Machine learning-based nutrient classification recommendation algorithm and nutrient suitability assessment questionnaire

JaHyung Koo[1], LanMi Hwang[1], HooHyun Kim[1], TaeHee Kim[1], JinHyang Kim[1], and HeeSeok Song[1*]

[1] SEERS, 1401, 8, Seongnam-daero 331beon-gil, Bundang-gu Seongnam-si, Gyenoggi-do, 13558. Republic of Korea
[e-mail: chris.koo@seerstech.com, sam.song@seerstech.com]
[*]Corresponding author: HeeSeok Song

Received November 1, 2022; revised November 29, 2022; accepted December 21, 2022; published January 31, 2023

Abstract

The elderly population is increasing owing to a low fertility rate and an aging population. In addition, life expectancy is increasing, and the advancement of medicine has increased the importance of health to most people. Therefore, government and companies are developing and supporting smart healthcare, which is a health-related product or industry, and providing related services. Moreover, with the development of the Internet, many people are managing their health through online searches. The most convenient way to achieve such management is by consuming nutritional supplements or seasonal foods to prevent a nutrient deficiency. However, before implementing such methods, knowing the nutrient status of the individual is difficult, and even if a test method is developed, the cost of the test will be a burden. To solve this problem, we developed a questionnaire related to nutrient classification twice, based upon which an adaptive algorithm was designed. This algorithm was designed as a machine learning based algorithm for nutrient classification and its accuracy was much better than the other machine learning algorithm.

Keywords: Machine Learning, Classification, Nutrient Deficiency, Smart HealthCare, Adaptive Algorithm, Questionnaire

http://doi.org/10.3837/tiis.2023.01.002 **ISSN : 1976-7277**

1. Introduction

As interest in nutrition has grown based on information regarding various ingredients and methods of intake, the consumption of healthy food has also significantly increased. However, some people consume large amounts of saturated fats through fast food and do not eat a sufficient amount of fruits and vegetables, which is far from a healthy diet, leading to an increase in obesity and hypercholesterolemia [1, 2]. In addition, according to data released in 2015 and 2020 by the Korean Statistical Information Service (KOSIS), the population over the age of 60 has increased by approximately 22.5% during the past 5 years, and the trend is occurring nationwide [3]. As the population continues to age, interest in life expectancy, disease prevention, and management-oriented health care is increasing. Based on these social problems and phenomena, in this paper, an algorithm for nutrient recommendations based on user questionnaire data is proposed, where the results of the user's nutrient deficiencies are found through an algorithm and a nutritional diet is thereby recommended. This paper is about thesis research results from a project titled "A medical interview-based personalized smart food recommendation service platform."

There are many studies [4-11], Korea health examination questionnaire [12], and commercial products [13, 14] recommending nutritional diets for subjects based on questionnaires. However, owing to a difficulty in organizing research data and a security concern regarding the handling of biometric information, some researches [4, 6-8] were conducted on specific areas rather than on all regions, ages, and genders. For research data, public data based on national surveys and data compiled through agreements are available [15]. This study recognizes this problem, and thus the subjects were asked to fill out a consent form for personal information use to ensure the suitability as research data. Moreover, an electronic consent form was prepared for the Macrogen saliva test [16], which is used as the reference data. The recommended nutrient algorithm is based on data gathered through consent forms.

In addition, we previously analyzed [5, 9-11] for reference in the research progress in recent papers on nutrients. It is a research method for classifying food groups through machine learning using the food exchange system [5], and since our study also recommends nutrients through questionnaires, the direction is similar and used as reference data. [9-11] are references to examine the current status of nutrient-based AI researches. Through these papers, based on questionnaire data, the development direction for the AI-based recommended nutrient classification algorithm design was progressed.

The structure of this paper is as follows. In section 2, the questionnaire development is divided into the design and verification stages, and in section 3, to verify the results of the questionnaire, we introduce reference data and describe the nutrient recommendation algorithm. Section 4 evaluates the similarities between the questionnaire and reference data based on the results of the proposed algorithm, and evaluates the possibility of applying the questionnaire developed in this study. Section 5 provides some concluding remarks and suggests future research directions.

2. Overview of Medical Examination

This section describes the development and verification process used for a medical examination (M.E). The development progressed from version (v) 1.0 to the final version, v2.4.

2.1 Introduction to M.E

An overall introduction to the questionnaire is described in this subsection. Starting with the design of the questionnaire, the process of selecting a suitable research subject for the questionnaire, the data collection, and the reference data selection process, the selection of detailed items for the questionnaire and finally the questionnaire verification methods are described.

2.1.1 Questionnaire design

In the early stage of designing the questionnaire, we examined surveys conducted by many different companies and their resulting recommended "customized nutrients" [13, 14]. For each company, the questions are sequentially structured, and nutrient recommendations are made possible through step-by-step inquiries. Through a direct examination of the questionnaires of other companies, the flow and gist of the questions were identified and analyzed. The various symptoms, diseases, lifestyles, and health concerns are unique to each individual. Most companies use methods in which the respondents answer questions through a simple questionnaire and deficient or excess nutrients are easily determined based on the answers [13, 14]. However, through the questionnaire, we found that the use of methods developed by other companies is limited in terms of recommending a diet based on nutrient recommendations founded on nutritional necessity. Rather than recommending nutrients that an individual wants, to understand what nutrients the individual is deficient in, questions that can help understand the general eating habits were added to our basic questionnaire.

2.1.2 Questionnaire selection process for research subjects

During the research subject selection process, the respondents were grouped through itemized questionnaires, and through additional questionnaire development, the groups were subdivided into regions (Jeju Island and mainland residents) based on gender and age. However, the classification of the subjects based on their characteristics was difficult, and finding subjects from a wide range of groups was even more challenging; thus, the subjects who could easily give feedback were selected. [3, 17-24].

2.1.3 Data collection and reference data selection process for questionnaire

The reference data used in this study are the personal direct to consumer (DTC) test data of Macrogen's MyGenome Story (MGDTC) [16]. MGDTC is a genome test conducted using saliva and provides a total of 72 types of results. We used the nutrient portion of the results and applied only six of the nutrient results for collaboration with other participating agencies, i.e., calcium, potassium, ferrous (iron), zinc, magnesium, and vitamin C. The DTC test results classify the degree of deficiency for each nutrient into safe, normal, and deficient, and we relabeled the categories as unnecessary, moderately recommended, and strongly recommended, respectively. Macrogen was chosen as the reference because it is the premier company in the country for genome testing, with numerous research papers [25-27] related to genome testing available, and is uniquely able to coordinate genome test results with other cooperating institutions.

2.1.4 Selection of detailed items for the questionnaire

The deficient nutrients were selected through v1.0 of this questionnaire and a sampling of the subjects. Reference (MGDTC) response results and questionnaire v1.0 were derived from the target population, based upon which, questionnaire v2.0 was updated by reflecting the symptoms and characteristics of nutrient deficiency through an expert data analysis. Although v2.0 was designed for simple nutrient recommendations, we determined that it is difficult to recommend a diet suitable for individuals by simply extracting the results of various nutrients. Therefore, based on the information on their daily eating habits, the recommendations regarding necessary nutrients for the respondents were generated daily. Each question on the questionnaire was created based on the National Health Statistics [28] and 2020 Nutrient Intake Standards for Koreans [1, 2] provided by the Ministry of Health and Welfare. After completion of the questionnaire v2.1, we inquired about the appropriateness of the questions and the validity of the nutrient recommendations by consulting an expert in the field of nutrition at Jeju National University [29-32]. From the consultation results, we updated to versions 2.2, 2.3, and 2.4. The questionnaire update was based on studying various references [2, 17-24]. For example, the nutrients necessary for pregnant women were referred to through the corrigendum [17] of the latest data from the Ministry of Health and Welfare. In the case of the questionnaire developed in this study, since it is conducted for general people, through data from the Ministry of Education and the Korea Centers for Disease Control and Prevention [18], nutrients essential for adolescents were analyzed and referenced so that they could be included in the questionnaire. For the recommendation of nutrients according to drug intake in the early stage of the development of the questionnaire, the elderly nutrition section of the related book [20] was referred. The final responses were from questionnaire v2.4.

2.1.5 Explanation of improving questionnaire version

This section summarizes the changes to the questionnaire used in this study. **Table 1** shows the direction in which each questionnaire progressed, and **Table 2** lists the revised contents corresponding to **Table 1**. A total of five modifications occurred from v2.0 to v2.1, i.e. two changes and three additions. For example, through [1] and [2], we changed and added relevant questions from the items related to the recommended nutrients desired by the existing users, focusing on the changes in their usual eating habits. A total of 8 areas of interest, including exercise, chronic disease, and stress, were subdivided into 15 categories and modified to be more applicable to the classification of the recommended nutrients. In **Table 1** and **Table 2**, the part of the questionnaire that occurred while progressing from v1.0 to v2.0 was not filled in because it was not applicable.

There were 10 modifications, 2 additions, 5 deletions, and 1 retention out of a total of 18 revisions in v2.2. Questionnaire v2.2 was carried out with the advice of a professor in the Department of Food and Nutrition, Jeju National University. Among the items added in v2.1, the question regarding eating habits, i.e., “the number of consumptions per week” was changed based on a consultation, related studies, and books [19, 24]. For example, 1 or 2 times a week was changed to 2 or more times a day, once a day, and 4–6 times a week. As an additional example, based on [23] and an advisory opinion that there is a need to differentiate according to an increase in the number of vegetarians in modern society, a vegetarian category was also inserted. Because nutrients are also included in vegetables, a question related to vegetable intake was added. Based on [22], in relation to the drinking intake question, the standard value for the examples was changed.

V2.3 modified the weight-related items. A 5-kg change can have a relative effect depending on the body weight of the subject, and thus it was corrected to 5% of body weight. In addition, because “a person who does not drink” could not be drawn from the answers related to drinking intake in v2.2, the above contents were changed into two contents related to drinking intake. Using this modified questionnaire, a questionnaire on the first-round feedback was conducted.

V2.4 added dietary guidelines for the elderly according to their birth year based on evidence from [21] and first-round feedback. This occurred because, although a classification according to gender is important, the types of nutrient recommendations and target users according to age differ.

Finally, **Fig. 1** is part of the screen of the questionnaire web and app proposed in this study, and the overall thesis research structure can be shown in **appendix A**.

2.1.6 Method for M.E verification

For the questionnaire verification, a machine learning method was used to calculate the degree of agreement between the questionnaire results and the reference results according to each questionnaire version. In questionnaire v1.0, among the different machine learning methods, verification was applied using a multiclass support vector machine (SVM) [33]. Questionnaire v2.4 was verified through the random forest [34]. Verification is described in greater detail in section 4.

Table 1. Status of the questionnaire version (Consultation: A = Used; N/A = Not used)

Questionnaire version	Questionnaire item modification type					Consultation
	Addition	Modification	Deletion	Retention	Total	
2.0 → 2.1	2	2	0	0	4	
2.1 → 2.2	2	10	5	1	18	A
2.2 → 2.3	4	9	3	0	16	
2.3 → 2.4	6	0	2	0	8	N/A

Table 2. Representative modifications according to the questionnaire version

Questionnaire Version	Content change according to questionnaire version		Reference	Consultation Reference
	Before modification	After modification		
2.0 → 2.1	Recommendation of nutrients that users want	Change the user's eating habits and add related questions	[1, 2]	[29]
	8 items of interest	15 items of interest		
2.1 → 2.2	None	Vegetarian related question	[23]	
	Alcohol (soju 1 bottle, beer 1.5 bottles)	Alcohol (soju 1 shot, beer 1 glass)	[22]	
	Times per week	Subdivided into day and week	[19, 24]	[30]
2.2 → 2.3	Alcohol-related question-Drink more	Do not drink	[22]	
	Body weight change of 5 kg or more	Body weight change of 5 % or more	N/A	
2.3 → 2.4	None	Added dietary guidelines for seniors by year of birth	[21]	N/A

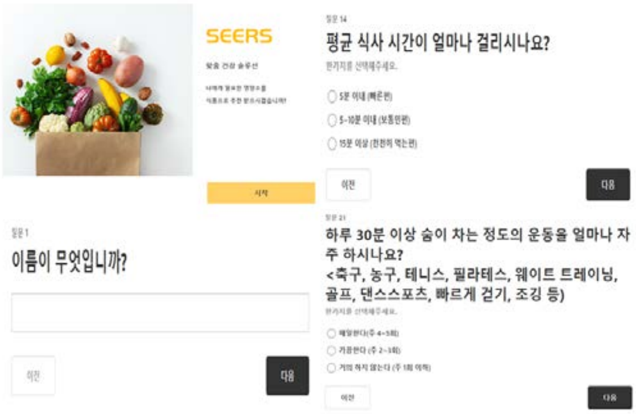


Fig. 1. Example image of questionnaire web and app used in this study

3. Proposed Method

3.1 Overview of proposed system

Fig. 2 is a schematic diagram of the overall system. The user performs a questionnaire, and the result of the questionnaire is used as input data. The input data is divided into those corresponding to 6 nutrients, and if a special condition is satisfied even if the same response according to the nutrient, a weight is given to the input data. Feature vectors are obtained through a machine learning model. Nutrient deficiency three scale values are calculated as feature vectors. The calculated results are expressed as 0, 1 and 2, where 0 means unnecessary, 1 means moderately recommended, and 2 means strongly recommended. The accuracy of the model is compared with the labeling data, Macrogen data, and the number of matches is obtained.

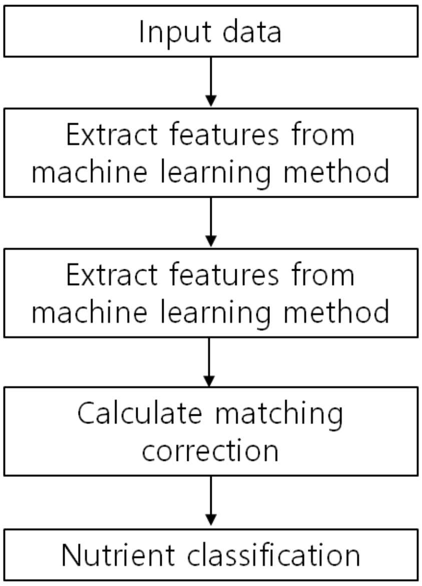


Fig. 2. Experimental flow chart of algorithm proposed in this study

3.2 Model for proposed method

The structure used in this study is based on random forest [34]. To obtain the performance by the optimal model for each of the six nutrients, the model in scikit-learn [35], a library provided by Python, was used, and to obtain the optimal performance and parameters were found through HalvingGridSearchCV. According to HalvingGridSearchCV, the optimal parameters for each of the six nutrients are not the same.

3.3 Implementation details of HalvingGridSearchCV

An additional explanation of HalvingGridSearchCV, which is a means to find the optimal parameter from machine learning methods, will be given. Random forest, decision tree, and SVM were used for the model using HalvingGridSearchCV, and logistic regression and k-Nearest Neighbor (k-NN) were not included in HalvingGridSearchCV, so the experiment was conducted with the randomly selected parameter values. Random forest parameter changed the values of max_depth, min_sample_split, and n_estimators, SVM changed gamma, C value, and kernel type, and in the case of decision tree, criterion, splitter, max_depth, and max_features were changed. Values related to parameters are set to be provided to scikit-learn.

4. Experiments and Analysis

4.1 Experimental data analysis

This section introduces a detailed description of the reference data and the data used in this experiment and describes a comparison and analysis of the results from the development of questionnaire v1.0 to v2.4.

4.1.1 Processing of experimental data

The data from the final experimentation are composed of users who answered questionnaire v2.4. This was achieved using questionnaire data from a total of 90 people, including 25 out of 30 users who had answered questionnaire v1.0, excluding 5 who had left the company; in addition, the corresponding reference data, i.e., Macrogen test data, were also used. **Table 3** divides the data used in the experiment into training and testing data and specifies the total data.

Table 4 is a composition of experimental data in this study. Most of the questionnaires are related to various nutrients rather than the case where the answer to one question is related to one nutrient. For example, if you look at “All” part common to six nutrients, you can see that vitamin takes up the most, and potassium has the least, with four. As such, **Table 4** shows that each of the six nutrients has a part in common with other nutrients, but can also have an independent effect. According to The Eighth Korea Health Statistics 2020 [28], when Koreans answered “not applicable” to a specific question about nutrients that are generally lacking in Korea, additional intake of calcium, zinc, vitamin C reflected so that it can be done [31, 32].

Table 3. Number of users who responded to the questionnaire by version

Data	M.E v1.0		M.E v2.4	
	M.E v1.0	MGDTC	M.E v2.4	MGDTC
Training data	24	24	63	63
Testing data	6	6	27	27
Total data	30	30	90	90

Table 4. Composition of experimental data

Nutrient	Answer type	Number of answer	Total	Total ratio (%)
Calcium	All	8	31	20.5
	Only calcium	7		
	[28, 31, 32]	10		
	Not recommended	6		
Potassium	All	4	10	6.6
	Sharing calcium	5		
	Only potassium	1		
Ferrous	All	8	27	17.8
	Sharing calcium	2		
	Only ferrous	2		
	Sharing vitamin C	1		
	[28, 31, 32]	10		
	Not recommended	4		
Magnesium	All	7	21	13.9
	Only magnesium	6		
	Sharing calcium	7		
	Sharing potassium	1		
Zinc	All	7	16	10.6
	Only zinc	5		
	Not recommended	4		
Vitamin C	All	21	46	30.5
	Only vitamin C	12		
	Sharing ferrous	1		
	[28, 31, 32]	10		
	Not recommended	2		

4.2 Experimental result analysis

The model developed in this study uses a model for each of the six nutrients. Considering the model complexity and computing speed, it was better to divide the models that classify the recommended nutrients according to a genetic test of the user rather than constructing a single model. For example, in the case of the calcium model, the part corresponding to calcium among the values for the individuals is input into the algorithm and compared with MGDTC, the labeled data, and the accuracy is calculated. The resulting values for these examples are shown in **Table 5**, and the results for **Table 5** are shown in a confusion matrix [36] in **Fig. 3**.

In **Table 5**, it can be seen that potassium and vitamin C are lower than the others. According to **Table 4**, the number of answers for potassium is 10, which is the smallest number among the six nutrients. Also, comparing the results of MGDTC, which is the reference data, with the user's questionnaire responses, it was confirmed that the responses for a total of 90 people and the results of MGDTC were different. We think that the majority of people respond because they are insensitive to the effects of potassium compared to other nutrients. However, vitamin C has a low performance, but shows the opposite pattern. Since vitamin C is the nutrient that occupies the largest portion of the 151 data, it can be considered that there is an advantage in classifying the recommended nutrients relatively. However, although the response sheet shows the direction of 2, most of MGDTC's results point to 1, and there is also 0, so it can be judged that the classification accuracy is low.

In **Fig. 3**, it can be seen that the result of the confusion matrix of calcium, magnesium, and zinc is overfitting to 1. It could be determined that the number of data of 0 and 2 was smaller than that of 1, and that the similarity between the reference and input data of 0 and 2 was lowered.

Table 5. Results for six nutrients by proposed method

Method	Accuracy	Total average
Calcium	0.852	0.723
Potassium	0.519	
Ferrous	0.704	
Magnesium	0.852	
Zinc	0.778	
Vitamin C	0.63	

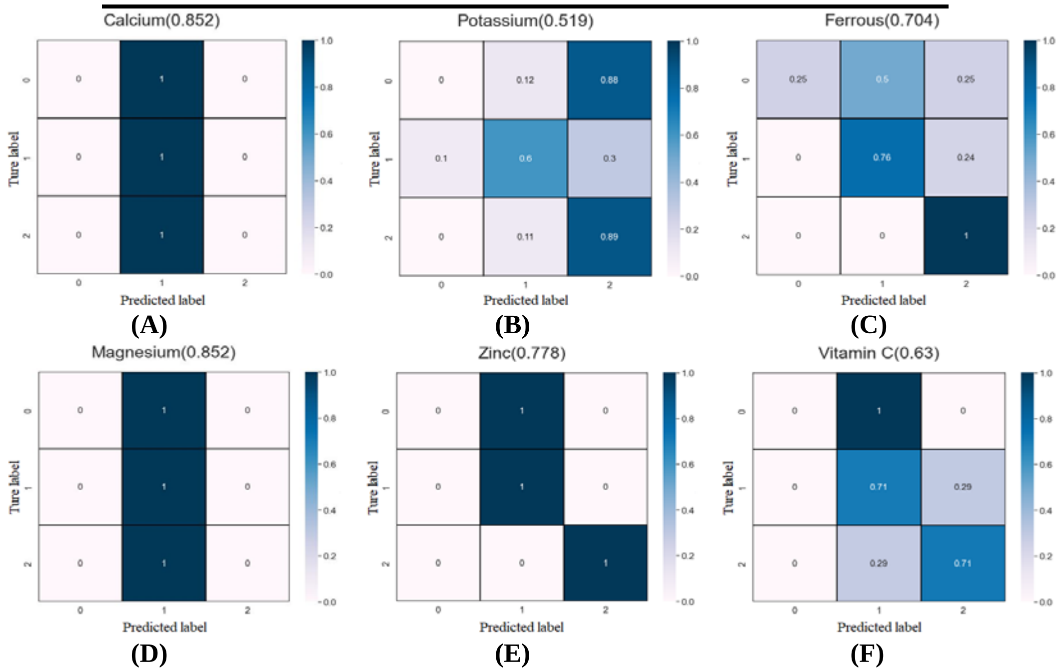


Fig. 3. Confusion matrix of the results from (A) calcium, (B) potassium, (C) ferrous, (D) magnesium, (E) zinc, and (F) vitamin C

4.2.1 Comparison among the proposed method and other machine learning approaches

Next, the results of the method proposed in this study and other existing machine learning approaches were compared and analyzed. Although the original nutrient analysis method uses a statistical analysis model from the "Statistical Package for Social Sciences" (SPSS) [37], because commercialization is essential for this study, we chose to implement other machine learning methods. **Table 6** summarizes and compares the results between the method proposed in this study and the existing machine learning approaches. From the table, one can check the classification accuracy for each of the six deficient nutrients and the overall average classification accuracy. For comparison, logistic regression (LR) [38], SVM [33], k-NN [39] and decision tree (DT) [40] were used. **Fig. 4** shows the results of the classification of the six nutrients in **Table 6** as a confusion matrix. **Fig. 5** displays the results for **Table 6** in a graph form, which indicates the superiority of the proposed approach over the other methods.

In **Table 6**, it can be seen that the results of the random forest method and decision tree proposed in this study are very similar. The reason is that the random forest model is similar to the decision tree, but in the case of a decision tree, there is one tree, whereas in the random forest, there is more than one. This is the reason why the random forest, which compensates for the shortcomings of the decision tree according to [41], was used as the model proposed in this study. **Fig. 4** shows the results of **Table 6** as a confusion matrix.

Table 6. Comparison of results between proposed method and other machine learning method

Method	Accuracy
SVM [32]	0.644
LR [38]	0.568
k-NN [39]	0.605
DT [40]	0.659
Proposed [33]	0.723

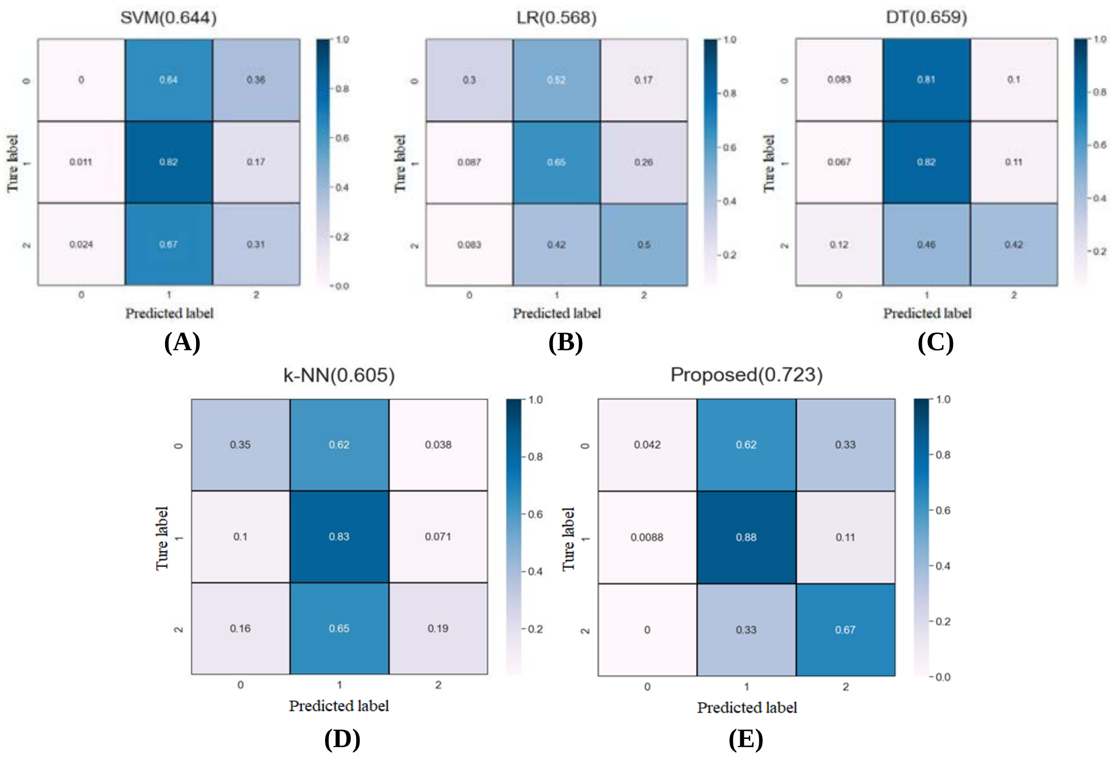


Fig. 4. Confusion matrix for total average accuracy results by using proposed method and other machine learning method. (A) SVM, (B) logistic regression (LR), (C) decision tree (DT), (D) k-NN, and (E) proposed

5. Conclusions

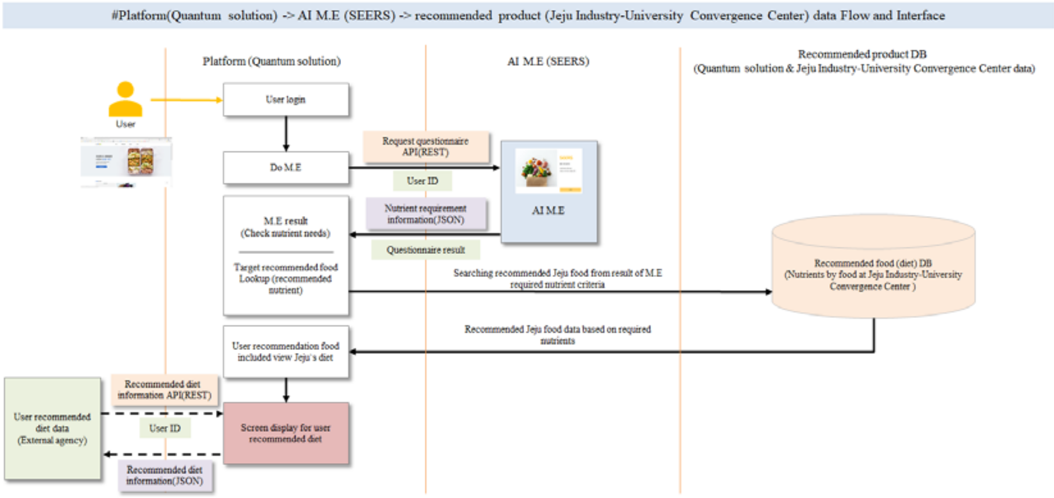
This paper is about of a thesis research result titled “A medical interview-based personalized smart food recommendation service platform.” During approximately 2 years, the above questionnaire evolved from v1.0 to v2.4. The questionnaire items were revised based on numerous studies on nutrient recommendations [2, 17-24] and consultations [29-32]. To verify whether the degree of nutrient deficiency was properly classified, a total of 90 users were recruited for the final questionnaire. To classify the degree of nutrient deficiency, Macrogen’s saliva test was used as reference data, and among the machine learning methods, the random forest-based model was used to design the algorithm applied to classification. Furthermore, to evaluate the performance of the model utilized in this study, decision tree, logistic regression, SVM, and k-NN methods were compared. From the performance comparison, it was confirmed that the proposed method in this study achieved better results. In addition, it was confirmed that the recall value of 1, which is moderately recommended among the three recommended methods, was almost close to 1. Through this, there will be an advantage in that the user can receive the recommended nutrients among the six nutrients through the our questionnaire. However, in the case of potassium or zinc, the unnecessary value of 0 is incorrectly judged as the strongly recommended value of 2. In conclusion, if the service equipped with the algorithm of this study is implemented, the our questionnaire response data will be accumulated, and if the algorithm is developed through them, this problem will be reduced. In addition, the questionnaire in this study was designed to be able to recommend nutrients according to alcohol intake based on [42]. Therefore, one of the characteristics is that it can affect the six nutrients recommended to the user according to the response of the related question items.

However, there are some limitations to this study. First, although some public data exist, they could not be used because the nature of the public data of the Ministry of Health and Welfare differs from the data used in this study, and data from other studies could not be applied because they contained specific personal information. Second, the reference data were limited. It was difficult to conduct other reference tests to prove the validity of the questionnaire results. This makes it difficult to find other test methods that can analyze the content and nutrient deficiencies, and it was therefore difficult to complete the task by the deadline and re-test a total of 90 people. As a result, two of the limitations can be considered as shortcomings of this study. To compensate for these shortcomings, various references and consultations were used.

Consequently, the platform service is the final product of the thesis research. Through this service, various users can check the nutritional information that is lacking through the questionnaire. It also provides a recommended diet for deficient nutrients. In the future, it is hoped that various services in a developed form will come out through what is mentioned in the research papers [43, 44] related to this study. we intend to conduct research on recommended nutrients for a specific age or a specific race like the study in [45].

Appendix

Appendix A. Total system of including SEERS nutrient questionnaire Web/App Structure



Acknowledgments

This research was financially supported by the Ministry of Trade, Industry and Energy, Korea, under the "Regional Innovation Cluster Development Program(R&D, P0016217)" supervised by the Korea Institute for Advancement of Technology(KIAT).

Reference

[1] The Korean Nutrition Society, "Energy and macronutrients of dietary reference intakes for Koreans 2020," Ministry of Health and Welfare, Sejong-si, Republic of Korea, 11-1352000-002852-01, Dec. 2020. [Online]. Available: http://www.mohw.go.kr/react/jb/sjb030301vw.jsp?PAR_MENU_ID=03&MENU_ID=032901&CONT_SEQ=362385

[2] The Korean Nutrition Society, "Application of dietary reference intakes for Koreans 2020," Ministry of Health and Welfare, Sejong–si, Republic of Korea, 11–1352000–003193–01, Jan. 2022.

[3] Statics Korea, KOSIS, 2020 population census-Seniors(over 60 years old) by gender/age/educational level; city, country, and gu. [Online]. Available: https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1PG2001&conn_path=I3

[4] B. J. Lee, "Prediction model of hypertension using sociodemographic characteristics based on machine learning," *KIPS Transactions on Software and Data Engineering*, vol. 10, no.11, pp. 541–546, Nov. 2021. Article(CrossRefLink)

[5] J. Y. Kim, S. M. Lee, H. J. Jeon, G. U. Kim, J. H. Kim, N. U. Park, C. G. Jin, J. Y. Kwon, and J. W. Kim, "Food exchange table organization model based on decision tree using machine learning," in *Proc. of the Korea Information Processing Society Conference*, pp. 680-684, 2020. Article (CrossRef Link)

[6] M. J. Choi and S. J. Yun, "The interest of female high school student's for weight control and nutrient intake status in the Daegu area," *J. East Asian Soc. Dietary Life*, vol. 17, no. 3, pp. 329–337, Jun. 2007. Article(CrossRefLink)

[7] H. S. Chang, “The study of nutrient intakes, blood lipids and bone density according to obesity degree among university students in Jeonbuk,” *Korean Journal of Human Ecology*, vol. 23, no, 4, pp. 743–757, Aug. 2014. Article(CrossRefLink)
[8] H. S. Park, D. K. Kim, S. Y. Lee, H. S. Kim, and H. C. Suh, “Validation of questionnaire for chewing and swallowing function of the elderly,” *The Korean Dysphagia Society*, vol. 7, no. 1, pp. 25–34, Jan. 2017. Article(CrossRefLink)
[9] J. Sak and M. Suchodolska, “Artificial intelligence in nutrients science research: A review,” *Nutrients*, vol. 13, no.2, pp. 322, Feb. 2021. Article(CrossRefLink)
[10] D. Kirk, C. Catal, and B. Tekinerdogan, “Precision nutrition: A systematic literature review,” *Comput. Biol. Med.*, vol. 133, pp. 104365, Jun. 2021. Article(CrossRefLink)
[11] N. V. Matusheski, A. Caffrey, L. Christensen, S. Mezgec, S. Surendran, M. F. Hjorth, and S. Péter, “Diets, nutrients, genes and the microbiome: Recent advances in personalised nutrition,” *Br. J. of Nutr.*, vol. 126, no. 10, pp. 1489–1497, Nov. 2021. Article(CrossRefLink)
[12] Korean Industrial Health Association, *Health Examination Questionnaire*. [Online]. Available: https://kiha21.or.kr/?page_id=4604
[13] K. J. Kim, E. J. Kim, Y. R. Song, Y.J. Kim, S. H. Chun, and J. Y. Kim, “Current Status of Personalized Dietary Supplement,” *Food Industry and Nutrition*, vol. 25(2), pp. 20-37, Dec. 2020. Article(CrossRefLink)
[14] Pilly, pillycare, [Online]. Available: https://pilly.kr/
[15] Ministry of Health and Welfare, *Korea Health Statistics 2017: The Seventh Korea National Health and Nutrition Examination Survey (KNHANES VII)*. [Online]. Available: https://knhanes.kdca.go.kr/knhanes/sub03/sub03_02_05.do
[16] Macrogen, My Genomestory The Plus kit, [Online]. Available: https://www.my-genomestory.com/order?selectLang=ko
[17] Ministry of Health and Welfare, “KDRIs erratum 3rd correction moisture dietary fiber correction,” Ministry of Health and Welfare, Sejong–si, Republic of Korea, Jan. 29, 2021. [Online]. Available: http://www.mohw.go.kr/react/jb/sjb030301vw.jsp?PAR_MENU_ID=03&MENU_ID=032901&CONT_SEQ=362385
[18] Korea Disease Control and Prevention Agency, “The 17th Korea youth risk behavior survey,” Ministry of Health and Welfare, Sejong–si, Republic of Korea, 11–1460736–000038–10, Apr. 2022.
[19] The Korean Nutrition Society, “Application of dietary reference intakes for Koreans 2020,” Ministry of Health and Welfare, Sejong–si, Republic of Korea, 11–1352000–003193–01, pp. 90–92, Jan. 2022.
[20] J. S. Han, *New Life Cycle Nutrition*. Paju–si, Gyeonggi–do, Republic of Korea: Jigu Culture Co. Ltd, 2016.
[21] Y.G. Park, H. S. Lim, D. R. Joo and M. H. Kim, “Dietary guide book designed for active senior from age 50 to 64: For health care personnel,” Ministry of Food and Drug Safety, Cheongju, Chungcheongbuk–do, Republic of Korea [Online]. Available: https://www.ganghwa.go.kr/open_content/clinic/bbs/bbsMsgDetail.do?msg_seq=277&bcd=nutrition&pgno=1, Accessed on: Aug. 17, 2021
[22] Ministry of Health and Welfare. Bugok National Hospital, Prevention Information. [Online]. Available: https://www.bgnmh.go.kr:2448/checkmehealme/bbs/bbsView.xx?catNo=2&idx=8
[23] U.S. Department of Agriculture and U.S. Department of Health and Human Services. *Dietary Guidelines for Americans, 2020–2025*. 9th Ed. December 2020. [Online]. Available: https://www.dietaryguidelines.gov.
[24] Korea Disease Control and Prevention Agency, Division of Health and Nutrition Survey and Analysis. *Average intake by nutrient (region, gender, age) reference value data, 1998~2020.* [Online]. Available: https://kosis.kr/statHtml/statHtml.do?orgId=177&tblId=DT_11702_N028&conn_path=I2 (accessed on 7 June 2022).
[25] J. I. Kim, Y. S. Ju, J. S. Seo, and et al., “A highly annotated whole-genome sequence of a Korean individual,” *Nature*, vol. 460, no. 7258, pp. 1011-1015, Aug. 2009. Article(CrossRefLink)

[26] J. S. Seo, J. Y. Shin, C.H. Kim, and et al., “De novo assembly and phasing of a Korean human genome,” *Nature*, vol. 538, no.7624, pp. 243-247, Oct. 2016. Article(CrossRefLink)
[27] H. S. Park, J. I. Kim, J. S. Seo, and et al., “Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing,” *Nat Genet*, vol. 42 no. 5, pp. 400-405, Apr. 2010. Article(CrossRefLink)
[28] Korea Disease Control and Prevention Agency, “The Eighth Korea Health Statistics 2020: Korea National Health and Nutrition Examination Survey (KNHANES VIII-2),” Ministry of Health and Welfare, Sejong–si, Republic of Korea, 11–1351159–000027–10, Jan. 2022.
[29] J. H. Kim, “1st expert advisor statement,” SEERS, Gyeonggi–do, Republic of Korea, Jun. 10, 2022.
[30] J. H. Kim, “2nd expert advisor statement,” SEERS, Gyeonggi–do, Republic of Korea, Jun. 22, 2022.
[31] J. H. Kim, “3rd expert advisor statement,” SEERS, Gyeonggi–do, Republic of Korea, Jul. 29, 2022.
[32] J. H. Kim, “4th expert advisor statement,” SEERS, Gyeonggi–do, Republic of Korea, Aug. 1, 2022.
[33] C. Hsu and C. Lin, “A comparison of methods for multiclass support vector machines,” *IEEE Trans. Neural Netw*., vol. 13, no. 2, pp. 415–425, Mar. 2002. Article(CrossRefLink)
[34] M. Belgiu and L. Drăguţ, “Random forest in remote sensing: A review of applications and future directions,” *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, April 2016. Article(CrossRefLink)
[35] F. Pedregosa, G. Varoquaux, A. Gramfort, and et al., “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825-2830, Nov. 2011. Article(CrossRefLink)
[36] J. T. Townsend, “Theoretical analysis of an alphabetic confusion matrix,” *Percept. Psychophys.*, vol. 9, no. 1, pp. 40–50. Jul. 1971. Article(CrossRefLink)
[37] A. Bryman and D. Cramer, “Quantitative data analysis with SPSS 14, 15 & 16: A guide for social scientists,” Taylor & Francis Group, London, UK: Routledge, 2008.
[38] L. G. Grimm and P. R, Yarnold, “Logistic regression,” in *Reading and Understanding Multivariate Statistics*, 1st ed., Washington D.C., USA: American Psychological Association, 1995, pp. 217–244.
[39] L. M. Zouhal and T. Denoeux, “An evidence-theoretic k-NN rule with parameter optimization,” *IEEE Trans. Syst. Man Cybern. Part C (Applications and Reviews)*, vol. 28, no. 2, pp. 263–271, May, 1998. Article(CrossRefLink)
[40] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, “An introduction to decision tree modeling,” *J. Chemom.*, vol. 18, pp. 275-285, Nov. 2004. Article(CrossRefLink)
[41] B. Talekar and S. Agrawal, “A Detailed Review on Decision Tree and Random Forest,” *Biosci. Biotechnol. Res. Commun*, vol. 13, no. 14, pp. 245–248, Dec. 2020. Article(CrossRefLink)
[42] J. W. Jung, “Alcoholism: Factors and policy options,” *Health and welfare policy forum*, vol. 229, pp.6–16, Nov. 2015. Article(CrossRefLink)
[43] K. J. Kim, Y. K. Lee, and J. Y. Kim, “Current scientific technology and future challenges for personalized nutrition service,” *Food Sci. Ind.*, vol. 54, no. 3, pp. 145-159, Sep. 2021. Article(CrossRefLink)
[44] Y. H. Kim, “Precision nutrition: approach for understanding intra-Individual biological variation,” *J. Nutr. Health*, vol. 55, no.1, pp.1-9, Feb. 2022. Article(CrossRefLink)
[45] A. Talukder and B. Ahammed, “Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh,” *Nutrition*, vol. 78, pp. 110861, Oct. 2020. Article(CrossRefLink)

JA HYUNG KOO received the B.S. degree and Ph.D degree in electronics and electrical engineering from Dongguk University, Seoul, South Korea, in 2016 and 2021, respectively. He is currently working for SEERS. His research interests include biometrics signal and deep learning.

LANMI HWANG received the M.S. degree in Radio Communication Engineering from the Chungbuk National University, South Korea, in 2016, and is currently working for SEERS. Her Current research interests include vital sign processing, Window application, machine learning and deep learning.

HOOHYUN KIM received the M.S. degree in Image Signal Processing from the Hanyang University, South Korea, in 2013, and is currently working for SEERS. His Current research interests include vital sign processing, Window application, backend system, machine learning and deep learning.

TAEHEE KIM used to major in Chemical engineering in U.C Berkeley till 2015 and received the associate degree in Medical Laboratory Technologist from Gwangyang Health Sciences University, Korea, in 2019. After graduation, she worked in Kangbuk Samsung Hospital. She is currently employed SEERS and doing ECG analysis and presenting the government's business and business model.

JINHYANG KIM is in the Master's Program in Computational Statistics, Jeju National University, and received her Bachelor's degree in Chemical Engineering from Soonchunhyang University in 2002. She is the head of the Jeju branch of Seers Technology Co., Ltd. and general manager of the external cooperation team, in charge of discovering and planning the medical service business.

HEESEOK SONG received the B.S. and M.S. degrees in electronics engineering from Sogang University, Seoul, South Korea, in 1997 and 2000, respectively. He is currently the Vice-President and the Chief Technical Officer of SEERS Technology Company Ltd., Seongnam-si, Republic of Korea. His research interests include remote patient monitoring systems using wearable medical devices and biosignal analysis algorithms.