



Article

A Predictive Model for Abnormal Bone Density in Male Underground Coal Mine Workers

Ziwei Zheng, Yuanyu Chen, Yongzhong Yang, Rui Meng, Zhikang Si, Xuelin Wang, Hui Wang and Jianhui Wu *

Key Laboratory of Coal Mine Health and Safety of Hebei Province, School of Public Health, North China University of Science and Technology, No. 21 Bohai Avenue, Caofeidian New Town, Tangshan 063210, China; zhengzw@stu.ncst.edu.cn (Z.Z.); chenyuanyu@stu.ncst.edu.cn (Y.C.); yangyz@stu.ncst.edu.cn (Y.Y.); mengrui@stu.ncst.edu.cn (R.M.); sizhikang@stu.ncst.edu.cn (Z.S.); wangxuelin@stu.ncst.edu.cn (X.W.); wanghui@stu.ncst.edu.cn (H.W.)

* Correspondence: wujianhui555@ncst.edu.cn

Abstract: The dark and humid environment of underground coal mines had a detrimental effect on workers' skeletal health. Optimal risk prediction models can protect the skeletal health of coal miners by identifying those at risk of abnormal bone density as early as possible. A total of 3695 male underground workers who attended occupational health physical examination in a coal mine in Hebei, China, from July to August 2018 were included in this study. The predictor variables were identified through single-factor analysis and literature review. Three prediction models, Logistic Regression, CNN and XG Boost, were developed to evaluate the prediction performance. The training set results showed that the sensitivity of Logistic Regression, XG Boost and CNN models was 74.687, 82.058, 70.620, the specificity was 80.986, 89.448, 91.866, the F1 scores was 0.618, 0.919, 0.740, the Brier scores was 0.153, 0.040, 0.156, and the Calibration-in-the-large was 0.104, 0.020, 0.076, respectively, XG Boost outperformed the other two models. Similar results were obtained for the test set and validation set. A two-by-two comparison of the area under the ROC curve (AUC) of the three models showed that the XG Boost model had the best prediction performance. The XG Boost model had a high application value and outperformed the CNN and Logistic regression models in prediction.

Keywords: male underground coal mine workers; bone density abnormalities; XG Boost



Citation: Zheng, Z.; Chen, Y.; Yang, Y.; Meng, R.; Si, Z.; Wang, X.; Wang, H.; Wu, J. A Predictive Model for Abnormal Bone Density in Male Underground Coal Mine Workers. *Int. J. Environ. Res. Public Health* **2022**, *19*, 9165. <https://doi.org/10.3390/ijerph19159165>

Academic Editor: Paul B. Tchounwou

Received: 29 May 2022

Accepted: 25 July 2022

Published: 27 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Bone mineral density (BMD), known as bone mineral density, is an important indicator of bone strength, reflects the degree of osteoporosis, and is an important predictor of fracture risk [1]. BMD was classified according to World Health Organization standards: normal bone mass, reduced bone mass, and osteoporosis [2]. Osteoporosis (OP) is a systemic disease characterized by low bone mass and destruction of bone structure, abnormal BMD was the main cause of OP. At present, the incidence of OP had jumped to the 7th in common disease, and more than 200 million people worldwide suffered from osteoporosis [3]. The results of the October 2018 China Osteoporosis Epidemiology Survey showed that the prevalence of OP among people aged 40–49 years in China was 3.2%, including 2.2% for men and 4.3% for women, and the highest incidence among people aged 65 years or older was 32.0%. Globally, one osteoporotic fracture occurs every 3 s, and 50% of first-time osteoporotic patients will have another osteoporotic fracture [4]. Abnormal bone density had become an increasingly serious public health problem due to the accelerated ageing of society's population, abnormal bone density in a large number of people, and a general lack of awareness of bone density.

The coal industry is one of the economic pillar industries in China. There are a great number of people engaged in coal industry, their health status directly related to the development of China's coal industry. Some studies had pointed out that the special environment of underground coal mines had a significant effect on bone metabolism of

people who worked underground for long years [5]. The vast majority of coal miners are male. They worked in the dark, damp and relatively narrow working environment deep underground for a long time, and exposed to occupational harmful factors such as shifts, their risk factors for abnormal bone density differ from those of the general population. A study by the World Health Organization reported that the health and diseases of the population were caused by a variety of factors, including behavioral lifestyle, environmental factors, biogenetic factors and the quality of health care services, among which behavioral lifestyle was the most important influencing factor, accounting for 60%, and miners mostly had bad habits such as smoking, alcohol consumption and high-salt diet [6], which together led to underground workers' bone metabolic alterations. Currently, domestic and international studies on BMD abnormalities focused on the risk factors and pathogenesis of BMD decline [7–9], and there were fewer studies on its early prevention and risk assessment. In addition, domestic studies on BMD had mainly focused on the elderly and menopausal women [10], with fewer studies on BMD in coal miners. If coal miners at risk of abnormal bone density can be identified early, and changing their own unhealthy lifestyles. The number of abnormal BMD can be effectively reduced and the incidence of OP and osteoporotic fractures can be reduced.

Data Mining is the process of extracting knowledge and information with potential application value from large databases and is a new type of information processing system that had developed rapidly in recent years. Disease risk prediction is a very important task in data mining, which is to take the precondition of multiple pathologies of diseases, select multiple influencing factors of diseases, and use suitable statistical analysis methods to construct models so as to predict the probability of occurrence of certain diseases in groups or individuals with certain characteristics [11]. Commonly used models include Logistic regression, neural networks, decision trees, support vector machine (SVM), and so on. Each of these methods had its own characteristics and had been widely and successfully used in the medical field [12]. In recent years, many scholars had used data mining risk prediction methods in the medical field, and all of them had obtained better results. For example, the heart disease risk prediction model based on Convolutional neural network (CNN) by Jian Wang [13] had high prediction accuracy (89.89%) and can accurately predict the risk of heart disease development; Chao-Wen Tan [14] et al. applied convolutional neural networks to effectively improve the robustness and accuracy of heart sound signal classification, which was expected to be applied to machine-aided auscultation. Sethuraman [15] used a feed-forward neural network for feature selection and reduced the number of attributes to 12, which also helped the prediction model achieve 89.4% training accuracy and 82.2% test accuracy. Heydari et al. [6] compared neural network, SVM, decision tree and Bayesian methods in type II diabetes diagnosis and found that the neural network model had the highest accuracy. Tian-Pei Su [16] developed a diabetes risk prediction model based on the eXtreme Gradient Boosting (XG Boost) algorithm, and found that XG Boost was overall a better fit and more accurate than random forest. Hong-Xia Zhang [17] et al. established a prediction model for type II diabetes based on the XG Boost algorithm, which had a good prediction effect with an accuracy of 86.6%. In addition, most of these models currently developed are for disease risk assessment in the general population, and ignore the special groups in the occupational population. There were a large number of coal miners in China, and their special occupational environment such as high temperature, noise, shift work, and other occupational exposures can cause or affect the development of chronic diseases [18–22]. Therefore, prediction models for BMD abnormalities in the general population are not applicable to coal miners. To improve the quality of life and health status of coal miners, there is an urgent need to develop a new predictive model for the risk of BMD abnormalities in coal miners.

Based on the information from the physical examination data of underground coal mine workers, we developed three bone density abnormality prediction models: Logistic regression, CNN, and XG Boost. Overall, our study includes two contributions.

1. Based on the data information of 3695 underground coal mine workers' physical examinations, the risk factors of their BMD abnormalities were screened to provide a basis for the development of early prevention strategies for BMD abnormalities in underground coal mine workers.
2. The XG Boost model had better predictive performance of the three models. The XG Boost model can be used to predict the risk of BMD abnormalities in underground coal mine workers, so as to achieve early prevention of BMD abnormalities in underground coal mine workers.

2. Materials and Methods

2.1. Study Subjects

A total of 3695 male on-the-job underground workers who participated in occupational health physical examination from July to August 2018 in Gequan and Dongpang mines of Hebei Jizhong Energy were selected for the study. Inclusion criteria: age greater than or equal to 18 years; more than 1 year of service. Exclusion criteria: age greater than or equal to 60 years; those with incomplete information; those with congenital metabolic diseases affecting bone metabolism. All study subjects signed an informed consent form. The study was conducted in accordance with the Declaration of Helsinki and was reviewed and approved by the Ethics Committee of North China University of Technology (approval number: 15006).

2.2. Information Collection

Face-to-face questionnaires were administered to the study subjects by uniformly trained investigators, and the information collected included (1) demographic information: age, education level, marital status, BMI, per capita monthly household income, etc.; (2) lifestyle habits: smoking, alcohol consumption, sleep, physical exercise, etc.; (3) physical and laboratory examinations: bone mineral density, blood pressure, blood glucose, lipids, etc.; (4) exposure to occupational hazards: length of service, shift work, work intensity, etc.

2.3. Laboratory Tests

Fasting venous blood was collected by the doctors early in the morning from the study subjects. Blood specimens were sent to the hospital's Laboratory Department for blood biochemistry testing using Myriad Automatic Biochemistry Analyzer (BS-800).

2.4. Diagnostic Criteria for Abnormal Bone Density

The heel bone density of underground coal mine workers was measured using a CM-200 ultrasonic bone densitometer (FURONO, Japan). Bone density abnormalities were classified according to WHO standards [2].

Normal bone mass: $T \geq -1$; reduced bone mass: $-2.5 < T < -1$; osteoporosis: $T \leq -2.5$.

2.5. Variable Definitions

1. Hypertension: blood pressure: systolic blood pressure ≥ 140 mmHg and/or diastolic blood pressure ≥ 90 mmHg, or a previous history of hypertension and current use of antihypertensive medication was defined as hypertension according to the classification criteria of the Chinese Guidelines for the Prevention and Treatment of Hypertension, 2018 Revised Edition [23].
2. Diabetes: according to the classification criteria for glucose metabolic status in the Chinese guidelines for the prevention and treatment of type 2 diabetes (2017 edition) [24], fasting blood glucose ≥ 7.0 mmol/L, or a previous history of diabetes mellitus currently undergoing diabetes treatment was defined as diabetes.
3. Smoking: according to the WHO definition of several terms for smoking [25], smoking status was classified in this study as never smoked, quit smoking and current smoking. Smoking: smoked at least 1 cigarette per day and smoked continuously for more than

- 6 months; current smoking: was smoking at the time of this survey; never smoked quit: used to smoke but had stopped smoking for at least 6 months at the time of this survey.
4. Drinking: according to the definition of drinking by the Chinese Center for Disease Control and Prevention [26], drinking was classified in this study as never drinking, having stopped drinking, and now drinking. Never drinking: drinking at least once a week and drinking continuously for more than 6 months; Now drinking: drinking at the time of this survey; Already abstaining: used to drink but had stopped drinking for at least 6 months at the time of this survey.
 5. Shift situation: a working hour system that requires 24 h continuous work in the production process and is ensured by one or several teams working in shifts. This study classifies the shift situation into the following three cases, never shift, once shift now not shift, and now shift.
 6. Years of shift work: the sum of years of shift work performed, this study divided the years of shift work into 5 groups, 0, 0~, 10~, 20~, and more than 30 years.
 7. Exercise: exercise more than three times a week, no less than 30 min each time.
 8. Body mass index: $BMI = \text{weight (kg)}/\text{height}^2 \text{ (m}^2\text{)}$. The normal range of body weight is $BMI < 24 \text{ kg/m}^2$, the overweight range is $24.0 \text{ kg/m}^2 \leq BMI < 28.0 \text{ kg/m}^2$, and the obese range is $BMI \geq 28.0 \text{ kg/m}^2$.
 9. Dyslipidemia: according to the Chinese guidelines for the prevention and treatment of dyslipidemia in adults (revised 2016) [27], serum total cholesterol $\geq 6.2 \text{ mmol/L}$, and/or triglycerides $\geq 2.3 \text{ mmol/L}$, and/or LDL cholesterol $\geq 4.1 \text{ mmol/L}$, and/or HDL cholesterol $< 1.0 \text{ mmol/L}$, or a previous A history of hyperlipidemia and current use of lipid-lowering drugs defined as dyslipidemia.
 10. High-intensity operations: The physical activity of workers was investigated using the International physical activity questionnaire (IPAQ) (long-volume version) [28], and weekly total physical activity levels $\geq 3000 \text{ MET-min/w}$ were defined as high-intensity operations.
 11. Medium-intensity work: weekly total force activity level $\geq 600 \text{ MET-min/w}$.

2.6. Sample Size Calculation

To ensure that the model could accurately predict the mean of the outcome events, the prevalence of abnormal bone density \emptyset was reviewed in the literature and was approximately 25% [29], with the margin of error δ set at 0.05, which was calculated to require at least 289 study subjects. As shown in Equation (1).

$$n = \left(\frac{1.96}{\delta} \right)^2 \emptyset(1 - \emptyset) \quad (1)$$

To control for the minimum mean error of all individual predicted values, the mean absolute error MAPE was set to 0.05, the expected contraction rate R^2_{CS} was set to 0.1, and the predictor variable P was approximately 24, which was calculated to require at least 951 study subjects. As shown in Equation (2).

$$n = \exp \left(\frac{-0.508 + 0.259 \ln(\emptyset) + 0.504 \ln(P) - \ln(\text{MAPE})}{0.544} \right)^2 \quad (2)$$

To ensure an expected contraction rate of 10% and to reduce model overfitting, S was set to 0.1 and the number of study variables P was approximately 24, which was calculated to require at least 2038 study subjects. As shown in Equation (3).

$$n = \frac{P}{(S - 1) \ln \left(1 - \frac{R^2_{CS}}{S} \right)} \quad (3)$$

To ensure the minimum difference between the developed model and the R^2_{CS} optimization adjustment value, R^2_{CS} in Equation (4) is 0.1 and $\max R^2_{CS}$ is 0.65, and S' is calculated to be 0.75, which is calculated to require at least 671 study subjects. As shown in Equations (4) and (5).

$$n = \frac{R^2_{CS}}{R^2_{CS} + \delta \max R^2_{CS}} \quad (4)$$

$$n = \frac{P}{(S' - 1) \ln \left(1 - \frac{R^2_{CS}}{S'} \right)} \quad (5)$$

It was calculated that a minimum of 2038 individuals needed to be included. A total of 3695 individuals were included in the study, and the sample size met the needs of the study.

2.7. Statistical Methods

Excel 2016 was used to establish the original database, and IBM SPSS24.0 was used for statistical analysis. Count data were described by rate or composition ratio, and χ^2 test was used for comparison between groups; Ordinal data were described by rate or constituent ratio, and the Kruskal-Wallis test was used for comparison between groups. measurement data obeying normal distribution were described by mean and standard deviation, and non-normally distributed data were expressed as median and quartiles; multi-factor unconditional Logistic was used for multivariate analysis of influencing factors. The test level $\alpha = 0.05$.

2.8. Software and Hardware Platform

The sample data are randomly selected in the ratio of 7:2:1 to divide the training set, test set and validation set. Dataset partition codes as shown in Supplementary Material S1.

2.8.1. Logistic Regression Model

The Logistic regression model was built using the sklearn package. The Logistic model codes as shown in Supplementary Material S2.

2.8.2. CNN Model

Convolutional neural network mainly consists of input layer, convolutional layer, pooling layer, fully connected layer and output layer. In this study, the CNN model was constructed by numpy package. The input layer included 2 nodes, both hidden layers contained 5 nodes, and the output layer included 1 node. The sigmoid function was used as the excitation function, softmax was used for probability normalization, cross-entropy loss function was used as the loss function, and stochastic gradient descent was used as the optimizer. The model was trained by updating the network parameters according to the computed gradients. The CNN model codes as shown in Supplementary Material S3.

2.8.3. XG Boost Model

The XG Boost model codes as shown in Supplementary Material S4. Based on the m eigenvalues and the n sample data, the XG boost prediction model was obtained by the following equation:

$$\hat{y}_i = \varphi(X_i) = \sum_{k=1}^K f_k(X_i), f_k \in F \quad (6)$$

In the formula:

i -number of total physical examination data samples

K -total number of trees

f_k -the k tree

Once the predicted values were obtained, the objective function was obtained by the following equation.

$$O_{bj}(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega f_k \tag{7}$$

In the formula:

$l(y_i, \hat{y}_i)$ -the training error of the sample x_i

Ωf_k -the regular term of the k tree

$$\Omega f_t = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \tag{8}$$

In the formula:

T-number of leaf nodes

w -score values of leaf nodes

γ -parameter to balance the complexity of the model

λ -Parameter to balance the complexity of the model

XG Boost was an additive model in which the objective function changes each time a tree was added to the model. With an additive strategy, the objective function can be written as:

$$O_{bj}(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^t \Omega f_k = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) + C \tag{9}$$

A Taylor expansion of the loss function via Equation (9) was viewed as, and the final loss function can be written as:

$$f(x + \Delta x) \approx f(x) + f'(x)\Delta x + \frac{1}{2} f''(x)\Delta x^2 \tag{10}$$

$$O_{bj}(\theta) \approx \sum_{i=1}^n \left(g_i f_i(x_i) + \frac{1}{2} h_i f_i(x_i)^2 \right) + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 + C \tag{11}$$

2.9. Model Evaluation

The model prediction effect was evaluated comprehensively in terms of both discrimination and calibration. As shown in Table 1.

Table 1. Model evaluation indexes.

Indicators	Meaning
Sensitivity	The percentage of study participants who actually had BMD and were accurately determined to have BMD by the risk prediction model.
Specificity	The percentage of study participants who did not actually have BMD and were accurately determined to not have BMD by the risk prediction model.
Youden index	Correctness Index, the model correctly determined the total capacity of BMD patients and non-patients.
F1 score	The adjusted mean values of precision and recall, used to evaluate the comprehensive performance of the model.
AUC	Area under the ROC curves.
Brier score	The quantitative score of the model calibration, ranging from 0 to 0.25, the smaller the value, the better the calibration of the model.
Log loss	The error between the true value of the response and the predicted value of the model.
Calibration-in-the-large	The intercept of the calibration curve.

2.10. Quality Control

The surveyors were trained by a unified induction, and the questionnaires were checked three times after recovery, and the entry system was double-entry to ensure the accuracy of the information. Measuring instruments were maintained and regularly calibrated by dedicated personnel. 1~2 workers were randomly selected, and the second measurement was made by the surveyor each day, and the results were compared to ensure the consistency of the measurement results. Ten workers were randomly selected to use CM-200 ultrasonic bone densitometer and QCT bone densitometer for bone density measurement, and the QCT bone densitometer measurement results were used to calibrate the CM-200 ultrasonic bone densitometer to ensure the accuracy of bone density measurement results.

3. Results

3.1. General Demographic Characteristics

The prevalence of BMD abnormalities was 28.25% among the 3695 male coal mine workers included in the study. Analysis of the age of the study subjects revealed that the age range of the study subjects was (19–59) years, with a mean age of (39.04 ± 8.41) years. The results showed that the differences in the prevalence of BMD abnormalities among male coal mine workers were statistically significant ($p < 0.05$) across age, education level, BMI, marital status, fracture, smoking status, drinking status, exercise and Sleep time (h), and not statistically significant ($p > 0.05$) across income, diabetes, hypertension and dyslipidemia groups. As shown in Table 2.

Table 2. General situation of male workers in coal mines.

General Information	Category	Number	Abnormal Bone Mineral Density		$\chi^2/H(K)$	p
			Number	Prevalence Rate (%)		
Age	<30	419	30	7.160	447.518 *	<0.001
	30~	1682	303	18.014		
	40~	945	354	37.460		
	50~	649	357	55.008		
Education level	Junior secondary school or lower	1647	507	30.783	14.956	0.001
	High school and secondary school	1083	308	28.440		
	College and above	965	229	23.731		
BMI (kg/m ²)	≤23.9	1321	516	39.061	180.677 *	<0.001
	24.0~	1470	411	27.959		
	28.0~	904	117	12.942		
Marital Status	Unmarried	126	24	19.048	6.312	0.043
	Married	3445	980	28.447		
	Other	124	40	32.258		
Family per capita monthly income (Yuan)	<1000	432	124	28.704	4.526 *	0.104
	1000~	2947	847	28.741		
	3000~	316	73	23.101		
Hypertension	No	2379	648	27.238	3.402	0.065
	Yes	1316	396	30.091		
Diabetes	No	3531	990	28.037	1.848	0.174
	Yes	164	54	32.927		
Dyslipidemia	No	2685	748	27.858	0.760	0.383
	Yes	1010	296	29.307		
Fracture	No	2817	694	24.636	76.564	<0.001
	Yes	878	350	39.863		

Table 2. Cont.

General Information	Category	Number	Abnormal Bone Mineral Density		$\chi^2/H(K)$	<i>p</i>
			Number	Prevalence Rate (%)		
Smoking status	No smoking	1460	301	20.616	69.733	<0.001
	Quit smoking	245	78	31.837		
	smoking	1990	665	33.417		
Drinking status	No drinking	724	148	20.442	32.053	<0.001
	Alcohol withdrawal	164	37	22.561		
	Drinking	2807	859	30.602		
Exercise	No	1574	501	31.830	17.291	<0.001
	Yes	2121	543	25.601		
Sleep time (h)	<7	1099	419	38.126	136.080 *	<0.001
	7~	1236	387	31.311		
	8~	1360	238	17.500		

* The K-W test was used for ordinal data.

3.2. Analysis of Occupational Hazardous Factors and Prevalence of Bone Density Abnormalities

The analysis of occupational hazardous factors in 3695 workers showed that the prevalence of BMD abnormalities in workers increased with working ages and shift length. There were statistically significant differences ($p < 0.05$) between groups of working ages, shift conditions, shift length, high intensity work and medium intensity work. As shown in Table 3.

Table 3. Analysis of occupational exposure characteristics of male workers in coal underground.

General Information	Category	Number	Abnormal Bone Mineral Density		$\chi^2/H(K)$	<i>p</i>
			Number	Prevalence Rate (%)		
Working ages	<10	1089	216	19.835	122.264 *	<0.001
	10~	1652	434	26.271		
	20~	539	216	40.074		
	30~	415	178	42.892		
Shift situations	Never	1341	204	15.213	254.064	<0.001
	Once	547	114	20.841		
	Now	1807	726	40.177		
Shift length	0	1341	204	15.213	228.767 *	<0.001
	<10	1098	384	34.973		
	10~	867	262	30.219		
	20~	229	111	48.472		
High intensity work	No	1502	356	23.702	25.876	<0.001
	Yes	2193	688	31.373		
Medium intensity work	No	437	70	16.018	36.606	<0.001
	Yes	3258	974	29.896		

* The K-W test was used for ordinal data.

3.3. Logistic Regression Analysis of Risk Factors for Abnormal Bone Density

It had been reported in the literature [30] that hypertension and diabetes had an effect on BMD, so these indicators were also included in the model building. The variable assignments are shown in Table 4.

Table 4. Assignment table for variables.

Variable	Variable Meaning	Assignment Method
Y	Bone mineral density	1 = normal, 2 = abnormal
X ₁	Age	1 = <30; 2 = 30~; 3 = 40~; 4 = ≥50
X ₂	Educational level	1 = Junior secondary school or lower; 2 = High school and secondary school; 3 = College and above
X ₃	BMI (kg/m ²)	1 = ≤23.9; 2 = 24.0~; 3 = ≥28.0
X ₄	Marital Status	1 = Unmarried; 2 = Married; 3 = Other
X ₅	Hypertension	1 = No; 2 = Yes
X ₆	Diabetes	1 = No; 2 = Yes
X ₇	Fracture	1 = No; 2 = Yes
X ₈	Smoking status	1 = No smoking; 2 = Quit smoking; 3 = smoking
X ₉	Drinking status	1 = No drinking; 2 = Alcohol withdrawal; 3 = Drinking
X ₁₀	Exercise	1 = No; 2 = Yes
X ₁₁	Sleep time (h)	1 = <7; 2 = 7~; 3 = ≥8
X ₁₂	Working age	1 = <10; 2 = 10~; 3 = 20~; 4 = ≥30
X ₁₃	Shift situation	1 = Never; 2 = Once; 3 = Now
X ₁₄	Shift length	1 = 0; 2 = <10; 3 = 10~; 4 = 20~; 5 = ≥30
X ₁₅	High intensity work	1 = No, 2 = Yes
X ₁₆	Medium intensity work	1 = No, 2 = Yes

All independent variables were tested for collinearity diagnostics and found to be free of multicollinearity, as shown in Table 5.

Table 5. Multicollinearity of independent variables.

Variable	Tolerance	VIF
Age	0.402	2.489
Educational level	0.806	1.241
BMI (kg/m ²)	0.879	1.138
Marital Status	0.945	1.058
Hypertension	0.899	1.112
Diabetes	0.937	1.067
Fracture	0.985	1.016
Smoking status	0.938	1.066
Drinking status	0.962	1.039
Exercise	0.932	1.073
Sleep time (h)	0.930	1.075
Working age	0.331	3.021
Shift situation	0.324	3.090
Shift length	0.279	3.589
High intensity work	0.890	1.124
Medium intensity work	0.908	1.101

The results of the multifactorial analysis showed that age, low level of education, diabetes, hypertension, fractures, smoking status, drinking status, shift situation, high intensity work and medium intensity work were all risk factors for abnormal bone mineral density, with BMI, physical activity and Sleep time (h) as protective factors. As shown in Table 6.

3.4. Bone Density Abnormalities Models Construction and Evaluation

Based on the results of the multifactorial analysis of factors, we included in the model 13 independent variables that were significant for the multifactorial analysis, including age, education, BMI, diabetes, hypertension, fracture, smoking, alcohol consumption, shift work status, heavy workload, moderate workload, exercise, and sleep duration. The sample data were partitioned, with 70% of the training set, 20% of the test set and 10% of the validation set, to construct Logistic, CNN and XG Boost models.

Table 6. Logistic regression analysis of the influencing factors of abnormal bone mineral density.

Variable	B	S.E	Wald	p	OR	95% CI for OR	
						Lower	Upper
Age							
<30			353.284	<0.001			
30~	2.272	0.261	75.697	<0.001	9.699	5.814	16.182
40~	4.145	0.299	191.881	<0.001	63.096	35.101	113.419
≥50	5.779	0.349	273.603	<0.001	323.510	163.112	641.635
Education level							
College and above			32.054	<0.001			
High school and secondary school	0.222	0.117	3.594	0.058	1.249	0.993	1.571
Junior secondary school or lower	0.775	0.137	32.003	<0.001	2.171	1.660	2.841
BMI (kg/m ²)							
≤23.9			353.668	<0.001			
24.0~	-1.784	0.124	207.232	<0.001	0.168	0.132	0.214
≥28.0	-2.978	0.165	324.298	<0.001	0.051	0.037	0.070
Hypertension	0.243	0.105	5.327	0.021	1.275	1.037	1.567
Diabetes	0.502	0.225	4.990	0.025	1.652	1.064	2.567
Fractures	0.736	0.109	45.982	<0.001	2.087	1.687	2.582
Smoking status							
No smoking			37.743	<0.001			
Quit smoking	0.601	0.194	9.620	0.002	1.825	1.248	2.668
Smoking	0.646	0.107	36.522	<0.001	1.908	1.547	2.353
Drinking status							
No drinking			43.725	<0.001			
Alcohol withdrawal	-0.136	0.277	0.243	0.622	0.872	0.507	1.501
Drinking	0.780	0.132	34.883	<0.001	2.182	1.684	2.827
Exercise	-0.322	0.100	10.294	0.001	0.725	0.595	0.882
Sleep time (h)							
<7			89.013	<0.001			
7~	-0.242	0.114	4.506	0.034	0.785	0.628	0.982
≥8	-1.159	0.128	82.117	<0.001	0.314	0.244	0.403
Shift situation							
Never			181.498	<0.001			
Once	-0.663	0.319	4.314	0.038	0.516	0.276	0.963
Now	1.356	0.271	25.079	<0.001	3.879	2.282	6.593
High intensity work	0.600	0.107	31.590	<0.001	1.822	1.478	2.245
Medium intensity work	1.020	0.176	33.715	<0.001	2.774	1.966	3.915
Constant quantity	-6.318	0.430	215.395	<0.001	0.002	-	-

The results of training set sample of 2586 cases (70%) showed that the sensitivity, Youden index, F1 score, AUC (95% CI), Brier score, Log loss, and Calibration-in-the-large of the XG Boost model were 82.058%, 0.715, 0.919, 0.858 (0.839~0.876), 0.040, 0.147, and 0.020, better than the other two models. The CNN model had a better specificity of 91.866%. The Logistic regression model performed worse. As shown in Table 7.

Table 7. Evaluation of three risk models.

Evaluation Indicator	Training Set			Test Set			Validation Set		
	Logistic	XG Boost	CNN	Logistic	XG Boost	CNN	Logistic	XG Boost	CNN
Sensitivity (%)	74.687	82.058	70.620	71.749	76.555	68.447	73.529	76.724	68.750
Specificity (%)	80.986	89.448	91.866	80.814	88.302	76.923	77.239	85.827	74.818
Youden index	0.557	0.715	0.625	0.526	0.649	0.454	0.508	0.626	0.436
F1 Score	0.618	0.919	0.740	0.631	0.753	0.571	0.583	0.787	0.600
AUC (95% CI)	0.778 (0.757~0.799)	0.858 (0.839~0.876)	0.812 (0.792~0.833)	0.763 (0.723~0.802)	0.824 (0.787~0.861)	0.727 (0.685~0.769)	0.754 (0.696~0.811)	0.813 (0.762~0.864)	0.718 (0.656~0.779)
Brier Score	0.153	0.040	0.156	0.333	0.107	0.172	0.153	0.040	0.156
Log Loss	0.540	0.147	0.492	1.124	0.358	0.538	0.540	0.147	0.494
Calibration-in-the-large	0.104	0.020	0.076	0.104	0.019	0.071	0.146	0.019	0.077

The results of test set sample of 739 cases (20%) showed that the XG Boost model had a sensitivity, specificity, Youden index, F1 score, AUC (95% CI), Brier score, Log loss, and Calibration-in-the-large of 76.555%, 88.302%, 0.649, 0.753, and 0.824(0.787~0.861), 0.107, 0.358, and 0.019, better than the other two models. As shown in Table 7.

The results of test set sample of 370 cases (10%) showed that the sensitivity, specificity, Youden index, F1 score, AUC (95% CI), Brier score, Log loss, and Calibration-in-the-large of the XG Boost model were 76.555%, 85.827%, 0.626, 0.787, and 0.813 (0.762~0.864), 0.040, 0.147, and 0.019, better than the other two models. As shown in Table 7.

A two-by-two comparison of the area under the ROC curves (AUC) of the three models showed that the XG Boost model had the best prediction performance, followed by the CNN model, and the Logistic model had the worst prediction performance, with the differences being statistically significant ($p < 0.017$). The test set results showed that the XG Boost model had the best prediction performance, and the differences were all statistically significant ($p < 0.017$). The results of the validation set showed that the XG Boost model outperformed the CNN model, and the differences were statistically significant ($p < 0.017$); the results of the test and validation sets showed that the differences in prediction performance between the Logistic and CNN models were not statistically significant ($p > 0.017$). As shown in Table 8 and Figure 1.

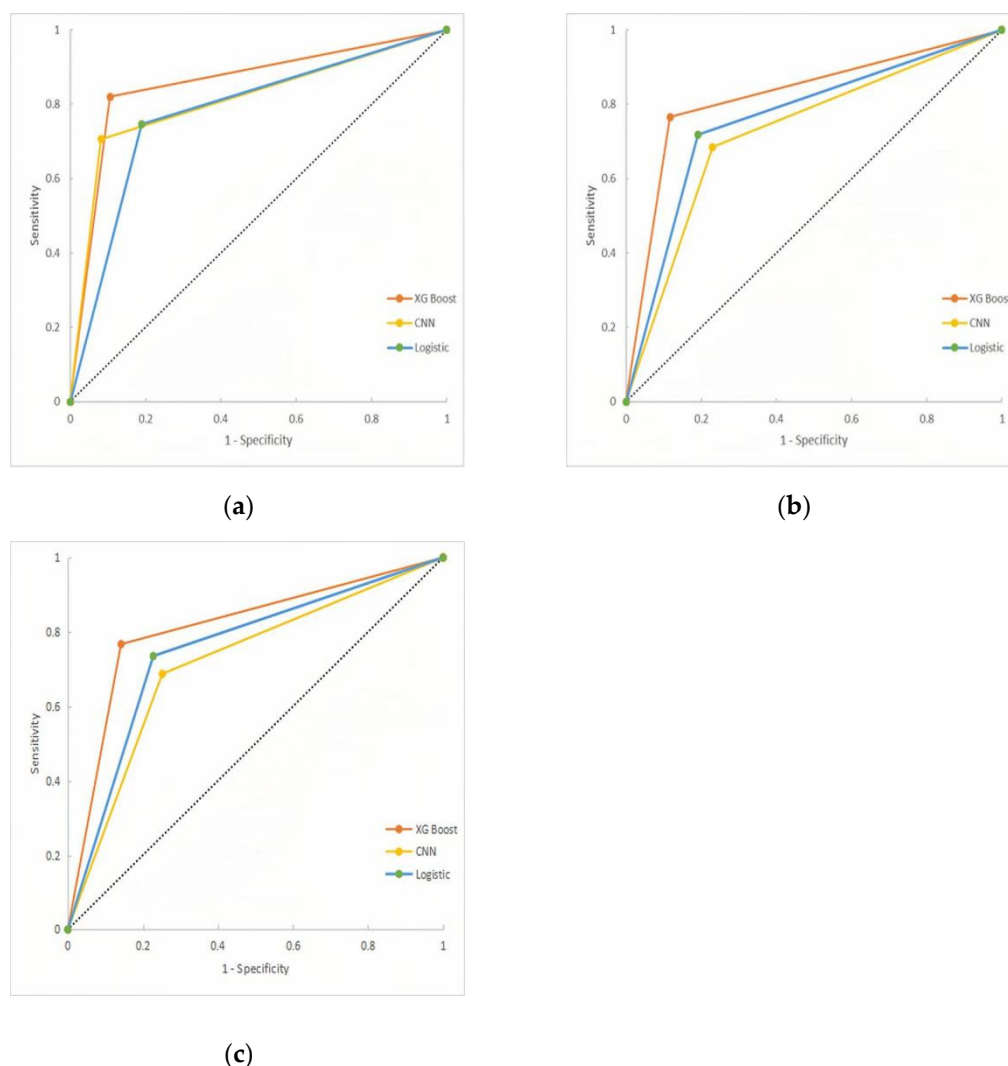


Figure 1. ROC curves of three models (a) Training set; (b) Test set; (c) Validation set.

Table 8. AUC comparison of three models.

Data Set	Model	Difference Value of AUC	SE	95% CI		χ^2	<i>p</i>
				Lower	Upper		
training set	Logistic and XG Boost	0.071	0.008	0.055	0.087	8.715	0.001
	Logistic and CNN	0.019	0.007	0.006	0.032	2.886	0.004
	XG Boost and CNN	0.052	0.008	0.036	0.068	6.256	0.001
test set	Logistic and XG Boost	0.074	0.016	0.042	0.106	4.545	<0.001
	Logistic and CNN	0.022	0.019	−0.016	0.060	1.154	0.248
	XG Boost and CNN	0.096	0.020	0.058	0.135	4.923	<0.001
validation set	Logistic and XG Boost	0.039	0.025	−0.009	0.088	1.580	0.114
	Logistic and CNN	0.047	0.022	0.003	0.090	2.110	0.035
	XG Boost and CNN	0.086	0.020	0.047	0.125	4.281	<0.001

The XG Boost model Brier Score, Log Loss, and Calibration-in-the-large metrics all outperformed the CNN and Logistic regression models. The calibration curves for the training, test and validation sets were all close to the diagonal, with no serious deviations in the results, and the calibration curves for the Logistic regression model were more deviant. As show in Figure 2a–c.

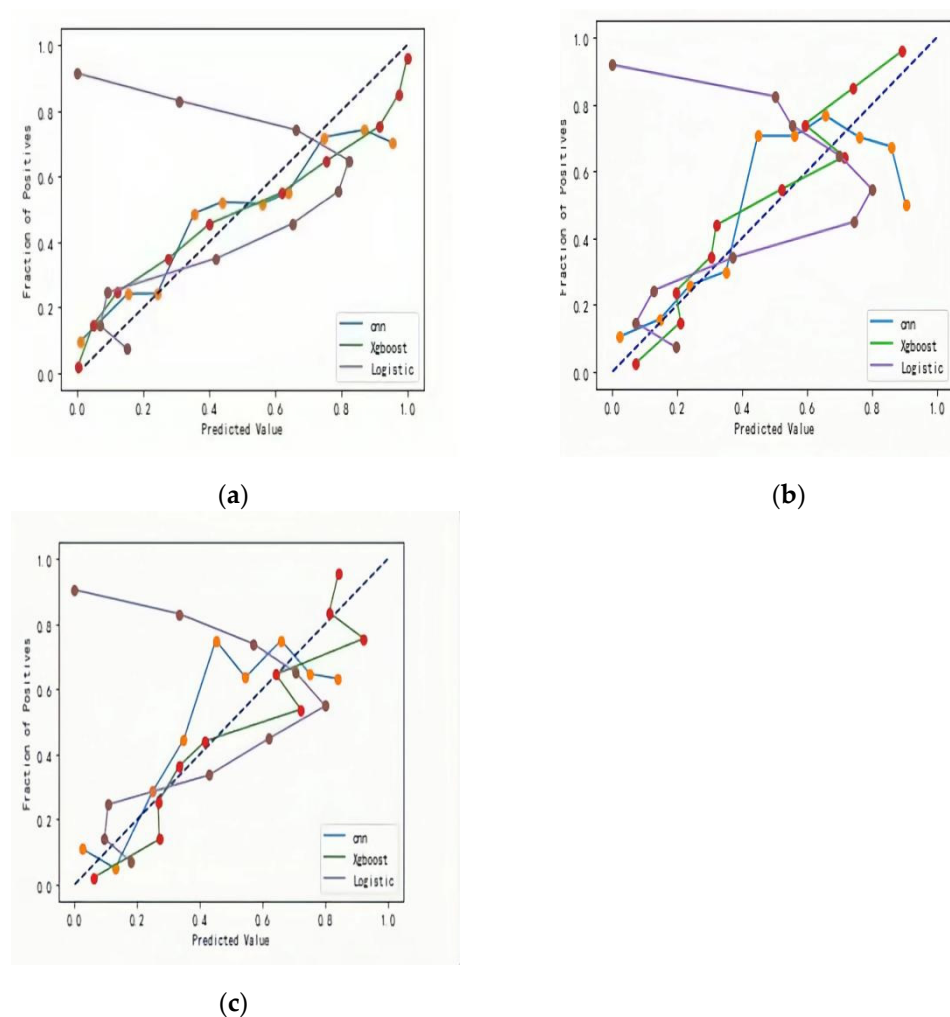


Figure 2. Calibration curves of three models (a) Training set; (b) Test set; (c) Validation set.

In summary, the XG Boost model had excellent performance in the training set, test set and validation set, its predicted risk was in good agreement with the actual occurrence

of risk, and each evaluation indexes were significantly better than the CNN model and the Logistic regression model. The XG boost model was the optimal model for this study.

4. Discussion

Building risk prediction models was important for early identification and intervention of diseases. Early detection, diagnosis and treatment can contribute to tertiary prevention strategies for the disease. Machine learning had shown advantages in disease models. Meng D et al. conducted a machine learning study on the incidence of hand, foot and mouth disease in all provinces of mainland China and found that the predictive ability of the XG Boost model was generally better than that of the random forest model [31]. Li Z et al. constructed a novel predictive model integrating GCN, CNN and squeeze inspired network (GCSENet) for identifying miRNA-disease associations. By applying the three models together, Li Z et al. obtained an AUC of 0.950 and an F1 score of 0.864, which satisfactorily predicted miRNA disease relevance [32]. Workers working underground in coal mines were susceptible to the effects of their environment on bone metabolism [33], and the factors affecting abnormal bone density were different from those in the general population. Therefore, early identification of high-risk groups and strict control of their influencing factors can reduce the incidence of BMD abnormalities. In this study, 3695 male underground coal mine workers were investigated and their BMD abnormalities were found to be influenced by various factors. By constructing Logistic regression, CNN and XG boost risk prediction models, a comparison of the prediction performance of the three models revealed that the XG boost risk prediction model was the best prediction performance model in this study.

The results of this study showed that the rate of abnormal bone density in a coal mine worker was 28.25%. Based on the importance of the predictor variables in the three models developed, we found that the top four variables were age, BMI, shift work and sleep duration, indicating that these four factors play a very important role in the occurrence of BMD abnormalities. In this study, advanced age was found to be a risk factor for BMD abnormalities, which was consistent with previous findings [34]. The reasons for this might be: decrease in estrogen production with age, which in turn affected parathyroid hormone levels, affecting bone reconstruction and loss of bone mass; the level of secretion decreases with age, inducing increased osteoclast activity and reduced bone content. However, due to the limitation of research conditions, retired workers over 60 years of age were not included in the scope of the study, which may lead to deviations in the research results. Subsequent research can be conducted on retired workers to evaluate the effect of age on bone mineral density of coal mine workers. The detection rate of abnormal bone density was lowest when BMI was between 24.0 and 27.9, which was consistent with previous studies [35]. BMI affected bone density probably because: adipose tissue increased the body's estrogen content, which favored bone formation; the increased mechanical load on the skeleton at higher BMI can promote bone formation. The risk of abnormal bone density was 1.356 times higher for shift workers compared to workers without shift work. Possible reasons for this were: shift work broke the normal work and rest schedule of workers, making them prone to circadian rhythm disorders and metabolic disorders, and shift work was also associated with the development of diseases such as sleep, hypertension, diabetes and obesity, which can indirectly affect bone density [36]. In terms of sleep duration, this study found that the risk of abnormal BMD in the study participants in the ≥ 8 h sleep group was 0.561 times lower than that in the < 7 h sleep group, which was consistent with the results of previous studies [37]. This may be due to the reduced sleep time of workers, especially at night due to shift work, and exposure to light at night, which reduced melatonin secretion, which can reduce bone mass and also disrupted body metabolism, which in turn affected BMD [38]. The present study found that physical activity was a protective factor for abnormal BMD in male coal mine workers with an OR of 0.725 (95% CI: 0.595, 0.882), similar to the findings of Anupama DS et al. [39]. Hauger AV et al. found that high physical activity was positively associated with total hip bone mineral density compared to a sedentary lifestyle [40]. The

present study found that the risk of BMD occurrence was lower in workers with high intensity work than in workers with moderate intensity work, which may be due to skeletal muscle contraction activating bone biomodulation mechanisms that enhance BMD to adapt to exercise load [41]. The results of numerous studies had concluded that smoking had a negative impact on human bone. The results of the present study showed that the risk of abnormal BMD was 1.908 (95CI: 1.547, 2.353) times higher in workers who smoked compared to those who never smoked, which was consistent with the results of previous studies [42–44]. The possible reasons for this were that tobacco can affect the production and metabolism of estrogen and androgen, affect the activity of osteoblasts and osteoclasts, and inhibit the vitamin D-parathyroid hormone axis, which can have a negative impact on bones [45]. Adequate vitamin D increases intestinal calcium absorption, promotes bone mineralization, maintains muscle strength, improves balance, and reduces the risk of falls. Vitamin D deficiency can lead to secondary hyperparathyroidism, which increases bone resorption and thus causes or exacerbates osteoporosis. Concurrent calcium and vitamin D supplementation may reduce the risk of osteoporotic fractures. Vitamin D insufficiency also affects the efficacy of other anti-osteoporosis drugs [46]. The current study found that the risk of BMD abnormalities in workers who consumed alcohol was 2.182 (95% CI: 1.684, 2.827) times higher than in non-drinkers, suggesting that alcohol consumption was a risk factor for BMD abnormalities, which was consistent with the findings of some studies [47]. However, some studies had also suggested that moderate alcohol consumption was a protective factor for BMD abnormalities and that it was heavy alcohol consumption that led to BMD abnormalities [48]. Previous studies have been inconsistent regarding the relationship between education level and BMD, and Yan Ren's analysis of hip fracture risk factors in middle-aged and elderly Chinese found that elderly people with lower education levels were at high risk of fracture [49]. In contrast, Chen et al. [50] observed a positive association between education level and the risk of hip fracture in postmenopausal women in Taiwan. The current study found a high prevalence of BMD in workers with higher education levels. The relationship between education level and BMD needs to be further explored. No consensus conclusion had been reached regarding the effect of alcohol consumption on BMD, which may be related to factors such as study population selection, alcohol intake, frequency of alcohol consumption and type of alcohol consumption.

Logistic regression models were widely used in the field of risk factor screening and disease prediction. It was easy to use and had clear parameter meanings, but the predictive power of Logistic regression models decreased when the data do not meet the requirements [51]. Yan X et al. applied Logistic regression to build an osteoporosis risk model and performed internal and external validation, and the results showed that the C-index was 0.947 for internal validation and 0.946 for external validation, and the calibration curve showed a good agreement between predicted and actual probabilities [52]. In this study, there were limitations in applying the Logistic regression model to BMD prediction for underground workers in coal mines, and all three data sets performed poorly in terms of calibration index, suggesting that the consistency between the predicted and actual results of the Logistic regression model was not high, and that it was prone to bias when used for BMD risk prediction. The CNN model was a deep learning method of multi-layered networks, including input layer, convolutional layer, pooling layer, fully connected layer and output layer. The number of network parameters was effectively reduced and the computational complexity was greatly reduced. It had been used as a neural network model to predict the risk of various diseases in recent years [53,54], but the prediction effect of CNN on different diseases was unstable. For example, Dai G et al. used CNN model to explore the effect of hypertension on the retinal microvascular system, and the results were not satisfactory, with a sensitivity of 60.94%, a specificity of 51.54% and an AUC of 0.6506, which may be due to the fact that the model construction needed to be further improved [55]. Jiang J et al. used a CNN model to identify the degree of left atrial enlargement and showed that the AUCs for normal, mild and moderately severe left atrial enlargement ECGs were 0.942, 0.951 and 0.998 respectively [56]. In this study, the CNN

model performed better than the Logistic regression model in terms of calibration metrics, but its ability to distinguish between abnormal and non-abnormal BMD was inferior to the XG Boost model and was not the preferred choice for predicting the risk of abnormal BMD in underground coal mine workers. XG Boost was an improvement of the boosting algorithm based on GBDT, which performed a second-order loss function Taylor expansion, which allowed for higher accuracy. The inclusion of a regular direction in the objective function made the trained model simpler and can effectively combat over-fitting [57]. In this study, the XG Boost model not only had a high ability to distinguish between BMD anomalies and non-anomalies but also had the highest agreement between the prediction results and the actual results, and had the best fit with the BMD anomaly data information of underground workers in coal mines, which can be used for the prediction of BMD anomalies of underground workers in coal mines.

In addition, there were limitations in this study. Our study did not measure vitamin and D levels in underground coal mine workers, did not investigate the type and location of fractures, and did not take into account the medication status of the study subjects. This study only established and completed the internal validation of the risk prediction model for abnormal bone density in coal miners, and no external validation was conducted. There were many methods and sites of bone density measurement, and only one was selected for this study. Moreover, this study was a cross-sectional study, and only information on the prevalence of BMD abnormalities in coal miners was obtained, and the causality argument was less effective, and further cohort studies could be conducted for discussion.

5. Conclusions

In this study, the data related to abnormal BMD in male underground coal mine workers were analyzed and found that age ≥ 30 years, Junior secondary school or lower, diabetes, hypertension, fracture, smoking, drinking, shift work, BMI ≥ 28 kg/m², high intensity work and medium intensity work were risk factors. Exercise and sleep time ≥ 7 h were protective factors for bone density abnormalities.

The XG Boost model outperformed the CNN and Logistic regression models in prediction.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijerph19159165/s1>, Material S1: The dataset partition codes; Material S2: The Logistic model codes; Material S3: The CNN model codes; Material S4: The XG Boost model codes.

Author Contributions: Data collection, Z.Z., Y.C. and Y.Y.; statistical processing of data, Z.Z., R.M., Z.S., X.W. and H.W.; writing-original manuscript preparation, Z.Z.; quality control and revision, J.W. All authors had read and agreed to the published version of this manuscript.

Funding: This work was supported by the National Key R&D Program of China (No.2016YFC0900605).

Institutional Review Board Statement: The study was reviewed and approved by the Ethics Committee of North China University of Technology (approval number: 15006).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data available on request due to restrictions privacy. The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the data are not readily available.

Acknowledgments: This study acknowledges all researchers in the design and operation.

Conflicts of Interest: This article has no conflict of interest.

References

1. Lewiecki, E.M. Assessment of Skeletal Strength: Bone Density Testing and Beyond. *Endocrinol. Metab. Clin. N. Am.* **2021**, *50*, 299–317. [[CrossRef](#)] [[PubMed](#)]
2. Meng, X.W. *Clinical Diagnosis and Treatment Guidelines, Osteoporosis and Bone Mineral Diseases*; People's Health Publishing House: Beijing, China, 2006; pp. 1–7.
3. Zhang, B.; Xu, C.; Peng, Y. Research progress on prevention of osteoporosis based on Mediterranean diet. *Chin. J. Endocrinol. Metab.* **2021**, *37*, 584–588.
4. Ling, X.; Huang, G.Y.; Wu, W. Osteoporosis and bone mineral Disease Branch of Chinese Medical Association. The results of the epidemiological survey of osteoporosis in China and the special action of “healthy bones” were released. *Chin. J. Osteoporos. Bone Miner. Dis.* **2019**, *12*, 317–318.
5. Bai, M.H.; Ge, B.F.; Liu, J.M. Quantitative ultrasound of calcaneus and changes of bone metabolism in underground coal miners. *China Bone Inj.* **2005**, *05*, 291–293.
6. Heydari, F.; Rafsanjani, M.K. A Review on Lung Cancer Diagnosis Using Data Mining Algorithms. *Curr. Med. Imaging* **2021**, *17*, 16–26. [[CrossRef](#)]
7. Kim, B.; Cho, Y.J.; Lim, W. Osteoporosis therapies and their mechanisms of action (Review). *Exp. Ther. Med.* **2021**, *22*, 1379. [[CrossRef](#)] [[PubMed](#)]
8. He, Y.; Chen, Y. The potential role of lncRNAs in osteoporosis. *J. Bone Miner. Metab.* **2021**, *39*, 341–352. [[CrossRef](#)]
9. Zhu, X.; Bai, W.; Zheng, H. Twelve years of GWAS discoveries for osteoporosis and related traits: Advances, challenges and applications. *Bone Res.* **2021**, *9*, 23. [[CrossRef](#)]
10. Fischer, V.; Haffner-Luntzer, M. Interaction between bone and immune cells: Implications for postmenopausal osteoporosis. *Semin. Cell Dev. Biol.* **2022**, *123*, 14–21. [[CrossRef](#)]
11. Zheng, L.Q.; Zhang, R. Research progress of fitting evaluation method of disease incidence risk prediction model. *China Health Stat.* **2015**, *32*, 544–546.
12. Qiu, Y.; Lu, G.; Wu, Y. Coexpression of PBX1 and EMP2 as Prognostic Biomarkers in Estrogen Receptor-Negative Breast Cancer via Data Mining. *J. Comput. Biol.* **2020**, *27*, 1509–1518. [[CrossRef](#)]
13. Li, X.Q. Research on Heart Disease Prediction Method Based on Convolutional Neural Network. Master's Thesis, Northeast Forestry University, Harbin, Heilongjiang, 2019.
14. Tan, Z.W.; Wang, W.L.; Zong, R. Classification algorithm of heart sound signal of congenital heart disease based on convolutional neural network. *J. Biomed. Eng.* **2019**, *36*, 728–736.
15. Sethuraman, R.R.; Athisayam, J.S.K. An Improved Feature Selection Based on Naive Bayes with Kernel Density Estimator for Opinion Mining. *Arab. J. Sci. Eng.* **2021**, *46*, 4059–4071. [[CrossRef](#)]
16. Su, T.P. Traffic Flow Prediction and Research Based on Multi Model Fusion. Master's Thesis, Guangdong University of Technology, Guangzhou, China, 11 July 2019.
17. Zhang, H.X.; Guo, H.; Wang, J.X. Accurate prediction model of type 2 diabetes mellitus based on XGBoost algorithm. *Chin. Exp. Diagn.* **2018**, *22*, 408–412.
18. Wu, J.; Qin, S.; Wang, J. Develop and Evaluate a New and Effective Approach for Predicting Dyslipidemia in Steel Workers. *Front. Bioeng. Biotechnol.* **2020**, *8*, 839. [[CrossRef](#)] [[PubMed](#)]
19. Li, X.; Xu, X.; Song, Y. An association between cumulative exposure to light at night and the prevalence of hyperuricemia in steel workers. *Int J. Occup. Med. Env. Health.* **2021**, *34*, 385–401. [[CrossRef](#)] [[PubMed](#)]
20. Li, X.; Cui, S.; Wu, J. Job category differences in the prevalence and associated factors of insomnia in steel workers in China. *Int. J. Occup. Med. Env. Health.* **2020**, *33*, 215–233. [[CrossRef](#)] [[PubMed](#)]
21. Kong, N.; Chen, G.; Wang, H. Blood leukocyte count as a systemic inflammatory biomarker associated with a more rapid spirometric decline in a large cohort of iron and steel industry workers. *Respir. Res.* **2021**, *22*, 254. [[CrossRef](#)]
22. Nyarubeli, I.P.; Tungu, A.M.; Moen, B.E.; Bråtveit, M. Prevalence of Noise-Induced Hearing Loss Among Tanzanian Iron and Steel Workers: A Cross-Sectional Study. *Int. J. Env. Res. Public Health* **2019**, *16*, 1367. [[CrossRef](#)]
23. Revision Committee of the Guidelines for the Prevention and Treatment of Hypertension in China. The 2018 revision of the Chinese guidelines for the prevention and treatment of hypertension. *Chin. J. Cardiovasc. Med.* **2019**, *19*, 1–44.
24. Diabetes Branch of Chinese Medical Association. Guidelines for the Prevention and Treatment of Type 2 Diabetes in China (2020 Edition). *Int. J. Endocrinol. Metabolism* **2021**, *41*, 482–548.
25. Menchaca, J.R. Tobacco or health: A global status report. *Ann. Saudi Med.* **1997**, *18*, 195–205.
26. Chinese Center for Disease Control and Prevention. *Surveillance Report on Chronic Diseases and Their Risk Factors in CHINA (2010)*; Military Medical Science Press: Beijing, China, 2012.
27. Joint Committee on Revision of Guidelines for the Prevention and Treatment of Dyslipidemia in Adults in China. Guidelines for the prevention and treatment of dyslipidemia in adults in China (2016 revision). *Chin. J. Cardiol.* **2016**, *44*, 833–853.
28. Lou, X.W.; He, Q. Validity and Reliability of the International Physical Activity Questionnaire in Chinese Hemodialysis Patients: A Multicenter Study in China. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **2019**, *25*, 9402–9408. [[CrossRef](#)]
29. Li, L.L.; Liu, Y.; Yang, H.Y. Analysis of bone mineral density test results and low bone mass factors in taxi drivers in Beijing. *Ind. Hyg. Occup. Dis.* **2022**, *48*, 105–109. [[CrossRef](#)]
30. Bledsoe, L.; Alessi, K.; Toro, J.B. Fragility Fractures: Diagnosis and Treatment. *Am. J. Orthop.* **2018**, *47*, 12. [[CrossRef](#)]

31. Meng, D.L.; Xu, J.; Zhao, J.J. Analysis and prediction of hand, foot and mouth disease incidence in China using Random Forest and XGBoost. *PLoS ONE* **2021**, *16*, e0261629. [[CrossRef](#)]
32. Li, Z.; Jiang, K.; Qin, S. GCSENet: A GCN, CNN and SENet ensemble model for microRNA-disease association prediction. *PLoS Comput. Biol.* **2021**, *17*, e1009048. [[CrossRef](#)]
33. Zhu, X.; Wu, G. Changes of quantitative ultrasound and bone metabolism of calcaneus in miners under coal mine environment. *Chin. Community Physician (Med. Spec.)* **2010**, *12*, 150.
34. Cai, D.; Ge, J.; Chen, F. Analysis of bone mineral density in 3006 adults in eastern Fujian. *Fujian Med. J.* **2018**, *40*, 123–127.
35. Yu, H.; Guo, W.; Huang, W. Correlation analysis between FRAX score, bone mineral density and BMI in patients with type 2 diabetes. *Chin. J. Osteoporos.* **2017**, *23*, 761–763.
36. Hardy, R.S.; Zhou, H.; Seibel, M.J. Glucocorticoids and Bone: Consequences of Endogenous and Exogenous Excess and Replacement Therapy. *Endocr. Rev.* **2018**, *39*, 519–548. [[CrossRef](#)] [[PubMed](#)]
37. Lin, J. Epidemiological study on female sleep quality and human composition. Ph.D. Thesis, Zhejiang University, Hangzhou, Zhejiang, 2018.
38. Lee, S.; Le, N.H.; Kang, D. Melatonin alleviates oxidative stress-inhibited osteogenesis of human bone marrow-derived mesenchymal stem cells through AMPK activation. *Int. J. Med. Sci.* **2018**, *15*, 1083–1091. [[CrossRef](#)] [[PubMed](#)]
39. Anupama, D.S.; Norohna, J.A.; Acharya, K.K.; George, A. Effect of exercise on bone mineral density and quality of life among postmenopausal women with osteoporosis without fracture: A systematic review. *Int. J. Orthop. Trauma Nurs.* **2020**, *39*, 100796. [[CrossRef](#)] [[PubMed](#)]
40. Hauger, A.V.; Holvik, K.; Bergland, A. Physical capability, physical activity, and their association with femoral bone mineral density in adults aged 40 years and older: The Tromsø study 2015–2016. *Osteoporos. Int.* **2021**, *32*, 2083–2094. [[CrossRef](#)] [[PubMed](#)]
41. Qi, H.; Sheng, Y.; Chen, S. Bone mineral density and trabecular bone score in Chinese subjects with sarcopenia. *Aging Clin. Exp. Res.* **2019**, *31*, 1549–1556. [[CrossRef](#)] [[PubMed](#)]
42. Trevisan, C.; Alessi, A.; Girotti, G. The Impact of Smoking on Bone Metabolism, Bone Mineral Density and Vertebral Fractures in Postmenopausal Women. *J. Clin. Densitom.* **2020**, *23*, 381–389. [[CrossRef](#)] [[PubMed](#)]
43. Kopiczko, A.; Gryko, K.; Łopuszańska-Dawid, M. Bone mineral density, hand grip strength, smoking status and physical activity in Polish young men. *Homo* **2018**, *69*, 209–216. [[CrossRef](#)]
44. Lee, J.H.; Hong, A.R.; Kim, J.H. Amount of smoking, pulmonary function, and bone mineral density in middle-aged Korean men: KNHANES 2008–2011. *J. Bone Miner. Metab.* **2018**, *36*, 95–102. [[CrossRef](#)]
45. Akkawi, I.; Zmerly, H. Osteoporosis: Current Concepts. *Joints* **2018**, *6*, 122–127. [[CrossRef](#)] [[PubMed](#)]
46. Xia, W.B.; Zhang, Z.L.; Lin, H. Guidelines for the diagnosis and treatment of primary osteoporosis (2017). *Chin. J. Osteoporos.* **2019**, *25*, 281–309.
47. Horai, T.; Hishimoto, A.; Otsuka, I. A cross-sectional study exploring useful indicators for low bone mineral density in male alcoholic patients. *Neuropsychiatr. Dis. Treat.* **2018**, *14*, 663–669. [[CrossRef](#)] [[PubMed](#)]
48. Redondo, N.; Nova, E.; Díaz-Prieto, L.E.; Marcos, A. Effects of moderate beer consumption on health. *Nutr. Hosp.* **2018**, *35*, 41–44. [[CrossRef](#)]
49. Ren, Y.; Hu, J.; Zhou, W.J. Prevalence and risk factors of hip fracture in a middle-aged and older Chinese population. *Bone* **2019**, *122*, 143–149. [[CrossRef](#)]
50. Chen, F.P.; Fu, T.S.; Yu, C. Risk factors and quality of life for the occurrence of hip fracture in postmenopausal women. *Biomed. J.* **2018**, *41*, 202–208. [[CrossRef](#)]
51. Wang, Q.Q.; Yu, S.C.; Qi, X. Overview of logistic regression model analysis and application. *Chin. J. Prev. Med.* **2019**, *53*, 955–960. [[CrossRef](#)]
52. Yan, X.; Xu, Z.; Li, S.; Yan, L.; Lyu, G.; Wang, Z. Establishment and verification of an osteoporosis risk model in patients with rheumatoid arthritis: A valuable new model. *Arch. Osteoporos.* **2021**, *16*, 3. [[CrossRef](#)]
53. Raheja, S.; Kasturia, S.; Cheng, X.; Kumar, M. Machine learning-based diffusion model for prediction of coronavirus-19 outbreak. *Neural Comput. Appl.* **2021**, *1*, 20. [[CrossRef](#)]
54. Kim, Y.D.; Noh, K.J.; Byun, S.J. Effects of Hypertension, Diabetes, and Smoking on Age and Sex Prediction from Retinal Fundus Images. *Sci. Rep.* **2020**, *10*, 4623. [[CrossRef](#)]
55. Dai, G.; He, W.; Xu, L. Exploring the effect of hypertension on retinal microvasculature using deep learning on East Asian population. *PLoS ONE* **2020**, *15*, e0230111. [[CrossRef](#)]
56. Jiang, J.; Deng, H.; Xue, Y.; Liao, H.; Wu, S. Detection of Left Atrial Enlargement Using a Convolutional Neural Network-Enabled Electrocardiogram. *Front. Cardiovasc. Med.* **2020**, *7*, 609976. [[CrossRef](#)] [[PubMed](#)]
57. Guan, X.; Zhang, B.; Fu, M. Clinical and inflammatory features based machine learning model for fatal risk prediction of hospitalized COVID-19 patients: Results from a retrospective cohort study. *Ann. Med.* **2021**, *53*, 257–266. [[CrossRef](#)] [[PubMed](#)]