International Journal of Nursing & Clinical Practices
Short Communication Open Access

Prediction of Serum Vitamin D Levels in Japanese Older Adults Using XGBoost Algorithm and Logistic Regression

Noboru Hasegawa[1]*, Seiji Tsuchiya[2], Minatsu Kobayashi[3], Nobuko Shimizu[4], Yoshihito Tsubouchi[5], Takako Yamada[6], Mayumi Kato[7], Miyako Mochizuki[8], Hunsa Sethabouppha[9], Nattaya Suwankruhasn[9] and Chalinee Suvanayos[9]

[1]*Graduate School of Nursing, Doshisha Women's College, Kodo, Kyotanabe, Kyoto 610-0395, Japan*
[2]*Doshisha University, 1-3 Tatara Miyakodani, Kyotanabe-shi, Kyoto 610-0394, Japan*
[3]*Otsuma Women's University, 12 Sanban-cho, Chiyoda-ku, Tokyo 102-8357, Japan*
[4]*Toyama Prefectural University, 2-2-78 Nishinagae, Toyama 930-0975, Japan*
[5]*Naragakuen University, 3-15-1 Nakatomigaoka, Nara 631-8524, Japan*
[6]*Bukkyo University, 7 Higashitoganoo-cho, Nishinokyo, Nakagyo-ku, Kyoto 604-8418, Japan*
[7]*Aichi Medical College for Physical and Occupational Therapy, 519 Ichiba, Kiyosu-City, Aichi 452-0931, Japan*
[8]*Kyoto Bunkyo Junior College, 80 Senzoku, Makishima-cho, Uji, Kyoto 611-0041, Japan*
[9]*Chiang Mai University, Chiang Mai, 50200, Thailand*

Abstract

Background: Machine learning has unique advantages in dealing with complex interactions and nonlinear relationships. In this study, an XGBoost algorithm and logistic regression prediction model were established to predict the serum 25OHD levels in elderly people.

Methods: A total of 70 healthy adults age ≥65 years, taken from among adult day-care center clients, were included in the study. Body fat percentage (%Fat), serum 25OHD levels, total cholesterol (TC), triglycerides (TG), HDLcholesterol (HDL), LDL cholesterol (LDL) and blood sugar (BS) were measured. The daily intake of carbohydrate (CH), lipid and protein were assessed. The 25OHD levels in these subjects were classified as either deficient (<20) or insufficient (>=20 to 29.9). We tested the XGBoost algorithm to find features related to serum 25OHD levels using clinical data, daily intake and %Fat as predictors. A logistic regression model was used to analyze the factors associated with 25OHD levels. The receiver operating characteristic (ROC) was used to analyze the predictive value of the model and the area under the ROC curve (AUC) was calculated.

Results: The XGBoost algorithm's top 4 essential characteristics were "CH," "BS," "TG," and "TC," respectively. Protein intake is not a good predictor. The logistic regression model showed positive and negative coefficients. Positive values for the "BS," "CH," "LDL" and "Lipid" features decrease the risk of 25OHD insufficiency, and negative values for the "TG", "TC" and "%Fat" increase the risk of 25OHD insufficiency. The AUC of the logistic regression model was 0.76.

Conclusion: Our results indicate that by considering lipid intake and metabolism, the 2 algorithms are a good predictive tool for maintaining serum 25OHD levels.

Publication History:

Received: April 14, 2023
Accepted: May 07, 2023
Published: May 09, 2023

Keywords:

XGBoost, Logistic regression, Machine learning, serum 25OHD, Body fat percentage, Dairy intake, Clinical data

Introduction

Vitamin D is a secosteroid associated with peripheral calcium homeostasis and nervous system function, cancer, cardiovascular problems, autoimmune diseases, respiratory infections and allergies [1, 2]. Vitamin D is available in two forms, vitamin D2 from plants and D3 from animals. Both vitamin D2 and D3 are biologically inert and require activation through two hydroxylation processes involving 25-hydrooxylase (CYP2R1) and 1α-hydroxylase (CYP27B1), which are located in the liver and the kidney, respectively [2]. 1, 25-Dihydroxyvitamin D is a biologically active metabolite produced by two hydroxylation reaction steps [2].

Low 25OHD levels have been associated with skeletal muscle strength and physical performance [3]. In a previous study, we also showed that 25OHD supplementation was associated with improved serum 25OHD levels and possibly improved 4-m gait speed [4]. However, the features that influence blood 25OHD levels are not well understood.

Machine learning has unique advantages in dealing with complex interactions and nonlinear relationships [5, 6]. In recent years, the use of the XGBoost algorithm in medical treatment has also gradually increased [7-10].

Therefore, in this study, an XGBoost algorithm and logistic regression prediction model were established to predict the serum 25OHD levels in elderly subjects.

Materials &Methods

Subjects and Setting

Prior to conducting this study, approval was obtained from the ethics committee of the Aichi Medical University Ethics Review Board (2017-M052) in Japan. A total of 70 healthy adults age ≥65 years were included in the study, taken from among adult day-care center clients in Uji city (n=23, Kyoto), Eiheiji-cho (n=21, Fukui) and

***Corresponding Author:** Prof. Noboru Hasegawa, Graduate School of Nursing, Doshisha Women's College, Kodo, Kyotanabe, Kyoto 610-0395, Japan, Tel: +81-774-65-8855, Fax: +81-774-65-8820; E-mail: nhasegaw@dwc.doshisha.ac.jp

Citation: Hasegawa N, Tsuchiya S, Kobayashi M, Shimizu N, Tsubouchi Y, et al. (2023) Prediction of Serum Vitamin D Levels in Japanese Older Adults Using XGBoost Algorithm and Logistic Regression. Int J Nurs Clin Pract 10: 373. doi: https://doi.org/10.15344/2394-4978/2023/373

Nanao city (n=26, Ishikawa). Study researchers were present at the adult day-care centers to ensure the proper management of safety and confidentiality in the study. The managers of the adult day-care centers invited clients to participate in the study, and subjects were enrolled from August 2019 to June 2022. After obtaining informed consent from a family member belonging to the same household, we enrolled 28 Japanese men (age: 71.8 ± 4.9) and 42 Japanese women (age: 77.6 ± 7.9) in this study.

Body fat percentage

Body fat percentage (%Fat) was measured using Inbody 430 (Inbody Japan, Tokyo).

Daily intake

The Short Self-Administered Food Frequency Questionnaire was used to assess the dairy intake of carbohydrate (CH), lipid and protein (Education Software Co., Ltd., Tokyo).

Diaper changes performed by nurses using double gloving

This method is used by nurses when performing diaper changes daily in nursing practice. They wear 2 gloves before a diaper change, and if they find the second (outer) glove contaminated after the diaper change, they remove the contaminated glove, and use the first (inner) glove.

Clinical data

Blood was collected by venipuncture and serum 25OHD concentration, total cholesterol (TC), triglycerides (TG), HDLcholesterol (HDL), LDL cholesterol (LDL), and blood sugar (BS) were measured by Kyoto Biken Laboratories Inc. (Kyoto, Japan), Nikken Igaku Co. (Fukui, Japan) and Falco Holdings Co. (Kyoto, Japan). In all of these subjects, 25OHD levels were found to be either deficient (<20 mg/mL) or insufficient (>20 to 29.9 mg/mL).

XGBoost algorithm and logistic regression

In the experiment, we tested the XGBoost algorithm to find features related to serum 25OHD using clinical data, daily intake and %Fat as predictors. A total of 70 elderly with 25OHD that converted into "1" (<20 ng/mL) and "22" (≥20 ng/mL) were randomly divided into a training set (n=35) and a test set (n=35). The training set was used to construct the prediction model based on the XGBoost algorithm and the test set was used to evaluate the prediction effect of the model. The logistic regression model was used to analyze the factors associated with 25OHD. The receiver operating characteristic (ROC) was used to analyze the predictive value of the model and the area under the ROC curve (AUC) was calculated.

Results

Study Subjects

Obesity was defined as a BMI of ≥25.0 kg/m^2. The prevalence of obesity determined by BMI was 38.5% in males and 35.8% in females. This showed a tendency to obesity in comparison with the standard for all 65 - 74-year-old Japanese (21.5 – 24.9 kg/m^2) [11]. Serum 250HD was classified as normal (>30 ng/ml), insufficient (>20 to 29.9 ng/m), or deficient (<20 ng/m). In the subjects of this study, the level was either deficient or insufficient.

Serum 25OHD model on the XGBoost algorithm

The XGBoost model demonstrated that multiple factors were related to serum 25OHD levels. The critical feature importance results were obtained through calculation (Figure 1). The top 4 essential characteristics for predicting 25OHD levels sufficiency were CH, BS, TG, and TC, respectively, with a precision of 60%. Protein intake is not a good predictor.

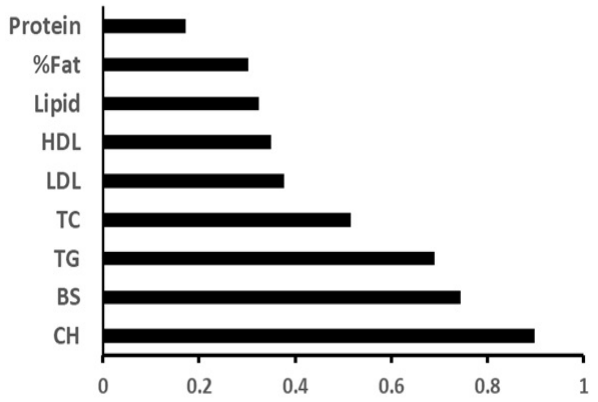

Figure 1: Feature importance of prediction index based on the XGBoost algorithm.

Multivariate logistic regression analysis

The logistic regression model showed positive and negative coefficients (Figure 2). Positive values for the "BS," "CH," "LDL" and "Lipid" features decreased the risk of Vitamin D insufficiency, and negative values for the "TG", "TC" and "%Fat" increased the risk of Vitamin D insufficiency. The AUC was 0.76, as shown in Figure 3.

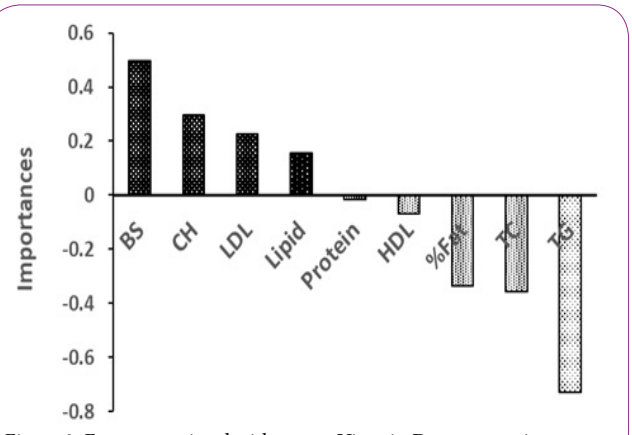

Figure 2: Factor associated with serum Vitamin D concentration.

Discussion

In this study, a prediction model was established based on XGBoost and logistic regression to predict vitamin D levels.

The results showed that a moderate intake of food (CH and lipid) containing vitamin D and TC are necessary. Vitamin D is a liposoluble pleiotropic hormone and vitamin D3 is synthesized in the skin from TC precursors upon exposure to solar UVB radiation [2]. Our findings support the need for TC.

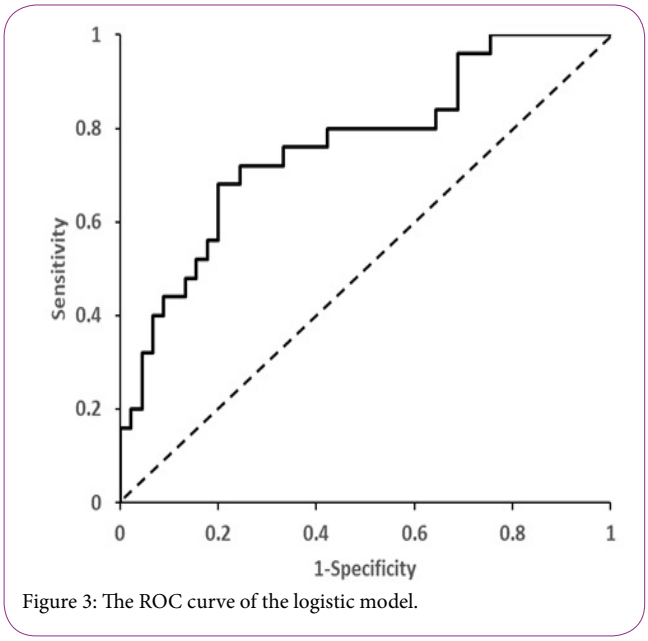

Figure 3: The ROC curve of the logistic model.

The prediction model constructed in this paper also found negative values for "TG," "TC" and "%Fat". It is thought that increased body fat from excessive lipid intake (TG, TC) may be related to adsorption of 25OHD. These results are effective for the maintenance of 25OHD levels by decreasing excessive accumulation of body fats through overeating.

In addition, these facts indicate the need to adopt active sunbathing in day-care services in winter or at institutions located at higher latitudes.

Although we only used a small number of cases, the model with the algorithm were satisfactory for the prediction task. Thus, our findings should be able to serve as a foundation for larger prospective studies.

Conclusion

Lipid intake and metabolism conditions become good predictors for 25OHD levels. By considering lipid intake and deposition, this model is good predictive tool for maintaining serum 25OHD levels. This model could serve as a tool to aid nurses in clinical decision making processes.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

Dr. Hasegawa was responsible for the study conception, design, interpretation of data, and drafting of the manuscript.

Dr. Tsuchiya was responsible for the machine learning approach and checking the manuscript.

Dr. Kobayashi was responsible for checking the manuscript.

Dr. Tsubouchi was responsible for data acquisition and checking the manuscript.

Dr. Yamada was responsible for data acquisition and checking the manuscript.

Dr. Shimizu was responsible for data acquisition and checking the manuscript.

Ms. Kato was responsible for data acquisition and checking the manuscript

Ms. Mochizuki was responsible for data acquisition and checking the manuscript.

Dr. Sethabouppha, Dr. Suwankruhasn, and Dr. Suvanayos were responsible for the research plan.

Funding

This work was supported by KAKENHI (grant number 22K11220).

Refrences

1. Christakos S, Dhawan P, Verstuyf A, Verlinden L, et al. (2016) Vitamin D: metabolism, molecular mechanism of action, and pleiotropic effects. Physiol Rev 96: 365-408.
2. Sarfraz Z (2015) Power of Vitamin D (3rd edition): A Vitamin D Book that Contains the Most Scientific, Useful and Practical Information About Vitamin D - Hormone D. CreateSpace Independent Publishing Platform, USA.
3. Kamwa V, Hassan-Smith ZK (2019) The inter-relationship between marginal vitamin D deficiency and muscle. Curr Opin Endocrinol Diabetes Obes 26: 322-328.
4. Hasegawa N, Mochizuki M, Kato M, Shimizu N, Yamada T (2018) Vitamin D3 supplementation improved physical performance in healthy older adults in Japan: A pilot study. Health 10: 1200-1209.
5. Li T, Fong S, Wong KKL, Wu Y, Yang X-S et al. (2020) Fusing wearable and remote sensing data streams by fast incremental learning with swarm decision table for human activity recognition. Inf Fusion: 41–64.
6. Zhang M., Zhu L, Sun Y, Niu D, Liu J (2022) Computed tomography of ground glass nodule image based on fuzzy C-means clustering algorithm to predict invasion of pulmonary adenocarcinoma. J Rad Res Appl Sci 15:152–158.
7. Hou N, Li M, He L, Xie B, Lin Wanget al. (2020) Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGBoost. J Transl Med 18:462-474.
8. Paleczek A, Grochala D, Rydosz A (2021) Artificial breath classification using XGBoost algorithm for diabetes detection. Sensors 21:4187-4202.
9. Huang Y, Chen C, Miao Y (2022) Prediction model of bone marrow infiltration in patients with malignant lymphoma based on logistic regression and XGBoost algorithm. Comput Math Methods Med 2022: 9620780.
10. Yan Z, Wang J, Dong Q, Zhu L, Lin W, Jiang X (2022) XGBoost algorithm and logistic regression to predict the postoperative 5-year outcome in patients with glioma. Ann Transl Med 10: 860-868.
11. Ministry of Health, Labor and Welfare, Japan (2019) The National Health and Nutrition Survey in Japan, 2019.