB26 are only suitable for the PMHB26 questionnaire. Generalizing these parameters to other surveys might be biased because the sample consisted only of patients at one specific study hospital. Additional studies are needed to reexamine whether the psychometric properties of PMHB26 are similar to our study.

5. Conclusions

The contributions in this study include the following: overcoming the problem of missing responses that affect CNN computation; introducing CNN availability in Microsoft Excel and incorporating MCAT in classification; and demonstrating an application incorporating MCAT with numerous parameters in CNN. The

CAT–MCT PMHB26 model is recommended for mimicking similar studies on improving the accuracy of determining individual classes. An application developed for helping patients self-assess their perceptions about satisfaction in several domains is required in the future.

Acknowledgments

We thank the Ministry of Science and Technology for the support of grand funds (MOST 106-2410-H-217-001-SSS, MOST 107-2410-H-217-001-SSS) in 2017 and 2018 to accomplish the research.

Author contributions

CF conceived and designed the study, YH and TW interpreted the data, and YH monitored the process and the manuscript. TW drafted the manuscript. All authors read the manuscript and approved the final manuscript.

Conceptualization: Chen-Fang Hsu.

Data curation: Chen-Fang Hsu, Yu-Hua Yan.

Methodology: Tsair-Wei Chien.

References

- [1] Ryan P, Sawin KJ. The individual and family self-management theory: background and perspectives on context, process, and outcomes. *Nurs Outlook* 2009;57:217–25.
- [2] Turvey C, Klein D, Fix G, et al. Blue Button use by patients to access and share health record information using the Department of Veterans Affairs' online patient portal. *J Am Med Inform Assoc* 2014;21:657e63.
- [3] Hogan TP, Nazi KM, Luger TM, et al. Technology-assisted patient access to clinical information: an evaluation framework for blue button. *JMIR Res Protoc* 2014;3:e18.
- [4] Pearce C, Bainbridge M. A personally controlled electronic health record for Australia. *J Am Med Inform Assoc* 2014;21:707e13 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24650635>. Accessed on March 1, 2016
- [5] Greenhalgh T, Stramer K, Bratan T, Byrne E, Russell J, Mohammad Y, et al. Summary Care Record early adopter programme: An independent evaluation by university college London. PDF available at <http://discovery.ucl.ac.uk/6602/1/6602.pdf>, accessed on February 15, 2016.
- [6] Greenhalgh T, Wood GW, Bratan T, Stramer K, Hinder S. Patients' attitudes to the summary care record and HealthSpace: qualitative study. *BMJ* 2008;336:1290.
- [7] Huang SK, Pan YT, Chen MS. My Health Bank 2.0-making a patron saint for people's health. *J Formos Med Assoc* 2017;116:69–71.
- [8] National Health Insurance Administration. My health bank promotes people's self-health management, in line with world trends. 2016. Available at: http://www.mohw.gov.tw/CHT/Ministry/DM2_P.aspx?f_list_no=334&doc_no=53991. Accessed December 1, 2016.
- [9] National Health Insurance Administration (2019). My health bank break through millions, and the sooner you log in, the more you like. 2019. Available at: https://www.nhi.gov.tw/News_Content.aspx?n=FC05EB85BD57C709&cs=53E47AF9482689F3. Accessed February 21, 2019.
- [10] Van Alstyne MW, Parker GG, Choudary SP. Pipelines, platforms, and the new rules of strategy. *Harvard Bus Rev* 2016;94:54–62.
- [11] Barrett M, Oborn E, Orlikowski W. Creating value in online communities: the sociomaterial configuring of strategy, platform, and stakeholder engagement. *Inform Syst Res* 2016;27:704–23.
- [12] Iyer B, Davenport TH. Reverse engineering Google's innovation machine (cover story). *Harvard Bus Rev* 2008;86:58–68.
- [13] Pinho N, Beirão G, Patrício L, Fisk RP. Understanding value cocreation in complex services with many actors. *J Serv Manage* 2014;25:470–93.
- [14] Banerjee I, Ling Y, Chen MC, et al. Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification. *Artif Intell Med* 2019;97:79–88.

- [15] Lee YL, Chou W, Chien TW, Chou PH, Yeh YT, Lee HF. An app developed for detecting nurse burnouts using the convolutional neural networks in microsoft excel: population-based questionnaire study. *JMIR Med Inform* 2020;8:e16528.
- [16] Ma SC, Chou W, Chien TW, et al. An app for detecting bullying of nurses using convolutional neural networks and web-based computerized adaptive testing: development and usability study. *JMIR Mhealth Uhealth* 2020;8:e16747.
- [17] Yan YH, Chien TW, Yeh YT, Lin CJ, Chou W, Hsing SC. An app for classifying personal mental illness at workplace using fit statistics and Convolutional Neural Networks(CNN): a survey-based quantitative study. *JMIR Mhealth Uhealth* 2020;8:e17857.
- [18] Hsu CF, Chien TW, Chow JC, Yeh YT, Chou W. An app for identifying children at risk for developmental problems using multidimensional computerized adaptive testing: development and usability study. *JMIR Pediatr Parent* 2020;3:e14632.
- [19] Lee YL, Lin KC, Chien TW. Application of a multidimensional computerized adaptive test for a Clinical Dementia Rating Scale through computer-aided techniques. *Ann Gen Psychiatry* 2019;18:5.
- [20] Mzoughi H, Njeh I, Wali A, et al. Deep multi-scale 3D convolutional neural network (CNN) for MRI gliomas brain tumor classification [published online ahead of print, 2020 May 21]. *J Digit Imaging* 2020;10.1007/s10278-020-00347-9. doi:10.1007/s10278-020-00347-9.
- [21] Saha S, Pagnozzi A, Bourgeat P, et al. Predicting motor outcome in preterm infants from very early brain diffusion MRI using a deep learning convolutional neural network (CNN) model. *Neuroimage* 2020;215: 116807. doi:10.1016/j.neuroimage.2020.116807.
- [22] Wei R, Zhou F, Liu B, et al. Real-time tumor localization with single X-ray projection at arbitrary gantry angles using a convolutional neural network (CNN). *Phys Med Biol* 2020;65: 065012. Published 2020 Mar 19. doi:10.1088/1361-6560/ab66e4.
- [23] Liang B, Tian Y, Chen X, et al. Prediction of radiation pneumonitis with dose distribution: a Convolutional Neural Network (CNN) based model. *Front Oncol* 2020;9:1500.
- [24] Ohsaka A. Artificial intelligence (AI) and hematological diseases: establishment of a peripheral blood Convolutional Neural Network (CNN)-based digital morphology analysis system. *Rinsho Ketsueki* 2020;61:564-9.
- [25] Moitra D, Mandal RK. Automated AJCC (7th edition) staging of non-small-cell lung cancer (NSCLC) using a deep convolutional neural network (CNN) and recurrent neural network (RNN). *Health Inf Sci Syst* 2019;7:14.
- [26] Liu K, Zhong P, Zheng Y, Yang K, Liu M. P_VggNet: a convolutional neural network (CNN) with pixel-based attention map. *PLoS One* 2018;13: e0208497. Published 2018 Dec 12. doi:10.1371/journal.pone.0208497.
- [27] Muñoz-Organero M, Powell L, Heller B, Harpin V, Parker J. Automatic extraction and detection of characteristic movement patterns in children with ADHD based on a Convolutional Neural Network (CNN) and acceleration images. *Sensors (Basel)* 2018;18:3924.
- [28] Ravi D, Wong C, Deligianni F, et al. Deep learning for health informatics. *IEEE J Biomed Health Inform* 2017;21:4-21.
- [29] Sathyanarayana A, Joty S, Fernandez-Luque L, et al. Sleep quality prediction from wearable data using deep learning. *JMIR Mhealth Uhealth* 2016;4:e125.
- [30] Tobore I, Li J, Yuhang L, et al. Deep learning intervention for health care challenges: some biomedical domain considerations. *JMIR Mhealth Uhealth* 2019;7:e11966.
- [31] Kwon S, Hong J, Choi E, et al. Deep learning approaches to detect atrial fibrillation using photoplethysmographic signals: algorithms development study. *JMIR Mhealth Uhealth* 2019;7:e12770.
- [32] Xu JH, Zhou XM, Ma JL, et al. Application of convolutional neural network to risk evaluation of positive circumferential resection margin of rectal cancer by magnetic resonance imaging. *Zhonghua Wei Chang Wai Ke Za Zhi* 2020;23:572-7.
- [33] Li Q, Zhong L, Huang H, et al. Auxiliary diagnosis of developmental dysplasia of the hip by automated detection of Sharp's angle on standardized anteroposterior pelvic radiographs. *Medicine (Baltimore)* 2019;98:e18500.
- [34] Tarancon JEP. Which Programming Language Are Best for Artificial Neural Network Researchers? Additionally, in That Which Language Is Preferable? June 22, 2020. Available at: https://www.researchgate.net/post/Which_programming_language_are_best_for_artificial_neural_network_researchers_Also_in_that_which_language_is_preferable.
- [35] Chien TW, Wu HM, Wang WC, Castillo RV, Chou W. Reduction in patient burdens with graphical computerized adaptive testing on the ADL scale: tool development and simulation. *Health Qual Life Outcomes* 2009;7:39.
- [36] Allen M, Yen W. Introduction to Measurement Theory. Monterey, CA: Brooks/Cole; 1979. ISBN 0-8185-0283-5.
- [37] Stanley J. Thorndike RL. Reliability. Educational Measurement 2nd ed Washington, DC: American Council on Education; 1971.
- [38] Wainer H, Thissen D. Wainer H, Thissen D. True score theory: the traditional method. Test Scoring. Mahwah, NJ: Lawrence Erlbaum; 2001.
- [39] Ma S, Wang H, Chien T. A new technique to measure online bullying: online computerized adaptive testing. *Ann Gen Psychiatry* 2017;16:26.
- [40] Chien TW, Lin WS. Simulation study of activities of daily living functions using online computerized adaptive testing. *BMC Med Inform Decis Mak* 2016;16:130.
- [41] Chien TW, Lin WS. Improving inpatient surveys: web-based computer adaptive testing accessed via mobile phone QR Codes. *JMIR Med Inform* 2016;4:e8.
- [42] Djaja N, Janda M, Olsen CM, Whiteman DC, Chien TW. Estimating skin cancer risk: evaluating mobile computer-adaptive testing. *J Med Internet Res* 2016;18:e22.
- [43] Ma S, Chien T, Wang H, Li Y, Yui M. Applying computerized adaptive testing to the Negative Acts Questionnaire-Revised: Rasch analysis of workplace bullying. *J Med Internet Res* 2014;16:e50.
- [44] Dillman D. Constructing the Questionnaire, Mail and Internet Surveys. New York: John Wiley & Sons; 2000.
- [45] Fornell C, Larcker DF. Structural equation models with unobservable variables and measurement error. *J Mark Res* 1981;18:39-50.
- [46] Hair JF, Hult GTM, Ringle CM, Sarstedt MA. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). 2nd Edition. Thousand Oaks, CA: Sage Publications Inc.; 2017.
- [47] Kan WC, Kuo SC, Chien TW, et al. Therapeutic duplication in Taiwan hospitals for patients with high blood pressure, sugar, and lipids: evaluation with a mobile health mapping tool. *JMIR Med Inform* 2020;8:e11627.
- [48] Wang HY, Chou W, Shao Y, Chien TW. Comparison of Ferguson's and the Gini coefficient used for measuring the inequality of data related to health quality of life outcomes. *Health Qual Life Outcomes* 2020;18:111.
- [49] Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Chicago: University of Chicago Press; 1960.
- [50] Lincare JM. Winsteps Rasch measurement computer program (Version 4.0.1). Chicago, IL: Winsteps.com; 2006.
- [51] Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. 2nd ed. Mahwah NJ: Erlbaum; 2007.
- [52] Linacre LM. What do infit and outfit, mean-square and standardized mean? *Rasch Meas Trans* 2003;16:878.
- [53] Andrich D. A rating formulation for ordered response categories. *Psychometrika* 1978;43:561-73.
- [54] Wright BD, Mead RJ, Ludlow LH. KIDMAP: Person-By-Item Interaction Mapping. Research Memorandum No. 29. 1980;MESA Psychometric Laboratory, Department of Education, University of Chicago.
- [55] Chien TW, Wang WC, Wang HY, Lin HJ. Online assessment of patients' views on hospital performances using Rasch model's KIDMAP diagram. *BMC Health Serv Res* 2009;9:135.
- [56] Chien TW, Wang WC, Lin SB, Lin CY, Guo HR, Su SB. KIDMAP, a web based system for gathering patients' feedback on their doctors. *BMC Med Res Methodol* 2009;9:38.
- [57] Chien TW, Lee YL, Wang HY. Detecting hospital behaviors of upcoding on DRGs using Rasch model of continuous variables and online cloud computing in Taiwan. *BMC Health Serv Res* 2019;19:630.
- [58] Rashidi HH, Tran NK, Betts EV, Howell LP, Green R. Artificial intelligence and machine learning in pathology: the present landscape of supervised methods. *Acad Pathol* 2019;6: 2374289519873088. doi: 10.1177/2374289519873088.
- [59] Buehler L, Rashidi H. Bioinformatics Basics, Application in Biological Science and Medicine. 2nd ed. Philadelphia: Taylor and Francis Group; 2005.
- [60] Chen C, Luo J, Parker K. Image segmentation via adaptive K-mean clustering and knowledge-based morphological operations with biomedical applications. *IEEE Trans Image Process* 1998;7:1673-83.
- [61] Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas* 2002;3:85-106.

- [62] Chien TW. Rasch CAT used in this study. June 20, 2020. Available at: http://www.healthup.org.tw/irs/irsin_e.asp?type1=92.
- [63] Chien TW. Kidmap used for understanding the aberrant responses for Domain 1. June 20, 2020. Available at: <http://www.healthup.org.tw/irs/kidmapbmcats.asp?repno=92&kid=1001330506&mcat=1>.
- [64] Chien TW. Kidmap used for understanding the aberrant responses for Domain 2. Accessed on June 20, 2020. Available at <http://www.healthup.org.tw/irs/kidmapbmcats.asp?repno=92&kid=1001330506&mcat=2>.
- [65] Chien TW. Kidmap used for understanding the aberrant responses for Domain 3. Accessed on June 20, 2020. Available at: <http://www.healthup.org.tw/irs/kidmapbmcats.asp?repno=92&kid=1001330506&mcat=3>.
- [66] Chien TW. Articles with CNN in the title in Pubmed library. Accessed on November 4, 2020. Available at: <https://pubmed.ncbi.nlm.nih.gov/?term=%22convolutional+neural+network%22%5Btitle%5D&sort=pubdate>.
- [67] Chien TW, Wang WC, Huang SY, Lai WP, Chow JC. A web-based computerized adaptive testing (CAT) to assess patient perception in hospitalization. *J Med Internet Res* 2011;13:e61.
- [68] Wainer HW, Dorans NJ, Flaugher R, et al. *Computerized Adaptive Testing: A Primer*. Hillsdale, N.J.: L. Erlbaum Associates; 1990.
- [69] Embretson S, Reise S. *Item Response Theory for Psychologists*. Mahwah, N.J.: L. Erlbaum Associates; 2000. 158–186.
- [70] Weiss DJ. Improvement measurement quality and efficiency with adaptive testing. *Appl Psychol Measure* 1982;6:473–92.
- [71] Joiner KA, Lusch RF. Evolving to a new service-dominant logic for health care. *Innov Entrepr Health care* 2016;3:25–33.
- [72] Vargo SL, Lusch RF. Institutions and axioms: An extension and update of service-dominant logic. *J Acad Market Sci* 2016;44:5–23.
- [73] Lusch RF, Nambisan S. Service innovation: a service-dominant logic perspective. *MIS Quart* 2015;39:155–75.
- [74] McColl-Kennedy JR, Cheung L, Ferrier E. Cocreating service experience practices. *J Serv Manage* 2015;26:249–75.
- [75] Hulin CL, Drasgow F, Parsons C. *Item Response Theory: Applications to Psychological Measurement*. Homewood: Dow & Jones Irwin; 1983.
- [76] Wang H, Cui Z, Chen Y, Avidan M, Abdallah AB, Kronzer A. Predicting hospital readmission via cost-sensitive deep learning. *IEEE/ACM Trans Comput Biol and Bioinf* 2018;15:1968–78.
- [77] Mitchell SJ, Godoy L, Shabazz K, Horn IB. Internet and mobile technology use among urban African American parents: survey study of a clinical population. *J Med Internet Res* 2014;16:e9.
- [78] Nelder JA, Wedderburn RWM. Generalized linear models. *J Royal Stat Soc A* 1972;135:370.
- [79] Cox DR. The regression analysis of binary sequences. *J Royal Stat Soc B* 2018;20:215–32.
- [80] Szumilas M. Explaining odds ratios. *J Can Acad Child Adolesc Psychiatry* 2010;19:227–9. <http://europepmc.org/abstract/MED/20842279>.
- [81] Bayes T, Price R. An essay towards solving a problem in the doctrine of Chances. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, A.M.F.R.S. *Phil Trans Royal Soc London* 1763;53:370–418.
- [82] Domingos P, Pazzani M. On the optimality of the simple Bayesian classifier under zero-one loss. *Mach Learn* 1997;29:103–30.
- [83] Quinlan JR. Induction of decision trees. *Mach Learn* 1986;1:81–106.
- [84] Breiman L, Sammut C, Webb GI. *Random forests*. *Encyclopedia of Machine Learning and Data Mining*. Boston, Massachusetts: Springer; 2017;32.
- [85] Tin KH. *Random decision forests*. *Proceedings of 3rd International Conference on Document Analysis and Recognition*; 14–16 Aug; Montreal, Canada. 1995.
- [86] Friedman JH. Stochastic gradient boosting. *Comput Stat Data Analysis* 2002;38:367–78.
- [87] Freund Y, Mason L. *The Alternating Decision Tree Algorithm*. *Proceedings of the 16th International Conference on Machine Learning*; 26–27 June; Ithaca, New York. 1999.
- [88] Mehta A. *Digital Vidya*. 2019. Jan 25, [2020-02-12]. *A Comprehensive Guide to Types of Neural Networks*. Available at: <https://www.digitalvidya.com/blog/types-of-neural-networks/>.
- [89] Wu ML, Adams RJ, Wilson MR. *Acer ConQuest*. Melbourne: Australian Council for Educational Research press; 1998.
- [90] Adams RJ, Wilson MR, Wang WC. The multidimensional random coefficients multinomial logit model. *Appl Psychol Meas* 1997;21:1–23.