International Journal of
Nursing & Clinical Practices

Short Communication Open Access

Prediction of Serum Vitamin D Status and Related Factors in Female Older Adults in Thailand Using Based on a Machine Learning Study

Noboru Hasegawa[1*], Nobuko Shimizu[2], Takako Yamada[3], Tomohiro Umemura[4], Minatsu Kobayashi[5], Mayumi Kato[6], Kyosuke Yorozuya[7], Antonio Oliveira Nzinga Rene[8], Piyathorn Rengrew[9], Pattaranai Chaiprom[9], Pattana Nakatong[9] and Kamolthip Thipsungwan[9]

[1]*Doshisha Women's College, Kodo, Kyotanabe, Kyoto 610-0395, Japan*
[2]*Toyama Prefectural University, 2-2-78 Nishinagae, Toyama 930-0975, Japan*
[3]*Bukkyo University, 7 Higashitoganoo-cho, Nishinokyo, Nakagyo-ku, Kyoto 604-8418, Japan*
[4]*Aichi Medical University, 1-1 Yazakokarimata, Nagakue, Aichi, 480-1195, Japan*
[5]*Otsuma Women's University, 12 Sanban-cho, Chiyoda-ku, Tokyo 102-8357, Japan*
[6]*Aichi Medical College for Physical and Occupational Therapy, 519 Ichiba, Kiyosu-City, Aichi 452-0931, Japan*
[7]*Seijoh University, 2-172 Fukinodai, Tokai 476-8588, Aichi, Japan*
[8]*Toyama Prefectural University, 5180 Kurokawa, Imizu City, Toyama 939-0398, Japan*
[9]*Lampang Rajabhat University, 119 Lampang-Mae Ta Road, Chomphu Subdistrict, Mueang District, Lampang 52100, Thailand*

Abstract

Background: Machine learning has unique advantages in dealing with complex interactions and nonlinear relationships. In this study, a random forest algorithm and logistic regression a prediction model were established to develop a prediction model for serum 25OHD levels in elderly people in Thailand.

Methods: A total of 60 healthy adults age ≥60 years 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 carbohydrates (CH), lipids and proteins were assessed. The 25OHD levels in these subjects were classified as sufficient (>=30), insufficient (>=20 to 29.9) or deficient (<20). We tested the random forest 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 serum 25OHD levels. A cumulative gains chart (CGC) was used to analyze the predictive value of the model and the area under the ROC curve (AUC) was calculated.

Results: The random forest algorithm's top 4 essential characteristics were "TC," "BS," "HDL," and "Fat," respectively. Protein intake is not a good predictor. The logistic regression model showed positive and negative coefficients. Positive values for the "TC," "%Fat," and "LDL" decrease the risk of 25OHD insufficiency, and negative values for the "BS", "HDL," "CH" and "TG" increase the risk of 25OHD insufficiency. The AUC of the random forest algorithm and logistic regression model were 0.93 and 0.94, respectively.

Conclusion: Our results indicate that by considering animal fat intake and UV exposure, the 2 algorithms are a good predictive tool for maintaining serum 25OHD levels in Thailand.

Publication History:

Received: March 17, 2025
Accepted: April 16, 2025
Published: April 18, 2025

Keywords:

Random forest Algorithm, Logistic regression, Machine learning, Serum 25OHD, Body fat percentage, Daily intake, Clinical data, Female elderly in Thailand

Introduction

The role of vitamin D in bone health is well known. In addition, vitamin D is an important nutrient in the health of humans and vitamin D receptors are found in several tissues in the human body [1]. Vitamin D is a secosteroid associated with peripheral calcium homeostasis and nervous system function, cancer, cardiovascular problems, autoimmune diseases, respiratory infections and allergies [2, 3]. Vitamin D is biologically inert and requires activation through two hydroxylation processes involving 25-hydrooxylase (CYP2R1) and 1α-hydroxylase (CYP27B1), which are located in the liver and the kidney, respectively [3]. Through blood circulation, vitamin D reaches various organs [2]. 25OHD levels have been associated with skeletal muscle strength and physical performance [4]. In a previous study, we also showed that 25OHD supplementation was associated with improved serum 25OHD levels and possibly improved 4-m gait speed in Japanese elderly [5]. These results suggest that vitamin D is an important factor in the health of elderly people.

The major source of vitamin D is "sunshine," with vitamin D being produced in the skin from 7-dehydrocholesterol from cholesterol with exposure to ultraviolet rays. Another source of vitamin D is diet [3]. High levels of vitamin D are expected in a tropical country such as Thailand, where there is abundant sunshine throughout the year.

Machine learning has unique advantages in dealing with complex interactions and nonlinear relationships [6, 7]. In recent years, the use of machine learning algorithms in medical treatment has also gradually increased [8-11]. We developed a prediction model for serum vitamin D level and identified related factors in Japanese female older adults using XGBoost and Logistic regression [12]. However, the features that influence blood 25OHD levels are not well understood in elderly subjects in Thailand

Therefore, in this study, a random forest algorithm and logistic regression prediction model were established to predict the serum 25OHD levels in elderly subjects in Thailand.

***Corresponding Author:** Prof. Noboru Hasegawa, Doshisha Women's College, Kodo, Kyotanabe, Kyoto 610-0395, Japan

Citation: Hasegawa N, Shimizu N, Yamada T, Umemura T, Kobayashi M, et al. (2025) Prediction of Serum Vitamin D Status and Related Factors in Female Older Adults in Thailand Using Based on a Machine Learning Study. Int J Nurs Clin Pract 12: 414. doi: https://doi.org/10.15344/2394-4978/2025/414

Materials & Methods

Subjects and Setting

Prior to conducting this study, approval was obtained from the ethics committee of the Toyama Prefectural University Ethics Review Board (R6-19) in Japan and Human Research Ethics Committee of Boromarajonani College of Nursing in Lampang, Certificate number: E2567-055. A total of 60 healthy adults (age: 64.9 ± 3.7) were included in the study.Subjects were enrolled from September 2024 after obtaining informed consent in Lampang.

Body fat percentage

Body fat percentage (%Fat) was measured using ACCUNIQ (Wellcle, Thai).

Daily intake

The Short Self-Administered Food Frequency Questionnaire was used to assess the daily intake of carbohydrate (CH), lipid and protein (Education Software Co., Ltd., Tokyo). The calculated nutrient intakes were based on results validated by Dr, Kobayashi, who studies Thai nutritional intake.

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 MP LAB Medical Technical Clinic (Lampang, Thailand). The method used is the same as that used in Japan. In all of these subjects, 25OHD levels were found to be either deficient (<20 ng/mL), insufficient (>20 to 29.9 ng/mL) or sufficient (>30ng/mL).

Random forest algorithm and logistic regression

In the experiment, we used a random forest algorithm and logistic analysis model to find features related to serum 25OHD using clinical data, daily intake and %Fat as predictors. There is no missing data. A total of 60 elderly people were randomly divided into a training set (n=42) and a test set (n=18). Though there were no samples available for validation. This is a limitation of the study. The training set was used to construct the prediction model based on the random forest algorithm or logistic analysis model and the test set was used to evaluate the prediction effect of the models. The models were used to analyze the factors associated with 25OHD. A cumulative gains chart (CGC) was used to analyze the predictive value of both models and the area under the CGC curve (AUC) was calculated.

Results

Study subjects

Obesity was defined as a BMI of ≥25.0 kg/m^2. The prevalence of obesity judged from BMI was 22.7 ± 3.3 (SD) %. This showed a tendency to be in the normal range in comparison with the standard for all 65 - 74-year-old Japanese (21.5 – 24.9 kg/m^2) [13]. Serum 25OHD was classified as sufficient (>30 ng/ml), insufficient (>20 to 29.9 ng/ml), or deficient (<20 ng/m). The prevalence of sufficiency was higher in subjects in Thailand (Table 1).

25(OH)D	%	n
Sufficiency	18.3	11
Insufficiency	60.0	36
Deficiency	21.7	13

Table 1: Proportion of elderly with serum 25OHD sufficiency, insufficiency and deficiency in Thailand.

Serum 25OHD model with the random forest algorithm

The random forest 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 TC, BS, HDL, and Fat, respectively, with a precision of 60%. The CGC curve is shown in Figure 2, with an AUC of 0.93. The CGC curve shows that if we take the 20% of subjects with the highest probability of positive response using this model, we will get 95% of all the possible positive subjects (Figure 2).

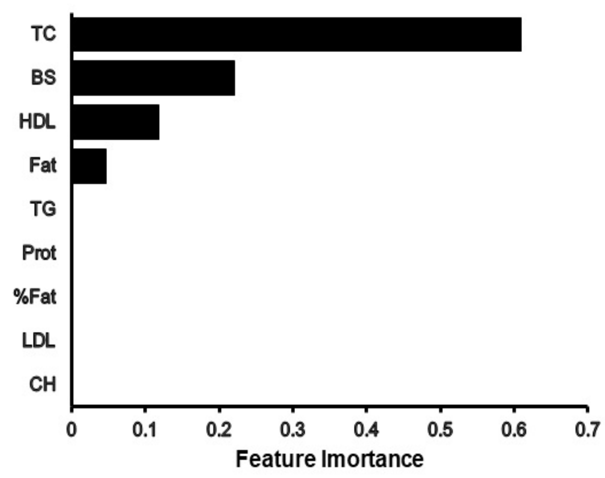

Figure 1: Feature importance of prediction index based on the random forest algorithm.

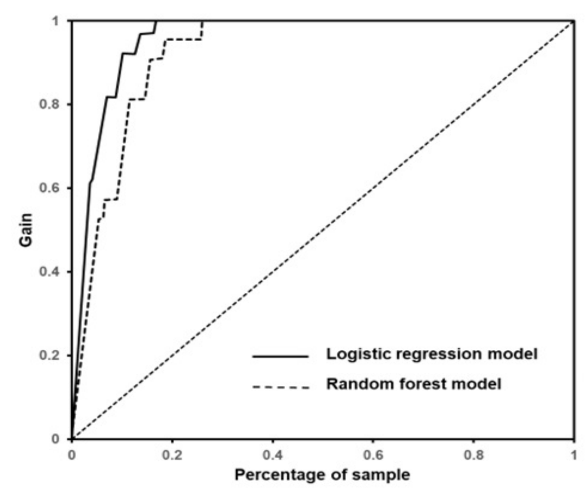

Figure 2: The CGC curves of the random forest and logistic regression model.

Citation: Hasegawa N, Shimizu N, Yamada T, Umemura T, Kobayashi M, et al. (2025) Prediction of Serum Vitamin D Status and Related Factors in Female Older Adults in Thailand Using Based on a Machine Learning Study. Int J Nurs Clin Pract 12: 414. doi: https://doi.org/10.15344/2394-4978/2025/414

Logistic regression analysis

The logistic regression model showed positive and negative coefficients (Figure 3). Positive values for the "TC," "%Fat" and "LDL" features decreased the risk of Vitamin D insufficiency, and negative values for the "BS," "HDL" and "%TG" increased the risk of Vitamin D insufficiency. The CGC curve is shown in Figure 3 with an AUC of 0.94. The CGC curve shows that if we will take the 10% of subjects with the highest probability of positive response using this model, we will get 95% of all the possible positive subjects.

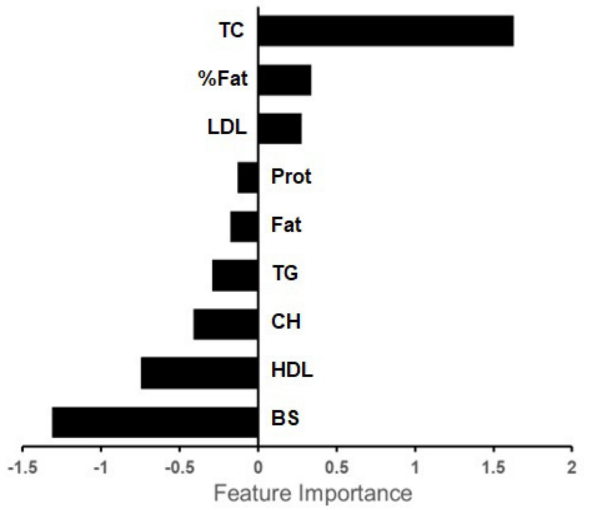

Figure 3: Factor associated with serum Vitamin D concentration.

Discussion

In this study, a prediction model was established based on a random forest and logistic regression to predict vitamin D levels. These models achieved good performance and may help to improve low serum 25OHD.

The results showed that a moderate intake of food containing vitamin D and TC and a certain amount of UV rays are necessary to maintain vitamin D levels. 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]. In Thailand, abundant sunshine can cause high levels of vitamin D. Our findings support the need for sufficient TC in order to maintain 25OHD levels in Thailand. It is thought that increased body fat from excessive lipid intake (TG, TC) may be related to adsorption of 25OHD. These results suggest the benefit of excessive accumulation of body fat for the maintenance of 25OHD levels.

The prediction model constructed in Japanese in a previous paper [12] showed positive values for "BS" and "CH" in Japanese with lower vitamin D levels than in Thailand. Vitamin D deficiency may be one of the factors accelerating the development of insulin resistance [14]. Japan is among the top 11 countries with the highest number of adults aged 18 and over living with diabetes [15]. These results suggest that it is desirable for Japanese people to maintain sufficient vitamin D to prevent diabetes.

These facts also 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 this algorithm was 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 and UV exposure conditions are good predictors for 25OHD levels. By considering lipid intake and deposition, these models are good predictive tools for maintaining serum 25OHD levels. These models 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. Shimizu was responsible for data acquisition and checking the manuscript.
Dr. Kobayashi was responsible for data acquisition and checking the manuscript.
Dr. Yamada was responsible for checking the manuscript.
Dr. Umemura was responsible for checking the manuscript.
Ms. Kato was responsible for checking the manuscript.
Dr. Yorozuya was responsible for checking the manuscript.
Dr. Antonio was responsible for checking the manuscript.
Dr. Piyathorn was responsible for the research plan.
Dr. Pattaranai was responsible for the research plan.
Dr. Pattana was responsible for the research plan.
Dr. Kamolthip was responsible for the research plan.

Funding

This work was supported by KAKENHI (grant number 22K11220 and 21K10969).

References

1. Ginde AA, Sullivan AF, Mansbach JM, Camargo CA Jr. (2010) Vitamin D insufficiency in pregnant women of childbearing age in the United States. Am J Obstet Gynecol 202: 436e1-436e8.
2. 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.
3. 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.
4. Kamwa V, Hassan-Smith ZK (2019) The inter-relationship between marginal vitamin D deficiency and muscle. Curr Opin Endocrinol Diabetes Obes 26: 322-328.
5. 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.
6. 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.
7. 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.
8. Hou N, Li M, He L, Bing Xie1, Lin Wang et 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.

9. Paleczek A, Grochala D, Rydosz A (2021) Artificial breath classification using XGBoost algorithm for diabetes detection. Sensors 21: 4187-4202.
10. 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:
11. 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
12. Hasegawa N, Tsuchiya S, Kobayashi M, Shimizu N, Tsucouchi 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: 373-375.
13. Ministry of Health, Labor and Welfare, Japan (2021) The National Health and Nutrition Survey in Japan, 2019.
14. Contrera-Boliva V, Garcia-Fontana B, Garcia-Fontana C, Munoz-Torres M. (2021) Mechanisms Involved in the Relationship between Vitamin D and Insulin Resistance: Impact on Clinical Practice. Nutrients 13: 3491-3516
15. NCD Risk Factor Collaboration (NCD-RisC) Worldwide trends in diabetes prevalence and treatment from 1990 to 2022: a pooled analysis of 1108 population-representative studies with 141 million participants. Lancet 404: 2077-2093.