

Original Paper

# Web-Based Skin Cancer Assessment and Classification Using Machine Learning and Mobile Computerized Adaptive Testing in a Rasch Model: Development Study

Ting-Ya Yang<sup>1\*</sup>, MD; Tsair-Wei Chien<sup>2\*</sup>, MBA; Feng-Jie Lai<sup>3\*</sup>, MD, PhD

<sup>1</sup>Department of Family Medicine, Chi Mei Medical Center, Tainan, Taiwan

<sup>2</sup>Department of Medical Research, Chi-Mei Medical Center, Tainan, Taiwan

<sup>3</sup>Department of Dermatology, Chi-Mei Medical Center, Tainan, Taiwan

\* all authors contributed equally

**Corresponding Author:**

Feng-Jie Lai, MD, PhD

Department of Dermatology

Chi-Mei Medical Center

901, Zhonghua Rd

Yongkang District

Tainan, 710

Taiwan

Phone: 886 6 2812811 ext 57109

Fax: 886 6 2203706

Email: [lai.fengjie@gmail.com](mailto:lai.fengjie@gmail.com)

## Abstract

**Background:** Web-based computerized adaptive testing (CAT) implementation of the skin cancer (SC) risk scale could substantially reduce participant burden without compromising measurement precision. However, the CAT of SC classification has not been reported in academics thus far.

**Objective:** We aim to build a CAT-based model using machine learning to develop an app for automatic classification of SC to help patients assess the risk at an early stage.

**Methods:** We extracted data from a population-based Australian cohort study of SC risk (N=43,794) using the Rasch simulation scheme. All 30 feature items were calibrated using the Rasch partial credit model. A total of 1000 cases following a normal distribution (mean 0, SD 1) based on the item and threshold difficulties were simulated using three techniques of machine learning—naïve Bayes, k-nearest neighbors, and logistic regression—to compare the model accuracy in training and testing data sets with a proportion of 70:30, where the former was used to predict the latter. We calculated the sensitivity, specificity, receiver operating characteristic curve (area under the curve [AUC]), and CIs along with the accuracy and precision across the proposed models for comparison. An app that classifies the SC risk of the respondent was developed.

**Results:** We observed that the 30-item k-nearest neighbors model yielded higher AUC values of 99% and 91% for the 700 training and 300 testing cases, respectively, than its 2 counterparts using the hold-out validation but had lower AUC values of 85% (95% CI 83%-87%) in the k-fold cross-validation and that an app that predicts SC classification for patients was successfully developed and demonstrated in this study.

**Conclusions:** The 30-item SC prediction model, combined with the Rasch web-based CAT, is recommended for classifying SC in patients. An app we developed to help patients self-assess SC risk at an early stage is required for application in the future.

(*JMIR Med Inform* 2022;10(3):e33006) doi: [10.2196/33006](https://doi.org/10.2196/33006)

**KEYWORDS**

skin cancer assessment; computerized adaptive testing; naïve Bayes; k-nearest neighbors; logistic regression; Rasch partial credit model; receiver operating characteristic curve; mobile phone

## Introduction

### Background

Skin cancer (SC) is the most common malignant neoplasm occurring in White populations, and it is mainly divided into (1) malignant melanoma (MM) and (2) nonmelanoma SCs (NMSCs), which include squamous cell carcinoma and basal cell carcinoma as the major subtypes. The global incidence of MM and NMSCs is well-established and on the rise. In Australia, SC accounts for most newly diagnosed cancers each year, with age-standardized incidence rates for MM of  $65.3 \times 10^{-5}$  and  $1878 \times 10^{-5}$  for NMSCs [1,2]. There are >434,000 people in a population of only 23 million who treat keratinocyte cancer each year in Australia [2], causing a substantial socioeconomic burden and impact on public health services.

There are several well-recognized risk factors that increase the potential for the development of SC and have been reported in previous literature, such as UV radiation, genetic susceptibility, smoking, ionizing radiation, and the use of photosensitizing drugs [3]. Among the aforementioned risk factors, excessive UV radiation exposure remains the major causative risk factor for SC [4]. Therefore, it is crucial to modify personal behaviors to reduce direct and excessive sun exposure, such as avoiding long-term sunbathing or the use of indoor tanning devices, appropriately applying sunscreens, using sun-protective cloth garments, and staying in the shade.

### Requirement for Prediction Model in Classification of SC

In practice, it is difficult to provide people with their individual risk of SC [1]. Given the lack of clear recommendations for organized SC screening, physical exploration, clinical history of lesion changes, and correlated family SC history continue to be key for detecting skin neoplasms. Assuming that a person has attributes that highly correlate with the underlying architecture of the skin, the potential risk of SC can be assessed through questions (ie, questionnaire items); for example, underlying pigmentation traits include hair color, eye color, the propensity to freckle and sunburn, skin phenotypes, and some personal behavior factors such as tanning attitudes and sunbed use. Accordingly, it is feasible to construct a unidimensional scale to measure these attributes using the responses to the unidimensional items and further calculate an overall SC risk score using an assessment tool (eg, web-based computerized adaptive testing [CAT] administrations [1]) or even classify the SC risk for patients in clinical settings.

### Predicting SC Risk and Classifying the SC Possibility

Statistical validity is based on the correlations among item measures (or scores) on a questionnaire and people's unobservable true status (eg, melanoma status—deemed latent traits that cannot be directly detectable in the real world) [5]. The Rasch model [6] is a mathematical modeling approach that has been used to assess how well the items measure the underlying latent traits [7-13], which are based on a unidimensional scale when the data fit the Rasch model's expectation (ie, all items can be added to a summation score) [10,13]. Nonetheless, no SC classifications that use machine

learning to predict SC risk have been illustrated and demonstrated in the literature. We are motivated to develop a prediction model for classifying SC in adults who are potentially at risk.

### CAT Assessment and Limitation in SC Classification

CAT is a tailored measure based on item response theory (IRT) [14,15] that can better align with each examinee's ability level [10,13,16]. The computer follows an IRT-based algorithm, and the difficulty of the next selected item depends mainly on all previously answered items. As such, each patient needs to answer the fewest possible items by dynamically selecting appropriate testing items, resulting in less respondent burden without compromising measurement precision and thereby making it possible to individualize each participant's assessment [1,10,13].

The limitation of CAT applied to machine learning is the missing responses (ie, unanswered items) in the data. Fortunately, generating the expected responses to endorse the answers in CAT has been resolved to overcome the drawback of not having all the items answered in CAT (ie, using the expected value to fill in the missing data, as done in previous studies [13,17,18]). As such, convolutional neural networks (CNNs) [19,20] combined with the expected responses to classify the groups of individual bullying levels [13] are applicable. Thus, we are interested in applying the expected responses to CAT to (1) reduce participant burden with more accurate outcomes [1,10,13,16] and (2) predict SC classification in patients.

### Web-Based Assessment Using Smartphones

With the advent of the era of digital technology, the advancement and maturation of mobile health and health communication technology have been rapidly increasing [21]. To date, smartphone apps for classifying SC using CAT-based machine learning for patients in health care settings are lacking when searching for publications in the PubMed library using the keywords *skin and cancer AND computerized adaptive testing AND CAT AND machine learning* as of December 5, 2021. It is not only the complexity of the CAT procedure with multimedia illustrations embedded into a web-based module but also the difficulty of the model's parameters that need to be transformed into the probability of classification types when SC is assessed on the web. A web-based CAT app incorporating machine learning and SC could provide patients with a better understanding of the SC classification and prediction of SC at risk before a serious SC problem occurs.

### Study Aims

The aims of this study are to (1) compare the prediction accuracy of SC between machine learning models in SC classification and (2) build a CAT-based SC assessment using machine learning to develop an app for automatic classification of SC to help patients assess SC risk at an early stage.

## Methods

### Data Source

On the basis of a previous study [1,22], we extracted data from a population-based Australian cohort study of SC risk (N=43,794) by simulating Rasch data [23], including 1000 virtual patients across 30 feature variables defined in the previous study [1] ([Multimedia Appendix 1](#)).

All data used in this study were simulated and extracted from the previous article [1]. Given that this study design uses simulation data, ethical approval was not required according to the Taiwan Ministry of Health and Welfare regulations.

### Characteristics of the Simulated Data

#### *The Original Survey Data*

The original data were retrieved from the baseline questionnaire in the QSkin Sun and Health study [22]. A population-based cohort study of 43,794 men and women aged 40 to 69 years

was randomly sampled from the population of Queensland, Australia [1], to obtain a calibration data set (two-thirds; 29,314/43,794, 66.94%) and a validation data set (one-third; 14,480/43,794, 33.06%). In the calibration data set, 24.61% (7213/29,314) of participants had a history of SC, and 75.39% (22,101/29,314) of participants did not.

#### *The Study Simulation Data*

For simplification, the 30-item difficulties calibrated in the previous study [1] ([Table 1](#)) using the Rasch partial credit model [24] were applied to yield 1000 virtual cases following a normal distribution (mean 0, SD 1; see the demonstration in [Multimedia Appendix 1](#) with an MP4 video). The suggested cutoff point was set at 0.88 logits [1] to determine the 2 groups of cancer and noncancer in the simulation data. As such, the data with 1000 people  $\times$  30 items and 1 label (ie, 1 and 0 for melanoma status defined as cancer and noncancer groups) were applied in this study with the following 2 sections (ie, 3 models and 3 tasks).

**Table 1.** Overall and threshold difficulties in logit (log odds) across the 30 items.

Number	Variable	Overall difficulty	Threshold difficulty			
			Step 1	Step 2	Step 3	Step 4
1	Gender (male as 1 and female as 0)	0.16	0.00	N/A <sup>a</sup>	N/A	N/A
2	Skin color on areas never exposed to the sun?	-2.32	-2.46	0.78	1.68	N/A
3	Your behavior in the strong sun for 30 minutes at noon?	-0.17	-1.51	0.41	1.10	N/A
4	Your behavior outdoors in the sun without protecting your skin?	-0.49	-0.85	-0.42	1.27	N/A
5	What color are your eyes?	-0.11	-0.04	0.60	1.55	-2.11
6	What was your natural hair color at the age of 21 years?	0.48	0.70	-0.83	-1.30	1.43
7	How many freckles were on your face at the age of 21 years?	0.72	-0.37	0.01	0.36	N/A
8	How many moles did you have on your skin at the age of 21 years?	0.76	-1.45	0.53	0.92	N/A
9	How many times in your whole life have you used sunbeds?	1.27	1.35	0.30	-0.75	-0.69
10	How many separate skin cancers have you ever had excised from your skin?	0.98	0.45	-1.36	1.30	-0.39
11	How many separate sunspots or skin cancers have you ever had frozen or burnt off on your skin?	0.53	-0.05	0.49	-0.22	-0.11
12	Have I been told that I have melanoma?	0.99	0.99	N/A	N/A	N/A
13	Will you get melanoma at some point in the future?	0.26	-1.14	-0.82	1.14	0.82
14	How many times were you sunburned so badly that you were sore for at least 2 days or your skin peeled as a child?	0.58	-1.41	0.37	0.11	0.36
15	How many times were you sunburned so badly that you were sore for at least 2 days or your skin peeled in your teenage years?	0.17	-2.40	0.35	0.27	0.74
16	How many times were you sunburned so badly that you were sore for at least 2 days or your skin peeled in adulthood?	0.58	-1.83	0.59	0.10	0.46
17	How many hours did you spend outdoors and in the sun from Monday to Friday in the past year?	0.29	-0.04	0.44	-0.39	N/A
18	How many hours did you spend outdoors and in the sun from Monday to Friday at the age of 10 to 19 years?	-0.51	-0.65	0.24	0.41	N/A
19	How many hours did you spend outdoors and in the sun from Monday to Friday at the age of 20 to 29 years?	-0.15	-0.46	0.41	0.05	N/A
20	How many hours did you spend outdoors and in the sun from Monday to Friday at the age of 30 to 39 years?	0.04	-0.29	0.42	-0.13	N/A
21	How many hours did you spend outdoors and in the sun during Saturday and Sunday in the past year?	-0.14	-0.42	0.23	0.19	N/A
22	How many hours did you spend outdoors and in the sun during Saturday and Sunday at the age of 10 to 19 years?	-0.94	-0.46	0.21	0.26	N/A
23	How many hours did you spend outdoors and in the sun during Saturday and Sunday at the age of 20 to 29 years?	-0.72	-0.60	0.18	0.43	N/A
24	How many hours did you spend outdoors and in the sun during Saturday and Sunday at the age of 30 to 39 years?	-0.45	-0.56	0.19	0.37	N/A
25	Routinely apply sunscreen to my face	-0.46	0.00	N/A	N/A	N/A
26	Routinely apply sunscreen to my hands and forearms	-1.80	0.00	N/A	N/A	N/A
27	Routinely apply sunscreen to other parts of my body	-2.56	0.00	N/A	N/A	N/A
28	Routinely apply sunscreen going out in the sun: no	-0.36	0.00	N/A	N/A	N/A
29	Whether applying sunscreen outside in the sun?	-0.31	-0.90	-0.16	1.06	N/A
30	How often have you been outside in the sun in the past year?	0.45	-0.77	-0.08	0.85	N/A
31	Melanoma status (label as cancer and noncancer group)	N/A	N/A	N/A	N/A	N/A

<sup>a</sup>N/A: not applicable.

### The 3 Models of Machine Learning Used in Microsoft Excel

#### The 3 Models Applied in This Study

Three models of machine learning—naïve Bayes (NB) [25], k-nearest neighbors (KNN) [26], and logistic regression (LR) [27-31]—were applied to compare the model accuracy of classifying SC in the 1000×30 rectangle data set. The 2 training (70%) and testing (30%) sets (ie, the hold-out validation) were separated to examine the model's accuracy with a proportion of 70:30, where the former was used to predict the latter.

We calculated the sensitivity, specificity, receiver operating characteristic curve (area under the curve [AUC]), and CIs along with the accuracy and precision across the 3 aforementioned models for comparison. In addition, k-fold cross-validation was performed for the 3 models using the Weka software (University of Waikato) [32]. If the Weka Explorer (graphical user interface) and the *Classify* tab are selected, we can find it by looking for the *Choose* button under the *Classify* tab. Once we navigate through the folders, the 3 classifiers are used (ie, NB classifiers→Bayes→NB; instance-based learner [IBk] classifiers→lazy→IBk; and classifiers→functions→logistic). For instance, once we select IBk for the KNN classifier, we click on the box immediately to the right of the button. This will open up a large number of options. If we then click on the button *More* in the *Options* window, we will see all the options explained. We can do this for all the classifiers to obtain additional information (eg, NB, logistic, or more; see the demonstration using an MP4 video in [Multimedia Appendix 2](#)). Meanwhile, more information about the 3 models is provided in [Multimedia Appendix 3](#).

#### Calculation of Model Accuracy

After the parameters in the selected model are estimated, the accuracy of a model in the training and testing sets can be obtained through the following equations [33,34]:

The accuracy was determined by observing the higher sensitivity, specificity, precision, accuracy, and AUC in the models. The definitions are as follows:

True positive (TP) = the number of predicted cancers to the true SCs (1)

True negative (TN) = the number of predicted non-SCs to the true noncancers (2)

False positive (FP) = the number of noncancers – the number of TN (3)

False negative (FN) = the number of cancers – the number of TP (4)

Sensitivity = TP rate = TP/(TP + FN) (5)

Specificity = TN rate = TN/(TN + FP) (6)

Precision = positive predictive value = TP/(TP + FP) (7)

Accuracy = (TP + TN)/N (8)

N = TP + TN + FP + FN (9)

AUC = (1 – specificity) × sensitivity/2 + (sensitivity + 1) × specificity/2 (10)

SE for AUC =  $\sqrt{(AUC \times [1 - AUC]/N)}$  (11)

95% CI = AUC ± 1.96 × SE for AUC (12)

Similarly, the confusion matrix can be made when the true conditions (ie, SC and non-SC) and the predictions (ie, positive and negative) are known in the predicted training set (or the testing data set) matched to the label (ie, 1 and 0 as cancer and noncancer groups) in the training set. Other indicators in equations (1) to (12) can be obtained accordingly.

It is worth noting that we made the model residual with the average values in the 2 groups (ie, average [range in the group of SC] + average [range in the group of non-SC]) to overcome the imbalance class data. As such, the AUC for sensitivity and specificity could be balanced in reports [35]. Details about the setting formula are provided in the Microsoft Excel module in [Multimedia Appendix 1](#).

### The 3 Tasks

#### Feature Variables Shown on a Forest Plot (Task 1)

The 30 variables [1] were shown on a forest plot [36-38] via the following steps: standardize each variable based on the mean (0) and SD (1) and compare the standardized mean difference on a forest plot [39].

The chi-square test was conducted to evaluate the heterogeneity between variables. Forest plots (CI plots) were drawn to display the effect estimates and their CIs for each indicator.

#### Comparing the Accuracies in Models (Task 2)

We calculated the sensitivity, specificity, AUC, and CIs along with the accuracy and precision across the proposed models in comparison using equations (1) to (12). Both AUCs in the training and testing sets were compared to assess the model accuracy and stability [34,35].

#### SC Risk and Classification (Task 3)

#### The Rasch Model and the First-Order Derivative in Calculus

In the Rasch model, the probability can be expressed as follows:

$$f(\theta_n - \delta_i) = \text{Pro}(X_{ni}|\theta) = \frac{e^{(\theta_n - \delta_i)}}{1 + e^{(\theta_n - \delta_i)}}$$

(13)

where  $\theta$  is the person's ability, and  $\delta$  is the item difficulty for person  $n$  and item  $i$ , respectively. The processes of the first-order derivative on  $\theta$  are described below:

$$\begin{aligned} f'(\theta_n - \delta_i) &= \left( \frac{e^{(\theta_n - \delta_i)}}{1 + e^{(\theta_n - \delta_i)}} \right)' \\ &= \left[ \frac{(e^{(\theta_n - \delta_i)})'(1 + e^{(\theta_n - \delta_i)}) - (e^{(\theta_n - \delta_i)})(1 + e^{(\theta_n - \delta_i)})'}{(1 + e^{(\theta_n - \delta_i)})^2} \right] \\ &= \left[ \frac{(e^{(\theta_n - \delta_i)})(1 + e^{(\theta_n - \delta_i)}) - (e^{(\theta_n - \delta_i)})(e^{(\theta_n - \delta_i)})}{(1 + e^{(\theta_n - \delta_i)})^2} \right] \\ &= \frac{(e^{(\theta_n - \delta_i)})(1 + e^{(\theta_n - \delta_i)}) - e^{(\theta_n - \delta_i)}e^{(\theta_n - \delta_i)}}{(1 + e^{(\theta_n - \delta_i)})^2} = \frac{(e^{(\theta_n - \delta_i)})}{(1 + e^{(\theta_n - \delta_i)})^2} \end{aligned}$$

(14)

#### The Newton-Raphson Iteration Method

The Newton-Raphson iteration method, one of the essential iteration techniques for parameter estimation, has been frequently mentioned in the methodology literature [40-43] and popularly used in practice with the Rasch model [44,45].

A revised estimated measure,  $\theta_m + 1$ , is obtained from the previous measure of  $\theta_m$  and the adjustment by the residual and the summed variance (defined by  $f'[\theta_m - \delta_i]$  across all answered items in equation 15):

$$\theta_m = \theta_{m-1} + \frac{\sum_{i=1}^m (O_{mi} - f(\theta_m - \delta_i))}{\sum_{i=1}^m f'(\theta_m - \delta_i)} \tag{15}$$

The CAT SE is defined by the following equation:

$$\text{CAT SE} = \sqrt{\frac{1}{\sum_{i=1}^m f'(\theta_m - \delta_i)}} \tag{16}$$

The next selected item is determined by the maximum information (variance =  $f'[\theta_m - \delta_i]$ ) of the item in all answered items shown in the following equation:

$$\text{Information}_i = f'(\theta_m - \delta_i) \tag{17}$$

**CAT Stop Criterion**

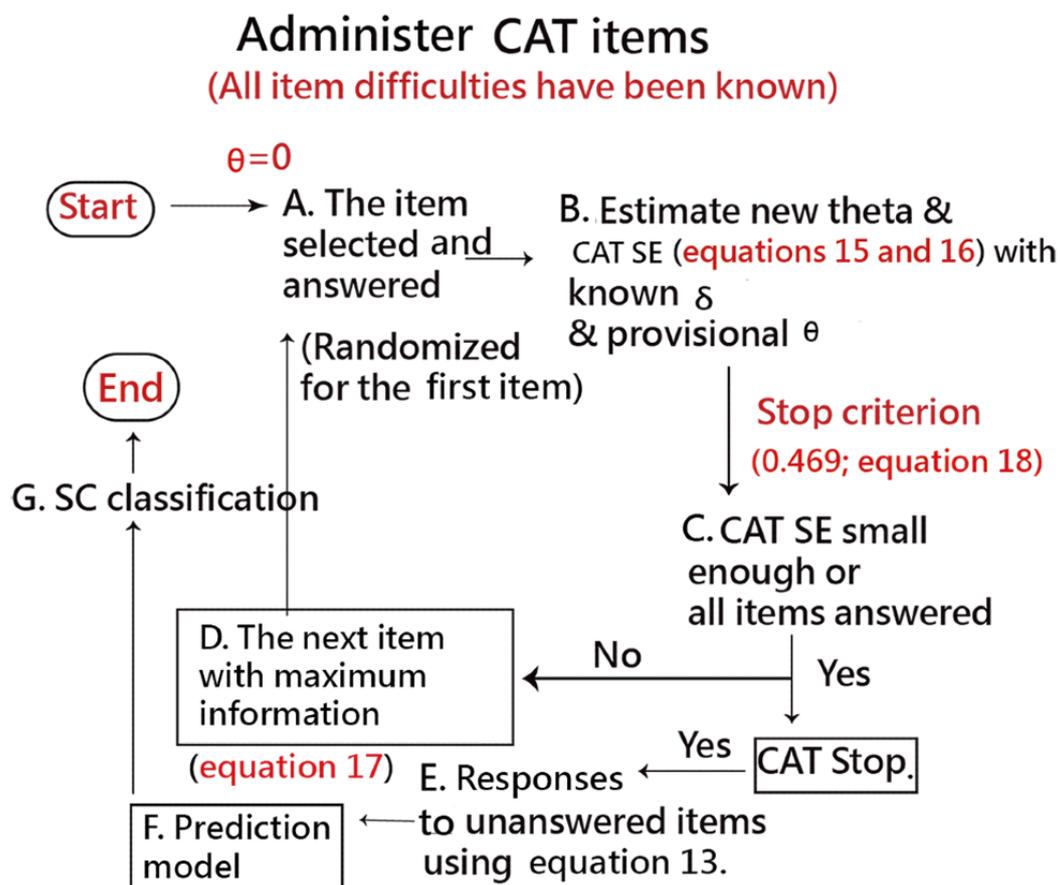
The CAT termination is set at the CAT SE smaller than the SE of measurement (SEM) [1,46].

$$\text{SEM} = \text{SD} \sqrt{(1 - \text{Rel})} \tag{18}$$

Rel is the Cronbach  $\alpha$  of the questionnaire. Therefore, if there is a test (or questionnaire) with an SD of 1.0 logits and a Cronbach  $\alpha$  of .78 [1], the SEM would be 0.469 ( $1 \times \sqrt{[1 - .78]}$ ).

If CAT is terminated, the responses to unanswered items are filled in with their expected values using equation (13) when the final measure is known. The SC classification is then performed (Figure 1).

**Figure 1.** SC-CAT process and SC classification using machine learning. CAT: computerized adaptive testing; SC: skin cancer.



**The Fit Statistics of the Mean Square Error**

The Rasch fit statistics of mean square errors (MNSQs), including infit and outfit [40,41], are shown on the SC CAT to represent the extent of the deviation from the expectation of the Rasch model for the examinee’s responses.

Infit MNSQ =

$$\frac{\sum_{i=1}^m (O_{ni} - E_{ni})^2}{\text{Variance}} = \frac{\sum_{i=1}^m (O_{ni} - E_{ni})^2}{\sum_{i=1}^m f'(\theta_m - \delta_i)} \tag{19}$$

Outfit MNSQ =

$$\sum_{i=1}^m \frac{(O_{ni} - E_{ni})^2}{f'(\theta_m - \delta_i)} \tag{20}$$

where  $O_{ni}$  is the observed response for person  $n$  on item  $i$ , and  $E_{ni}$  is the corresponding expected value in equation (13). The variance is referred to in equations (14) and (17).

Again, another way to judge a person’s responses depends on the Z score (denoted by  $Z$ ) in equation (21). According to the

Rasch model, these accumulated  $Z^2$  values ought to follow a chi-square distribution with 1 degree of freedom (denoted by  $df$ ) for each  $Z^2$  value minus the degree of freedom necessary to estimate the person measure  $\theta_n$  [47]. Any sum of  $Z^2$ , when divided by its  $df$ , should follow the mean square distribution in equation (22). This can conveniently be evaluated as the  $t$  statistic, which has approximately a unit normal distribution (ie,  $N[0,1]$ ) [46], shown in equation (23).

$$Zscore_{ni} = Z_{ni} = \frac{(O_{ni} - E_{ni})}{f'(\theta_n - \delta_i)^{0.5}} = \frac{(O_{ni} - E_{ni})}{SD_{ni}}$$

(21)

$$v_n = \sum_{i=1}^m \frac{(Zscore_i)^2}{df} = \sum_{i=1}^m \frac{(Zscore_i)^2}{m-1}$$

(22)

$$t_n = (\ln(v_n) + v_n - 1) \times \left(\frac{df}{8}\right)^{0.5}$$

(23)

### The Skin Cancer–Computerized Adaptive Testing Algorithm

Wright [48] suggests a simpler algorithm for classroom use, classification, and performance tracking in a low-stakes environment. This algorithm is easy to implement and could be successfully used at the end of each learning module to keep track of the persons' responses in the process [46]. Figure 1 shows the core steps of skin cancer–computerized adaptive testing (SC–CAT) needed for practical adaptive testing using the Rasch model:

1. Start with a patient at an initial  $\theta$  (SC score in logit) of 0.
2. Find a randomized item from the item poll via the SC–CAT.
3. Respond to the item with difficulty and the corresponding threshold  $\delta$  (difficulty; label A in Figure 1).
4. Calculate the provisional  $\theta$  in equation (15) based on the known item difficulties (label B).
5. Examine whether the CAT stop criterion (ie,  $SEM=0.469$ ) is reached in equations (16) and (18) (label C).
6. Select the next item in equation (17) if the SC–CAT continues (label D).

7. Return to Step 3.
8. Fill in the expected values of the unanswered items via equation (13) when the SC–CAT stops based on the final estimated  $\theta$  (label E).
9. Perform the prediction model (label F).
10. Obtain the classification (ie, SC or non-SC; label G).

### The App Developed in This Study

An app for the detection of SC in adults was designed and developed. A 30-item self-assessment app using mobile phones was designed to predict and classify SC using machine learning and model parameters. The model parameters were embedded in the computer module.

The results of the classification (ie, SC+ and SC–) instantly appear on smartphones. A visual representation displaying the classification effect is plotted using 2 curves (ie, one from the bottom left to the top right corner denotes the success [SC+] feature, and another from the top left to the bottom right is the failure [SC–] attribute). The visual dashboard with binary (ie, SC+ and SC–) category curves is shown on Google Maps.

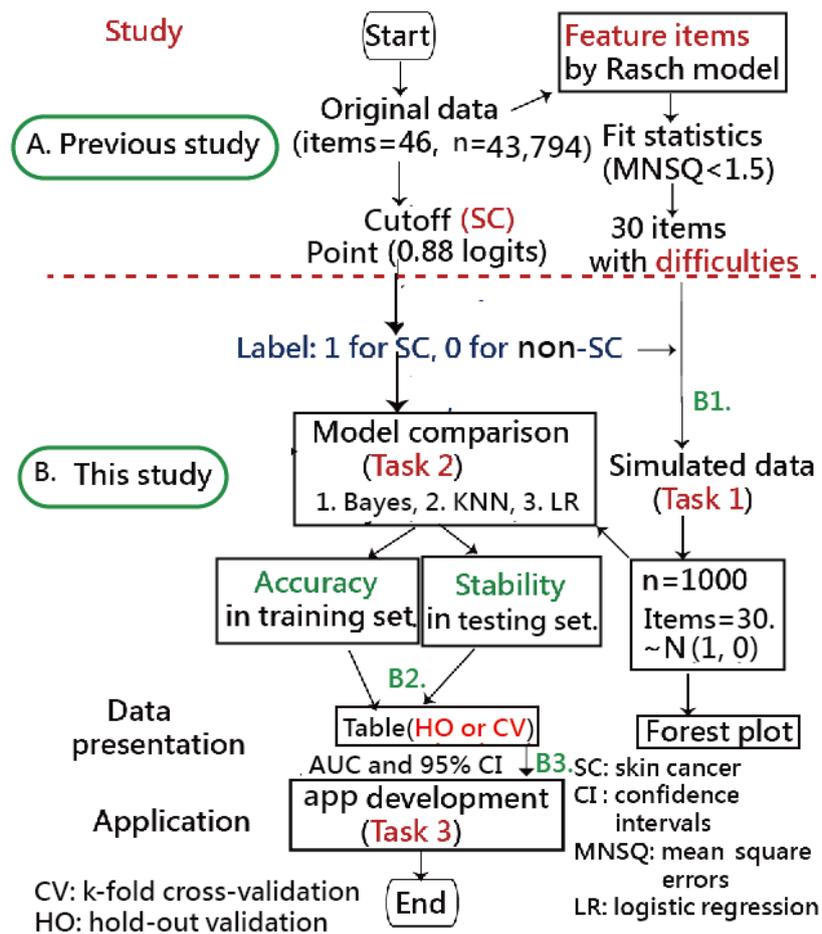
### Statistical Tools and Data Analysis

MedCalc 9.5.0.0 for Windows (MedCalc Software) was used to calculate the sensitivity, specificity, and the corresponding AUC using LR when the observed labels (ie, 0 for SC– and 1 for SC+) and the predicted probabilities (ie, the continuous variable in equation 13) were applied.

Author-made modules in Microsoft Excel were applied to compute the model prediction indicators expressed in equations (1) to (12). The three proposed models—NB, KNN, and LR—were performed using Microsoft Excel and Weka [32] (Multimedia Appendix 1 and 2). The web-based CAT was programmed using the classic active server pages.

The study flowchart (shown in Figure 2) comprises two parts: one is from the previous study [1] and another includes 3 models. A total of 3 tasks are elaborated in this study. The abstract video is provided in Multimedia Appendix 1 as well.

**Figure 2.** Two major parts are in the study flowchart (in the upper and bottom panels), and three tasks are in the bottom panel. AUC: area under the curve; KNN: k-nearest neighbors; MNSQ: mean square error; SC: skin cancer; HO: hold out validation.



### Ethics Approval and Consent to Participate

Not applicable. All data were simulated and extracted from a previous study [1].

### Availability of Data and Materials

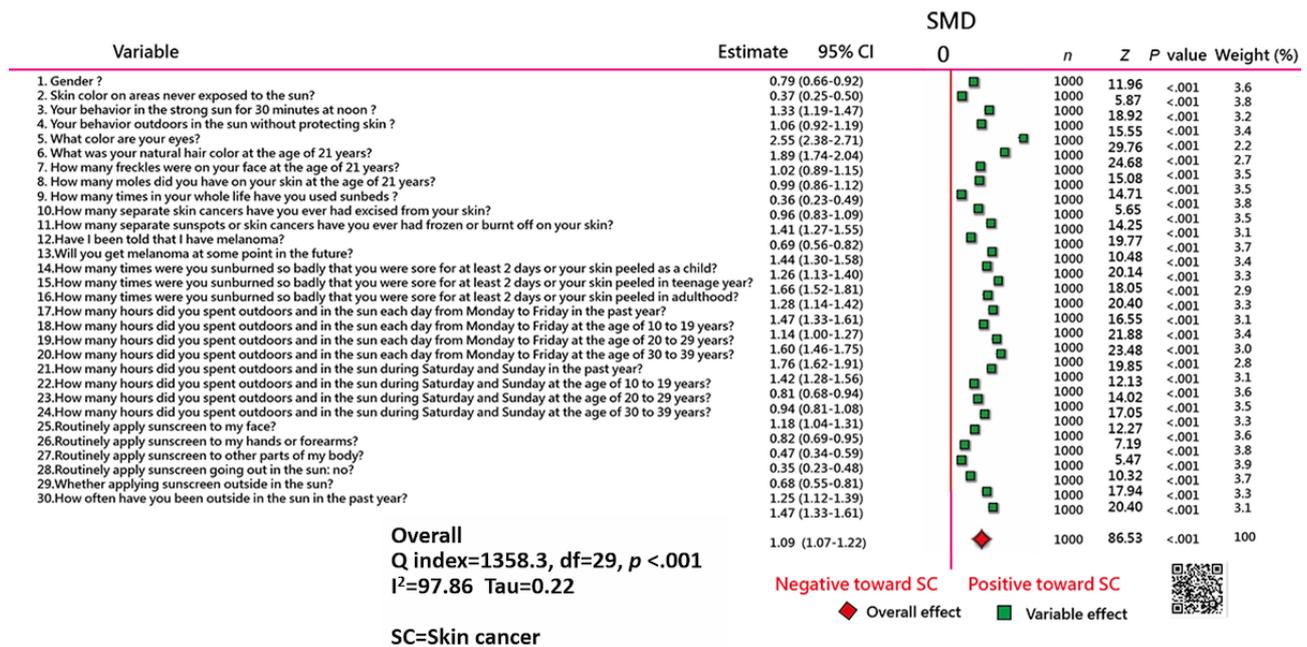
All data used in this study are available in the Multimedia Appendices.

## Results

### Task 1: Feature Variables Demonstrated on a Forest Plot

The 30 variables are presented in a forest plot (Figure 3). We can see that all green boxes are on the right side beyond the mean standardized mean difference (0), indicating that the variables are eligible ( $P<.05$ ) for discriminating the melanoma status (ie, SC and non-SC groups).

**Figure 3.** Using the forest plot to display feature variables on smartphones [49] or clicking the QR Code. SMD: standardized mean difference.



### Task 2: Comparing the Accuracies Between Models

A comparison of the model accuracies is shown in Table 2. We can see that all AUCs are >0.80 in models across the training and testing sets. The 30-item KNN model yielded higher AUC

values of 99% and 91% for the 700 training and 300 testing cases, respectively, far beyond the other 2 models (ie, NB and LR; Table 3). However, if k-fold cross-validation is performed, the 30-item KNN model yields lower AUC values of 85% (95% CI 83%-87%), shown in Table 4.

**Table 2.** Comparison of model accuracy and stability using simulation data (hold-out validation).

Study model	Training cases/testing cases, N	Accuracy ≥0.80 (training sets)					Stability ≥0.70 (testing sets)				
		Sensitivity	Specificity	Precision	Accuracy	AUC <sup>a</sup>	Sensitivity	Specificity	Precision	Accuracy	AUC
Naïve Bayes	700/300	0.92	0.89	0.82	0.90	0.90	0.79	0.98	0.97	0.91	0.89
KNN <sup>b</sup>	700/300	0.98	0.99	0.98	0.99	0.99	0.83	0.99	0.99	0.93	0.91
LR <sup>c</sup>	700/300	0.82	0.91	0.84	0.88	0.87	0.70	0.92	0.85	0.84	0.81

<sup>a</sup>AUC: area under the curve.

<sup>b</sup>KNN: k-nearest neighbors.

<sup>c</sup>LR: logistic regression.

**Table 3.** Comparison of model accuracy and stability using simulation data (95% CIs of the area under the curve [AUC] for hold-out validation)<sup>a</sup>.

Study model	Accuracy ≥0.80 (training sets)			Stability ≥0.70 (testing sets)		
	Training cases, N	AUC (95% CI)	Significant difference	Testing cases, N	AUC (95% CI)	Significant difference
Naïve Bayes (1)	700	0.90 (0.88-0.92)	1, 2	300	0.89 (0.85-0.93)	— <sup>b</sup>
KNN <sup>c</sup> (2)	700	0.99 (0.98-1.00)	1, 3	300	0.91 (0.88-0.94)	3
LR <sup>d</sup> (3)	700	0.87 (0.85-0.89)	1, 2	300	0.81 (0.77-0.85)	2

<sup>a</sup>The computation of the 95% CI for the AUC is referred to in equations (10) to (12).

<sup>b</sup>Data not available.

<sup>c</sup>KNN: k-nearest neighbors.

<sup>d</sup>LR: logistic regression.

**Table 4.** Comparison of model accuracy and stability using simulation data (k-fold cross-validation).

Study model	Training cases/testing cases, N	Accuracy $\geq 0.80$ (training sets)				Stability $\geq 0.70$ (testing sets)		
		Sensitivity	Specificity	Precision	Accuracy	AUC <sup>a</sup>	AUC (95% CI)	Significant difference
Naïve Bayes (1)	700/300	0.93	0.92	0.87	92.40	0.98	0.98 (0.97-0.99)	2
KNN <sup>b</sup> (2)	700/300	0.87	0.90	0.84	89.20	0.85	0.85 (0.83-0.87)	1, 2
LR <sup>c</sup> (3)	700/300	0.90	0.90	0.90	92.40	0.98	0.98 (0.97-0.98)	2

<sup>a</sup>AUC: area under the curve.

<sup>b</sup>KNN: k-nearest neighbors.

<sup>c</sup>LR: logistic regression.

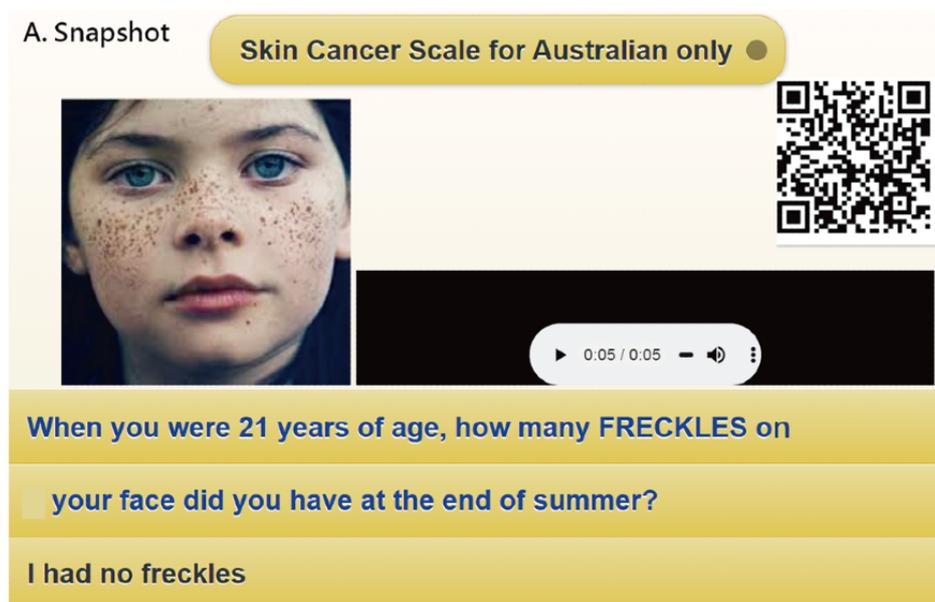
### Task 3: Developing an App for SC Classification

A screenshot obtained from a mobile phone used to respond to the questions is shown in Figure 4, the CAT process is shown in Figure 5, and the assessment results are shown in Figure 6. In this example, we can see that the item-by-item CAT process

is displayed in Figure 5, and the patient has a high probability (0.88) of developing SC, as shown in Figure 6.

Readers are invited to scan the QR code in Figure 4 and practice the web-based CAT on their own. The CAT process is shown in Figure 5. The assessment of the calibration plot is shown in Figure 6.

**Figure 4.** Snapshot of skin cancer assessment on smartphones from the web-based CAT model [50] or clicking the QR Code.



We developed the CAT-based app for classifying SC in adults. The CAT process was demonstrated item by item and is shown in the 3 panels of Figure 5. Person  $\theta$  is the provisional ability (eg, the third column in the top panel of Figure 5 or the blue line in the middle panel of Figure 5) estimated by the CAT module (equation 15).

The SEs (equation 16) are along the orange line in the middle panel of Figure 5 (or the dotted lines in the top panel of Figure 5). We can see that the more items responded to by a patient, the smaller the SEs will be. The SE was generated by the formula  $1/\sqrt{(\text{Information}[i])}$  (equation 17), where  $i$  refers to the CAT items responded to by a patient.

In addition, the item difficulties (shown in Table 1) are along the green line in the middle panel of Figure 5. The residual is derived from the difference (observed – expected; bottom panel of Figure 5). The Z score ( $i$ ) along the brown line is computed

using equation (21), which equals the squared variance ( $i$ ) shown in the bottom panel of Figure 5.

CAT will stop if the residual value is  $< 0.05$ . The correlation coefficient between the CAT estimated measures and the step series numbers using the last 5 estimated  $\theta$  values was computed. A flatter  $\theta$  trend indicates a higher probability of a person's measure converging to the final estimation.

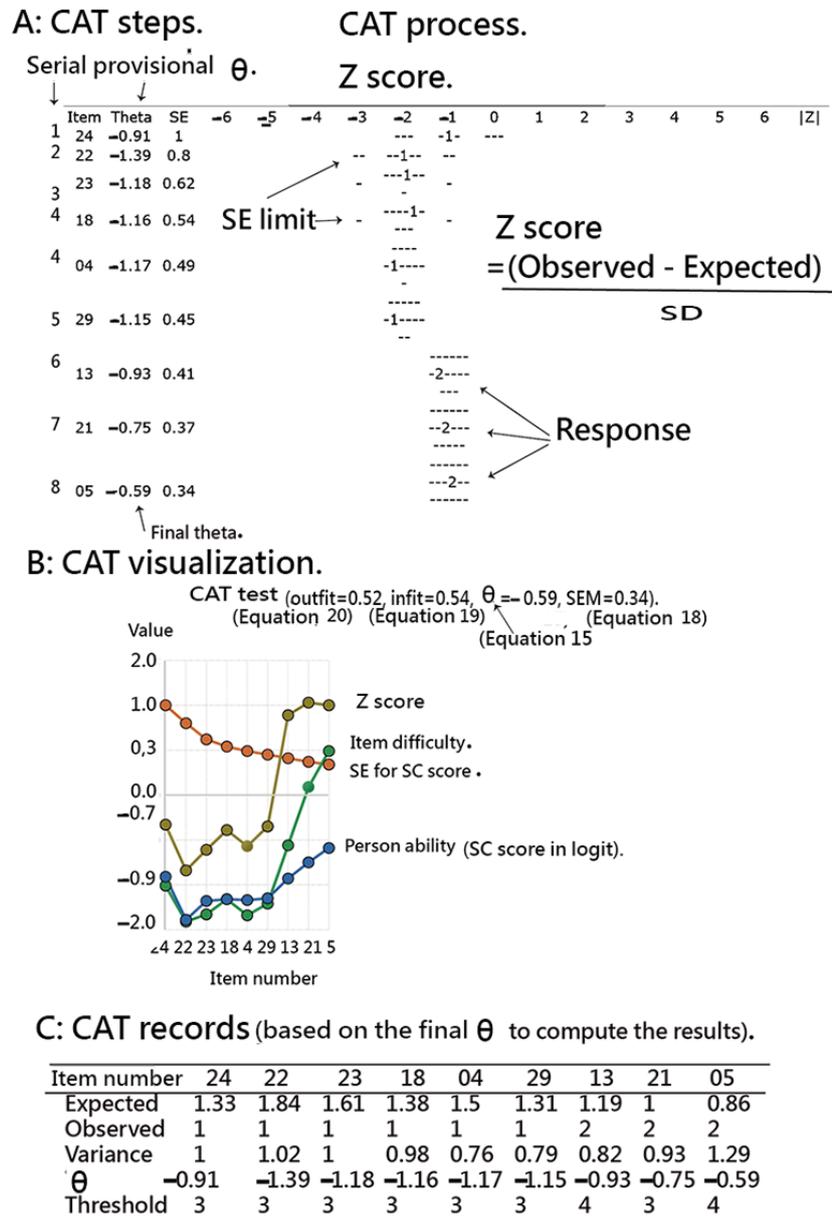
It is worth noting that a person's MNSQs (ie, infit and outfit at the top of the middle panel in Figure 5) are generated by the formula in equations (19) and (20). If the value of the outfit is  $> 2.0$  [51], the person's response pattern is significantly aberrant beyond the model's expectation. In the example shown in the middle panel of Figure 5, we can see that the patient's response pattern with outfit MNSQ (0.52, less than the cutoff point of 2.0) and the  $t$  statistic ( $-0.95 = [\ln(0.585) + 0.585 - 1] \times \sqrt{\frac{9-1}{8}}$ ),

where  $\nu = 0.52 \times 9/[9 - 1]$  based on equations (22) and (23) meets the expectation of the Rasch model rather well.

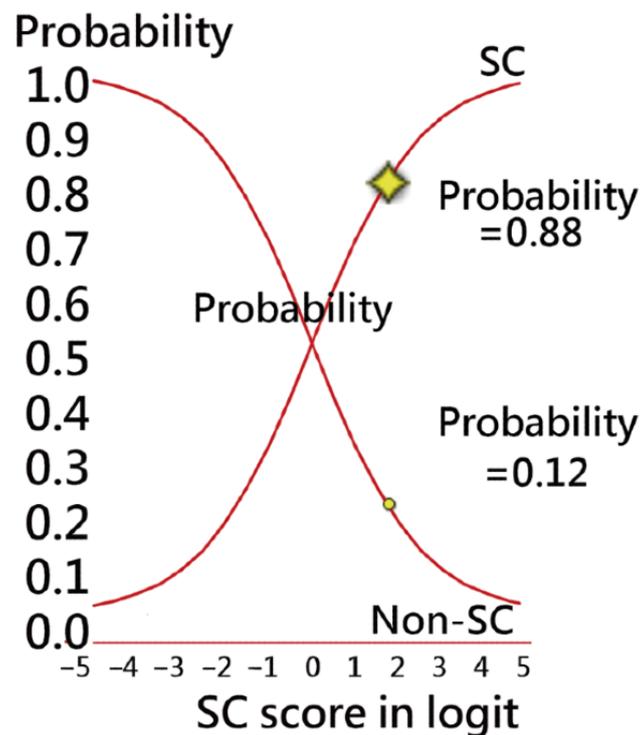
Once the CAT terminates, the resulting example is shown in Figure 6. We can see that the SC+ with a high probability (0.88) is shown on the curve of success from the bottom left to the top right corner. The sum of both probabilities (ie, SC+ and

SC-) equals 1.0. The odds can be computed by the formula  $p/(1 - p) = 0.88/0.12 = 7.33$ , indicating that the patient had an extremely high probability or tendency toward SC+. It is worth noting that CAT substantially reduces participant burden (ie, only 9 items were responded to in the CAT, and 70% [(30 - 9)/30] efficiency gains were from the CAT) without compromising measurement precision.

**Figure 5.** The process in SC-CAT on smartphones with three panels A, B, C denoted by steps, visualizations and records, respectively. CAT: computerized adaptive testing; SC: skin cancer; SEM: standard error of measurement.



**Figure 6.** The result of SC+ assessment with classification and probability on smartphones. SC: skin cancer.



### Web-Based Dashboards Shown on Google Maps

A total of 2 QR codes shown in Figures 3 and 4 (or links [49,50]) are provided for readers who can manipulate the dashboards on their own. In Figures 3 and 4, the animation-type dashboards make the data (eg, feature variables) and the app easier and clearer to understand once the QR Codes are clicked on.

## Discussion

### Principal Findings

We built a CAT-based model via a machine learning approach to develop an app to predict the classification of SC and help patients identify SC risk earlier to reduce participant burden and maintain acceptable measurement precision. A total of 1000 cases were simulated based on the item difficulties with a cutoff point of 0.88 logits to determine 2 groups (cancer and noncancer) using Rasch analysis addressed in a previous study [1]. A total of 3 types of machine learning (NB, KNN, and LR) were applied to compare the accuracy and stability of the models in SC classification. We observed that (1) the 30-item KNN model yielded higher AUC values of 99% and 91% for the 700 training and 300 testing cases, respectively, than its 2 counterparts using the hold-out validation but had lower AUC values of 85% (95% CI 83%-87%) in the k-fold cross-validation and (2) an app for patients that predicts SC classification was successfully developed and demonstrated in this study.

### Previous Research Using Computers to Diagnose SC Instead of Classifying SC

Melanoma is considered one of the fastest-growing and most aggressive SCs; it was first described as a “fatal black tumor” by Hippocrates in 5000 BC and was later recognized to have the propensity to metastasize by William Norris in 1820 [52]. It causes most of the deaths from SC. Therefore, timely and accurate recognition of melanoma combined with appropriate treatment regimens could optimize clinical outcomes and avoid potentially fatal metastasis. Although computer-based algorithms have been proposed to develop novel predictors of prognosis and improve the efficiency and diagnostic accuracy of cancer metastasis, significant challenges for SC prediction and classification still remain [52].

For instance, a report that sniffer dogs are able to detect MM at a curable stage was first described in the United Kingdom by William et al [53]. Thereafter, studies focusing on the utility of dog olfaction for screening or diagnosing different medical conditions, such as COVID-19, malignancies, diabetes, Parkinson disease, seizures, certain hormonal and enzymatic defects [54-67], and melanoma [53], ensued. Machine learning models based on CNNs were applied to extract the region of interest of the skin lesion data set and showed that training CNN models with the region of interest–extracted data set could improve the accuracy of the prediction [55-57].

A mobile CAT was developed to help people efficiently assess their SC risk [1]. However, no such classification of SC using machine learning was provided to readers before, as we did in Figure 4 of this study. This mobile assessment could be used to quickly estimate a person’s SC risk and educate patients about the need to implement skin protection and promote

self-examination of the skin [68-70]. In particular, patients with a history of SC had a higher mean score of responses than those without a history of SC [1].

### Animation-Type CAT Module to Increase Health Literacy for Patients

Patients' health literacy (eg, understanding their own SC risk) is increasingly considered a critical factor affecting patient-physician communication and health outcomes [71]. Populations with below-basic or basic health literacy are less likely to obtain health issue-related information from traditional printed sources such as newspapers, magazines, books, or brochures than those with higher health literacy [72]. A brief CAT, such as the one we developed in this study, could be used to inform people quickly about their potential risk of SC and help these individuals engage in sun-protective behaviors.

This CAT module is a practical tool that can efficiently identify suitable item subsets for each individual and, therefore, maximize the efficiency and precision of the entire testing process. Through CAT, it was found that it can save up to 42% (or more) of test length and achieve a very similar degree of measurement precision as a non-CAT. This is consistent with the literature [73-76].

The tool offers diagnostics that can help practitioners assess whether responses are distorted or abnormal. For example, outfit mean square values of  $\geq 2.0$  suggest an unusual response [51]. If responses do not fit well with the model's requirement, they can be highlighted for suspected cheating, careless responding, lucky guessing, creative responding, or random responding [74]. Otherwise, one can take follow-up action (eg, medical consultation) to recheck the reasons for unexpected responses to questions [8,77,78] if the result shows a high cancer risk. Readers are invited to run the SC-CAT mobile app through the QR code, as shown in Figure 4.

### Strengths and Features of This Study

There are two major forms of standardized assessments in clinical settings [79]: (1) a traditional self-administered questionnaire and (2) a rapid short-form scale [80]. Each has its own advantages and shortcomings. Traditional pencil-and-paper questionnaires require higher financial investment and have a substantial burden on respondents resulting from the following rationale: participants need to answer questions that do not provide additional information about their personal risk of certain diseases to achieve adequate precision measurement [20]. In contrast, by administering items that are most informative for the examinee, the CAT can provide precise measurement of an examinee's proficiency with the fewest possible items and then terminate at an appropriate number of items according to the required person reliability [1] (equation 18).

Second, not all questions were answered in the CAT. In contrast to those using the mean value [20] over the entire data set to fill in the missing values, we applied the expected value in the model for each unanswered question to fill in the missing data, as done in previous studies [13,24,25]. By doing so, the expected responses and model parameters can be applied to classify the SC groups. To date, we have not seen anyone using CAT

combined with machine learning to classify SC in the literature, which is a breakthrough and the second feature of this study.

Third, as with all forms of web-based technology, advances in mobile health and health communication technology are rapidly emerging [21]. The use of mobile web-based CAT is promising and worth implementing in many fields for the assessment of health issues. The CAT graphical representations shown in Figure 4 are modern and innovative in academics.

Few studies have used machine learning to perform NB, KNN, and LR on Microsoft Excel, as we did in this study. These modules are provided in Multimedia Appendix 1, which is the fourth feature of this study.

We applied the LN algorithm along with the model's parameters to design a routine on an app that is used to classify individual SCs (Figure 6), which is the fifth feature of this study. We have not seen any such SC-CAT combined with LN implemented on mobile phones before.

Different results were found when comparing the model accuracy of the AUC between the hold-out validation and the k-fold cross-validation (Tables 2, 3, and 4), which might be attributed to the small sample size (eg, 1000) used in this study. The evidence providing the k-fold cross-validation to improve the strength and confidence in the models' evaluation is the sixth feature of this study.

### Limitations and Future Studies

Our study has some limitations. First, although the psychometric properties of the 30-item SC assessment have been validated for measuring SC risk [1], there is no evidence to support that the 30-item SC assessment is suitable for users outside of Australia. We recommend additional studies using their own database of SC assessment to estimate the item parameters and see whether a difference exists.

Second, although the Bayesian model performed better than the other 2 models (KNN and LR), CAT was incorporated with LR instead of the Bayesian model. The reason for this is that LR requires less computation time than the Bayes and KNN algorithms, as the latter uses pair-to-pair comparison in the algorithm. Future studies are encouraged to compare the efficiency and time consumption in computation between different models.

Third, the study was based on an article [1] that used the 30-item SC-CAT module. All the model parameters (ie, item difficulties and step-threshold difficulties) were derived from this study [1]. If any environment or condition is changed (eg, other populations in the country and different ethnicities), the result (eg, the model's parameters) will be different from that of this study. The ethnicity of the study population was also a limitation. It is worth further verifying and investigating different populations and ethnic groups under the concept we used in this study.

Fourth, the SC assessment is a 1-dimensional construct. The item difficulties used to estimate a person's measure were calibrated using Rasch Winsteps software. Traditionally, a person's ability ( $\theta$ ) should be estimated by the CAT method, as previous studies have done [1,10,13,16]. In this study, the

SC group should be further classified (eg, transforming the log odds to probability in LR and determining the SC group by observing the probability greater or less than 0.5). Different models applied to CAT will use disparate classification schemes. Future studies should be cautious on this matter.

Fifth, readers are encouraged to access the app by scanning the QR code in [Figure 4](#). Professional practical apps should be further developed for Android and iOS systems in the future.

Finally, the study sample was retrieved from the baseline questionnaire in the QSkin Sun and Health study [22]. The data used in this study were simulated from item difficulties calibrated in a previous study [1]. The Rasch partial credit model [24] was used on the simulated data owing to the different number of categories across items. Further research should focus

on whether the psychometric properties of the SC assessment are similar to those of this study if other IRT models are applied.

## Conclusions

The contributions of this study are (1) overcoming the problem of missing responses that limit CAT development when applying the machine learning algorithm, (2) introducing 3 models available on Microsoft Excel and the k-fold cross-validation in Weka software, and (3) demonstrating an app that incorporates Rasch CAT with numerous parameters in LR.

The 30-item SC prediction model, combined with the Rasch web-based CAT, is recommended for classifying SC in adults. An app developed to help patients self-assess SC risk at an early stage is required for application in the future.

## Acknowledgments

The authors would like to thank Enago for the English language review of this manuscript. All authors declare no conflicts of interest.

## Authors' Contributions

TWC conceived and designed the study. TYY and TWC interpreted the data, and FJL monitored the process and the manuscript. TYY and TWC drafted the manuscript. All authors have read the manuscript and have approved the final manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Data deposited at OSF (Open Science Framework) research sharing platform.  
[\[DOCX File , 15 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

K-fold cross validation performed in Weka.  
[\[DOCX File , 14 KB-Multimedia Appendix 2\]](#)

## Multimedia Appendix 3

Detailed information about the three models used in this study.  
[\[DOCX File , 392 KB-Multimedia Appendix 3\]](#)

## References

1. Djaja N, Janda M, Olsen CM, Whiteman DC, Chien T. Estimating skin cancer risk: evaluating mobile computer-adaptive testing. *J Med Internet Res* 2016 Jan 22;18(1):e22 [[FREE Full text](#)] [doi: [10.2196/jmir.4736](https://doi.org/10.2196/jmir.4736)] [Medline: [26800642](https://pubmed.ncbi.nlm.nih.gov/26800642/)]
2. Australian Institute of Health and Welfare. URL: <http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=60129542353> [accessed 2022-02-06]
3. Narayanan D, Saladi R, Fox J. Ultraviolet radiation and skin cancer. *Int J Dermatol* 2010 Sep;49(9):978-986. [doi: [10.1111/j.1365-4632.2010.04474.x](https://doi.org/10.1111/j.1365-4632.2010.04474.x)] [Medline: [20883261](https://pubmed.ncbi.nlm.nih.gov/20883261/)]
4. Global Solar UV Index: A Practical Guide. Geneva: World Health Organization; 2002.
5. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Denmark: Danmarks Paedagogiske Institut; 1960.
6. Lerdal A, Kottorp A, Gay CL, Grov EK, Lee KA. Rasch analysis of the Beck Depression Inventory-II in stroke survivors: a cross-sectional study. *J Affect Disord* 2014 Apr;158:48-52 [[FREE Full text](#)] [doi: [10.1016/j.jad.2014.01.013](https://doi.org/10.1016/j.jad.2014.01.013)] [Medline: [24655764](https://pubmed.ncbi.nlm.nih.gov/24655764/)]
7. Forkmann T, Boecker M, Wirtz M, Eberle N, Westhofen M, Schauer P, et al. Development and validation of the Rasch-based Depression Screening (DESC) using Rasch analysis and structural equation modelling. *J Behav Ther Exp Psychiatry* 2009 Sep;40(3):468-478. [doi: [10.1016/j.jbtep.2009.06.003](https://doi.org/10.1016/j.jbtep.2009.06.003)] [Medline: [19589499](https://pubmed.ncbi.nlm.nih.gov/19589499/)]

8. Sauer S, Ziegler M, Schmitt M. Rasch analysis of a simplified Beck Depression Inventory. *Personal Individual Differences* 2013 Mar;54(4):530-535. [doi: [10.1016/j.paid.2012.10.025](https://doi.org/10.1016/j.paid.2012.10.025)]
9. Chien T, Wang W, Huang S, Lai W, Chow JC. A web-based computerized adaptive testing (CAT) to assess patient perception in hospitalization. *J Med Internet Res* 2011 Aug 15;13(3):e61 [FREE Full text] [doi: [10.2196/jmir.1785](https://doi.org/10.2196/jmir.1785)] [Medline: [21844001](https://pubmed.ncbi.nlm.nih.gov/21844001/)]
10. 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 Feb 17;16(2):e50 [FREE Full text] [doi: [10.2196/jmir.2819](https://doi.org/10.2196/jmir.2819)] [Medline: [24534113](https://pubmed.ncbi.nlm.nih.gov/24534113/)]
11. Djaja N, Youl P, Aitken J, Janda M. Evaluation of a skin self examination attitude scale using an item response theory model approach. *Health Qual Life Outcomes* 2014 Dec 24;12:189 [FREE Full text] [doi: [10.1186/s12955-014-0189-x](https://doi.org/10.1186/s12955-014-0189-x)] [Medline: [25539671](https://pubmed.ncbi.nlm.nih.gov/25539671/)]
12. Bjorner JB, Kreiner S, Ware JE, Damsgaard MT, Bech P. Differential item functioning in the Danish translation of the SF-36. *J Clin Epidemiol* 1998 Nov;51(11):1189-1202. [doi: [10.1016/s0895-4356\(98\)00111-5](https://doi.org/10.1016/s0895-4356(98)00111-5)] [Medline: [9817137](https://pubmed.ncbi.nlm.nih.gov/9817137/)]
13. Ma S, Chou W, Chien T, Chow JC, Yeh Y, Chou P, 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 May 20;8(5):e16747 [FREE Full text] [doi: [10.2196/16747](https://doi.org/10.2196/16747)] [Medline: [32432557](https://pubmed.ncbi.nlm.nih.gov/32432557/)]
14. Lord FM. Practical applications of item characteristic curve theory. *J Educ Measurement* 1977 Jun;14(2):117-138. [doi: [10.1111/j.1745-3984.1977.tb00032.x](https://doi.org/10.1111/j.1745-3984.1977.tb00032.x)]
15. Lord F. Applications of Item Response Theory To Practical Testing Problems. Milton Park, Abingdon-on-Thames, Oxfordshire United Kingdom: Taylor & Francis; 1980.
16. Ma S, Wang H, Chien T. A new technique to measure online bullying: online computerized adaptive testing. *Ann Gen Psychiatry* 2017;16:26 [FREE Full text] [doi: [10.1186/s12991-017-0149-z](https://doi.org/10.1186/s12991-017-0149-z)] [Medline: [28680455](https://pubmed.ncbi.nlm.nih.gov/28680455/)]
17. Lee Y, Chou W, Chien T, Chou P, Yeh Y, Lee H. An app developed for detecting nurse burnouts using the convolutional neural networks in Microsoft excel: population-based questionnaire study. *JMIR Med Inform* 2020 May 07;8(5):e16528 [FREE Full text] [doi: [10.2196/16528](https://doi.org/10.2196/16528)] [Medline: [32379050](https://pubmed.ncbi.nlm.nih.gov/32379050/)]
18. Chien T, Lin W. Simulation study of activities of daily living functions using online computerized adaptive testing. *BMC Med Inform Decis Mak* 2016 Oct 10;16(1):130 [FREE Full text] [doi: [10.1186/s12911-016-0370-8](https://doi.org/10.1186/s12911-016-0370-8)] [Medline: [27724939](https://pubmed.ncbi.nlm.nih.gov/27724939/)]
19. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep learning for health informatics. *IEEE J Biomed Health Inform* 2017 Jan;21(1):4-21. [doi: [10.1109/JBHI.2016.2636665](https://doi.org/10.1109/JBHI.2016.2636665)] [Medline: [28055930](https://pubmed.ncbi.nlm.nih.gov/28055930/)]
20. 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 Bioinf* 2018 Nov 1;15(6):1968-1978. [doi: [10.1109/tcbb.2018.2827029](https://doi.org/10.1109/tcbb.2018.2827029)]
21. 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 Jan 13;16(1):e9 [FREE Full text] [doi: [10.2196/jmir.2673](https://doi.org/10.2196/jmir.2673)] [Medline: [24418967](https://pubmed.ncbi.nlm.nih.gov/24418967/)]
22. Olsen CM, Green AC, Neale RE, Webb PM, Cicero RA, Jackman LM, QSkin Study. Cohort profile: the QSkin sun and health study. *Int J Epidemiol* 2012 Aug;41(4):929-92i. [doi: [10.1093/ije/dys107](https://doi.org/10.1093/ije/dys107)] [Medline: [22933644](https://pubmed.ncbi.nlm.nih.gov/22933644/)]
23. Lai P, Chien T. The determination of inflection curve on a given ogive curve using the second order derivative in calculus. *J Bibliographical Analyses Stat* 2021;18(3):31-33 [FREE Full text]
24. Masters GN. A rasch model for partial credit scoring. *Psychometrika* 1982 Jun;47(2):149-174. [doi: [10.1007/BF02296272](https://doi.org/10.1007/BF02296272)]
25. Tang X, Shu Y, Liu W, Li J, Liu M, Yu H. An optimized weighted naïve Bayes method for flood risk assessment. *Risk Anal* 2021 Dec;41(12):2301-2321. [doi: [10.1111/risa.13743](https://doi.org/10.1111/risa.13743)] [Medline: [33928661](https://pubmed.ncbi.nlm.nih.gov/33928661/)]
26. Viana Dos Santos Santana Á, Cm da Silveira A, Sobrinho A, Chaves E Silva L, Dias da Silva L, Santos DF, et al. Classification models for COVID-19 test prioritization in Brazil: machine learning approach. *J Med Internet Res* 2021 Apr 08;23(4):e27293 [FREE Full text] [doi: [10.2196/27293](https://doi.org/10.2196/27293)] [Medline: [33750734](https://pubmed.ncbi.nlm.nih.gov/33750734/)]
27. Golpour P, Ghayour-Mobarhan M, Saki A, Esmaily H, Taghipour A, Tajfard M, et al. Comparison of support vector machine, naïve Bayes and logistic regression for assessing the necessity for coronary angiography. *Int J Environ Res Public Health* 2020 Sep 04;17(18):6449 [FREE Full text] [doi: [10.3390/ijerph17186449](https://doi.org/10.3390/ijerph17186449)] [Medline: [32899733](https://pubmed.ncbi.nlm.nih.gov/32899733/)]
28. Gholizadeh P, Esmaili B. Developing a multi-variate logistic regression model to analyze accident scenarios: case of electrical contractors. *Int J Environ Res Public Health* 2020 Jul 06;17(13):4852 [FREE Full text] [doi: [10.3390/ijerph17134852](https://doi.org/10.3390/ijerph17134852)] [Medline: [32640549](https://pubmed.ncbi.nlm.nih.gov/32640549/)]
29. Nhu V, Shirzadi A, Shahabi H, Singh SK, Al-Ansari N, Clague JJ, et al. Shallow landslide susceptibility mapping: a comparison between logistic model tree, logistic regression, naïve bayes tree, artificial neural network, and support vector machine algorithms. *Int J Environ Res Public Health* 2020 Apr 16;17(8):2749 [FREE Full text] [doi: [10.3390/ijerph17082749](https://doi.org/10.3390/ijerph17082749)] [Medline: [32316191](https://pubmed.ncbi.nlm.nih.gov/32316191/)]
30. Choi Y, Boo Y. Comparing logistic regression models with alternative machine learning methods to predict the risk of drug intoxication mortality. *Int J Environ Res Public Health* 2020 Jan 31;17(3):897 [FREE Full text] [doi: [10.3390/ijerph17030897](https://doi.org/10.3390/ijerph17030897)] [Medline: [32023993](https://pubmed.ncbi.nlm.nih.gov/32023993/)]
31. Wu L, Deng F, Xie Z, Hu S, Shen S, Shi J, et al. Spatial analysis of severe fever with thrombocytopenia syndrome virus in china using a geographically weighted logistic regression model. *Int J Environ Res Public Health* 2016 Nov 11;13(11):1125 [FREE Full text] [doi: [10.3390/ijerph13111125](https://doi.org/10.3390/ijerph13111125)] [Medline: [27845737](https://pubmed.ncbi.nlm.nih.gov/27845737/)]

32. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I. The WEKA data mining software. SIGKDD Explor Newsl 2009 Nov 16;11(1):10-18 [[FREE Full text](#)] [doi: [10.1145/1656274.1656278](https://doi.org/10.1145/1656274.1656278)]
33. Rere LM, Fanany MI, Arymurthy AM. Metaheuristic algorithms for convolution neural network. Comput Intell Neurosci 2016;2016:1537325 [[FREE Full text](#)] [doi: [10.1155/2016/1537325](https://doi.org/10.1155/2016/1537325)] [Medline: [27375738](https://pubmed.ncbi.nlm.nih.gov/27375738/)]
34. Chou P, Chien T, Yang T, Yeh Y, Chou W, Yeh C. Predicting active NBA players most likely to be inducted into the basketball hall of famers using artificial neural networks in Microsoft excel: development and usability study. Int J Environ Res Public Health 2021 Apr 16;18(8):4256 [[FREE Full text](#)] [doi: [10.3390/ijerph18084256](https://doi.org/10.3390/ijerph18084256)] [Medline: [33923846](https://pubmed.ncbi.nlm.nih.gov/33923846/)]
35. Tey S, Liu C, Chien T, Hsu C, Chan K, Chen C, et al. Predicting the 14-day hospital readmission of patients with pneumonia using artificial neural networks (ANN). Int J Environ Res Public Health 2021 May 12;18(10):5110 [[FREE Full text](#)] [doi: [10.3390/ijerph18105110](https://doi.org/10.3390/ijerph18105110)] [Medline: [34065894](https://pubmed.ncbi.nlm.nih.gov/34065894/)]
36. Hamling J, Lee P, Weitkunat R, Ambühl M. Facilitating meta-analyses by deriving relative effect and precision estimates for alternative comparisons from a set of estimates presented by exposure level or disease category. Stat Med 2008 Mar 30;27(7):954-970. [doi: [10.1002/sim.3013](https://doi.org/10.1002/sim.3013)] [Medline: [17676579](https://pubmed.ncbi.nlm.nih.gov/17676579/)]
37. Chen C, Wang L, Kuo H, Fang Y, Lee H. Significant effects of late evening snack on liver functions in patients with liver cirrhosis: a meta-analysis of randomized controlled trials. J Gastroenterol Hepatol 2019 Jul;34(7):1143-1152. [doi: [10.1111/jgh.14665](https://doi.org/10.1111/jgh.14665)] [Medline: [30883904](https://pubmed.ncbi.nlm.nih.gov/30883904/)]
38. Lalkhen AG, McCluskey A. Statistics V: introduction to clinical trials and systematic reviews. Continuing Educ Anaesthesia Critical Care Pain 2008 Aug;8(4):143-146. [doi: [10.1093/bjaceaccp/mkn023](https://doi.org/10.1093/bjaceaccp/mkn023)]
39. Yan Y, Chien T. The use of forest plot to identify article similarity and differences in characteristics between journals using medical subject headings terms: a protocol for bibliometric study. Medicine (Baltimore) 2021 Feb 12;100(6):e24610 [[FREE Full text](#)] [doi: [10.1097/MD.00000000000024610](https://doi.org/10.1097/MD.00000000000024610)] [Medline: [33578568](https://pubmed.ncbi.nlm.nih.gov/33578568/)]
40. Wright BD, Douglas GA. Conditional versus unconditional procedures for sample-free item analysis. Educ Psychol Measurement 2016 Jul 02;37(3):573-586. [doi: [10.1177/001316447703700301](https://doi.org/10.1177/001316447703700301)]
41. Wright B, Douglas G. Estimating Rasch (person, ability, theta) measures with known dichotomous item difficulties: anchored maximum likelihood estimation (AMLE). Rasch Measurement Transactions. URL: <https://www.rasch.org/rmt/rmt102t.htm> [accessed 2022-02-06]
42. Ludlow L, Haley K. Newton: pinball wizard? Popular Measure 1999;2(1):5 [[FREE Full text](#)]
43. Wright BD, Stone MH. Measurement Essentials 2nd Edition. Wilmington, Delaware: Wide Range, Inc; 1999.
44. Chien T, Shao Y. Rasch analysis for continuous variables. Rasch Measurement Transact 2016;30(1):1574-1576.
45. Chien T, Shao Y, Kuo S. Development of a Microsoft Excel tool for one-parameter Rasch model of continuous items: an application to a safety attitude survey. BMC Med Res Methodol 2017 Jan 10;17(1):4 [[FREE Full text](#)] [doi: [10.1186/s12874-016-0276-2](https://doi.org/10.1186/s12874-016-0276-2)] [Medline: [28068901](https://pubmed.ncbi.nlm.nih.gov/28068901/)]
46. Linacre J. Computer-adaptive testing: a methodology whose time has come. MESA Memorandum. URL: <https://www.rasch.org/memo69.htm> [accessed 2022-02-06]
47. Wright B, Stone M. Best Test Design Rasch Measurement. Chicago, IL: Mesa Press; 1979.
48. Wright B. Practical adaptive testing. Rasch Measurement Transact 1988;2(2):21.
49. Chien T. iHELP system. URL: <http://www.healthup.org.tw/gps/skincancer2021.htm> [accessed 2022-02-06]
50. Web-based computerized adaptive testing model for skin cancer assessment on smartphones. iHELP. URL: [http://www.healthup.org.tw/irs/irsin\\_e.asp?type1=15](http://www.healthup.org.tw/irs/irsin_e.asp?type1=15) [accessed 2022-02-06]
51. Linacre J. Optimizing rating scale category effectiveness. J Appl Meas 2002;3(1):85-106. [Medline: [11997586](https://pubmed.ncbi.nlm.nih.gov/11997586/)]
52. Alix-Panabieres C, Magliocco A, Cortes-Hernandez LE, Eslami- S, Franklin D, Messina JL. Detection of cancer metastasis: past, present and future. Clin Exp Metastasis 2021 May 07 (forthcoming). [doi: [10.1007/s10585-021-10088-w](https://doi.org/10.1007/s10585-021-10088-w)] [Medline: [33961169](https://pubmed.ncbi.nlm.nih.gov/33961169/)]
53. Williams H, Pembroke A. Sniffer dogs in the melanoma clinic? Lancet 1989 Apr 01;1(8640):734. [doi: [10.1016/s0140-6736\(89\)92257-5](https://doi.org/10.1016/s0140-6736(89)92257-5)] [Medline: [2564551](https://pubmed.ncbi.nlm.nih.gov/2564551/)]
54. Eskandari E, Ahmadi Marzaleh M, Roudgari H, Hamidi Farahani R, Nezami-Asl A, Laripour R, et al. Sniffer dogs as a screening/diagnostic tool for COVID-19: a proof of concept study. BMC Infect Dis 2021 Mar 05;21(1):243 [[FREE Full text](#)] [doi: [10.1186/s12879-021-05939-6](https://doi.org/10.1186/s12879-021-05939-6)] [Medline: [33673823](https://pubmed.ncbi.nlm.nih.gov/33673823/)]
55. Boedeker E, Friedel G, Walles T. Sniffer dogs as part of a bimodal bionic research approach to develop a lung cancer screening. Interact Cardiovasc Thorac Surg 2012 May;14(5):511-515 [[FREE Full text](#)] [doi: [10.1093/icvts/ivr070](https://doi.org/10.1093/icvts/ivr070)] [Medline: [22345057](https://pubmed.ncbi.nlm.nih.gov/22345057/)]
56. Zanddzari H, Nguyen N, Zeinali B, Chang JM. A new preprocessing approach to improve the performance of CNN-based skin lesion classification. Med Biol Eng Comput 2021 May;59(5):1123-1131. [doi: [10.1007/s11517-021-02355-5](https://doi.org/10.1007/s11517-021-02355-5)] [Medline: [33904008](https://pubmed.ncbi.nlm.nih.gov/33904008/)]
57. Ningrum DN, Yuan S, Kung W, Wu C, Tzeng I, Huang C, et al. Deep learning classifier with patient's metadata of dermoscopic images in malignant melanoma detection. J Multidiscip Healthc 2021;14:877-885 [[FREE Full text](#)] [doi: [10.2147/JMDH.S306284](https://doi.org/10.2147/JMDH.S306284)] [Medline: [33907414](https://pubmed.ncbi.nlm.nih.gov/33907414/)]

58. Alheejawi S, Berendt R, Jha N, Maity SP, Mandal M. Automated proliferation index calculation for skin melanoma biopsy images using machine learning. *Comput Med Imaging Graph* 2021 Apr;89:101893. [doi: [10.1016/j.compmedimag.2021.101893](https://doi.org/10.1016/j.compmedimag.2021.101893)] [Medline: [33752078](https://pubmed.ncbi.nlm.nih.gov/33752078/)]
59. Welsh JS. Olfactory detection of human bladder cancer by dogs: another cancer detected by "pet scan". *BMJ* 2004 Nov 27;329(7477):1286-1287 [FREE Full text] [doi: [10.1136/bmj.329.7477.1286-b](https://doi.org/10.1136/bmj.329.7477.1286-b)] [Medline: [15564264](https://pubmed.ncbi.nlm.nih.gov/15564264/)]
60. Urbanová L, Vyhnančková V, Krisová S, Pacík D, Nečas A. Intensive training technique utilizing the dog's olfactory abilities to diagnose prostate cancer in men. *Acta Vet Brno* 2015 Mar 19;84(1):77-82. [doi: [10.2754/avb201585010077](https://doi.org/10.2754/avb201585010077)]
61. Lippi G, Cervellin G. Canine olfactory detection of cancer versus laboratory testing: myth or opportunity? *Clin Chem Lab Med* 2012 Mar;50(3):435-439. [doi: [10.1515/CCLM.2011.672](https://doi.org/10.1515/CCLM.2011.672)] [Medline: [21790506](https://pubmed.ncbi.nlm.nih.gov/21790506/)]
62. Elliker K, Williams H. Detection of skin cancer odours using dogs: a step forward in melanoma detection training and research methodologies. *Br J Dermatol* 2016 Nov;175(5):851-852. [doi: [10.1111/bjd.15030](https://doi.org/10.1111/bjd.15030)] [Medline: [27790682](https://pubmed.ncbi.nlm.nih.gov/27790682/)]
63. Willis CM, Church SM, Guest CM, Cook WA, McCarthy N, Bransbury AJ, et al. Olfactory detection of human bladder cancer by dogs: proof of principle study. *BMJ* 2004 Sep 25;329(7468):712 [FREE Full text] [doi: [10.1136/bmj.329.7468.712](https://doi.org/10.1136/bmj.329.7468.712)] [Medline: [15388612](https://pubmed.ncbi.nlm.nih.gov/15388612/)]
64. Kane E. Cancer-sniffing dogs: how canine scent detection could transform human medicine. *dvm360*. URL: <https://www.dvm360.com/view/cancer-sniffing-dogs-how-canine-scent-detection-could-transform-human-medicine> [accessed 2022-02-06]
65. McCulloch M, Jezierski T, Broffman M, Hubbard A, Turner K, Janecki T. Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. *Integr Cancer Ther* 2006 Mar;5(1):30-39 [FREE Full text] [doi: [10.1177/1534735405285096](https://doi.org/10.1177/1534735405285096)] [Medline: [16484712](https://pubmed.ncbi.nlm.nih.gov/16484712/)]
66. Ehmann R, Boedeker E, Friedrich U, Sagert J, Dippon J, Friedel G, et al. Canine scent detection in the diagnosis of lung cancer: revisiting a puzzling phenomenon. *Eur Respir J* 2012 Mar;39(3):669-676 [FREE Full text] [doi: [10.1183/09031936.00051711](https://doi.org/10.1183/09031936.00051711)] [Medline: [21852337](https://pubmed.ncbi.nlm.nih.gov/21852337/)]
67. Los EA, Ramsey KL, Guttman-Bauman I, Ahmann AJ. Reliability of trained dogs to alert to hypoglycemia in patients with type 1 diabetes. *J Diabetes Sci Technol* 2017 May;11(3):506-512 [FREE Full text] [doi: [10.1177/1932296816666537](https://doi.org/10.1177/1932296816666537)] [Medline: [27573791](https://pubmed.ncbi.nlm.nih.gov/27573791/)]
68. Robinson JK, Gaber R, Hultgren B, Eilers S, Blatt H, Stapleton J, et al. Skin self-examination education for early detection of melanoma: a randomized controlled trial of Internet, workbook, and in-person interventions. *J Med Internet Res* 2014 Jan 13;16(1):e7 [FREE Full text] [doi: [10.2196/jmir.2883](https://doi.org/10.2196/jmir.2883)] [Medline: [24418949](https://pubmed.ncbi.nlm.nih.gov/24418949/)]
69. Brady MS, Oliveria SA, Christos PJ, Berwick M, Coit DG, Katz J, et al. Patterns of detection in patients with cutaneous melanoma. *Cancer* 2000 Jul 15;89(2):342-347. [doi: [10.1002/1097-0142\(20000715\)89:2<342::aid-cnrc19>3.0.co;2-p](https://doi.org/10.1002/1097-0142(20000715)89:2<342::aid-cnrc19>3.0.co;2-p)]
70. Berwick M, Begg CB, Fine JA, Roush GC, Barnhill RL. Screening for cutaneous melanoma by skin self-examination. *J Natl Cancer Inst* 1996 Jan 03;88(1):17-23. [doi: [10.1093/jnci/88.1.17](https://doi.org/10.1093/jnci/88.1.17)] [Medline: [8847720](https://pubmed.ncbi.nlm.nih.gov/8847720/)]
71. Williams MV, Davis T, Parker RM, Weiss BD. The role of health literacy in patient-physician communication. *Fam Med* 2002 May;34(5):383-389. [Medline: [12038721](https://pubmed.ncbi.nlm.nih.gov/12038721/)]
72. Cutilli C, Bennett I. Understanding the health literacy of America: results of the National Assessment of Adult Literacy. *Orthop Nurs* 2009;28(1):27-32; quiz 33 [FREE Full text] [doi: [10.1097/01.NOR.0000345852.22122.d6](https://doi.org/10.1097/01.NOR.0000345852.22122.d6)] [Medline: [19190475](https://pubmed.ncbi.nlm.nih.gov/19190475/)]
73. Pedersen PM, Jørgensen HS, Nakayama H, Raaschou HO, Olsen TS. Comprehensive assessment of activities of daily living in stroke. The Copenhagen stroke study. *Archives Physical Med Rehab* 1997 Feb;78(2):161-165. [doi: [10.1016/s0003-9993\(97\)90258-6](https://doi.org/10.1016/s0003-9993(97)90258-6)]
74. Wainer H, Dorans N, Flaughner R, Green B, Mislevy R. *Computerized Adaptive Testing A Primer*. Milton Park, Abingdon-on-Thames, Oxfordshire United Kingdom: Taylor & Francis; 1990.
75. Weiss DJ, McBride JR. Bias and information of bayesian adaptive testing. *Applied Psychol Measure* 2016 Jul 27;8(3):273-285. [doi: [10.1177/014662168400800303](https://doi.org/10.1177/014662168400800303)]
76. Chien T, Wu H, Wang W, Castillo R, Chou W. Reduction in patient burdens with graphical computerized adaptive testing on the ADL scale: tool development and simulation. *Health Qual Life Outcomes* 2009 May 05;7:39 [FREE Full text] [doi: [10.1186/1477-7525-7-39](https://doi.org/10.1186/1477-7525-7-39)] [Medline: [19416521](https://pubmed.ncbi.nlm.nih.gov/19416521/)]
77. Eack SM, Singer JB, Greeno CG. Screening for anxiety and depression in community mental health: the beck anxiety and depression inventories. *Community Ment Health J* 2008 Dec;44(6):465-474. [doi: [10.1007/s10597-008-9150-y](https://doi.org/10.1007/s10597-008-9150-y)] [Medline: [18516678](https://pubmed.ncbi.nlm.nih.gov/18516678/)]
78. Shear MK, Greeno C, Kang J, Ludewig D, Frank E, Swartz HA, et al. Diagnosis of nonpsychotic patients in community clinics. *Am J Psychiatry* 2000 Apr;157(4):581-587. [doi: [10.1176/appi.ajp.157.4.581](https://doi.org/10.1176/appi.ajp.157.4.581)] [Medline: [10739417](https://pubmed.ncbi.nlm.nih.gov/10739417/)]
79. Ramirez Basco M, Bostic JQ, Davies D, Rush AJ, Witte B, Hendrickse W, et al. Methods to improve diagnostic accuracy in a community mental health setting. *Am J Psychiatry* 2000 Oct;157(10):1599-1605. [doi: [10.1176/appi.ajp.157.10.1599](https://doi.org/10.1176/appi.ajp.157.10.1599)] [Medline: [11007713](https://pubmed.ncbi.nlm.nih.gov/11007713/)]
80. De Beurs DP, de Vries AL, de Groot MH, de Keijser J, Kerkhof AJ. Applying computer adaptive testing to optimize online assessment of suicidal behavior: a simulation study. *J Med Internet Res* 2014 Sep 11;16(9):e207 [FREE Full text] [doi: [10.2196/jmir.3511](https://doi.org/10.2196/jmir.3511)] [Medline: [25213259](https://pubmed.ncbi.nlm.nih.gov/25213259/)]

## Abbreviations

**AUC:** area under the curve  
**CAT:** computerized adaptive testing  
**CNN:** convolutional neural network  
**FN:** false negative  
**FP:** false positive  
**IBk:** instance-based learner  
**IRT:** item response theory  
**KNN:** k-nearest neighbors  
**LR:** logistic regression  
**MM:** malignant melanoma  
**MNSQ:** mean square error  
**NB:** naïve Bayes  
**NMSC:** nonmelanoma skin cancer  
**SC:** skin cancer  
**SC-CAT:** skin cancer-computerized adaptive testing  
**SEM:** standard error of measurement  
**TN:** true negative  
**TP:** true positive

*Edited by C Lovis; submitted 18.08.21; peer-reviewed by Á Sobrinho, IS Tzeng; comments to author 03.10.21; revised version received 08.11.21; accepted 10.01.22; published 09.03.22*

*Please cite as:*

*Yang TY, Chien TW, Lai FJ*

*Web-Based Skin Cancer Assessment and Classification Using Machine Learning and Mobile Computerized Adaptive Testing in a Rasch Model: Development Study*

*JMIR Med Inform 2022;10(3):e33006*

*URL: <https://medinform.jmir.org/2022/3/e33006>*

*doi: [10.2196/33006](https://doi.org/10.2196/33006)*

*PMID: [35262505](https://pubmed.ncbi.nlm.nih.gov/35262505/)*

©Ting-Ya Yang, Tsair-Wei Chien, Feng-Jie Lai. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org/>), 09.03.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.