

RESEARCH

Open Access



Machine learning-assisted rapid determination for traditional Chinese Medicine Constitution

Wen Sun¹ , Minghua Bai², Ji Wang², Bei Wang³, Yixing Liu^{1*}, Qi Wang^{2*} and Dongran Han^{3*}

Abstract

The aim of this study was to develop a machine learning-assisted rapid determination methodology for traditional Chinese Medicine Constitution. Based on the Constitution in Chinese Medicine Questionnaire (CCMQ), the most applied diagnostic instrument for assessing individuals' constitutions, we employed automated supervised machine learning algorithms (i.e., Tree-based Pipeline Optimization Tool; TPOT) on all the possible item combinations for each subscale and an unsupervised machine learning algorithm (i.e., variable clustering; varclus) on the whole scale to select items that can best predict body constitution (BC) classifications or BC scores. By utilizing subsets of items selected based on TPOT and corresponding machine learning algorithms, the accuracies of BC classifications prediction ranged from 0.819 to 0.936, with the root mean square errors of BC scores prediction stabilizing between 6.241 and 9.877. Overall, the results suggested that the automated machine learning algorithms performed better than the varclus algorithm for item selection. Additionally, based on an automated machine learning item selection procedure, we provided the top three ranked item combinations with each possible subscale length, along with their corresponding algorithms for predicting BC classification and severity. This approach could accommodate the needs of different practitioners in traditional Chinese medicine for rapid constitution determination.

Keywords Automated machine learning (AutoML), Unsupervised machine learning, Constitution in Chinese Medicine Questionnaire (CCMQ), Tree-based Pipeline Optimization Tool (TPOT), Variable clustering (varclus)

Introduction

The Yellow Emperor's Classic of Medicine introduced the concept of "weibing," which refers to the state preceding the onset of disease, essentially describing a condition of subhealth. Within conventional medicine paradigms, the state between health and disease is referred to as the "third state". However, the scope of the "third state" is broad, and its mechanisms are unclear, which complicates targeted intervention strategies. In traditional Chinese medicine (TCM), there are systematic theories and intervention methods specifically designed to address subhealth issues. One of the common theoretical frameworks in TCM for subhealth is described using body constitution (BC), which is relatively stable across an individual's lifespan.

*Correspondence:

Yixing Liu
yixingliu1987@gmail.com

Qi Wang
wangqi710@126.com

Dongran Han
handongr@gmail.com

¹ School of Management, Beijing University of Chinese Medicine, Beijing 100029, China

² School of Traditional Chinese Medicine/National Institute of TCM Constitution and Preventive Medicine, Beijing University of Chinese Medicine, Beijing 100029, China

³ School of Life Sciences, Beijing University of Chinese Medicine, Beisanhuan East Road No. 11, Chaoyang District, Beijing 100029, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

BC arises from a combination of inherent genetic factors (e.g., race, family genetics, prenatal development, etc.) and acquired traits (e.g., dietary nutrition, lifestyle habits, mental state, disease damage, drug treatment, etc.). BC manifests as an individual's physical morphology, physiological functions, psychological conditions, and interactions with environmental factors [1, 2]. Wang systematically analyzed the theoretical origins, formation, and development of body constitution theory based on 108 ancient Chinese traditional classic works and 168 modern documents. Combining current clinical research, Wang identified nine constitutional types: Gentleness Constitution (GTC: calm mentality, strong adaptability to the environment and resilience to illness), Qi-deficiency Constitution (QDC: shortness of breath and low energy), Yang-deficiency Constitution (YaDC: low tolerance to cold weather), Yin-deficiency Constitution (YiDC: insufficient body fluid), Phlegm-dampness Constitution (PDC: tendency to have abdominal obesity), Damp-heat Constitution (DHC: tendency of excessive humidity, overheating, sweating of hands and feet, yellow urine, and loose stools), Blood-stasis Constitution (BSC: tendency towards increased blood viscosity), Qi-stagnation Constitution (QSC: tendency to have depression, anxiety and chest pain), and Special-diathesis Constitution (SDC: tendency to have allergy) [1]. BC is determined by people's observed characteristics (e.g., physiological, psychological, and reactive states) and their connections with the nature, occurrence, development, and outcome of different diseases [3]. Thus, the identification of BC may aid in understanding the pathological mechanisms of certain diseases, guiding therapeutic interventions and prognostic evaluations [4, 5].

TCM practitioners typically assess a patient's constitution through four examinations: observing, listening and smelling, questioning, and palpating. The accuracy of this approach largely depends on practitioners' expertise and experience. To standardize the BC type identification procedure, Wang developed the Constitution in Chinese Medicine Questionnaire (CCMQ), which quantitatively measures the extent to which an individual exhibits a specific BC [6]. The CCMQ has been widely used in clinical research and practice. Furthermore, it has been translated into several languages, including English, Korean, Japanese, and Vietnamese, enhancing its accessibility and applicability across different cultural contexts [7–10].

The CCMQ and its simplified versions have demonstrated robust reliability and validity in assessing body constitutions within the field of TCM [11–14]. However, scoring methods for such questionnaires generally rely on linear addition of items within a subscale, a process that fails to capture the nonlinear relationships inherent in TCM [15]. Moreover, even when a linear relationship

is assumed, these traditional scoring methods often overlook the need to account for the relative weights of different items. To further improve the efficiency of the determination of BC classifications and scores, this study introduces machine learning-assisted methods for their rapid determination. These methods enable automatic classifications of BC and calculation of BC scores using selected subsets of CCMQ items. Utilizing both linear and nonlinear algorithms, machine learning-based approaches incorporate the relative weights of items, thereby refining the scoring and classification processes for BC types.

Machine learning methodologies can be broadly categorized into two primary types: supervised and unsupervised machine learning. During the process of supervised machine learning-assisted rapid determination, items are utilized as predictors, and each BC classification or score is used as the target to construct predictive models. A subset of core items that contribute most to the predicted outcome are chosen in the training process, and then they are used as predictors to obtain the outcome (i.e., BC classification or score) in the testing process. Common supervised machine learning algorithms for item selection include support vector machines, elastic net, extreme gradient boosting, gradient boosting, k-nearest neighbors, random forests, and extremely randomized trees. These algorithms have been proven effective in various studies [16, 17] and are considered potential supervised learning options in this study. In contrast, unsupervised machine learning does not involve predefined targets but instead aims to discover intrinsic data patterns and groupings based on the relationships among all items, potentially transcending the limitations of predefined dimensions. Using unsupervised methods, the most representative items for inherent groupings were retained. These items are subsequently utilized as features in supervised machine learning algorithms to predict BC classifications and scores. The dimensional structure and item formulation of the CCMQ are primarily designed based on the experiential insights of experts in TCM, lacking consistent structural validity supported by empirical data. As mentioned previously, supervised and unsupervised machine learning algorithms each offer distinct advantages in analyzing the relationships between the CCMQ dimensions and items. Supervised algorithms operate on the premise of acknowledging existing dimensions, while unsupervised algorithms do not. In the process of predicting BC classifications or scores with a limited number of items, comparing the selections made by these two approaches can help identify consistencies and discrepancies in item choice, which further reveals items that are crucial for the stability of the CCMQ dimensions.

In summary, this study illustrated the application of various machine learning methods to rapidly determine BC classifications and calculate BC scores. Specifically, we utilized an automated machine learning algorithms (AutoML) known as the Tree-based Pipeline Optimization Tool (TPOT), as well as unsupervised machine learning through variable clustering analysis (varclus). Using machine learning methods (i.e., model selection and parameter tuning), the predicted BC classifications or scores from a subset of core items are expected to be highly correlated with the original classifications or scores; thus, a subset of core items can be used instead of the whole set to improve the test efficacy of the CCMQ.

Materials and methods

Data source and collection

Between August 26, 2015, and October 12, 2017, a survey was conducted via a web-based platform using the CCMQ to assess the BC of individuals aged 15–64 across China. A total of 94,718 questionnaires were collected. However, 3573 responses were discarded due to fictitious or inconsistent answers, yielding 91,145 valid questionnaires for analysis.

Measurement instrument

This study employed the 60-item adult version of the CCMQ developed by Wang et al. (2006). The questionnaire is organized into nine subscales, each containing 6–8 items. Questions are presented in a question format (e.g., “Were you energetic?”) Responses are measured using a five-point Likert scale ranging from “1. Not at all” to “5. Very much.” Specific items associated with the GTC subscale include the following: “(2) Did you become tired easily?”, “(7) Was your voice weak when talking?”, “(8) Did you feel in low spirits and depressed?”, “(21) Did you feel more vulnerable to the cold than others (winter coldness, air conditioners, fans, etc.)?”, “(54) Did you easily experience insomnia?”, and “(27) Did you forget things easily?” are reverse scored. In contrast, these items are positively scored in other BCs as well as the remaining items. The formula for calculating the converted score is as follows:

$$\text{Converted Score} = \frac{\sum \text{Raw Score} - \text{Number of Items}}{\text{Number of Items} \times 4} \times 100$$

BC types were classified based on the criteria established in the “Classification and Determination of Chinese Medicine Constitution” [18]. Specifically, the criterion for identifying GTC requires a minimum conversion score of 60, whereas the conversion scores for the remaining biased BCs are less than 30. For the eight biased BCs, a converted score above 40 indicates their presence, and scores ranging from 30 to 40 suggest a predisposition toward a specific constitution.

In this study, the Cronbach’s alpha coefficients for the subscales were as follows: GTC at 0.664, QDC at 0.742, YaDC at 0.778, YiDC at 0.673, PDC at 0.706, DHC at 0.649, BSC at 0.660, QSC at 0.775, and SDC at 0.721.

Methods

Supervised machine learning

AutoML was employed to select the optimal supervised machine learning algorithm for BC classifications and scores, facilitating the implementation of traditional machine learning model design strategies in an automated, data-driven manner. The Tree-based Pipeline Optimization Tool (TPOT), a prevalent AutoML algorithm, automatically designs and optimizes machine learning pipelines for specific problem domains without human intervention [19]. TPOT primarily comprises two components: model construction via genetic programming and optimal model selection using Pareto efficiency. First, predictive models were built using all possible CCMQ item combinations as predictor variables, ranging from combinations with single item to all items, with the BC classifications and scores from the original subscales as the predicted outcomes. During the model construction process, new pipeline configurations are generated through crossover and mutation operations within the genetic algorithm. However, a challenge arises because maximizing model predictive performance (the model with the highest accuracy for classification and the model with the smallest mean squared error (MSE) for regression) often results in increased model complexity. By leveraging genetic programming in conjunction with Pareto optimality, the model selected by TPOT effectively balances predictive performance against complexity. Based on the model selected by the TPOT, the item combination that demonstrated the best performance across all item combinations with the same number of items was selected using the area under the curve (AUC) or R-squared (R^2). Then, given the best item combinations for a specific number of items, we may need to further determine the most appropriate number of items for predicting BC classifications or scores. With this aim, we calculated the AUC or R^2 improvement for the predictive performance of an item combination with at least two items over the item combination with one fewer item in predicting the original BC classifications or scores. In this study, the item combinations that show obvious improvements in AUC and R^2 are recommended for rapidly assessing BC classifications and scores. When the improvement progresses steadily, it is recommended to use a threshold of 0.8 for AUC and R^2 , which are generally considered indicators of excellent predictive performance [20, 21], to determine the

minimum number of items to be retained when predicting BC classifications or scores. Finally, the predictive performance will be assessed by AUC, accuracy, F1 score, R^2 , root mean square error (RMSE) and mean absolute percentage error (MAPE).

Unsupervised machine learning

Variable clustering analysis (varclus) was used in this study as the unsupervised machine learning method, which groups similar items into clusters. Each cluster can be represented by a single item. This approach reduces data complexity while ensuring interpretability. Based on the clusters, representative items are selected, and the scale items are simplified accordingly.

First, the 60 items of the CCMQ were normalized to form a single cluster. This cluster was then iteratively segmented, continuing until the second eigenvalue within a cluster did not exceed the threshold (shown in Fig. 1). In this study, the thresholds were dynamically adjusted to capture clusters of all possible quantities (i.e., 1–60). Next, representative items were selected from each cluster. The selection of representative items is based on the following formula:

$$r = \frac{\lambda_{ij}}{\lambda_j^{\max, \text{other}}}$$

λ_{ij} represents the loading of the j th variable within the i th cluster on the first principal component. $\lambda_j^{\max, \text{other}}$

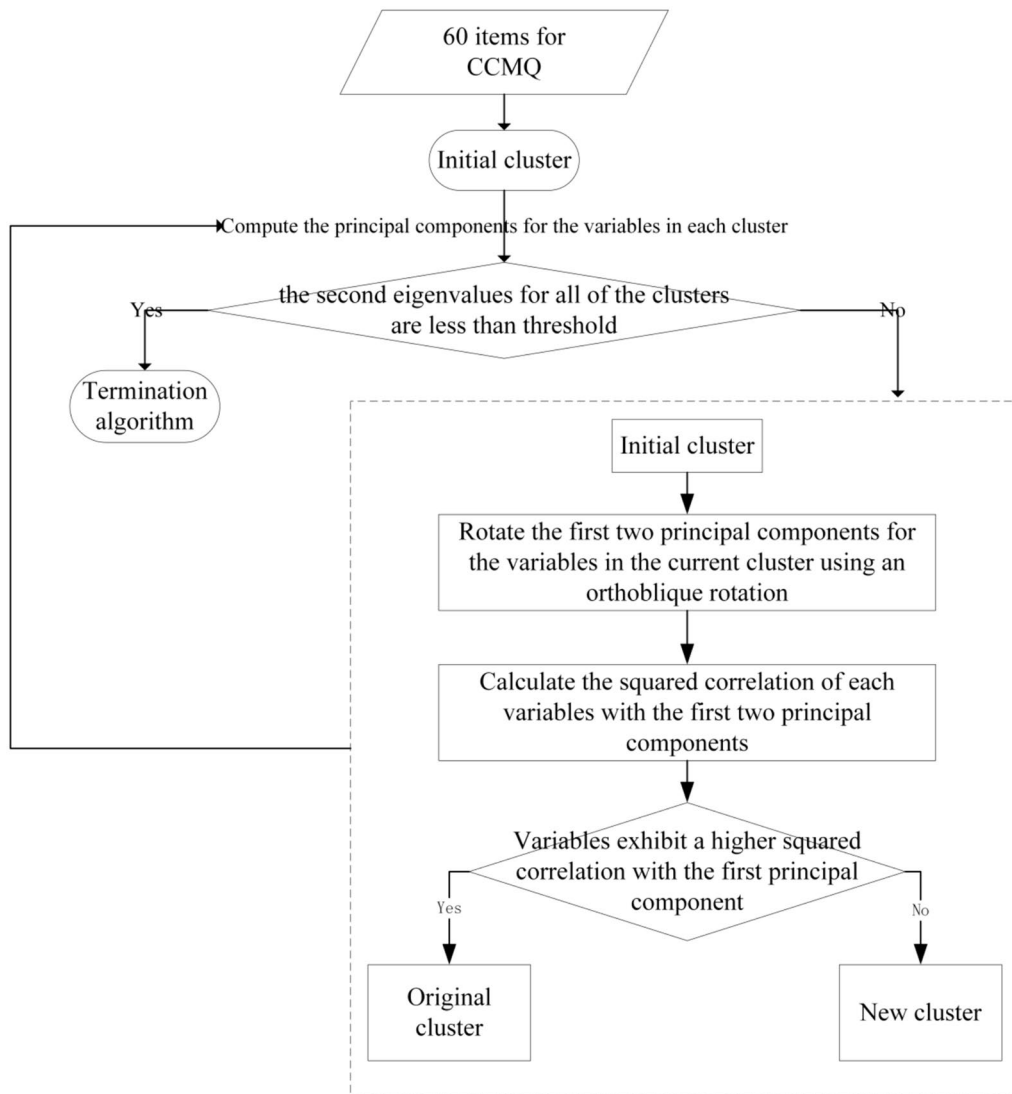


Fig. 1 Clustering process

represents the *j*th variable’s maximum loading on the first principal component of all other clusters, excluding the cluster to which it belongs. The item with the highest ratio is chosen as the representative item for a given cluster.

In other studies, the number of principal components is typically determined using a cumulative variance contribution rate of 70–90% [22]. In this study, we adopt the average variance contribution rate at a median level within this range to determine the number of clusters. We selected the least number of clusters based on the criterion that the average within-cluster variance explained by the representative item is 80%. The items selected by varclus were used to predict BC classifications or scores by selecting the appropriate supervised machine learning methods using the TPOT algorithms, and their predictive capacities were evaluated using AUC, accuracy, and F1 score (or R^2 , RMSE, and MAPE).

Results

Distribution of scores among different BCs

Figure 2 shows that the scores for the eight biased BCs exhibit a pronounced left-skewed distribution. Among the seven biased BCs (i.e., QDC, YaDC, YiDC, PDC, DHC, BSC, and QSC), there was a greater number of individuals in the 20–50 score range.

Only 13.66% of the population exhibited GTC, 36.15% were classified as QDC, 30.34% as YaDC, 20.02% as YiDC, 25.07% as PDC, 27.83% as DHC, 22.66% as BSC, 27.08% as QSC, and 8.51% as SDC. It should be noted that some individuals may exhibit multiple biased BCs simultaneously.

Item selection based on automated machine learning

The optimal performance of the supervised machine learning item selection models, utilizing BC classifications as the dependent variable across various item combinations, is presented in Table 1, with the corresponding

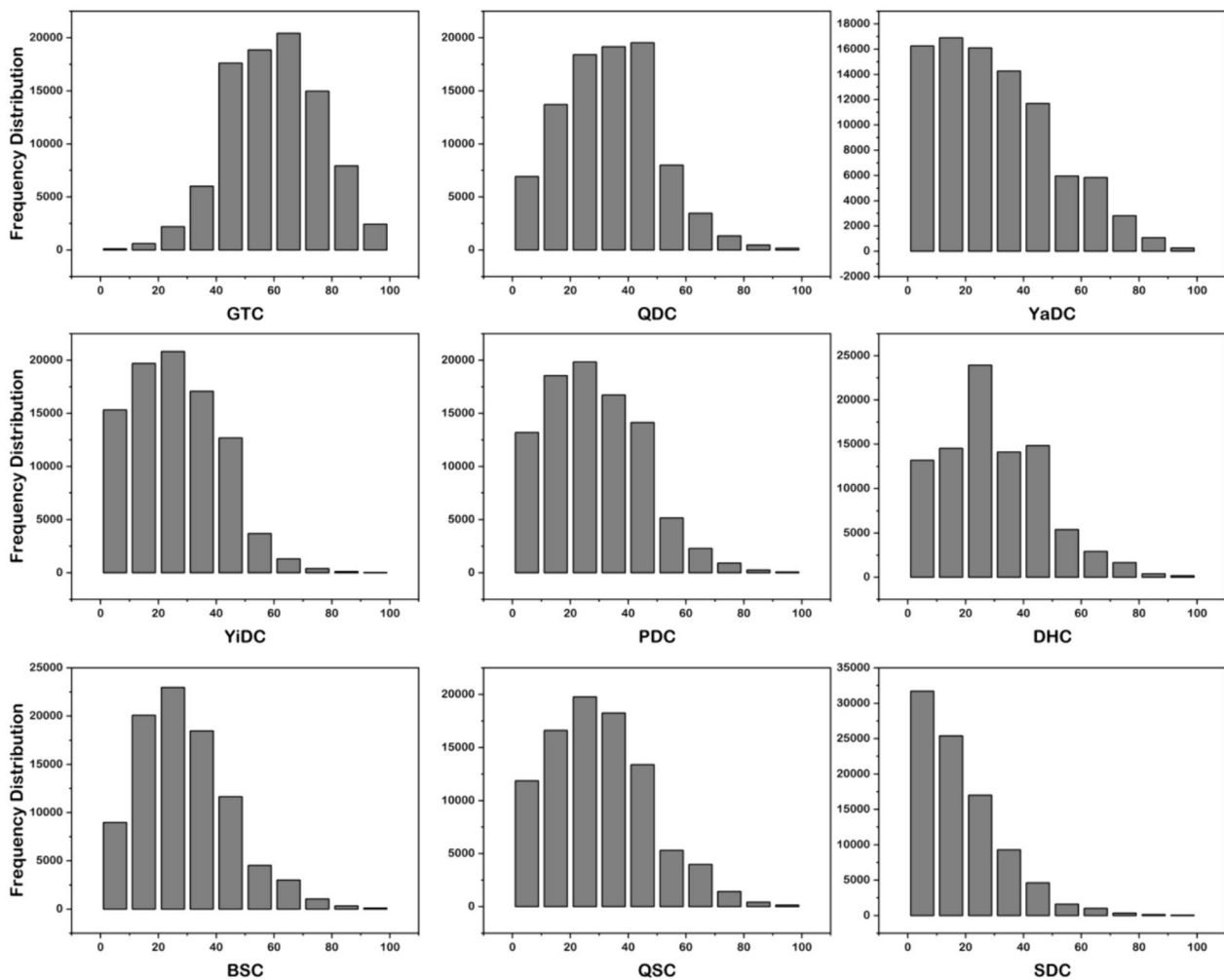


Fig. 2 Distribution of scores for the BCs

Table 1 The optimal item combinations for BC classifications as the dependent variable and their corresponding algorithms

	GTC	QDC	YaDC	YiDC	PDC	DHC	BSC	QSC	SDC
1 item	2 GaussianNB	3 MLPClassifier	19 MLPClassifier GaussianNB	20 MLPClassifier	49 MLPClassifier	60 MLPClassifier	40 MLPClassifier	9 GaussianNB	30 GaussianNB
2 items	2, 21 XGBClassifier	3, 6 XGBClassifier	19, 52 XGBClassifier	20, 35 XGBClassifier	49, 50 MLPClassifier	39, 59 ExtratreesClassifier	40, 43 MLPClassifier	9, 14 XGBClassifier	24, 31 Gradient-BoostingClassifier
3 items	2, 8, 21 Polynomial-Features XGBClassifier	3, 5, 7 XGBClassifier	18, 19, 52 XGBClassifier	20, 35, 44 MLPClassifier	15, 49, 50 RobustScaler MLPClassifier	39, 56, 60 StandardScaler MaxAbsScaler MLPClassifier	37, 40, 43 MLPClassifier	9, 10, 14 ExtratreesClassifier	24, 30, 31 FectureUnion XGBClassifier
4 items	2, 8, 21, 27 RobustScaler XGBClassifier	3, 5, 7, 26 RobustScaler	17, 18, 19, 52 XGBClassifier	16, 35, 38, 44 Standard-Scaler MLPClassifier	13, 28, 49, 50 Gradient-Boosting-Classifier MLPClassifier	39, 48, 56, 60 MLPClassifier	37, 40, 43, 45 MLPClassifier	9, 10, 12, 47 GaussianNB MLPClassifier	24, 30, 31, 34 MLPClassifier
5 items	2, 7, 8, 21, 27 MLPClassifier	3, 5, 6, 22, 26 MinMax-Scaler MLPClassifier	17, 18, 21, 22, 52 RobustScaler	16, 38, 44, 46, 57 MLPClassifier	13, 28, 50, 51, 58 RobustScaler	39, 41, 48, 56, 60 GaussianNB	36, 37, 40, 43, 45 MLPClassifier	9, 10, 12, 14, 47 RobustScaler	24, 25, 30, 31, 34 XGBClassifier
6 items	2, 7, 8, 21, 27, 54 MLPClassifier	3, 5, 6, 7, 22, 26 ZeroCount MLPClassifier	17, 18, 21, 22, 52, 55 RobustScaler MLPClassifier	16, 20, 29, 44, 46, 57 ZeroCount MLPClassifier	13, 28, 49, 50, 51, 58 MLPClassifier	39, 41, 48, 56, 59, 60 MLPClassifier	33, 36, 37, 40, 43, 45 RobustScaler MLPClassifier	9, 10, 11, 12, 14, 47 StandardScaler MLPClassifier	24, 25, 30, 31, 32, 34 RobustScaler GaussianNB MLPClassifier
7 items	2, 7, 8, 21, 27, 53, 54 MLPClassifier	3, 4, 5, 6, 7, 22, 26 GaussianNB Standard-Scaler MLPClassifier	17, 18, 19, 21, 22, 52, 55 MLPClassifier	16, 29, 35, 38, 44, 46, 57 RobustScale MLPClassifier	13, 28, 42, 49, 50, 51, 58 RobustScale MLPClassifier		33, 36, 37, 40, 43, 27, 45 MLPClassifier	8, 9, 10, 11, 12, 14, 47 MLPClassifier	23, 24, 25, 30, 31, 32, 34 LinearSVC
8 items	1, 2, 7, 8, 21, 27, 53, 54 ZeroCount MLPClassifier	2, 3, 4, 5, 6, 7, 22, 26 MLPClassifier		16, 20, 29, 35, 38, 44, 46, 57 MLPClassifier	13, 15, 28, 42, 49, 50, 51, 58 LinearSVC				

The specific meanings represented by the entries are found in the supplementary materials

algorithms detailed therein. Figure 3 illustrates the improvement in the AUC for the optimal item combinations for predicting each BC classification.

For all models except the QSC model, the AUC plots revealed elbow points for either two or four items: GTC: item 2 (i.e., tiredness; abbreviated form throughout; items, questions, and question abbreviations are provided in Table S1), item 8 (i.e., depression), item 21 (i.e., cold intolerance), item 27 (i.e., forgetfulness); QDC: item 3 (i.e., breathlessness), item 6 (i.e., quietude); YaDC: item 19 (i.e., cold aversion), item 52 (i.e., cold sensitivity); YiDC: item 20 (i.e., localized hotness), item 35 (i.e.,

dryness); PDC: item 49 (i.e., sticky mouth), item 50 (i.e., flabby abdomen); DHC: item 39 (i.e., oily skin), item 59 (i.e., urethral heat); BSC: item 40 (i.e., hyperpigmentation), item 43 (i.e., dark circles); and SDC: item 24 (i.e., chronic rhinitis), item 31 (i.e., urticaria). Remarkably, the QSC model achieved the predefined threshold of AUC=0.8 with the inclusion of only one item (item 9: anxiety). The selected items achieved AUC values ranging from 0.857 to 0.946 (shown in Table 3). In general, predictive models are considered excellent when their AUC values fall between 0.8 and 0.9 and outstanding when they exceed 0.9 [21]; therefore, all these models

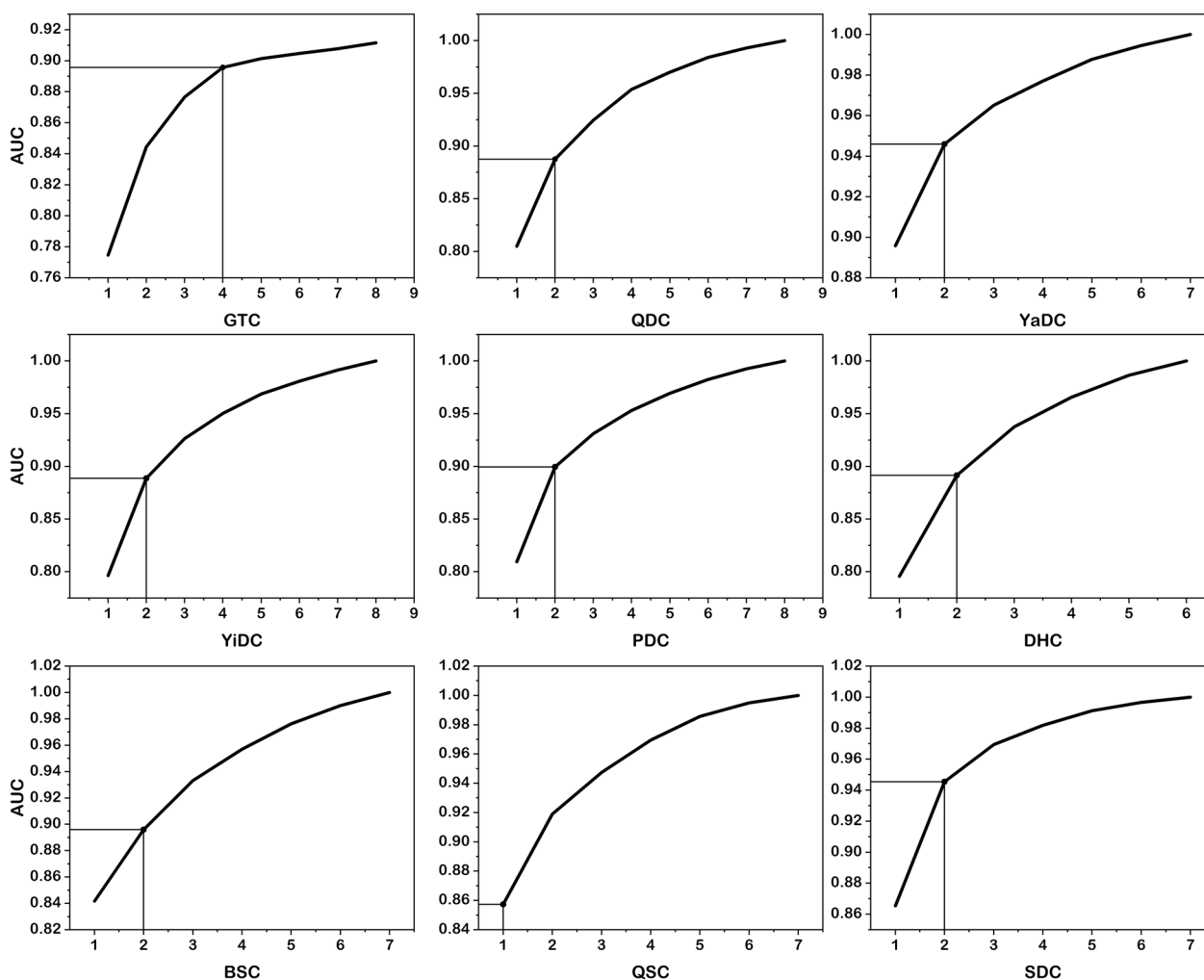


Fig. 3 AUC for different item combinations. Note. For the GTC, QDC, YaDC, YiDC, PDC, DHC, BSC and SDC, elbow points were selected where the item combination maximized the improvement in the AUC. For QSC, the curve was relatively smooth, so we selected the item combination with the fewest number of items when the AUC exceeded 0.8

demonstrated excellent predictive performance, with some even reaching outstanding levels. Additionally, the accuracy and F1 scores of these models' predictions were calculated, with accuracy values ranging from 0.819 to 0.936 and F1 scores ranging from 0.417 to 0.807.

Table 2 presents the optimal performance of the supervised machine learning item selection models with the BC score as the dependent variable. Figure 4 demonstrates the improvement in R^2 performance for these optimal models.

In Fig. 4, a clear elbow point is shown for the YaDC model with a two-item combination (item 19: cold aversion and item 52: cold sensitivity). Other BC score predictive models did not exhibit significant elbow points. Accordingly, the item combinations for the final BC scores prediction were selected from the models with an R^2 exceeding 0.8 that used the fewest

predictors. Specifically, the models for GTC, QDC, YiDC, PDC, DHC, and BSC included the first four items (GTC: item 2: tiredness, item 8: depression, item 21: cold intolerance, item 53: adaptability; QTC: item 3: breathlessness, item 5: dizziness, item 6: quietude, item 26: hyperhidrosis; YiDC: item 20: localized hotness, item 44: dry eyes, item 46: thirstiness, item 57: constipation; PDC: item 15: lethargy, item 28: oily T-zone, item 49: sticky mouth, item 50: flabby abdomen; DHC: item 39: oily skin, item 48: bitter mouth, item 56: sticky stools, item 60: wet scrotum/yellowing leukorrhoea; BSC: item 27: forgetfulness, item 37: pain, item 40: hyperpigmentation, item 43: dark circles, respectively). Meanwhile, the QSC and SDC models incorporated the first three items (QSC: item 9: anxiety, item 10: vulnerability, item 14: sighing; SDC: item 24: chronic rhinitis, item 30: allergies, item 34:

Table 2 The optimal item combinations for the BC score as the dependent variable and their corresponding algorithms

	GTC	QDC	YaDC	YiDC	PDC	DHC	BSC	QSC	SDC
1 item	8 Decision-treeRegressor	3 Decision-TreeRegressor	19 RandomForestRegressor	35 XGBRegressor	49 Gradient-BoostingRegressor	60 RandomForestRegressor	40 ExtraTreesRegressor	9 Gradient-BoostingRegressor	31 XGBRegressor
	8, 21 RandomForestRegressor	3, 6 RandomForestRegressor	19, 52 ExtraTreesRegressor	20, 35 RandomForestRegressor	49, 50 Gradient-BoostingRegressor	39, 59 RandomForestRegressor	37, 40 Decision-TreeRegressor	9, 10 RandomForestRegressor	24, 31 Decision-TreeRegressor
3 items	1, 8, 21 Decision-TreeRegressor	3, 6, 26 ElasticNetCV	17, 19, 52 ExtraTreesRegressor	20, 35, 44 RidgeCV	15, 49, 50 RandomForestRegressor	39, 56, 59 RandomForestRegressor	37, 40, 43 ExtraTreesRegressor	9, 10, 14 ZeroCount	24, 30, 34 ElasticNetCV
	Decision-TreeRegressor	RandomForestRegressor		ExtraTreesRegressor				ExtraTreesRegressor	RandomForestRegressor
	Decision-TreeRegressor								
4 items	2, 8, 21, 53 XGBRegressor	3, 5, 6, 26 ExtraTreesRegressor	17, 18, 19, 52 LassoLarsCV	20, 44, 46, 57 XGBRegressor	15, 28, 49, 50 RobustScaler	39, 48, 56, 60 Gradient-BoostingRegressor	27, 37, 40, 43 RidgeCV	9, 10, 14, 47 RobustScaler	24, 30, 31, 34 RidgeCV
		RobustScaler			RBFsampler		XGBRegressor	RBFsampler	XGBRegressor
		ElasticNetCV			RidgeCV			MinMaxScaler	
					ElasticNetCV			RidgeCV	
5 items	2, 8, 21, 53, 54 ElasticNetCV	3, 5, 6, 22, 26 RobustScaler	17, 18, 21, 22, 52 MaxAbsScaler	16, 38, 44, 46, 57 RidgeCV	15, 28, 50, 51, 58 LinearSVR	39, 41, 48, 56, 60 RidgeCV	27, 37, 40, 43, 45 LinearSVR	9, 10, 12, 14, 47 RobustScaler	24, 25, 30, 31, 34 XGBRegressor
	RandomForestRegressor	RobustScaler	XGBRegressor	ElasticNetCV		Polynomial-Features		LinearSVR	RidgeCV
		RidgeCV		RandomForestRegressor		RidgeCV			
6 items	2, 7, 21, 27, 53, 54 MinMaxScaler	2, 3, 5, 6, 22, 26 ElasticNetCV	17, 18, 21, 22, 52, 55 AdaBoostRegressor	16, 35, 38, 44, 46, 57 Gradient-BoostingRegressor	13, 15, 28, 50, 51, 58 RidgeCV	39, 41, 48, 56, 59, 60 LassoLarsCV	27, 36, 37, 40, 43, 45 XGBRegressor	9, 10, 11, 12, 14, 47 Decision-TreeRegressor	23, 24, 25, 30, 31, 34 RobustScaler
	ZeroCount	Gradient-BoostingRegressor	RidgeCV				RidgeCV	StandardScaler	ElasticNetCV
	Gradient-BoostingRegressor							RidgeCV	RandomForestRegressor
7 items	1, 7, 8, 21, 27, 53, 54 ElasticNetCV	2, 3, 5, 6, 7, 22, 26 RidgeCV	17, 18, 19, 21, 22, 52, 55 LassoLarsCV	16, 20, 29, 35, 44, 46, 57 LinearSVR	13, 15, 28, 42, 50, 51, 58 RobustScaler		27, 33, 36, 37, 40, 43, 45 FunctionUnion	8, 9, 10, 11, 12, 14, 47 LinearSVR	23, 24, 25, 30, 31, 32, 34 SGDRegressor
	RandomForestRegressor				ZeroCount		RidgeCV		StandardScaler
					LassoLarsCV				RidgeCV
8 items	1, 2, 7, 8, 21, 27, 53, 54 RidgeCV	2, 3, 4, 5, 6, 7, 22, 26 LassoLarsCV		16, 20, 29, 35, 38, 44, 46, 57 RidgeCV	13, 15, 28, 42, 49, 50, 51, 58 SGDRegressor				

The specific meanings represented by the items are found in the supplementary materials

dermatographism, respectively), aligning with this criterion. For predicting BC scores, the models yielded R² values ranging from 0.785 to 0.879 (shown in Table 3).

According to general guidelines that R² values of 0.75 and 0.50 are considered substantial and moderate, respectively, of predictive performance [20], all of

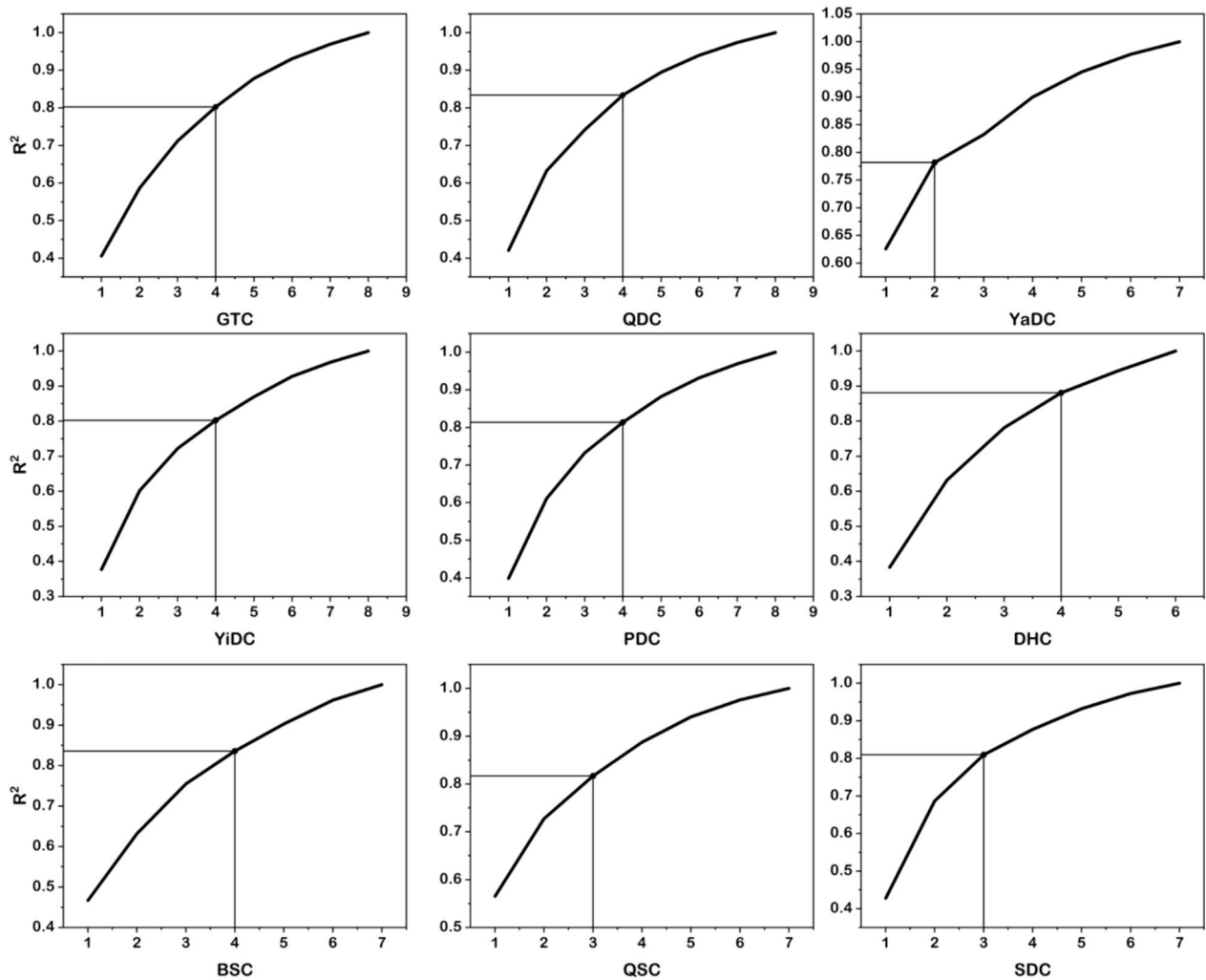


Fig. 4 R^2 for different item combinations. For YaDC, an elbow point was selected where the item combination maximized the improvement in the AUC. For GTC, QDC, YiDC, PDC, DHC, BSC, QSC and SDC, the curves were relatively smooth, so we selected the item combination with the fewest number of items when the AUC exceeded 0.8

these models had excellent predictive capacities. The range of RMSE values for these models was between 6.241 and 9.877, and the MAPE values ranged from 10.868 to 39.337.

Item selection based on the unsupervised machine learning algorithm

For all possible numbers of clusters, the average variance contribution of all representative items within their respective clusters is illustrated in Fig. 5. When the number of clusters was set to 29, the representative items accounted for 80% of the variance within each cluster, on average. To further assess the predictive capability of these items for BC classification or score, the appropriate supervised machine learning models were constructed by

TPOT, with these 29 items as independent variables and BC classifications or scores as dependent variables. The specific predictive algorithms and their corresponding performances are detailed in Table 4.

In terms of predicting BC classifications, the selected items achieved AUC values ranging from 0.847 to 0.965, prediction accuracies ranging from 0.820 to 0.950 and F1 scores from 0.294 to 0.858. For predicting BC scores, the models yielded R^2 values from 0.549 to 0.888, RMSE values from 5.602 to 12.135 and MAPE values from 14.470 to 48.239 (shown in Table 4). According to the criteria mentioned above [20, 21], most models exhibited outstanding predictive performance, while a few were considered to have moderate predictive capability.

Table 3 Evaluation of the automated machine learning results based on the appropriate supervised machine learning methods

Subscale	The BC classification as the target variable					The BC score as the target variable				
	Items	Algorithms	AUC	Accy	F1 score	Items	Algorithms	R ²	RMAE	MAPE
GTC	4, 8, 21, 27	RobustScale XGBClassifier	0.891	0.878	0.417	2, 8, 21, 53	XGBRegressor	0.799	7.162	10.868
QDC	3, 6	XGBClassifier	0.888	0.823	0.741	3, 5, 6, 26	ExtraTreesRegressor RobustScaler ElasticNetCV	0.835	6.749	21.513
YaDC	19, 52	XGBClassifier	0.946	0.884	0.804	19, 52	ExtratreesRegressor	0.785	9.877	39.337
YiDC	20, 35	XGBClassifier	0.890	0.857	0.561	20, 44, 46, 57	XGBRegressor	0.804	6.740	27.560
PDC	49, 50	MLPClassifier	0.890	0.853	0.685	15, 28, 49, 50	RobustScaler RBFSampler RidgeCV ElasticNetCV	0.815	7.147	28.133
DHC	39, 59	ExtraTreesClassifier	0.894	0.835	0.693	39, 48, 56, 60	GradientboostingRegressor	0.879	6.241	22.208
BSC	40, 43	MLPClassifier	0.897	0.857	0.641	27, 37, 40, 43	RidgeCV XGBRegressor	0.836	6.720	21.582
QSC	9	GaussianNB	0.857	0.819	0.569	9, 10, 14	ZeroCount ExtraTreesRegressor	0.814	7.805	29.207
SDC	24, 31	GradientBoostingClassifier	0.942	0.936	0.603	24, 30, 34	ElasticNetCV RandomForestRegressor	0.811	6.575	34.079

The specific meanings represented by the items are found in the supplementary materials

Comparison between automated machine learning and unsupervised machine learning

We conducted three types of item selection procedures from the CCMQ: items selected by automated machine learning algorithms with classifications as the target variables, items selected by automated machine learning algorithms with scores as the outcome variables, and representative items selected by unsupervised learning. Figure 6A illustrates the frequency of items selected based on the three types of item selection procedures. As shown in Fig. 6A, item 2 (i.e., tiredness), item 6 (i.e.,

quietude), item 19 (i.e., cold aversion), item 20 (i.e., localized hotness), item 27 (i.e., forgetfulness), item 39 (i.e., oily skin), item 43 (i.e., dark circles), item 50 (i.e., flabby abdomen), and item 52 (i.e., cold sensitivity) were consistently chosen by these three types of item selection procedures, indicating that they contain multiple pieces of information, including associations with BC classifications and scores, as well as relationships with other items. Additionally, the supervised machine learning algorithms consistently selected item 3 (i.e., breathlessness), item 8 (i.e., depression), item 9 (i.e., anxiety), item 21 (i.e., cold intolerance), item 24 (i.e., chronic rhinitis), item 40 (i.e., hyperpigmentation), and item 49 (i.e., ticky mouth) for predicting BC classifications and scores. Different from the items selected by the supervised machine learning algorithms, the varclus algorithm selected item 16 (i.e., palmar-plantar hot), item 22 (i.e., susceptibility to colds), item 23 (i.e., frequent sneezing), item 25 (i.e., sensitivity cough), item 29 (i.e., reddened lips), item 33 (i.e., unexplained bruising), item 36 (i.e., facial telangiectasia), item 51 (i.e., excess phlegm), item 55 (i.e., cold-induced diarrhea), and item 58 (i.e., thick tongue coating) to predict the BC classifications or scores.

In the prediction of classifications, the items selected based on TPOT and those selected by varclus, using the appropriate prediction models, demonstrated similar AUCs, accuracies and F1 scores (Fig. 6B). However, except for the GTC model, TPOT selected fewer items for the biased BC classifications compared to varclus.

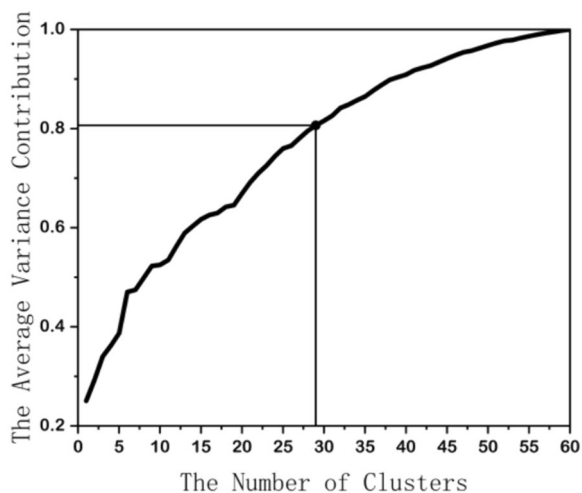


Fig. 5 The average variance contribution of the representative items

Table 4 Evaluation of varclus results based on the appropriate supervised machine learning methods

Subscale	Items	The BC classification as the target variable			The BC score as the target variable				
		Algorithms	AUC	Accy	F1 score	Algorithms	R ²	RMAE	MAPE
GTC	2, 27, 53	XGBClassifier	0.847	0.865	0.294	ZeroCount RandomForestRegressor	0.676	9.143	14.470
QDC	2, 4, 6, 22, 26	RobustScaler MLPClassifier MLPClassifier	0.965	0.899	0.858	PolynomialFeatures LassoLarsCV	0.888	5.602	17.021
YaDC	19, 22, 52, 55	MaxAbsScaler RobustScaler MLPClassifier	0.960	0.904	0.841	GradientBoostingRegressor RidgeCV	0.866	8.608	35.056
YiDC	16, 20, 29, 44, 57	MLPClassifier	0.960	0.909	0.754	PolynomialFeatures LassoLarsCV	0.817	6.593	29.433
PDC	50, 51, 58	MLPClassifier	0.924	0.867	0.719	RandomForestRegressor	0.694	9.188	38.755
DHC	39, 48, 60	ExtraTreesClassifier	0.934	0.874	0.765	RobustScaler RBFsampler LassoLarsCV	0.769	8.616	31.898
BSC	27, 33, 36, 37, 43	XGBClassifier MLPClassifier	0.951	0.902	0.773	PolynomialFeatures RidgeCV	0.843	6.567	23.705
QSC	10	MLPClassifier	0.855	0.820	0.611	AdaBoostRegressor	0.549	12.135	48.239
SDC	23, 25, 31	ZeroCount MLPClassifier	0.963	0.950	0.672	XGBRegressor	0.777	7.206	42.381

The specific meanings represented by the items are found in the supplementary materials

In the prediction of scores for BC, such as GTC, PDC, DHC, QSC, and SDC, the items selected based on TPOT demonstrated notable advantages in comparison with those selected by varclus (Fig. 6C). However, in the prediction of scores for BC types such as QDC and YaDC, selections made by varclus and its corresponding predictive algorithms exhibited superior performance. For the YiDC and BSC score predictions, the performances of the items chosen by TPOT and varclus were similar. Overall, while the number of items selected by TPOT and varclus remains similar, the selections made by TPOT yield a higher average R² and lower average RMSE and MAPE.

Discussion

Self-report questionnaires are recognized in clinical practice as effective tools for quantifying abstract concepts, aiding in assessments of disease risk [23–25]. As the use of questionnaires grows, there is an increased demand for them to possess strong predictive capabilities and to facilitate rapid disease determination. In this study, we utilized machine learning techniques for rapidly determining BCs.

The comparative analysis between AutoML and supervised machine learning in terms of item selection for BC classifications and scores revealed a slight advantage for AutoML. The reason might be that the predictive targets were predefined dimensions of the CCMQ. Thus, AutoML are recommended for predicting original BC classifications or scores.

There was consistency among the different item selection procedures. Items 2 (i.e., tiredness), 6 (i.e., quietude), 19 (i.e., cold aversion), 20 (i.e., localized hotness), 27 (i.e., forgetfulness), 39 (i.e., oily skin), 43 (i.e., dark circles), 50 (i.e., flabby abdomen), and 52 (i.e., cold sensitivity) were selected by both supervised machine learning algorithms and varclus. These items were consistently selected because they may encompass information in other items within a given BC type. Given that constitutions are closely related to the development of certain diseases [26–29] special attention may need to be given to these factors for rapid disease prediction. For example, items 2 (i.e., tiredness), 27 (i.e., forgetfulness), 19 (i.e., cold aversion), 52 (i.e., cold sensitivity), and 20 (i.e., localized hotness), which are related to the QDC, YaDC and YiDC, can be core predictors of chronic fatigue syndrome (CFS), generalized anxiety disorder, depression, and anemia [26, 30–32]. Items 39 (i.e., oily skin) and 50 (i.e., flabby abdomen) are indicators of DHC and PDC, respectively, and can be used to predict the occurrence of metabolic diseases and polycystic ovary syndrome [33–35].

Previous research simplified scales by calculating feature importance using supervised machine learning algorithms for predicting total scores with item scores and retaining items with high feature importance [16, 17, 36]. However, this method faces the issue that different machine learning algorithms may assign items different weights [37]. We may select the appropriate supervised machine learning algorithms before

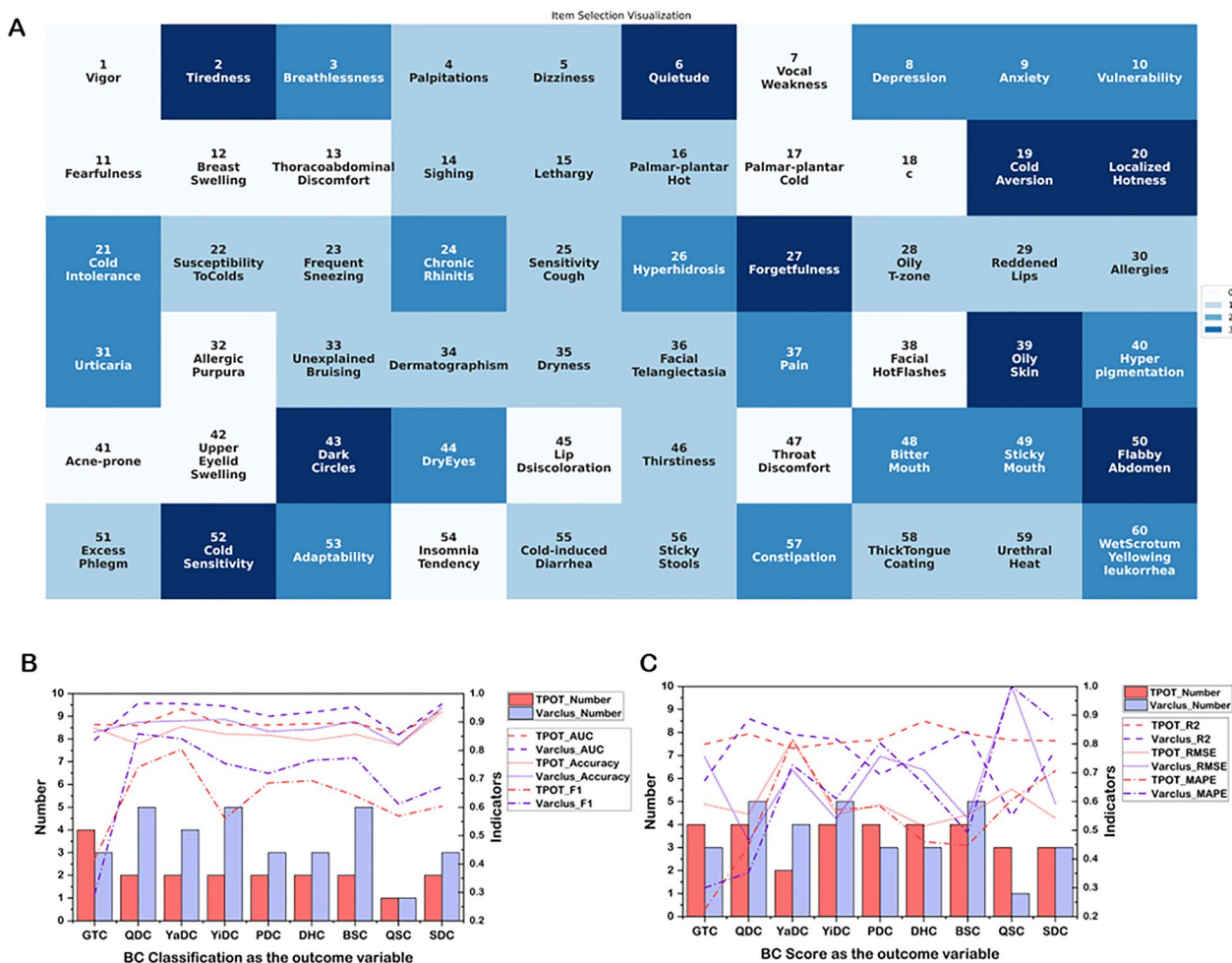


Fig. 6 Comparison of prediction performance using the appropriate supervised machine learning for items selected based on TPOT and varclus. **A** The frequency of items selected based on TPOT and varclus. **B** Performance of items selected based on TPOT and varclus in predicting BC classifications using the appropriate supervised machine learning method. **C** Performance of items selected based on TPOT and varclus in predicting BC scores using the appropriate supervised machine learning method. In **C**, the RMSE measure is represented as RMSE/RMSEmax, and the MAPE is represented as MAPE/MAPEmax

calculating feature importance, but this chosen algorithm may not perform best when using the selected items to predict the original scores. In this study, we innovatively selected the best-performing model from those built using all possible combinations of items as input variables. Thus, the absolute improvement in the models' predictive performance as the number of items used to predict the original BC classifications or scores increases can be calculated. For each possible number of items, the top three ranked item combinations in terms of predictive effectiveness, along with their corresponding algorithms, are listed in the supplemental material (Figure S1 and S2). Consequently, we can comprehensively understand the predictive capabilities

of all possible item combinations and their corresponding algorithms and assist practitioners in selecting item combinations based on their understanding of TCM theories and specific needs in scale development (e.g., required reliability and validity, test efficiency).

In summary, this study has significant implications for the BC identification in clinical practice. Firstly, the machine learning algorithms proposed in this study enable rapid BC identification based on a subset of items. Secondly, by comparing different supervised and unsupervised machine learning algorithms, it is possible to gain deeper insights into how different items contribute to the various BC dimensions, thereby assisting clinical practitioners in achieving a more thorough understanding of these dimensions.

A major limitation of this study is the computational cost arising from the complexity of the algorithm combined with the large volume of data. Therefore, future research should focus on optimizing algorithms to enhance their processing speed and efficiency in big data environments. Also, in the future, we can further optimize the rapid body constitution identification process by integrating multimodal data, such as tongue and pulse diagnostics.

Conclusion

The items in the CCMQ were shown to have varying information weights. The use of highly important items may assist in the rapid determination of BCs. The use of supervised machine learning algorithms with all the possible item combinations for predicting BC classifications or scores achieved acceptable and stable predictive performance. The top item combinations obtained by supervised machine learning algorithms for predicting BC classifications or scores were identified so that other researchers can make selections according to their needs.

Abbreviations

CCMQ	Constitution of Chinese Medicine Questionnaire
BC	Body constitution
TCM	Traditional Chinese Medicine
AutoML	Automated machine learning
TPOT	Tree-based pipeline optimization tool
varclus	Variable clustering
GTC	Gentleness constitution
QDC	Qi-deficiency constitution
YaDC	Yang-deficiency constitution
YiDC	Yin-deficiency constitution
PDC	Phlegm-dampness constitution
DHC	Damp-heat constitution
BSC	Blood-stasis constitution
QSC	Qi-stagnation constitution
SDC	Special-diathesis constitution
AUC	Area under the curve
R ²	R-squared
RMSE	Root mean square error.
MAPE	Mean absolute percentage error

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13020-024-00992-0>.

Supplementary Material 1

Acknowledgements

We would like to extend our gratitude to all the participants who completed the CCMQ.

Author contributions

Dongran Han: conceptualization, funding acquisition, supervision. Qi Wang: funding acquisition. Yixing Liu: conceptualization, supervision, writing—review and editing. Minghua Bai: data curation. Ji Wang: data curation. Wen Sun: investigation, formal analysis, writing—review and editing. Bei Wang: writing—review and editing.

Funding

The work was supported by National Natural Science Foundation of China (Grant No: 82205315) awarded to Prof Dongran Han and National Natural Science Foundation of China (Grant No: T2341006) awarded to Prof Qi Wang.

Availability of data and materials

All data included in this article are available from the Prof Qi Wang upon request.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Beijing University of Chinese Medicine (Grant No. 2020BHYLL0102).

Consent for publication

Not applicable.

Competing interests

The authors declare that there are no competing interests.

Received: 16 June 2024 Accepted: 30 August 2024

Published online: 15 September 2024

References

- Wang Q. Classification and diagnosis basis of nine basic constitutions in Chinese medicine. *J Beijing Univ Tradit Chin Med.* 2005;28(4):1–8. <https://doi.org/10.3321/j.issn:1006-2157.2005.04.001>.
- Wang Q. On Chinese constitutional theory from three aspects. *J Beijing Univ Tradit Chin Med.* 2008;31(10):653–5. <https://doi.org/10.3321/j.issn:1006-2157.2008.10.001>.
- Wang J, Li Y, Ni C, Zhang H, Li L, Wang Q. Cognition research and constitutional classification in Chinese medicine. *Am J Chin Med.* 2011;39(04):651–60. <https://doi.org/10.1142/S0192415X11009093>.
- Jiang Y. Relationship between damp-heat constitution and syndrome of damp-heat. *J Beijing Univ Tradit Chin Med.* 2005;28(4):16–8. <https://doi.org/10.3321/j.issn:1006-2157.2005.04.004>.
- Chen AL, Zhao LC, Ruan JY, Liu CY, Wang WY, Liu YY, Liu KT, Yu Y, Wang C. Epidemiological investigation of life style on body constitution. *Shanxi Tradit Chin Med.* 2009;25(3):42–3. <https://doi.org/10.3969/j.issn.1000-7156.2009.03.027>.
- Wang Q, Zhu YB, Xue HS, Li S. Preliminary development of the constitution in Chinese Medicine Questionnaire. *Chin J Tissue Eng Res.* 2006;10(3):15–7.
- Jing HR, Wang J, Wang Q. Preliminary compiling of English version of constitution in Chinese Medicine Questionnaire. *J Anhui Univ Chin Med.* 2015;34(5):21–5. <https://doi.org/10.3969/j.issn.2095-7246.2015.05.007>.
- Kim S, Kim J. A development of the Korean version of the constitutions in Chinese Medicine Questionnaire. *J Korea Acad-Ind Coop Soc.* 2016;17(3):652–60. <https://doi.org/10.5762/KAIS.2016.17.3.652>.
- Zhu YB, Origasa H, Uebaba K, Xu FH, Wang Q. Development and validation of the Japanese version of the constitution in Chinese Medicine Questionnaire (CCMQ). *Kampo Med.* 2008;59(6):783–92. <https://doi.org/10.3937/kampomed.59.783>.
- Nguyen DTH, Le TT, Tang HK, Le LB, Lam TC, Le LTH. The Vietnamese version of the constitution in Chinese Medicine Questionnaire (CCMQ): validity and reliability. *Med Pharm Res.* 2022;6(2):18–27. <https://doi.org/10.32895/UMPMR.6.2.3>.
- Zhu YB, Wang Q, Xue HS, Hideki S. Preliminary assessment on performance of constitution in Chinese Medicine Questionnaire. *Chin J Clin Rehabil.* 2006;10(3):15–7. <https://doi.org/10.3321/j.issn:1673-8225.2006.03.010>.
- Bai MH, Wong W, Hou SJ, Zhang YF, Li QR, Li ZQ, Zhang Y, Li LR, Wang J, Wang Q. Development and evaluation of short form of constitution in Chinese Medicine Questionnaire: a national epidemiological survey data of 21948 case. *J Tradit Chin Med.* 2011;39(04):651–60. <https://doi.org/10.1142/S0192415X11009093>.

13. Zhu YB, Wang Q, Yu XH, Shi HM. Construct validity and responsiveness of the 41-item short version of constitution in Chinese Medicine Questionnaire. *Chin Gen Pract.* 2017;20(26):3282. <https://doi.org/10.3969/j.issn.1007-9572.2017.07.y15>.
14. Zhu YB, Wang Q, Shi HM, Yu XH. Formulation and evaluation on short version of Chinese medical constitution questionnaire with 30 items. *J Tradit Chin Med.* 2018;59(18):1554–9.
15. Meng QG, Wang LX, Zhao SC. The important role of nonlinearity in Chinese medicine research. *Chin Arch Tradit Chin.* 2005;23(9):1557–8. <https://doi.org/10.3969/j.issn.1673-7717.2005.09.005>.
16. Froelich W, Deja R. Selection a group of features based on machine learning algorithms to simplify psycho-technical examination. *Proced Comput Sci.* 2022;207:319–26. <https://doi.org/10.1016/j.procs.2022.09.065>.
17. Lee W, Kim H, Shim J, Kim D, Hyeon J, Joo E, Joo BE, Oh J. The simplification of the insomnia severity index and epworth sleepiness scale using machine learning models. *Sci Rep.* 2023;13(1):6214. <https://doi.org/10.1038/s41598-023-33474-8>.
18. China Association of Chinese Medicine. Classification and determination of Chinese medicine constitution (ZYYXH/T157–2009). *World J Integr Tradit West Med.* 2009;4(04):303–4.
19. Olson RS, Urbanowicz RJ, Andrews PC, Lavender NA, Kidd LC, Moore JH. Automating biomedical data science through tree-based pipeline optimization. In: Squillero G, Burelli P, editors. *Applications of evolutionary computation: 19th European Conference, EvoApplications 2016 on Springer International Publishing: 30 Mar–1 Apr, 2016, Porto, Portugal; 2016.* p. 123–37. <https://doi.org/10.1007/978-3-319-31153-1>.
20. Hair JF Jr, Sarstedt M, Hopkins L, Kuppelwieser VG. Partial least squares structural equation modeling (PLS-SEM): an emerging tool in business research. *Eur Bus Rev.* 2014;26(2):106–21. <https://doi.org/10.1108/EBR-10-2013-0128>.
21. Hosmer DW. Assessing the fit of the model. In: Hosmer DW, Lemeshow S, Sturdivant RX, editors. *Applied logistic regression.* 2nd ed. New York: John Wiley and Sons; 2000. p. 160–4. <https://doi.org/10.1002/sim.1236>.
22. Jolliffe IT. Choosing a subset of principal components or variables. In: Jolliffe IT, editor. *Principal component analysis.* 2nd ed. New York: Springer; 2002. p. 111–49. https://doi.org/10.1007/0-387-22440-8_6.
23. Avram C, Gligor A, Roman D, Soylu A, Nyulas V, Avram L. Machine learning based assessment of preclinical health questionnaires. *Int J Med Inform.* 2023;180:105248. <https://doi.org/10.1016/j.ijmedinf.2023.105248>.
24. Hermanns N, Kulzer B, Krichbaum M, Kubiak T, Haak T. How to screen for depression and emotional problems in patients with diabetes: comparison of screening characteristics of depression questionnaires, measurement of diabetes-specific emotional problems and standard clinical assessment. *Diabetologia.* 2006;49:469–77. <https://doi.org/10.1007/s00125-005-0094-2>.
25. Coulehan JL, Schulberg HC, Block MR. The efficiency of depression questionnaires for case finding in primary medical care. *J Gen Intern Med.* 1989;4:541–7. <https://doi.org/10.1007/BF02599556>.
26. Zhang ZX, Zhang Y, Wu LL, Chen M, Zhou CL, Shen H. A survey on the relationship between the constitutional patterns of traditional Chinese Medicine and Chronic Fatigue Syndrome (CFS). *Liaoning J Tradit Chin Med.* 2011;38(1):22–6.
27. Chen JZ, Wang F, Zhang Y, Nie JH. A review on the relationship between hypertension and phlegm-dampness constitution. *Clin J Chin Med.* 2022;14(20):127–30.
28. Chen J, Li JW, Chen SF, Wang G, Shen CC, Liu XW. Analysis of acne under appearance of “dampness heat” based on constitution theory. *J Basic Chin Med.* 2023;29(8):1272–4.
29. Pan LW, Wang XM, Yang XZ, Li GF. Analysis of the susceptible diseases of phlegmatic hygrois constitution. *J Jilin Chin Med.* 2017;37(8):761–5.
30. Afari N, Buchwald D. Chronic fatigue syndrome: a review. *Am J Psychiatry.* 2003;160(2):221–36. <https://doi.org/10.1176/appi.ajp.160.2.221>.
31. Basted AC, Marshall LM. Review of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: an evidence-based approach to diagnosis and management by clinicians. *Rev Environ Health.* 2015;30(4):223–49. <https://doi.org/10.1515/reveh-2015-0026>.
32. Cascio MJ, DeLoughery TG. Anemia: evaluation and diagnostic tests. *Med Clin.* 2017;101(2):263–84. <https://doi.org/10.1016/j.mcna.2016.09.003>.
33. Brunzell JD, Hokanson JE. Dyslipidemia of central obesity and insulin resistance. *Diabetes Care.* 1999;22:C10–3.
34. Sinikumpu SP, Jokelainen J, Tasanen K, Huilaja L. Cardiovascular and metabolic profile of subjects with acne in a cohort of middle-aged patients: a general population study of 1932 subjects. *Acta Dermato-Venereol.* 2023. <https://doi.org/10.2340/actadv.v103.5921>.
35. Goodarzi MO, Dumesic DA, Chazenbalk G, Azziz R. Polycystic ovary syndrome: etiology, pathogenesis and diagnosis. *Nat Rev Endocrinol.* 2011;7(4):219–31. <https://doi.org/10.1038/nrendo.2010.217>.
36. Feng L, Hancock R, Watson C, Bogley R, Miller ZA, Gorno-Tempini ML, Briggs-Gowan MJ, Hoeft F. Development of an abbreviated adult reading history questionnaire (ARHQ-Brief) using a machine learning approach. *J Learn Disabil.* 2022;55(5):427–42. <https://doi.org/10.1177/00222194211047631>.
37. Zien A, Krämer N, Sonnenburg S, Rättsch G. The feature importance ranking measure. In: Frasconi P, Landwehr N, Manco G, Vreeken J, editors. *Machine Learning and Knowledge Discovery in Databases: European Conference on Springer Berlin Heidelberg: 7–11 September 2009, Bled, Slovenia; 2009.* p. 694–709. https://doi.org/10.1007/978-3-642-04174-7_45.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.