A Study on Prediction of Osteoporosis Risk in Middle-Aged and Elderly People Based on Lipid Indicators

Miao Zhang¹,a, Yiran Wan¹,a, Jialu Yang¹,a, Pingping Yu²,b, Xun Lei¹,c,*

¹School of Public Health, Chongqing Medical University, Chongqing, 401331, China
²Department of Health Management Center, The Second Affiliated Hospital, Chongqing Medical University, Chongqing, 400010, China
a202110613@stu.cqmu.edu.cn, b465430809@qq.com, cleixun521@cqmu.edu.cn
*Corresponding author

Abstract: Osteoporosis is a common health problem in the middle-aged and elderly population, and its risk may be related to lipid levels. This study aimed to explore the effect of lipids on osteoporosis in middle-aged and elderly population and to develop a predictive model for osteoporosis. Middle-aged and elderly people over 45 years old who had undergone DEXA in several hospitals in Chongqing City from 2016 to 2020 as study subjects. After data cleaning and feature selection, the prediction model was established using the random forest (RF), light gradient boosting ma-machine (LightGBM), extreme gradient boosting (XGBoost) and classification boosting (CatBoost) algorithms. In the test set, the accuracy (0.83571) of the XGBoost model is significantly higher than that of RF (0.82142), CatBoost (0.80714), and LightGBM (0.79285). Lipid indicators could be used in osteoporosis prediction studies, and the established XGBoost model had the best prediction ability. The model needs to be further validated by a larger external sample.

Keywords: Osteoporosis; Middle-aged and elderly; Lipid indicators; Prediction models; Machine learning

1. Introduction

Osteoporosis is a common chronic systemic disease, which is usually characterized by a decrease in the amount of bone per unit volume and a decrease in bone density; and then the destruction of the microstructure of the bone tissue and an increase in the brittleness of the bones; ultimately leading to the occurrence of fractures and other serious consequences of the patients under the slightest external force[4], which increases the disability rate and mortality rate of the patients, and poses a greater threat to the patients' lives and health[5]. There are more than 200 million osteoporosis patients worldwide, and the number of osteoporosis patients in China has increased dramatically with the aging of China's population, and is now as high as 90 million people, more than the combined total of Europe, the Americas, and Japan (75 million) [3], and osteoporosis has become a serious danger to the lives and health of the Chinese people. The medical treatment and care of osteoporosis and bone fracture requires a large amount of human and material resources, which greatly depletes medical resources and creates a heavy burden on families and society.

Osteoporosis can occur at any age, and its level of incidence tends to increase with age [4], making middle-aged and older adults the most susceptible to osteoporosis and bone loss. According to the results of the 2018 Chinese residents' osteoporosis epidemiology survey, the prevalence rate of osteoporosis among the elderly over 50 years old was 19.2%, and the prevalence rate among people over 65 years old was 32.0% [5]. Meanwhile, residents' knowledge of osteoporosis is generally insufficient, with only 7.0% of patients over 50 years of age; the detection rate of osteoporosis needs to be improved urgently, with only 3.7% of people over 50 years of age receiving bone density tests [5]. Therefore, focusing on the risk factors of osteoporosis in middle-aged and old-aged people and taking targeted interventions in advance, moving the gate of osteoporosis primary prevention (etiologic prevention) forward, delaying bone loss, and reducing the destruction of the skeletal system have crucial clinical value and practical significance for improving the quality of life and health of middle-aged and old-aged people.

There may be a correlation between osteoporosis and blood lipid levels [6]. Osteoporosis and
dyslipidemia share many risk factors, such as aging, female menopause, genetics, smoking and drinking, and lifestyle [7][8]. Lipid profiles are useful in assessing the risk of cardiovascular disease [9]. There is also evidence from animal studies that lipids are involved in the development of osteoporosis [10]. Most studies have shown a negative correlation between lipid markers and BMD [11][12]. Li GH et al. who found a negative correlation between LDL-C levels and BMD in a combined epidemiological observational analysis and Mendelian randomization study [13]. Kan B et al. found that higher levels of TC and TG were associated with a greater risk of osteoporosis in a cross-sectional study [14]. However, a cross-sectional study by Ghadiri-Anari A found no significant correlation between lipid levels and femur and lumbar spine bone density in postmenopausal women [15]. In conclusion, as far as the current study is concerned, the relationship between lipid markers and osteoporosis remains somewhat controversial.

Traditional studies of influencing factors can provide ideas for the prevention of osteoporosis. However, for the middle-aged and elderly population, we are often more concerned about their risk of developing cardiovascular and cerebrovascular diseases, so much so that we neglect the development of osteoporosis and seldom undergo bone density testing. Therefore, we attempted to construct a prediction model for osteoporosis in middle-aged and elderly people by utilizing lipid testing indicators. The risk prediction of osteoporosis can be realized through routine physical examination without the need for expensive specialized tests.

With the rapid development of big data technologies, predictive models using machine learning and deep learning are widely used. Many studies have developed models to predict osteoporosis. Sadatsafavi M et al. [16] used artificial neural network for prediction of osteoporosis in Iranian postmenopausal women, and the results showed that the prediction performance of the artificial neural network model was higher than that of the traditional regression methods and the currently recognized decision rules. Zhi YJ et al. [17] studied the prediction of severe osteoporosis in postmenopausal women based on decision tree, artificial neural network, and logistic regression model, and by comparing the ROC curves of the three, they found that the fit of neural network and logistic regression was better, which showed that artificial neural network and logistic regression could be considered in the prediction study of severe osteoporosis. Tae Keun Yoo et al. [18] predicted osteoporosis based on a popular machine learning model and compared it with four traditional osteoporosis risk assessment tools. The results showed that the support vector machine model outperformed the other models when age, height, weight, body mass index, duration of menopause, duration of breastfeeding, estrogen therapy, hyperlipidemia, hypertension, osteoarthritis, and diabetes mellitus were entered as variables.

In this study, our main objective was to develop a prediction model for osteoporosis in middle-aged and elderly people. The model, which relies only on demographic characteristics and lipid testing indicators, not only provides a research idea but also a convenient tool for osteoporosis risk prediction in the middle-aged and elderly population. This is the main innovation and contribution of our study. In addition, as a retrospective study, our study also conducted a correlation analysis between osteoporosis and lipid indicators in middle-aged and elderly populations in Chongqing, China.

2. Materials and Methods

2.1. Study Subjects

The subjects of this study are all from the medical data platform "YIDUCLOUD", which contains the clinical data of patients from dozens of hospitals in Chongqing. The platform can access, manage, and analyze in-depth the multi-source heterogeneous medical data from multiple healthcare institutions in the city that have accessed the data platform according to the authorization. In the case system of the medical data platform, the study subjects were screened by conditional search with the following criteria:

Inclusion criteria: (1) age greater than or equal to 45 years old; (2) application of dual-energy X-ray absorptiometry (DEXA) for bone mineral density (BMD) measurement of L1-L4 lumbar vertebrae and hips of the subjects, and the measurement of lipid-related indexes was also performed; (3) limiting the time of the examination to the period from 2016 to 2021.

Exclusion criteria: (1) a history of metabolic disease-related conditions, such as diabetes mellitus, thyroid disease, chronic kidney disease, etc.; (2) hypertension, rheumatoid arthritis; (3) tumors, coronary artery disease, stroke, epilepsy, Parkinson's disease, asthma, or COPD; (4) gastrointestinal disorders or disorders of the gastrointestinal tract (e.g., chronic enteritis, gastric ulcers, ulcerative colitis, etc.); (5) fracture surgeries; and (6) glucocorticoid use.
This study was approved by the Medical Research Science Committee of Chongqing Medical University (20230601).

2.2. Data Collection

Based on the inclusion and exclusion criteria, the case data exported from the system contained data on a total of 15,554 patients. Data without records of osteoporosis diagnosis and lipid-related index tests results (n=8866), and data with missing key values (e.g. BMI) or logical errors (n=5991) were removed according to the requirements of this study, resulting in the inclusion of 697 individuals in the study. (Figure.1)

![Flowchart](image)

**Figure 1: The inclusion process of participants in this study.**

2.2.1. General Information

All case test data come from the platform's electronic medical records, and general information is obtained by querying the records, including gender, age, body mass index (BMI), and whether or not they smoke or drink alcohol, etc., while matching the bone densitometry results and lipid test results according to the medical record number.

2.2.2. Bone Densitometry and Diagnosis of Osteoporosis

BMD was measured by DEXA, which is considered to be a direct indicator of osteoporosis, with the more accurate DEXA BMD results as the "gold standard" recommended by the WHO. According to the 1994 WHO diagnostic criteria for osteoporosis: T≥1, BMD not less than 1 standard deviation (SD) from the normal average value for young adults of the same sex is considered normal bone mass; -2.5<T<1, BMD between -1 and -2.5 SD from the normal average value for young adults of the same sex is considered bone loss; T≤-2.5, BMD equal to 1 SD: -2.5, BMD equal to 1 SD is considered bone loss; T≤-2.5, BMD equal to 1 SD; and T≤-2.5, BMD equal to 1 SD is considered bone loss. -2.5, BMD equal to or less than 2.5 standard deviations below the mean value of normal young adults of the same sex is considered osteoporosis, and BMD reflected in the form of standard deviation values and T-values is used to define bone loss and osteoporosis. In this study, the diagnosis of osteoporosis was based on the fact that any of the T-values of the BMD values of the lumbar spine and the hip met the above criteria, and the diagnosis was divided into the osteoporosis group and the non-osteoporosis group.

2.2.3. Lipid Indicators testing

The blood lipids of the study subjects were also derived from the electronic medical record system, and the selected lipid indicators mainly included total cholesterol (TC), triglycerides (TG), low-density lipoprotein cholesterol (LDL-C), and high-density lipoprotein cholesterol (HDL-C). These indicators are important parameters for measuring the state of lipid metabolism in the human body.

2.3. Statistical Analysis

Continuous variables were described as mean ± standard deviation (SD) and analyzed analysis of
variance (ANOVA). Categorical variables were presented by number (percentage) and compared using the chi-square test. Two-sided P-value < 0.05 was statistically significant. All data analyses were performed using the IBM SPSS (version 26.0).

Before modeling the variables need to be screened, we performed a multicollinearity analysis of the characteristic variables, calculated the spearman correlation coefficients, and based on the results, the variables with strong correlation among them were removed.

The eligible participants (n=697) were randomly split into training set and test set in the ratio of 8:2. The data in the training set were used for model construction, in which the machine learning algorithms used were algorithms of Random Forest (RF), Light Gradient Boosting Machine (LightGBM), extreme gradient boosting (XGBoost) and categorical boosting (CatBoost). 5-fold cross-validation was used to train the model and grid optimization search was used to obtain the optimal parameters of the model. Accuracy, precision, specificity, F1-score, recall and area under the curve (AUC) of receiver operating characteristics were used to evaluate the model prediction performance.

Python (version 3.7) was used as the model development tool, the pandas library was utilized for data preprocessing, the sklearn library for algorithmic model construction.

3. Results

3.1. Demographic Characteristics and Lipid Indicators

A total of 697 middle-aged and elderly medical records were included in this study. The average age of the participants was 63.78±11.0 (years); the average BMI was 23.43±3.29 (kg/m²); the average TC was 4.71±1.08 (mmol/L); the average TG was 1.63±1.14 (mmol/L); the average HDL-C was 1.32±0.36 (mmol/L); the average LDL-C was 2.68±0.84 (mmol/L). There were 516 female participants, accounting for 74.03% of the study subjects; 129 participants who were still smoking, accounting for 18.51%; 107 participants who were still drinking, accounting for 15.35%. (Table 1)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (n=697, %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>63.78±11.00</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>23.43±3.29</td>
</tr>
<tr>
<td>TC (mmol/L)</td>
<td>4.71±1.08</td>
</tr>
<tr>
<td>TG (mmol/L)</td>
<td>1.63±1.14</td>
</tr>
<tr>
<td>HDL-C (mmol/L)</td>
<td>1.32±0.36</td>
</tr>
<tr>
<td>LDL-C (mmol/L)</td>
<td>2.68±0.84</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>516 (74.03)</td>
</tr>
<tr>
<td>Male</td>
<td>181 (25.97)</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>129 (18.51)</td>
</tr>
<tr>
<td>No</td>
<td>568 (81.49)</td>
</tr>
<tr>
<td>Drinking</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>107 (15.35)</td>
</tr>
<tr>
<td>No</td>
<td>590 (84.65)</td>
</tr>
</tbody>
</table>

3.2. Comparison of Characteristics Between Osteoporosis and Non-Osteoporosis Groups

Of all 697 subjects, a total of 228 cases (32.71%) were diagnose osteoporosis and 469 cases (67.29%) were diagnose non-osteoporosis. Compared to the non-osteoporosis group, middle-aged and elderly in the osteoporosis group had higher age (70.57±9.71 vs. 60.48±10.03), TC (4.84±1.06 vs. 4.64±1.09), and HDL-C (1.42±0.38 vs. 1.27±0.34). While BMI (22.51±3.39 vs. 23.88±3.16), TG (1.48±0.78 vs.1.70±1.27) were lower than that of the non-osteoporosis group. The female rate (92.1% vs. 65.2%) was higher in the osteoporosis group compared to the non-osteoporosis group. The smoking rate (13.2% vs. 23.9%) and drinking rate (11.2% vs. 20.3%) were lower in the osteoporosis group compared to the non-osteoporosis group. There were no significant differences between two groups in terms of diastolic blood pressure (DBP). There were no significant differences between two groups in terms of LDL-C (P>0.05). (Table 2)
Table 2: Comparison of Characteristics Between Osteoporosis and Non-Osteoporosis Groups.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Osteoporosis (n=228, %)</th>
<th>Non-Osteoporosis (n=469, %)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>70.57±9.71</td>
<td>60.48±10.03</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>22.51±3.39</td>
<td>23.88±3.16</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>TC (mmol/L)</td>
<td>4.84±1.06</td>
<td>4.64±1.09</td>
<td>0.027*</td>
</tr>
<tr>
<td>TG (mmol/L)</td>
<td>1.48±0.78</td>
<td>1.70±1.27</td>
<td>0.017*</td>
</tr>
<tr>
<td>HDL-C (mmol/L)</td>
<td>1.42±0.38</td>
<td>1.27±0.34</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>LDL-C (mmol/L)</td>
<td>2.72±0.84</td>
<td>2.66±0.83</td>
<td>0.365</td>
</tr>
<tr>
<td>Sex (Female)</td>
<td>210 (92.1)</td>
<td>306 (65.2)</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Smoking (Yes)</td>
<td>17 (13.2)</td>
<td>112 (23.9)</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Drinking (Yes)</td>
<td>12 (11.2)</td>
<td>95 (20.3)</td>
<td>&lt;0.0001***</td>
</tr>
</tbody>
</table>

* P-value < 0.05 ** P-value < 0.01 *** P-value < 0.001.

3.3. Screening of Modeling Feature Variables

In order to avoid the problem of multicollinearity of modeling variables, we calculated the spearman correlation coefficient between the feature variables and generated a heat map (Figure. 2). It can be seen that the spearman coefficients of gender and smoking, gender and drinking, smoking and drinking, and LDL-C and TC are greater than 0.5, and the correlation is strong. Based on the analysis of multiple dimensions, we removed the variables with strong correlation among them, and finally smoking, drinking and TC were excluded.

3.4. Lipid-based Prediction Models for Osteoporosis in Middle-Aged and Elderly People

The study subjects were divided into a training set (n = 559) and a test set (n = 138) according to a ratio of 8:2. The diagnose of osteoporosis was used as the dependent variable; sex, age, BMI, TG, HDL-C and LDL-C were used as independent variables. The optimal parameters of each model were obtained by grid optimization search, and the seed value was set to 1 to obtain the final four prediction models. Table 3 compared four prediction models through accuracy, precision, specificity, F1-score, recall and AUC. In the test set, the accuracy (0.83571) of the XGBoost model is significantly higher than that of RF (0.82142), CatBoost (0.80714), and LightGBM (0.79285), in addition to which the precision, specificity, F1-score and recall are also significantly higher than other models. The area under the curve (AUC) for LightGBM was 0.88320, slightly higher than the other three models, but the improvement was very insignificant. Figure. 3(a) illustrates the receiver operating characteristic (ROC) curves for the four models and Figure. 3(b) shows their calibration curves. Therefore, the XGBoost model was selected as the optimal model.
### Table 3: Comparison of Characteristics Between Osteoporosis and Non-Osteoporosis Groups.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Specificity</th>
<th>F1-score</th>
<th>Recall</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>0.82142</td>
<td>0.80555</td>
<td>0.92473</td>
<td>0.69879</td>
<td>0.61702</td>
<td>0.87851</td>
</tr>
<tr>
<td>LightGBM</td>
<td>0.79285</td>
<td>0.8</td>
<td>0.93548</td>
<td>0.62337</td>
<td>0.51063</td>
<td>0.88320</td>
</tr>
<tr>
<td>XGBoost</td>
<td>0.83571</td>
<td>0.85294</td>
<td>0.94623</td>
<td>0.71604</td>
<td>0.61702</td>
<td>0.87897</td>
</tr>
<tr>
<td>CatBoost</td>
<td>0.80714</td>
<td>0.8125</td>
<td>0.93548</td>
<td>0.65822</td>
<td>0.55319</td>
<td>0.87966</td>
</tr>
</tbody>
</table>

### Figure 3: (a) ROC Curves of Four Models. (b) Calibration Curves of Four Models

#### 4. Discussion

Osteoporosis seriously jeopardizes the life and health of the population, and its early clinical symptoms are not obvious, not easy to be detected and the course of the disease is complex, subject to the combined effects of multiple risk factors such as genetics and the environment. Because of this, clinically appropriate risk prediction mechanisms for osteoporosis are needed, with a focus on identifying risk factors for osteoporosis and low bone mass, screening high-risk groups, and achieving “early detection, early diagnosis, and early treatment”. The middle-aged and elderly population, as a high-risk group for osteoporosis, needs to be evaluated for risk. As a major city in central and western China, Chongqing’s sample of relevant medical studies is indeed partially representative of the general situation in such large and medium-sized cities, especially in terms of similar demographics and relevant features of the disease spectrum. Our study used clinical data from several hospitals in Chongqing, which is representative of the whole of China.

This is a novel study to develop a prediction model for osteoporosis based on simple personal information and lipid quadruple test. Our study used a total of four machine learning algorithms, CatBoost, Random Forest, XGBoost and LightGBM, to incorporate personal information and lipid quadruple test into the model to construct an osteoporosis risk prediction model for middle-aged and elderly people. By comparing the prediction results of the four models, the XGBoost algorithm has better discrimination and calibration, with an accuracy of 0.83571, and the calibration curve shows that the calibration curve of the XGBoost model is closest to the ideal curve. The order of importance of independent predictor variables based on the XGBoost algorithm is age, gender, BMI, LDL, TG and HDL-C. Lipid-related indicators played a part in the model. The XGBoost algorithm based on intrinsic interpretability also allowed for more accurate prediction of the risk of developing osteoporosis in individual study subjects.

The significance of the model is that the probability of developing osteoporosis can be obtained by using age, gender, BMI, TG, HDL-C, and LDL-C as input variables. Among these input variables, personal related information is easily available, and the indicators of blood lipids are frequently tested in routine medical checkups for middle-aged and elderly populations. Compared with the expensive DEXA method for bone density, the cheaper lipid test is more easily accepted by the middle-aged and elderly population. The model can be used as a tool for the initial determination of osteoporosis in the middle-aged and elderly population.

We found that the influencing factors of osteoporosis are gender, age, BMI and HDL-C, respectively. Numerous studies have shown that postmenopausal women are more susceptible to osteoporosis and that low BMD is not related to race [19][20]. The reason for this difference may be that men’s bone size tends to be larger than women’s as they grow, whereas women have better internal microstructure and less
opportunity for bone reconstruction. In addition, sex hormones play an important and dominant role in the physiological processes of bone in direct or indirect mechanisms, and the decrease in estrogen production during menopause in women ultimately leads to bone loss and osteoporosis. Moreover, the age of bone loss tends to be later in men compared to women. The older you are, the more likely you are to develop osteoporosis. The results of a study by Hiremath found that the risk increased by 20% for every 5 years increase in age. In addition, Nguyen TV et al. found that there was an age-related decline in BMD at the femoral neck in both sexes and at the lumbar spine in women. Between the ages of 60 and 80, the decrease in BMD at the femoral neck among women was 18.9%, which is almost twice the decrease in BMD among men (10.1%). In this study, a positive correlation was found between BMI and bone mineral density, and the greater the BMI, the less likely to suffer from osteoporosis. Some Meta-analyses showed that people with BMI below 18.5 had lower bone mineral density and higher risk of fracture. This may be due to the lack of fat which affects estrogen levels and accelerates bone loss, which in turn leads to osteoporosis.

There are some shortcomings in this study: the medical data platform from which the data were obtained has a lot of missing data, and the number of cases with complete information after data cleaning is only 697, which is representative but a small amount of data. In this study, four machine learning algorithms were used for predictive modeling of osteoporosis in middle-aged and elderly people, mainly considering the data situation and the interpretability of the model, other more and newer algorithms can be considered, which need to be further investigated.

5. Conclusion

The study found an association between lipid indices and osteoporosis in middle-aged and elderly people. Machine learning-based osteoporosis prediction model facilitates early identification of osteoporosis patients and mining risk factors, and timely preventive and control measures can reduce the incidence of the disease. Further studies are needed to validate our model for predicting the risk of osteoporosis in adults.

Acknowledgements

This work was supported by National Natural Science Foundation of China (No. 72174033), the Science and Technology Research Program of Chongqing Municipal Education Commission (No. KJZD-K202100403), and Program for Youth Innovation in Future Medicine, Chongqing Medical University (No. W0013)

References


