

The BiomolBiomed publishes an “Advanced Online” manuscript format as a free service to authors in order to expedite the dissemination of scientific findings to the research community as soon as possible after acceptance following peer review and corresponding modification (where appropriate). An “Advanced Online” manuscript is published online prior to copyediting, formatting for publication and author proofreading, but is nonetheless fully citable through its Digital Object Identifier (doi®). Nevertheless, this “Advanced Online” version is NOT the final version of the manuscript. When the final version of this paper is published within a definitive issue of the journal with copyediting, full pagination, etc., the final version will be accessible through the same doi and this “Advanced Online” version of the paper will disappear.

**RESEARCH
ARTICLE****TRANSLATIONAL AND
CLINICAL RESEARCH**

He et al.: Predicting FNBMD in elderly men

Development and evaluation of interpretable machine learning regressors for predicting femoral neck bone mineral density in elderly men using NHANES data

Wen He^{1#}, Song Chen^{2#*}, Xianghong Fu^{1#}, Licong Xu^{1#}, Jun Xie^{2*}, and Jinxing Wan^{3*}

¹Reproductive Medicine Center, Quzhou People’s Hospital, The Quzhou Affiliated Hospital of Wenzhou Medical University, Quzhou, China.

²Department of Orthopedics, Quzhou People’s Hospital, The Quzhou Affiliated Hospital of Wenzhou Medical University, Quzhou, China.

³Department of Endocrinology, Quzhou People’s Hospital, The Quzhou Affiliated Hospital of Wenzhou Medical University, Quzhou, China.

#These authors have contributed equally to this work and share first authorship.

*Corresponding authors: Song Chen; Email: chensong89@126.com; Jun Xie; Email: xiejun572572@wmu.edu.cn; Jinxing Wan; Email: Qrynfm661@wmu.edu.cn.

DOI: <https://doi.org/10.17305/bb.2024.10725>

Submitted: 09 May 2024/ **Accepted:** 28 June 2024/ **Published online:** 05 July 2024

Conflicts of interest: Authors declare no conflicts of interest.

Funding: This work was supported by the Natural Science Foundation of Zhejiang Province, China (grant number Q24H040021), the Science and Technology Programme of Quzhou, Zhejiang Province, China (grant number 2023K105, 2023K111), and the Medical and Healthcare Talent Research Initiation Fund Category G (grant numbers KYQD 2022-29, KYQD 2022-30).

License: © The Author(s) (2024). This work is licensed under a Creative Commons Attribution 4.0 International License.

EARLY ACCESS

ABSTRACT

Osteoporotic femoral neck fractures (OFNFs) pose a significant orthopedic challenge in the elderly population, accounting for up to 40% of all osteoporotic fractures and leading to considerable health deterioration and increased mortality. In addressing the critical need for early identification of osteoporosis through routine screening of femoral neck bone mineral density (FNBMD), this study developed a user-friendly prediction model aimed at men aged 50 years and older, a demographic often overlooked in osteoporosis screening. Utilizing data from the National Health and Nutrition Examination Survey (NHANES), the study involved outlier detection and handling, missing value imputation via the K nearest neighbor (KNN) algorithm, and data normalization and encoding. The dataset was split into training and test sets with a 7:3 ratio, followed by feature screening through the least absolute shrinkage and selection operator (LASSO) and the Boruta algorithm. Eight different machine learning algorithms were then employed to construct predictive models, with their performance evaluated through a comprehensive metric suite. The random forest regressor (RFR) emerged as the most effective model, characterized by key predictors such as age, body mass index (BMI), poverty income ratio (PIR), serum calcium, and race, achieving a coefficient of determination (R^2) of 0.218 and maintaining robustness in sensitivity analyses. Notably, excluding race from the model resulted in sustained high performance, underscoring the model's adaptability. Interpretations using Shapley additive explanations (SHAP) highlighted the influence of each feature on FNBMD. These findings indicate that our predictive model

effectively aids in the early detection of osteoporosis, potentially reducing the incidence of OFNFs in this high-risk population.

Keywords: Femoral neck bone mineral density; osteoporotic femoral neck fractures; NHANES; machine learning; random forest regressor.

EARLY ACCESS

INTRODUCTION

Osteoporotic femoral neck fractures (OFNFs) represent a prevalent orthopedic challenge among the elderly, comprising up to 40% of all osteoporotic fractures and significantly impairing health and increasing mortality [1-3]. Experts anticipate a 2.7-fold increase in the incidence of such fractures in Eastern Asia, projecting a rise from 18,388 cases in 2010 to 50,421 by 2035 [4]. Early detection of individuals with osteoporosis and the implementation of anti-osteoporotic treatments are crucial measures for preventing these fractures [5]. Femoral neck bone mineral density (FNBMD) is acknowledged as a valuable predictor of OFNFs and is recommended for diagnosing osteoporosis and assessing low bone mass [6]. Although Dual-energy X-ray absorptiometry (DXA) is the standard method for evaluating FNBMD, its high cost, limited accessibility, and associated radiation exposure restrict its widespread use in community screenings [7]. Consequently, developing a simple, cost-effective, and reliable alternative for routine FNBMD assessment is imperative.

Machine learning (ML), a branch of artificial intelligence, excels at handling extensive heterogeneous datasets and capturing intricate relationships between features [8-10]. It has demonstrated significant potential in predicting healthcare outcomes and complications, aiding clinicians in making informed decisions, and improving patient care [11-13]. This opens innovative avenues for creating accurate and reliable models for the real-time prediction of FNBMD, which is crucial for assessing patients' bone health and facilitating early osteoporosis diagnosis. Currently, ML algorithms have been effectively utilized to screen for osteoporosis with satisfactory outcomes. [14-16]. Nevertheless, most studies have primarily focused on the

models' ability to identify patients with osteoporosis, rather than on the real-time prediction of BMD values. Moreover, while osteoporosis is more prevalent among women aged 50 and above, leading to a concentration of research within this demographic, fewer studies have addressed men of the same age group. There are increasing calls for routine BMD screenings for men aged 50 and older [17, 18]. Addressing this research gap, the application of ML algorithms for real-time FNBMD prediction in this male population holds substantial clinical significance.

The objective of this study is to develop and validate a straightforward, cost-effective, and reliable ML prediction model for FNBMD in elderly men. This involves comparing the predictive performance of various ML algorithms. The hypothesis of this study is that the optimal model can accurately predict FNBMD using only a limited set of readily accessible features. This capability is crucial for the early identification of osteoporosis in older men, enabling the implementation of targeted preventive strategies aimed at reducing the incidence of OFNFs.

MATERIALS AND METHODS

This study adhered to the STROBE reporting guidelines. The overall design is depicted in Figure 1.

Data source

The data for this study was sourced from the continuous National Health and Nutrition Examination Survey (NHANES), conducted by the National Center for Health Statistics of the Centers for Disease Control and Prevention (<https://wwwn.cdc.gov/nchs/nhanes/>). NHANES

data were gathered from a nationally representative sample of the civilian, noninstitutionalized US population, utilizing a multistage probability design. The dataset includes questionnaire responses, laboratory testing results, and physical examination data.

Study population

We incorporated data from the NHANES covering the period from January 2005 to March 2020, excluding the 2011–2012 and 2015–2016 cycles due to the unavailability of FNBMD data. To ensure a comprehensive and representative dataset, while avoiding potential biases and redundancies, we merged data from the specified periods and took careful measures to include each individual only once. The exclusion criteria were as follows: (1) female participants; (2) individuals under the age of 50; (3) participants with hypoglycemia (fasting plasma glucose ≤ 50 mg/dl or 2.8 mmol/L) or diabetes (fasting plasma glucose ≥ 110 mg/dl or 7.0 mmol/L) at baseline; and (4) individuals with conditions including cancer, thyroid disorders, chronic renal failure, inflammatory arthritis, and chronic liver disease, as well as those using medications for dyslipidemia, corticosteroids, sex hormones, and diuretics. Participants with incomplete or invalid FNBMD data were excluded from the final analysis. Additionally, cases with outliers, defined as data points exceeding 3 times the interquartile range (IQR), were removed, as they constituted only 1% of the dataset.

BMD examination

Bone mineral density measurements, expressed in grams per square centimeter (gm/cm^2), were conducted using dual-energy X-ray absorptiometry (DXA) with Hologic QDR 4500A fan-beam densitometers (Hologic Inc., Bedford, MA, USA) [1, 2]. All measurements were

conducted by NHANES radiological technologists, who had undergone extensive training and certification. In the present research, FNBMD data were selected as the outcome because the femoral neck is often proposed as the reference skeletal site for defining osteoporosis in epidemiological research [3].

Features

Based on the literature [4-6], the following features were included: age, body mass index (BMI), race, education level, marital status, drinking status, smoking status, physical activity, family history of osteoporosis (FHOS), poverty income ratio (PIR), systolic blood pressure (SBP), diastolic blood pressure (DBP), high-density lipoprotein (HDL), fasting plasma glucose (FPG), TC, triglycerides, low-density lipoprotein (LDL), serum calcium, serum phosphorus, and serum 25-hydroxyvitamin D3 (25(OH)D3). Race was categorized as “Mexican American”, “non-Hispanic White”, “non-Hispanic Black”, or “other race”; education level was classified as “less than high school”, “high school”, or “college or above”; marital status was divided into “married or living with partner” or “single”. Regarding health-related behaviors, smoking status was defined based on a lifetime history of smoking at least five packs of cigarettes (equivalent to 100 cigarettes), categorized as “yes” for those meeting this criterion and “no” for those who smoked fewer than five packs and were currently nonsmoking. Alcohol consumption was binary, with individuals drinking at least once a month over the past year classified as “yes” and all others as “no”. Family history of osteoporosis was specifically noted if a parent had been diagnosed with the condition. Physical activity was quantified in metabolic

equivalent hours per week (MET-min/week) of moderate-to-vigorous physical activity, with categories set at “less active (< 600 MET min/week)” and “active (\geq 600 MET min/week)” [7].

Missing data

The details of the missing data are presented in Table S1. To enhance statistical power and reduce bias, the k-nearest neighbor (KNN) imputation [26] with k equal to 10 was employed to address missing values in eligible cases. To assess the imputation effect, we performed separate outlier tests on the imputation data and compared them with the raw data to identify any between-group differences, as detailed in Table S1. Following the acquisition of the qualified interpolated data, we applied Z-score standardization to quantitative features and one-hot encoding to qualitative features. Then, the data was divided into the training set and test set in a ratio of 7:3 for subsequent analysis.

Feature selection

We implemented a stringent feature selection process to pinpoint the most pertinent predictors for constructing the prediction model, using only the training cohort to prevent data leakage. Initially, a pairwise Pearson or Spearman correlation matrix was utilized to evaluate the continuous features for collinearity, setting a correlation threshold of $r > 0.8$. Collinearity, which occurs when two or more predictor variables are highly correlated, can obscure the unique contribution of each variable to the outcome. Consequently, we selected the most readily available variables among the collinear ones for further analysis. Next, we employed a two-step approach using both the Boruta algorithm ⁸ and the least absolute shrinkage and

selection operator (LASSO) [9]. We then took the intersection of the predictors identified by both algorithms to ensure only the most relevant and robust variables were included in the development of our prediction model. This combined methodology aims to enhance the model's accuracy and generalizability while minimizing the risk of overfitting or incorporating irrelevant predictors.

Model development and validation

Common supervised machine learning algorithms for regression encompass linear models such as Linear Regression, foundational tree-based methods like Decision Trees and ensemble-based Random Forests, Support Vector Machines which excel in complex function mapping with kernel function selection, Neural Networks renowned for their adaptability and proficiency in managing noisy data, and Gaussian Process Models, prized for their probabilistic approach and ability to estimate prediction uncertainty [10]. (1) A linear model, specifically Multivariable Linear Regression (MLR), posits a linear relationship between input and output variables. It is straightforward and establishes a baseline for comparison. While predicting phenomena such as FNBMD might surpass the complexity MLR can handle, it remains an excellent benchmark to gauge the degree of enhancement provided by more sophisticated non-linear machine learning algorithms. (2) Tree-based models such as Decision Tree Regression (DTR), Extreme Gradient Boosting Regression (XGBR), Random Forest Regression (RFR), and Extra-Trees Regression (ETR) are highly effective in managing complex medical datasets. Renowned for their interpretability and capability to process both numerical and categorical data, these models excel at identifying non-linear relationships and intricate interactions among

variables, making them particularly valuable in health research [11]. (3) Support Vector Regression (SVR), a variant of the Support Vector Machine, is frequently selected for its effectiveness in high-dimensional spaces. SVR is particularly skilled at navigating the complex patterns prevalent in medical datasets, establishing it as a robust choice for intricate data analysis [12]. (4) Gaussian Process Regression (GPR), a non-parametric method, is highly valued for its ability to provide uncertainty measures alongside predictions. This feature is particularly advantageous for analyzing medical data, which often involves considerable uncertainty [13]. (5) Multi-layer Perceptron Regression (MLPR), a neural network approach, is adept at capturing the intricate and often non-linear patterns found in large health datasets [14]. By incorporating a diverse array of machine learning approaches, our study offers a comprehensive comparison of FNBM prediction across various models.

All models were using “sklearn”, “xgboost”, “numpy”, and “pandas” packages, with grid search utilized to optimize the hyperparameters for each model. This method systematically explores a broad spectrum of hyperparameter values, enhancing the probability of discovering the most effective global solution for all critical parameters. This approach facilitates thorough yet efficient tuning, which is particularly advantageous given the limited size of our dataset. To assess the effectiveness of the predictive machine learning models, we employed five specific metrics for regression issues: R^2 (coefficient of determination), MAE (mean absolute error), MSE (mean squared error), MAPE (mean absolute percentage error), and RMSE (root mean squared error).

Model explainability

Machine learning models often face challenges in terms of explainability, particularly as the complexity and accuracy of the models increase, potentially reducing interpretability [15, 16].

To address this issue, we utilized SHAP values from the game theory-based “shap” package to plot feature importance for global explainability. This approach enhanced our understanding of the decision-making processes within the model that demonstrated the best performance [16].

To ascertain the most critical features for predicting FNBMD, we employed the permutation importance method. This technique evaluates feature importance by assessing the impact of randomly permuting (shuffling) the values of a feature on the model’s predictive performance.

To minimize error and stabilize the results, we conducted 1000 permutations for each feature across all constructed models, thereby generating 1000 importance values per feature. We then calculated the mean of these importance values and ranked the features based on these averages.

Sensitivity analysis

To ensure the robustness of our main findings, we implemented three sensitivity analyses. Firstly, we removed racial features from the optimal model, and then evaluated the new model’s performance in both the training and test sets. This was undertaken to develop a model independent of racial features, thereby expanding its applicability. Secondly, recognizing the documented reliability of the triglyceride blood glucose-body mass index (TyG-BMI) as a predictor of BMD [17, 18], we computed this index from imputed data and integrated it into

the selected algorithm for comparison with our optimal model. Details on the TyG-BMI calculation are provided in the Supplementary Material. Lastly, we conducted restricted cubic spline analyses [19] for TyG-BMI and each quantitatively selected feature post-engineering, setting five knots at the 5th, 35th, 50th, 65th, and 95th percentiles to flexibly model its relationship with FNBMD. This analysis was aimed at verifying the plausibility of the interpretations suggested by the SHAP values.

Sample size calculation

The R package “pmsampsize,” version 1.1.2, was employed to calculate the required minimum sample size for training [20]. We selected 22 candidate predictor parameters to construct a multivariable prediction model for the continuous outcome. We based our calculations on the assumption that an existing prediction model in the same field has an adjusted R-squared of 0.8 [6, 17], and that FNBMD values in present population have a mean of 0.81 and a standard deviation of 0.13. Consequently, the minimum sample size needed for the training cohort was determined to be 256 cases. Additionally, following Richard’s recommendations for external validation of a prognostic model [21], we used a confidence interval width of ≤ 5 for the calibration-in-the-large (considered precise on an outcome scale of 0 to 100) and ≤ 0.3 for the calibration slope to calculate the sample size, which indicated a minimum of 235 cases are required. This analysis confirms that the eligible population is adequate for both model construction and validation.

Ethical statement

According to the Helsinki Declaration of 1975, as revised in 2000, all procedures followed the ethics of the responsible committee on human experimentation. All participants gave written informed consent to participate in this study that was approved by the Institutional Review Board of the National Center for Health Statistics (NHANES). In this study, all methods were carried out according to relevant guidelines and regulations.

Statistical analysis

Continuous data were evaluated for normality using the Shapiro–Wilk test and presented as mean with standard deviation (SD) for normally distributed data, or median with interquartile range (IQR) for non-normally distributed data. The outliers were detected by using boxplots and Grubbs' test. The homogeneity of variance across groups was assessed using the Levene test. For data following a Gaussian distribution, parametric tests such as the unpaired two-tailed Student's t-test or Welch's t-test were used for comparisons between two groups. For non-Gaussian data, the Mann-Whitney U test was employed for two-group comparisons. Categorical data were expressed as counts and percentages, and analyzed using the Chi-squared test or Fisher's exact test, the latter being applied when more than 20% of cells had expected frequencies less than 5. Statistical significance was determined by a two-sided P value of less than 0.05. All statistical analyses were performed using R software, version 4.1.0.

RESULTS

Participant characteristics

A total of 1,182 eligible participants from the NHANES data collected between January 2005 and March 2020 were included in the analysis. The exclusion criteria are detailed in Figure 1. Missing values were found in BMI, smoking status, drinking status, and nearly all laboratory test-related features. Comparative analyses between the raw data and the imputed data showed no statistically significant differences (Table S1). No outliers were detected in either the complete dataset (data with all missing values removed) or the imputed data.

In the imputed dataset, the median age of participants was 61 years (IQR, 55–68), and the median BMI was 33.4 kg/m² (IQR, 29.7–36.5). The majority were non-Hispanic White (42.8%), had at least a bachelor's degree (48.5%), were not single (71.3%), and reported drinking alcohol (75.9%). The median FNBMD of these participants was 0.80 (IQR, 0.71-0.89).

Of all eligible participants, 829 were assigned to the training group and 353 to the test group. were similar across both cohorts, with no significant differences ($P > 0.05$) noted (Table 1). The median TyG-BMI values were 228 (IQR, 200-260) for the training group and 231 (IQR, 206-265) for the test group. The median FNBMD values were consistent across both groups, each registering at 0.80 (IQR, 0.71-0.89).

Feature selection

As depicted in Figure 2, Pearson correlation values for LDL and TC exceeded 0.8, signaling the presence of collinearity. opted to select LDL for the subsequent feature selection phase.

The Boruta (Figure 2B) and LASSO (Figures 2D and E) algorithms identified a total of six features as significant predictors of the outcome. These encompassed age, BMI, non-Hispanic Black, non-Hispanic White, PIR, and serum calcium. The chosen features were then integrated into eight machine learning regressors to cultivate predictive models.

Hyperparameters tuning

Table S2 lists the optimized hyperparameters for each algorithm. Hyperparameter tuning was not conducted on the MLR model, as it does not involve hyperparameters.

Development and validation of prediction models

The identified six predictors and optimized hyperparameters were incorporated into the FNBMD prediction regressors. Within the training cohort, the RFR model demonstrated superior performance, achieving the highest R^2 at 0.712, the lowest MSE at 0.005, and a RMSE of 0.072. Upon removing the racial features from the model, the performance of the adjusted model (referred to as ABPC-RFR, which includes age, BMI, PIR, and serum calcium) improved significantly, with R^2 rising to 0.841 and reductions in MAE, MSE, MAPE, and RMSE reaching the most favorable values of 0.043, 0.003, 0.054, and 0.057, respectively. This model outperformed all others, including the TyG-BMI-RFR model, which solely incorporates TyG-BMI. In the test cohort, the ABPC-RFR model's performance was comparable to that of the original RFR model, which integrates six features, and it surpassed the performance of all other models examined (as detailed in Table 2).

In Figure 3, the predictive behavior of the models RFR, ABPC-RFR, and TyG-BMI-RFR varies with the true value of FNBMD. Specifically, when the true FNBMD value is below 0.8,

the predicted values from these models are consistently higher than the actual values. Conversely, when the true FNBMD value exceeds 0.8, the predictions fall below the true values. This pattern suggests a decline in model performance when predicting extreme values. The scatter density plots (Figure 3A, 3C, and 3E) show that the ABPC-RFR model's fitted line deviates the least from the ideal line, signifying it has the best predictive performance among the models. Similarly, in the scatter plots for the training set (Figures 3B, 3D, and 3F), the ABPC-RFR model again shows the smallest angle of deviation from the perfect line, indicating superior performance in the training phase. In the test set, the fitted lines of the RFR and ABPC-RFR models exhibit similar angles of deviation from the perfect line, suggesting that their performance in the test set is comparable. These observations align with the findings presented in Table 2.

Model explainability

The SHAP summary plot (Figure 4) illustrates the impact of each feature across the RF, ABPC-RFR, and TyG-BMI-RFR models. SHAP values above zero suggest higher FNBMD values, whereas values below zero indicate lower FNBMD values. For instance, a higher BMI (depicted in red) correlates with larger SHAP values, suggesting that individuals with a higher BMI tend to have greater FNBMD values. Conversely, as age increases, the SHAP values typically decrease, indicating an association between advancing age and bone loss. These trends are depicted in Figures 5A and 5B. Additionally, a higher TyG-BMI value generally results in a larger SHAP value, demonstrating a positive relationship between this index and FNBMD (Figure 4C).

Permutation feature importance analysis identified the pivotal features for predicting FNBMD. To evaluate the relative importance of these features across all eight models, a ranking score system was employed. The most critical feature in each model was assigned a ranking of 1, with rankings progressively decreasing down to 6 for the least significant variable. As detailed in Table 3, BMI emerged as the most influential predictor, achieving a mean ranking of 1.0 ± 0.0 and a median ranking of 1.0 (IQR, 1.0-1.0). It was followed by non-Hispanic Black, age, PIR, serum calcium, and non-Hispanic White in terms of their impact on FNBMD prediction.

Sensitivity analysis

In the RCS analysis depicted in Figure 5, age demonstrates a negative linear correlation with FNBMD, as evidenced by a non-significance for non-linearity ($P = 0.879$). Serum calcium displays a positive linear relationship with FNBMD (P for non-linear = 0.239). BMI and TyG-BMI both exhibit positive non-linear associations with FNBMD, with P values for non-linearity being less than 0.001 and 0.030, respectively. PIR also shows a distinct non-linear association with FNBMD (P for non-linear < 0.001). These findings align with the interpretations provided by the SHAP values, confirming the consistency of the results across different analytical approaches.

DISCUSSION

Osteoporotic fractures of the proximal femur significantly impact the health and increase the mortality rate among the elderly. Early detection of osteoporosis and the initiation of anti-osteoporotic treatments are critical in preventing such fractures. The FNBMD is utilized both for diagnosing osteoporosis and as a key predictor of OFNFs. However, the methods currently

available for assessing FNBMD are not suitable for routine osteoporosis screening. Therefore, it is essential to develop a straightforward, cost-effective, and reliable alternative method for assessment. In this study, we created and validated a machine learning-based prediction model for FNBMD specifically for men aged 50 years and older. Through meticulous feature selection, six key predictors were identified: age, BMI, non-Hispanic Black, non-Hispanic White, PIR, and serum calcium. Among the eight models evaluated, the RFR model showed the best performance across all metrics. When race-related variables were excluded, the adjusted RFR model (ABPC-RFR) incorporating the remaining predictors performed comparably to the full RFR model and significantly outperformed the TyG-BMI-RFR model. The simplified model required only three demographic factors (age, BMI, and PIR) and one laboratory test variable (serum calcium) for making predictions. This simplicity enhances its usability for routine osteoporosis checks among older men of diverse races, aligning well with the study's hypothesis.

Our importance analysis revealed that BMI was the most valuable predictor of FNBMD in this study population (Table 3). Indeed, BMI is not only a simple and widely used health indicator but also a significant predictor of bone tissue structure, closely associated with BMD [22-24]. Compared to the BMD of the lumbar vertebral body, FNBMD is more stable and less influenced by factors such as degeneration, osteophytes, and sclerosis [25]. Several studies have consistently shown a positive correlation between BMI and the absolute values of FNBMD, exhibiting a synergistic increasing trend [22-26]. In research involving 900 elderly individuals, Asuman et al. [27] confirmed that femoral BMDs increased with rising BMI levels,

noting statistically significant differences in femoral BMD among men across different BMI categories, with the highest levels observed in the obese group (BMI of 30 kg/m² and above) compared to those with ideal body weight. Similarly, Kirchengast et al. [28] found that BMD increased with both weight and BMI in both sexes, with higher BMD values in overweight individuals compared to those of ideal body weight, aligning with the findings of our study (Figures 5A, 5B, and 6B). Through combining model interpretative analysis with Restricted Cubic Spline (RCS) analysis, we observed that changes in FNBMD values with increasing BMI were more pronounced when BMI was below 30 and diminished when BMI exceeded 30. This pattern may be attributed to both mechanical effects—where heavier loads on the skeleton induce bone-specific deformations that stimulate osteoblast activity, enhancing the synthesis and expression of osteoblast-related genes, thereby increasing bone density, and enabling the skeleton to adapt to applied stress [29, 30]—and hormonal effects due to increased estrogen production in adipose tissue. Weight gain and enhanced adipose tissue may promote the conversion of androgens to estrogens, improving bone mass in both men and women while maintaining healthy levels of insulin and regulatory factors like insulin-like growth factor-1, leptin, and lipocalin [25, 31]. These observations suggest potential variations in FNBMD between non-obese and obese populations, possibly necessitating distinct diagnostic criteria for osteoporosis in each group. Additionally, to address the modifiable factor of BMI, we recommend promoting exercise, particularly strength training, to increase lean body mass and stimulate bone remodeling to better accommodate loading. However, it is crucial to note that while a higher BMI can be protective against osteoporosis, maintaining a BMI around or

slightly below 30 is advisable to prevent the increased risks of falls, degenerative changes, and systemic diseases associated with obesity in the elderly [25, 27].

Age is another important predictor of FNBMD. In this research, we used SHAP values to understand the machine learning models (RFR, ABPC-RFR) along with RCS analysis to show how age and FNBMD are related. The results showed that age is inversely related to FNBMD, i.e., BMD decreases significantly with age. This finding is consistent with the existing literature [24, 25]. In a study of adult men in Kosovo, Rexhep et al. [24] identified a negative relationship between age and FNBMD, attributing this natural bone mass loss to the aging process. This insight is crucial for guiding local BMD assessments and initiating early preventive strategies against osteoporosis and related fractures in older men. Similarly, Jiang et al. [25] studied the FNBMD of 358 Chinese males aged 50 and older, arriving at the same conclusion. They interpreted this trend as a decline in bone mass beginning around the age of 50, due to osteoblast dysfunction and an increase in osteoclast resorption. The application of machine learning in this research offers a detailed understanding of how age interacts with other factors to influence bone health. Importantly, incorporating age as a predictor in our model underscores the necessity of early intervention and tailored preventive measures for the elderly. The data-driven insights provided by this study showcase the potential of machine learning to enhance predictive models for FNBMD, enabling healthcare providers to more effectively identify and manage individuals at high risk for osteoporosis.

The association between the PIR and FNBMD observed in our study adds a critical socioeconomic dimension to the understanding of bone health. Our analysis revealed that a

higher PIR (≥ 4 in the RCS analysis), indicative of greater socioeconomic status, correlates with improved FNBMD. This aligns with research by Du et al. [32], who reported similar findings and suggested that individuals with higher socioeconomic status have better access to healthcare resources, nutrition, and lifestyle choices conducive to bone health. A meta-analysis encompassing eight epidemiological studies demonstrated that most population-based research supports the observation that individuals with higher income levels are more likely to exhibit higher BMD [33]. This finding was further validated in another cross-sectional study involving 11,075 representative participants from the United States [34]. The positive impact of socioeconomic status on FNBMD emphasizes the potential barriers faced by lower-income populations in maintaining bone health, possibly due to limited access to nutritious food, healthcare services, and opportunities for physical activity [34]. Addressing these disparities could significantly improve bone health outcomes across different population segments. Therefore, our findings advocate for targeted public health interventions and policies that enhance access to bone health resources in economically disadvantaged communities, potentially reducing the prevalence of osteoporosis-related complications.

Serum calcium, a critical element in bone metabolism, was positively correlated with FNBMD in our study, supporting the hypothesis that adequate calcium levels are essential for optimal bone density. This observation is in concordance with the work of Pan et al. [35], who noted that calcium plays a pivotal role not only in bone formation but also in maintaining the structural integrity of the bone matrix. The implications of these findings suggest that monitoring and managing serum calcium levels could be a key strategy for preventing bone

density deterioration, especially in populations at risk for osteoporosis [36]. Additionally, our models reinforce the importance of integrating nutritional and metabolic factors into comprehensive assessments of bone health, advocating for a holistic approach to osteoporosis prevention and treatment strategies that encompass dietary calcium intake and its metabolic management.

The TyG-BMI, a novel metabolic marker explored in our study, demonstrated a significant predictive value for FNBMD. This finding underscores the intertwined roles of metabolic health and bone density. Notably, our analysis indicates that higher TyG-BMI values correlate with increased FNBMD, suggesting that metabolic efficiency and body composition collectively influence bone health. This relationship mirrors the results presented by Zhan et al. [37], who found that metabolic markers like the TyG index provide insight into the risk of metabolic bone diseases beyond traditional lipid and glucose measurements. The association between TyG-BMI and FNBMD enhances our understanding of how composite indices, which encapsulate multiple metabolic risks, can serve as effective tools for assessing bone health. Emphasizing the TyG-BMI in clinical evaluations could offer a more comprehensive assessment strategy, aiding in the early identification of individuals at risk for osteoporosis, thus facilitating timely intervention strategies.

Leveraging demographic (age, BMI, and PIR), laboratory test (serum calcium) characteristics, and the novel metabolic index (TyG-BMI) associated with FNBMD, we selected the RFR algorithm to construct three models: the RFR, ABPC-RFR, and TyG-BMI-RFR. Among these, the ABPC-RFR model emerged as the simplest and most practical. The computation of TyG-

BMI is relatively cumbersome, and the predictive efficacy based on this index is significantly lower compared to the ABPC-RFR model. Furthermore, the fact that race-related characteristics do not influence the ABPC-RFR model's performance suggests that race may not be a prominent factor in our specific dataset. These findings indicate that the ABPC-RFR model can effectively predict FNBMD in men over 50 years of age and holds potential for routine community screening. This facilitates the early detection of osteoporosis and the initiation of anti-osteoporosis treatments to prevent OFNFs. It is noteworthy that the present study did not encompass races from Asia and other regions. This limitation could influence the generalizability of our findings. Future analyses should aim to include more diverse racial groups to better understand the impact of race on FNBMD predictions and ensure the model's applicability across different populations.

Our study has several strengths. First, we utilized the large, nationally representative NHANES dataset, which provides a diverse and statistically significant sample, enhancing the generalizability of our findings. Meticulous data processing techniques were also employed, including the handling of outliers and the KNN estimation of missing data, ensuring data integrity. Further, our study featured rigorous feature selection and a comprehensive model development and validation process. Multiple machine-learning algorithms were utilized to pinpoint the best-performing model, optimizing predictive accuracy. Moreover, we incorporated various evaluation metrics and model interpretability techniques, such as SHAP values, to ensure transparency and facilitate the interpretation of results. Lastly, several

sensitivity analyses were conducted to test the robustness of our findings, ensuring the reliability of our conclusions in different scenarios and settings.

The present study has several limitations that warrant consideration. First, the exclusion of female participants and certain age groups limits the generalizability of our findings to the entire population. Women, particularly post-menopausal women, are at a high risk for osteoporosis, and the dynamics of bone mineral density may differ significantly between genders. Our study also excluded patients with hypoglycemia, diabetes, and those on medications for chronic diseases. These exclusions were made to control for potential confounding factors that could introduce bias into the model. However, we recognize that these patients represent a significant subset of the population, and their exclusion may limit the applicability of our findings to broader populations. Future research should aim to validate our models in more diverse populations that include these important subsets to ensure the generalizability and robustness of the predictive models and detect changes in the ranking of feature importance. Additionally, our reliance on cross-sectional data from the NHANES prevents us from establishing causality between the predictors and FNBMD; longitudinal studies would be required to confirm the directionality of these associations. Another limitation is the potential for residual confounding due to variables not included or not available in the NHANES dataset. Lastly, while machine learning models offer potent predictive capabilities, generalizing these models to different healthcare settings can be challenging. Future research should focus on further validating the model across diverse populations to ensure its broader applicability and effectiveness in varying healthcare contexts.

CONCLUSION

In conclusion, our study demonstrates that FNBMD can be easily and accurately assessed by integrating four readily accessible features (age, BMI, PIR, and serum calcium) through advanced machine learning techniques and outperforms the novel metabolic index (TyG-BMI). Our model interpretation results provide a deeper and more nuanced understanding of femoral neck bone health risks. These insights pave the way for more targeted and effective interventions to prevent and manage OFNFs in the development of femoral neck osteoporosis, especially in populations traditionally underrepresented in bone health studies, such as older men. Future research should aim to validate these predictive models in a broader range of populations to enhance their applicability and effectiveness. By continually refining these models and addressing the limitations outlined, we can better meet the healthcare needs of diverse populations and improve outcomes in bone health management.

ACKNOWLEDGMENTS

We acknowledge the data from the National Health and Nutrition Examination Survey (NHANES).

Data availability

The datasets analyzed during the current study are available on the NHANES website:

<https://www.cdc.gov/nchs/index.htm>.

REFERENCES

1. Hinton BJ, Fan B, Ng BK, Shepherd JA. Dual energy X-ray absorptiometry body composition reference values of limbs and trunk from NHANES 1999-2004 with additional visualization methods. *PLoS One*. 2017;12(3):e0174180. doi:10.1371/journal.pone.0174180
2. Kelly TL, Wilson KE, Heymsfield SB. Dual energy X-Ray absorptiometry body composition reference values from NHANES. *PLoS One*. Sep 15 2009;4(9):e7038. doi:10.1371/journal.pone.0007038
3. Looker AC, Melton LJ, 3rd, Harris TB, Borrud LG, Shepherd JA. Prevalence and trends in low femur bone density among older US adults: NHANES 2005-2006 compared with NHANES III. *J Bone Miner Res*. Jan 2010;25(1):64-71. doi:10.1359/jbmr.090706
4. Zhuo M, Chen Z, Zhong ML, et al. Association of insulin resistance with bone mineral density in a nationwide health check-up population in China. *Bone*. May 2023;170:116703. doi:10.1016/j.bone.2023.116703
5. Caeran G, de Almeida LL, Ilha T, et al. Insulin Resistance and Its Association With Osteoporosis in People Living With HIV. *J Endocr Soc*. Oct 26 2022;6(12):bvac148. doi:10.1210/jendso/bvac148

6. Yoon JH, Hong AR, Choi W, Park JY, Kim HK, Kang HC. Association of Triglyceride-Glucose Index with Bone Mineral Density in Non-diabetic Koreans: KNHANES 2008-2011. *Calcif Tissue Int.* Feb 2021;108(2):176-187. doi:10.1007/s00223-020-00761-9
7. Janssen I, Carson V, Lee IM, Katzmarzyk PT, Blair SN. Years of life gained due to leisure-time physical activity in the U.S. *Am J Prev Med.* Jan 2013;44(1):23-9. doi:10.1016/j.amepre.2012.09.056
8. Speiser JL, Miller ME, Tooze J, Ip E. A Comparison of Random Forest Variable Selection Methods for Classification Prediction Modeling. *Expert Syst Appl.* Nov 15 2019;134:93-101. doi:10.1016/j.eswa.2019.05.028
9. Vasquez MM, Hu C, Roe DJ, Chen Z, Halonen M, Guerra S. Least absolute shrinkage and selection operator type methods for the identification of serum biomarkers of overweight and obesity: simulation and application. *BMC Med Res Methodol.* Nov 14 2016;16(1):154. doi:10.1186/s12874-016-0254-8
10. Yousefi E, Pronzato L, Hainy M, Muller WG, Wynn HP. Discrimination between Gaussian process models: active learning and static constructions. *Stat Pap (Berl).* 2023;64(4):1275-1304. doi:10.1007/s00362-023-01436-x

11. Hu L, Li L. Using Tree-Based Machine Learning for Health Studies: Literature Review and Case Series. *Int J Environ Res Public Health*. Dec 1 2022;19(23)doi:10.3390/ijerph192316080
12. Zavaneli N, Lee SH, Guess M, Yeo WH. Continuous real-time assessment of acute cognitive stress from cardiac mechanical signals captured by a skin-like patch. *Biosens Bioelectron*. Mar 15 2024;248:115983. doi:10.1016/j.bios.2023.115983
13. Kim B, Manchuri AR, Oh GT, et al. Experimental analysis and prediction of radionuclide solubility using machine learning models: Effects of organic complexing agents. *J Hazard Mater*. May 5 2024;469:134012. doi:10.1016/j.jhazmat.2024.134012
14. Uttsha MM, Haque A, Banna TT, Deowan SA, Ariful Islam M, Hasan Babu HM. Enhancing agricultural automation through weather invariant soil parameter prediction using machine learning. *Heliyon*. Apr 15 2024;10(7):e28626. doi:10.1016/j.heliyon.2024.e28626
15. Allgaier J, Mulansky L, Draelos RL, Pryss R. How does the model make predictions? A systematic literature review on the explainability power of machine learning in healthcare. *Artif Intell Med*. Sep 2023;143:102616. doi:10.1016/j.artmed.2023.102616
16. Ali S, Akhlaq F, Imran AS, Kastrati Z, Daudpota SM, Moosa M. The enlightening role

of explainable artificial intelligence in medical & healthcare domains: A systematic literature review. *Comput Biol Med.* Oct 4 2023;166:107555. doi:10.1016/j.compbimed.2023.107555

17. Tian N, Chen S, Han H, Jin J, Li Z. Association between triglyceride glucose index and total bone mineral density: a cross-sectional study from NHANES 2011-2018. *Sci Rep.* Feb 20 2024;14(1):4208. doi:10.1038/s41598-024-54192-9

18. Wen Z, Li Y, Xu L, et al. Triglyceride Glucose-Body Mass Index Is a Reliable Indicator of Bone Mineral Density and Risk of Osteoporotic Fracture in Middle-Aged and Elderly Nondiabetic Chinese Individuals. *J Clin Med.* Sep 26 2022;11(19)doi:10.3390/jcm11195694

19. Gauthier J, Wu QV, Gooley TA. Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians. *Bone Marrow Transplant.* Apr 2020;55(4):675-680. doi:10.1038/s41409-019-0679-x

20. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ.* Mar 18 2020;368:m441. doi:10.1136/bmj.m441

21. Riley RD, Snell KIE, Archer L, et al. Evaluation of clinical prediction models (part 3): calculating the sample size required for an external validation study. *BMJ.* Jan 22 2024;384:e074821. doi:10.1136/bmj-2023-074821

22. Palermo A, Tuccinardi D, Defeudis G, et al. BMI and BMD: The Potential Interplay between Obesity and Bone Fragility. *Int J Environ Res Public Health*. May 28 2016;13(6)doi:10.3390/ijerph13060544
23. Lloyd JT, Alley DE, Hawkes WG, Hochberg MC, Waldstein SR, Orwig DL. Body mass index is positively associated with bone mineral density in US older adults. *Arch Osteoporos*. 2014;9:175. doi:10.1007/s11657-014-0175-2
24. Hoxha R, Islami H, Qorraj-Bytyqi H, Thaci S, Bahtiri E. Relationship of weight and body mass index with bone mineral density in adult men from kosovo. *Mater Sociomed*. Oct 2014;26(5):306-8. doi:10.5455/msm.2014.26.306-308
25. Jiang Y, Zhang Y, Jin M, Gu Z, Pei Y, Meng P. Aged-Related Changes in Body Composition and Association between Body Composition with Bone Mass Density by Body Mass Index in Chinese Han Men over 50-year-old. *PLoS One*. 2015;10(6):e0130400. doi:10.1371/journal.pone.0130400
26. Dorilleau C, Kanagaratnam L, Charlot I, et al. "The least significant change on bone mineral density scan increased in patients with higher degrees of obesity". *Aging Clin Exp Res*. Apr 23 2024;36(1):98. doi:10.1007/s40520-024-02736-4

27. Dogan A, Nakipoglu-Yuzer GF, Yildizgoren MT, Ozgirgin N. Is age or the body mass index (BMI) more determinant of the bone mineral density (BMD) in geriatric women and men? *Arch Gerontol Geriatr.* Nov-Dec 2010;51(3):338-41. doi:10.1016/j.archger.2010.01.015
28. Kirchengast S, Peterson B, Hauser G, Knogler W. [Significance of body weight status for bone density in the elderly and very old]. *Z Gerontol Geriatr.* Aug 2001;34(4):313-8. Zur Bedeutung des Gewichtsstatus für die Knochendichte bei Betagten und Hochbetagten. doi:10.1007/s003910170055
29. Kang D, Liu Z, Wang Y, et al. Relationship of body composition with bone mineral density in northern Chinese men by body mass index levels. *J Endocrinol Invest.* Apr 2014;37(4):359-67. doi:10.1007/s40618-013-0037-6
30. Wee J, Sng BY, Shen L, Lim CT, Singh G, Das De S. The relationship between body mass index and physical activity levels in relation to bone mineral density in premenopausal and postmenopausal women. *Arch Osteoporos.* 2013;8:162. doi:10.1007/s11657-013-0162-z
31. Pinar-Gutierrez A, Garcia-Fontana C, Garcia-Fontana B, Munoz-Torres M. Obesity and Bone Health: A Complex Relationship. *Int J Mol Sci.* Jul 27 2022;23(15)doi:10.3390/ijms23158303

32. Du Y, Zhao LJ, Xu Q, Wu KH, Deng HW. Socioeconomic status and bone mineral density in adults by race/ethnicity and gender: the Louisiana osteoporosis study. *Osteoporos Int.* May 2017;28(5):1699-1709. doi:10.1007/s00198-017-3951-1
33. Brennan SL, Pasco JA, Urquhart DM, Oldenburg B, Wang Y, Wluka AE. Association between socioeconomic status and bone mineral density in adults: a systematic review. *Osteoporos Int.* Feb 2011;22(2):517-27. doi:10.1007/s00198-010-1261-y
34. Zhang Y, Tan C, Tan W. BMI, socioeconomic status, and bone mineral density in U.S. adults: Mediation analysis in the NHANES. *Front Nutr.* 2023;10:1132234. doi:10.3389/fnut.2023.1132234
35. Pan K, Tu R, Yao X, Zhu Z. Associations between serum calcium, 25(OH)D level and bone mineral density in adolescents. *Adv Rheumatol.* Mar 10 2021;61(1):16. doi:10.1186/s42358-021-00174-8
36. Hodges JK, Cao S, Cladis DP, Weaver CM. Lactose Intolerance and Bone Health: The Challenge of Ensuring Adequate Calcium Intake. *Nutrients.* Mar 28 2019;11(4)doi:10.3390/nu11040718

37. Zhan H, Liu X, Piao S, Rong X, Guo J. Association between triglyceride-glucose index and bone mineral density in US adults: a cross sectional study. *J Orthop Surg Res.* Oct 30 2023;18(1):810. doi:10.1186/s13018-023-04275-6

EARLY ACCESS

TABLES AND FIGURES WITH LEGENDS

TABLE 1. Baseline characteristics of training and test datasets.

Characteristics	Training set (N = 829)	Test set (N = 353)	<i>P</i> value
FNBMD (gm/cm ²), median (IQR)	0.80 (0.71, 0.89)	0.80 (0.71, 0.89)	0.725
TyG-BMI, median (IQR)	228 (200, 260)	231 (206, 265)	0.135
Age (yrs), median (IQR)	61.0 (55.0, 68.0)	61.0 (55.0, 70.0)	0.225
BMI (kg/m ²), median (IQR)	32.8 (29.7, 36.5)	33.4 (29.9, 36.5)	0.421
Race, N (%)			0.084
Mexican American	108 (13.0)	47 (13.3)	
Other Hispanic	93 (11.2)	25 (7.08)	
Non-Hispanic White	353 (42.6)	153 (43.3)	
Non-Hispanic Black	202 (24.4)	83 (23.5)	
Other race	73 (8.81)	45 (12.7)	
Education, N (%)			0.130
Less than high school	227 (27.4)	115 (32.6)	
High school	197 (23.8)	70 (19.8)	
College or above	405 (48.9)	168 (47.6)	
Marital status, N (%)			0.114
Married or living with partner	580 (70.0)	263 (74.5)	
Single	249 (30.0)	90 (25.5)	
Drinking status, N (%)			0.447
No	205 (24.7)	80 (22.7)	
Yes	624 (75.3)	273 (77.3)	
Smoking status, N (%)			0.292
No	571 (68.9)	254 (72.0)	
Yes	258 (31.1)	99 (28.0)	
Physical activity, N (%)			0.779
Less active	369 (44.5)	154 (43.6)	
Active	460 (55.5)	199 (56.4)	
FHOS, N (%)			0.367
No	765 (92.3)	331 (93.8)	
Yes	64 (7.72)	22 (6.23)	
PIR, median (IQR)	2.70 (1.40, 4.28)	2.59 (1.35, 4.04)	0.364
SBP (mmHg), median (IQR)	127 (117, 139)	128 (119, 139)	0.598
DBP (mmHg), median (IQR)	74.0 (67.0, 80.0)	74.0 (67.0, 81.0)	0.71
HDL (mmol/L), median (IQR)	1.29 (1.09, 1.55)	1.29 (1.09, 1.60)	0.751
FPG (mmol/L), median (IQR)	5.66 (5.33, 6.05)	5.77 (5.38, 6.11)	0.171
TC (mmol/L), median (IQR)	4.99 (4.40, 5.61)	4.97 (4.32, 5.56)	0.323
Triglycerides (mmol/L), median (IQR)	1.09 (0.81, 1.43)	1.09 (0.79, 1.46)	0.522
LDL (mmol/L), median (IQR)	3.10 (2.53, 3.60)	3.03 (2.46, 3.60)	0.346
Serum calcium (mmol/L), median (IQR)	2.33 (2.28, 2.38)	2.33 (2.28, 2.38)	0.609

Serum phosphorus (mmol/L), median (IQR)	1.10 (1.00, 1.20)	1.07 (0.97, 1.16)	0.565
Serum 25(OH)D3 (ng/mL), median (IQR)	59.7 (52.4, 67.6)	58.9 (51.0, 67.0)	0.253

P values between groups were assessed by the Chi-square and Mann–Whitney U tests. IQR:

Inter-quartile range; FNBMD: Femoral neck bone mineral density; TyG-BMI: Triglyceride and

glucose-body mass index; FHOS: Family history of osteoporosis; PIR: Poverty income ratio;

SBP: Systolic blood pressure; DBP: Diastolic blood pressure; HDL: High-density lipoprotein;

FPG: Fasting plasma glucose; TC: Total cholesterol; LDL: Low-density lipoprotein.

EARLY ACCESS

TABLE 2. FNBMD prediction results for each model with metrics.

Model	R ²		MAE		MSE		MAPE		RMSE	
	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
MLR	0.264	0.168	0.092	0.095	0.014	0.014	0.116	0.121	0.116	0.120
DTR	0.252	0.150	0.092	0.095	0.014	0.015	0.117	0.121	0.117	0.121
SVR	0.298	0.204	0.090	0.093	0.013	0.014	0.114	0.119	0.114	0.117
GPR	0.691	0.099	0.058	0.106	0.006	0.019	0.072	0.134	0.075	0.138
XGBR	0.174	0.115	0.096	0.097	0.015	0.015	0.119	0.123	0.123	0.123
RFR ^a	0.712	0.218	0.053	0.092	0.005	0.013	0.067	0.118	0.072	0.116
RFR ^b	0.821	0.199	0.043	0.092	0.003	0.014	0.054	0.119	0.057	0.117
RFR ^c	0.158	0.073	0.097	0.100	0.015	0.017	0.124	0.127	0.123	0.130
ETR	0.708	0.212	0.052	0.093	0.005	0.014	0.064	0.118	0.073	0.117
MLPR	0.135	0.094	0.098	0.097	0.016	0.016	0.119	0.119	0.126	0.125

Note: The metrics of the best-performing model are bolded. ^aThe model included predictors of age, BMI, PIR, non-Hispanic Black, non-Hispanic White, and serum calcium. ^b The model included predictors of age, BMI, PIR, and serum calcium. ^c The model employed TyG-BMI as a predictor. FNBMD: Femoral neck bone mineral density; R²: Coefficient of determination; MAE: Mean absolute error; MSE: Mean squared error; MAPE: Mean absolute percentage error; RMSE: Root mean squared error; MLR: Multivariable linear regression; DTR: Decision tree regression; SVR: Support vector regression; GPR: Gaussian process regression; XGBR: Extreme gradient boosting regression; RFR: Random forest regression; ETR: Extra-trees regression; MLPR: Multi-layer perceptron regression; TyG-BMI: Triglyceride and glucose-body mass index.

TABLE 3. Predictor importance ranking in each model.

Model	BMI		Non-Hispanic Black		Age		PIR		Serum calcium		Non-Hispanic White	
	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
MLR	1	1	2	2	3	3	6	4	4	5	5	6
DTR	1	1	2	2	3	3	4	4	5	5	6	6
SVR	1	1	2	2	3	3	5	4	4	5	6	6
GPR	1	1	6	5	2	2	4	3	3	6	5	4
XGBR	1	1	3	2	2	3	5	5	6	6	4	4
RFR	1	1	2	2	3	3	4	4	5	5	6	6
ETR	1	1	2	2	3	3	4	4	5	5	6	6
MLPR	1	1	2	3	3	4	4	2	6	6	5	5
Mean±SD	1.0±0.0	1.0±0.0	2.6±1.4	2.5±1.1	2.8±0.5	3.0±0.5	4.5±0.8	3.8±0.9	4.8±1.0	5.4±0.5	5.4±0.7	5.4±0.9
Median (IQR)	1.0(1.0, 1.0)	1.0(1.0, 1.0)	2.0(2.0, 2.3)	2.0(2.0, 2.3)	3.0(2.8, 3.0)	3.0(3.0, 3.0)	4.0(4.0, 5.0)	4.0(3.8, 4.0)	5.0(4.0, 5.3)	5.0(5.0, 6.0)	5.5(5.0, 6.0)	6.0(4.8, 6.0)

BMI: Body mass index; PIR: Family income to poverty ratio; MLR: Multivariable linear regression; DTR: Decision tree regression; SVR: Support vector regression; GPR: Gaussian process regression; XGBR: Extreme gradient boosting regression; RFR: Random forest regression; ETR: Extra-trees regression; MLPR: Multi-layer perceptron regression; SD: Standard deviation; IQR: Inter-quartile range.

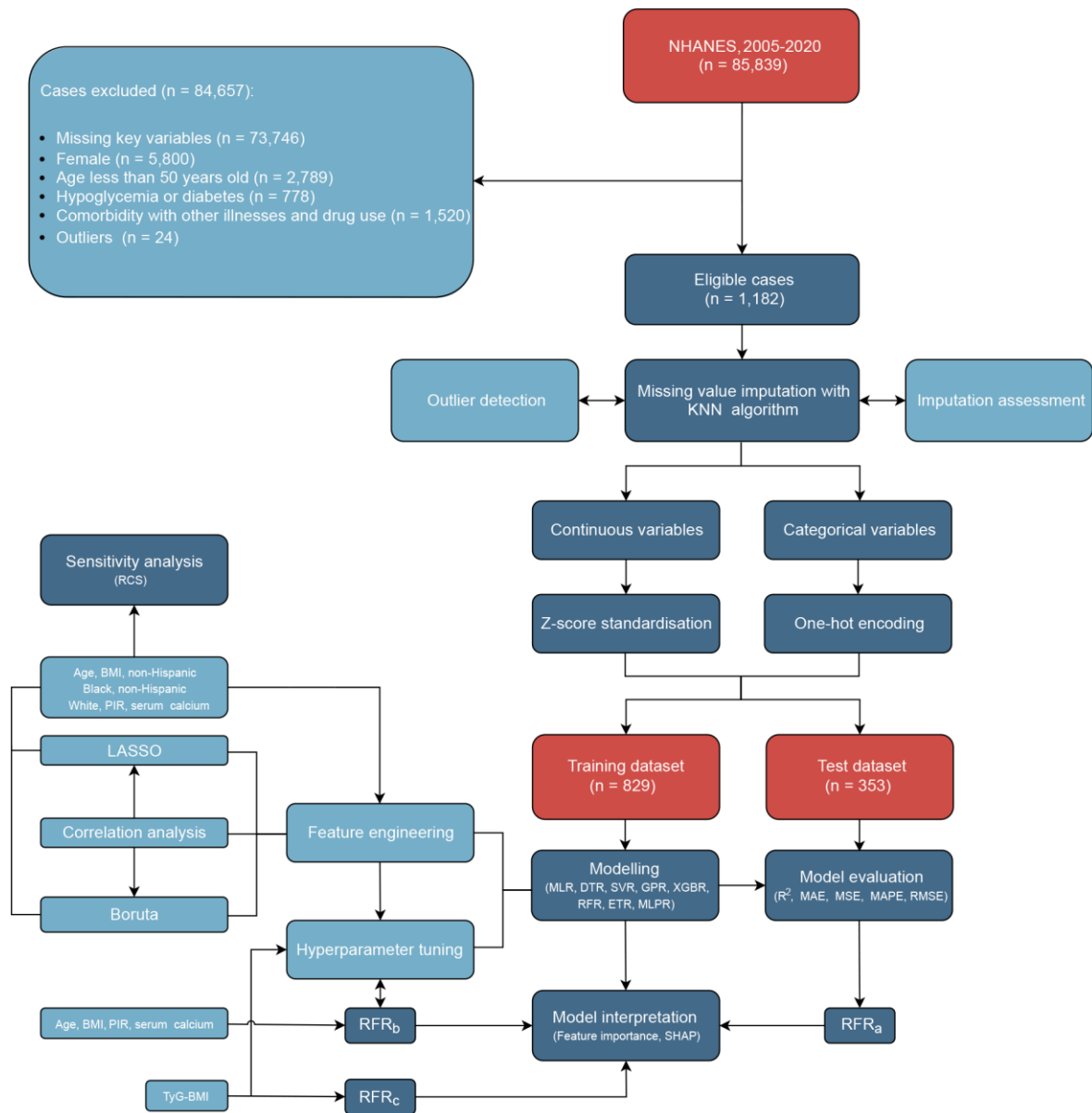


FIGURE 1. Flow chart of the study design. NHANES: National health and nutrition examination survey; KNN: K-nearest neighbors; MLR: Multivariable linear regression; DTR: Decision tree regression; SVR: Support vector regression; GPR: Gaussian process regression; XGBR: Extreme gradient boosting regression; RFR: Random forest regression; ETR: Extra-trees regression; MLPR: Multi-layer perceptron regression; R²: Coefficient of determination; MAE: Mean absolute error; MSE: Mean squared error; MAPE: Mean absolute percentage error; RMSE: Root mean squared error; TyG-BMI: Triglyceride and glucose-body mass index; PIR: Poverty income ratio; LASSO: Least absolute shrinkage and selection operator; RCS: Restricted cubic spline.

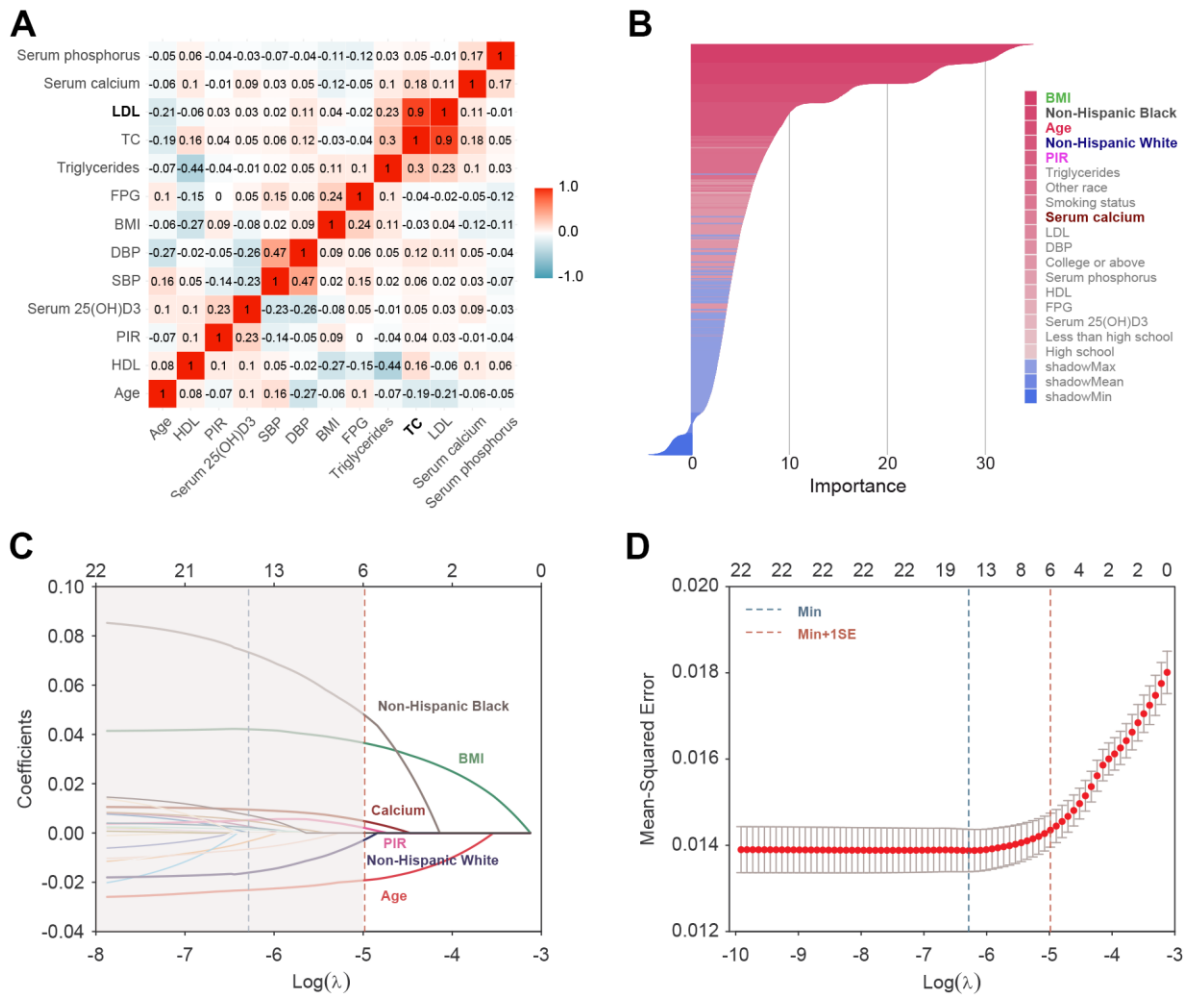


FIGURE 2. Feature engineering. (A) Spearman or Pearson correlation matrix of continuous variables. Feature pairs with correlation coefficients greater than 0.8 have been bolded. (B) Feature selection with the Boruta algorithm. The features highlighted in bold and color are identified by intersecting the results from the Boruta and LASSO algorithms. (C-D) feature selection was done by using the LASSO algorithm. HDL, high-density lipoprotein; PIR, poverty income ratio; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; FPG, fasting plasma glucose; TC, total cholesterol; LDL, low-density lipoprotein.

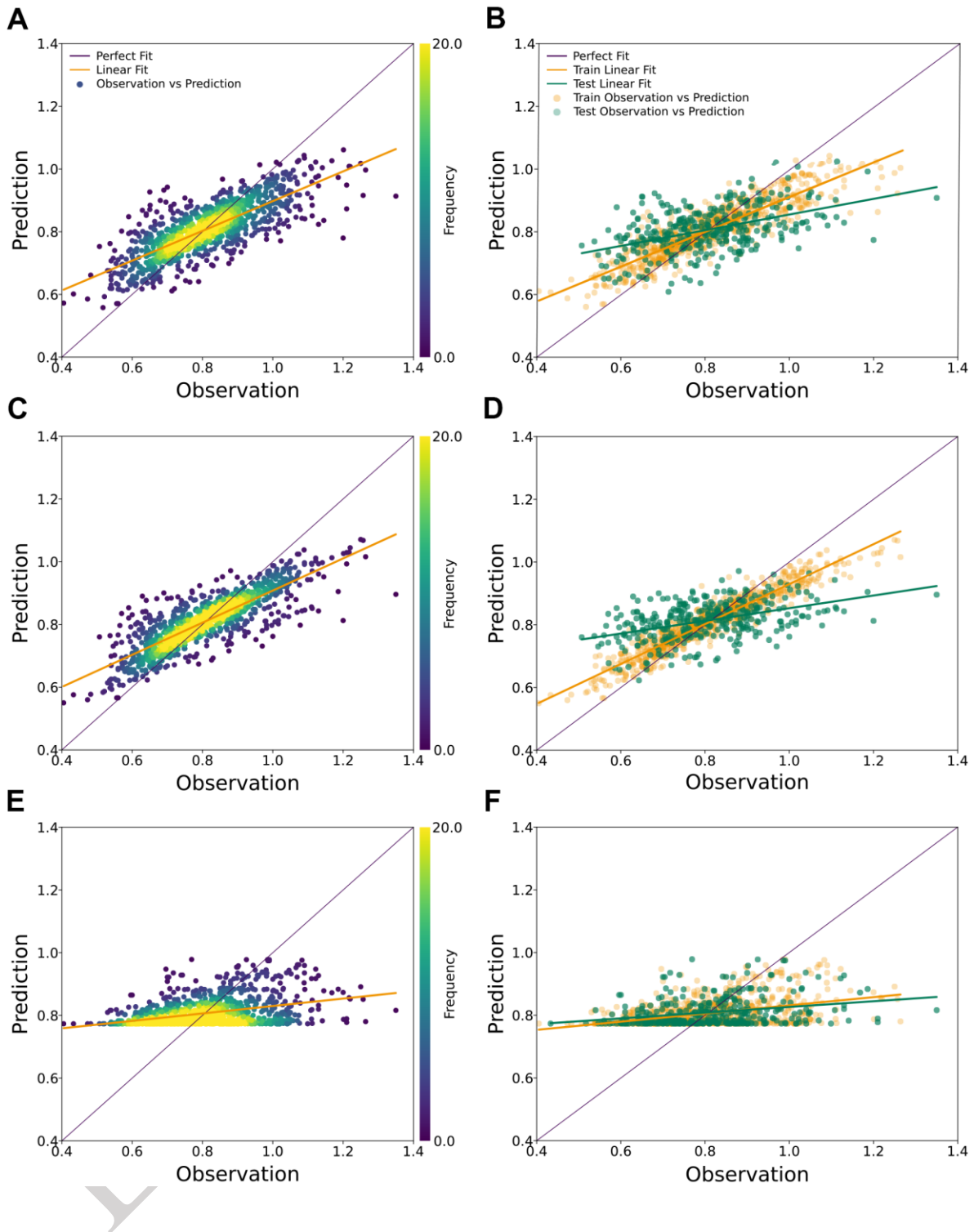


FIGURE 3. Scatter plots for FNBMD prediction in RFRs. (A-B) scatter density plots (left) and scatter training/test plots (right) for the RFR model, incorporating age, BMI, non-Hispanic Black and White, PIR, and serum calcium. (C-D) plots for the RFR model including age, BMI, PIR, and serum calcium. (E-F) plots for the RFR model solely containing the TyG-BMI index. Scatter density plots comprise all data points, whereas the scatter plots are separately composed of training and test sets. FNBMD, femoral neck bone mineral density; RFR, random forest regressor; BMI, body mass index; PIR, poverty income ratio; TyG-BMI, triglyceride and glucose-body mass index.

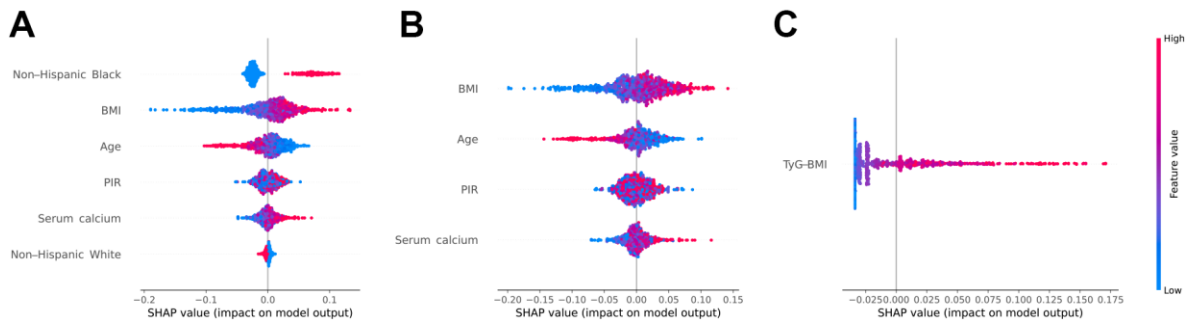


FIGURE 4. SHAP value of features and the impact on the model output for FNBMD prediction in RFRs. (A) plots for the RFR model, incorporating age, BMI, non-Hispanic Black and White, PIR, and serum calcium. (B) plots for the RFR model including age, BMI, PIR, and serum calcium. (C) plots for the RFR model solely containing the TyG-BMI index. SHAP, shapley additive exPlanations; FNBMD, femoral neck bone mineral density; BMI, body mass index; PIR, poverty income ratio; TyG-BMI, triglyceride and glucose-body mass index.

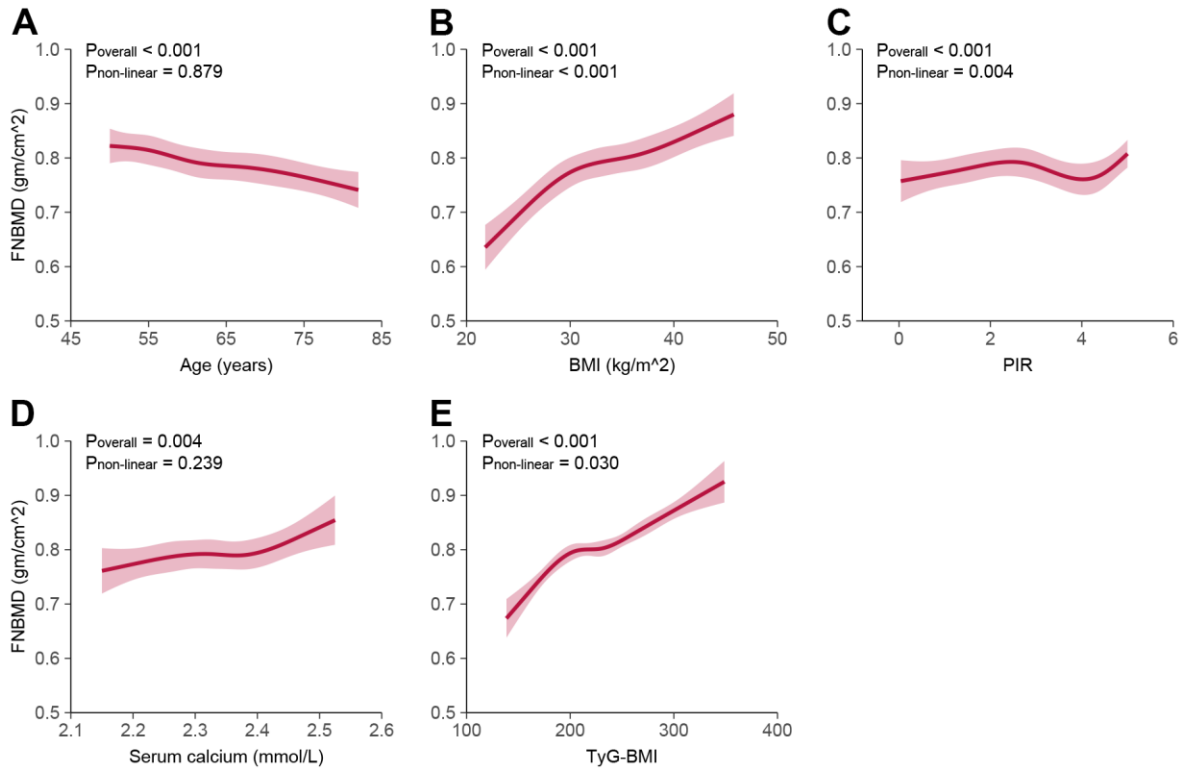


FIGURE 5. Restricted cubic spline analysis. (A-E) association between continuous features and FNBMD. Each curve employs five knots located at the 5th, 35th, 50th, 65th, and 95th percentiles. Shaded regions denote the 95% confidence intervals. FNBMD, femoral neck bone mineral density; BMI, body mass index; PIR, poverty income ratio; TyG-BMI, triglyceride and glucose-body mass index.

SUPPLEMENTAL DATA

Data S1. Triglyceride and glucose-body mass index (TyG-BMI)

Body Mass Index (BMI) was calculated using the formula weight divided by the square of height (weight/height²). After an overnight fast, venous samples were collected using aseptic techniques for the analysis of phosphorus, total calcium, plasma glucose, and lipid levels.

The definitions of the TyG-BMI terms were determined as follows:

1. The TyG index was calculated using the formula: $\text{Ln} [\text{triglycerides (TG) in mg/dL} \times \text{fasting plasma glucose (FPG) in mg/dL} / 2]$.
2. TyG-BMI was derived by multiplying the TyG index by BMI, expressed in kg/m^2 .

TABLE S1. Comparisons between raw and imputation datasets.

Characteristics	Missing values N (%)	Raw dataset (N = 1182)	Imputation dataset (N = 1182)	<i>P</i> value
FNBMD (gm/cm ²), median (IQR)		0.80 (0.71, 0.89)	0.80 (0.71, 0.89)	1.000
Age (yrs), median (IQR)		61.0 (55.0, 68.0)	61.0 (55.0, 68.0)	1.000
BMI (kg/m ²), median (IQR)	20 (1.7)	32.8 (29.7, 36.6)	33.4 (29.7, 36.5)	0.917
Race, N (%)				1.000
Mexican American		155 (13.1)	155 (13.1)	
Other Hispanic		118 (10.0)	118 (10.0)	
Non-Hispanic White		506 (42.8)	506 (42.8)	
Non-Hispanic Black		285 (24.1)	285 (24.1)	
Other race		118 (10.0)	118 (10.0)	
Education, N (%)				1.000
Less than high school		342 (28.9)	342 (28.9)	
High school		267 (22.6)	267 (22.6)	
College or above		573 (48.5)	573 (48.5)	
Marital status, N (%)				1.000
Married or living with partner		843 (71.3)	843 (71.3)	
Single		339 (28.7)	339 (28.7)	
Drinking status, N (%)	276 (23.4)			0.365
No		235 (25.9)	285 (24.1)	
Yes		671 (74.1)	897 (75.9)	
Smoking status, N (%)	426 (36.0)			0.056
No		487 (64.4)	825 (69.8)	
Yes		269 (35.6)	357 (30.2)	
Physical activity, N (%)				1.000
Less active		523 (44.2)	523 (44.2)	
Active		659 (55.8)	659 (55.8)	
FHOS, N (%)				1.000
No		1096 (92.7)	1096 (92.7)	
Yes		86 (7.28)	86 (7.28)	
PIR, median (IQR)	118 (10.0)	2.44 (1.29, 4.42)	2.64 (1.39, 4.21)	0.355
SBP (mmHg), median (IQR)		127.0 (117.0, 140.0)	127.0 (117.0, 139.0)	0.745
DBP (mmHg), median (IQR)	70 (5.9)	74.0 (67.0, 81.0)	74.0 (67.0, 80.8)	0.979
HDL (mmol/L), median (IQR)	24 (2.0)	1.29 (1.09, 1.55)	1.29 (1.09, 1.55)	0.670
FPG (mmol/L), median (IQR)	4 (0.3)	5.72 (5.33, 6.11)	5.72 (5.33, 6.11)	0.947
TC (mmol/L), median (IQR)	(1.4)	4.97 (4.37, 5.61)	4.97 (4.37, 5.61)	0.847
Triglycerides (mmol/L), median (IQR)	67 (5.7)	1.07 (0.79, 1.45)	1.09 (0.80, 1.44)	0.404
LDL (mmol/L), median (IQR)	21 (1.8)	3.08 (2.51, 3.60)	3.08 (2.51, 3.60)	0.764
Serum calcium (mmol/L), median (IQR)	28 (2.4)	2.33 (2.28, 2.38)	2.33 (2.28, 2.38)	0.985
Serum phosphorus (mmol/L), median (IQR)	14 (1.2)	1.07 (1.00, 1.20)	1.07 (1.00, 1.18)	0.986
Serum 25(OH)D3 (ng/mL), median (IQR)	668 (56.5)	59.3 (42.0, 75.1)	59.4 (51.9, 67.5)	0.545

P values between groups were assessed by the Chi-square and Mann–Whitney U tests. IQR, inter-quartile range; FNBMD, femoral neck bone mineral density; BMI, body mass index; FHOS, family history of osteoporosis; PIR, poverty income ratio; SBP, systolic blood pressure; DBP, diastolic blood pressure; HDL, high-density lipoprotein; FPG, fasting plasma glucose; TC, Total cholesterol; LDL, low-density lipoprotein.

TABLE S2. Optimization of hyperparameters using grid search.

Model	Hyperparameter	Range of hyperparameter in grid search	Selected value
MLR	-	-	-
DTR	max_depth	50, 200, 500, 800, 1000, 1200	1000
	min_samples_split	2, 10, 18, 26, 34, 42	10
	min_samples_leaf	3, 11, 19, 27, 35, 43	43
	max_features	0.8, 0.9, 1.0	0.8
SVR	gamma	10^{-6} , 10^{-5} , 10^{-4} , 10^{-3} , 10^{-2} , 10^{-1} , 10^0	10^{-1}
	C	10^{-3} , 10^{-2} , 10^{-1} , 10^0 , 10^1 , 10^2 , 10^3 , 10^4	10^{-1}
	epsilon	10^{-3} , 10^{-2} , 10^{-1} , 10^0 , 10^1 , 10^2	10^{-1}
GPR	alpha	10^{-10} , 10^{-9} , 10^{-8} , 10^{-7} , 10^{-6} , 10^{-5} , 10^{-4} , 10^{-3} , 10^{-2} , 10^{-1} , 10^0 , 10^1	10^{-1}
XGBR	n_estimators	100, 200, 400	100
	max_depth	3, 5, 7, 10	7
	learning_rate	0.1, 0.2, 0.3	0.3
	gamma	0, 5, 10, 20	0
	reg_alpha	5, 10, 15	5
	reg_lambda	0, 0.1, 0.2, 0.3	0
RFR ^a	n_estimators	100, 200, 400	100
	criterion	'squared_error', 'absolute_error', 'poisson'	'absolute_error'
	max_depth	10, 50, 100, 200, 400	100
	min_samples_split	2, 4, 8, 16	2
	min_samples_leaf	1, 2, 4, 8	2
	max_features	0.8, 0.9, 1.0	0.8
RFR ^b	n_estimators	100, 200, 400	200
	criterion	'squared_error', 'absolute_error', 'poisson'	'absolute_error'
	max_depth	10, 50, 100, 200, 400	400
	min_samples_split	2, 4, 8, 16	4
	min_samples_leaf	1, 2, 4, 8	1
	max_features	0.8, 0.9, 1.0	0.9
RFR ^c	n_estimators	100, 200, 400	100
	criterion	'squared_error', 'absolute_error', 'poisson'	'poisson'
	max_depth	10, 50, 100, 200, 400	10
	min_samples_split	2, 4, 8, 16	8
	min_samples_leaf	1, 2, 4, 8	2
	max_features	0.8, 0.9, 1.0	0.9
ETR	n_estimators	100, 200, 400	200
	criterion	'squared_error', 'absolute_error'	'absolute_error'
	max_depth	10, 50, 100	100
	min_samples_split	2, 4, 8, 16	2
	min_samples_leaf	1, 2, 4, 8	1
	max_features	0.6, 0.8, 1.0	0.8
	max_samples	0.8, 0.9, 1.0	1.0
	ccp_alpha	0, 1, 10, 100, 1000	0
MLPR	hidden_layer_sizes	(10,), (20,), (10, 10), (20, 20), (10, 20, 10)	(10, 10)
	activation	'identity', 'tanh', 'relu'	'identity'
	solver	'sgd', 'adam'	'adam'
	alpha	10^{-4} , 10^{-3} , 10^{-2} , 10^{-1}	10^{-2}
	learning_rate	'constant', 'adaptive'	'constant'
	learning_rate_init	10^{-3} , 10^{-2} , 10^{-1}	10^{-1}

^a The model included predictors of age, BMI, PIR, non-Hispanic Black, non-Hispanic White, and serum calcium. ^b The model included predictors of age, BMI, PIR, and serum calcium. ^c The model employed TyG-BMI as a predictor. MLR, multivariable linear regression; DTR, decision tree regression; SVR, support vector regression; GPR, gaussian process regression; XGBR, extreme gradient boosting regression; RFR, random forest regression; ETR, extra-trees regression; MLPR, multi-layer perceptron regression; TyG-BMI, triglyceride and

glucose-body mass index.

EARLY ACCESS