Original Article

2026-07-16

ChinaRxiv

Abstract

Background Hypertension is a known major risk factor for cardiovascular and cerebrovascular diseases. In plateau regions, unique natural environmental and sociocultural factors may make the epidemiological status and priorities for prevention and control of hypertension different from those in plain areas. However, there is currently a lack of specific hypertension risk identification tools for older Tibetan adults in the Tibet Autonomous Region. Objective Based on data from the National Health Services Survey, this study aimed to construct a hypertension risk prediction model applicable to this population, so as to provide a potential auxiliary tool for preliminary community-based hypertension risk screening and targeted health management in Tibet. Methods A cross-sectional study design was adopted, including a total of 2 140 Tibetan older adults aged $ 60yearsinTibet.Variableswerefirstscreenedbyunivariateanalysis, andthenfinalpredictorsweredetermined P=0.972). Multivariable Logistic regression analysis ultimately identified 8 predictors significantly associated with hypertension status: different prefectures/cities, BMI category, age, insomnia symptoms within the past two weeks, number of other chronic diseases, frequency of tooth brushing per day, social medical insurance, and employment status. Model comparison showed that the NaiveBayes model performed best on the test set, with an AUC of 0.640, accuracy of 59.8%, sensitivity of 54.9%, and specificity of 63.5%. A nomogram was constructed based on the final 8 factors. Conclusion Based on cross-sectional data, this study constructed a hypertension risk prediction model for older Tibetan adults incorporating 8 factors. The NaiveBayes model achieved an AUC of 0.640, accuracy of 59.8%, sensitivity of 54.9%, and specificity of 63.5% in the test set, demonstrating limited discriminative ability. The model and nomogram tool may provide a reference for community-based hypertension screening in this population, but its diagnostic performance and practicality still require further validation in prospective studies.

Full Text

Original Article

Development and Analysis of a Hypertension Risk Prediction Model for Tibetan Older Adults in the Tibet Region: Based on Data from the National Health Services Survey in Tibet

ZHANG Shijie1, YONGZHU Lacuo1, Luosang Danba1, CIDAN Zhuoga1,2, ZHAXI Cuomu1,2, YANG Zong1,2, ZHAXI Dawa1,2,3*

1. Medical College, Xizang University, Lhasa 850000, Xizang Autonomous Region, China

2. Key Laboratory of Public Health Safety and Health Policy Research, Lhasa, Xizang Autonomous Region (Xizang University), Lhasa 850000, Xizang Autonomous Region, China

3. Plateau Health Science Research Center, Xizang University, Lhasa 850000, Xizang Autonomous Region, China
*Corresponding author: Zhaxi Dawa; E-mail: zhaxi0891@qq.com

【Abstract】 **Background** Hypertension is one of the well-known major risk factors for cardiovascular and cerebrovascular diseases. In plateau regions, unique natural environmental and sociocultural factors may make the epidemiological status and priorities for prevention and control of hypertension different from those in plain areas. However, there is currently a lack of specific hypertension risk identification tools for Tibetan older adults in the Xizang Autonomous Region. **Objective** Based on data from the National Health Services Survey, to develop a hypertension risk prediction model suitable for this population, with the aim of providing a possible auxiliary tool for community-based preliminary screening of hypertension risk and targeted health management in the Xizang region. **Methods** A cross-sectional study design was adopted, including a total of 2 140 Tibetan older adults aged
$ 60yearsintheXizangregion.Variableswerefirstscreenedbyunivariateanalysis, andfinalpredictorswerether
learningmodels—LASSOregression, RandomForest, XGBoost, LightGBM, andNaiveBayes—wereconstru
∗Results∗∗Amongthe2140Tibetanolderadultsincluded, 926hadhypertension, withahypertensionprevalenceof
43.30 ^2=0.001$, $P = 0.972$). Through multivariable Logistic regression analysis, 8 predictors significantly associated with hypertension status were ultimately identified: prefecture/city, BMI category, age, insomnia symptoms within the past two weeks, number of other chronic diseases, daily toothbrushing frequency, social medical insurance, and employment status. Model comparison showed that the NaiveBayes model performed best in the test set, with an AUC of 0.640, accuracy of 59.8%, sensitivity of 54.9%, and specificity of 63.5%. A nomogram was constructed based on the 8 final factors. **Conclusion** Based on cross-sectional data, this study developed a hypertension risk prediction model for Tibetan older adults that included 8 factors. In the test set, the NaiveBayes model had an AUC of 0.640, accuracy of 59.8%, sensitivity of 54.9%, and specificity of 63.5%, demonstrating limited discriminative ability. This model and the nomogram tool may provide a reference for community screening of hypertension in this population, but their diagnostic performance and utility still need to be further validated in prospective studies.

【Keywords】 Hypertension; Machine learning; Predictive learning model; Tibetan; Older adults; Xizang

【Chinese Library Classification Number】 R 544.1 **【Document Code】** A DOI: 10.12114/j.issn.1007-9572.2025.0549

Development and Analysis of a Hypertension Risk Prediction Model for Tibetan Older Adults in the Tibet Region: Based on Data from the National Health Services Survey in Tibet

ZHANG Shijie1, YONGZHU Lacuo1, Luosang Danba1, CIDAN Zhuoga1,2, ZHAXI Cuomu1,2, YANG Zong1,2, ZHAXI Dawa1,2,3*

1.Medical College, Xizang University, Lhasa 850000, China
2.Lhasa Key Laboratory of Public Health Safety and Health Policy Research (Xizang University), Lhasa 850000, China
3.Plateau Health Science Research Center, Xizang University, Lhasa 850000, China
*Corresponding author: Zhaxi Dawa; E-mail: zhaxi0891@qq.com

【Abstract】 Background Hypertension is a well-recognized major risk factor for cardiovascular and cerebrovascular diseases. In plateau regions, unique environmental and sociocultural factors may lead to differences in the epidemiology, prevention, and control priorities of hypertension compared with plain areas. However, there is currently a lack of a specific

Funding project: Expanded-site subproject of the Seventh Health Services Survey of Xizang Autonomous Region (18080278)

Citation: Zhang Shijie, Yongzhu Lacuo, Luosang Danba, et al. Development and analysis of a hypertension risk prediction model for Tibetan older adults in the Tibet region: based on data from the National Health Services Survey in Tibet[J]. Chinese General Practice, 2026. DOI: 10.12114/j.issn.1007-9572.2025.0549. [Epub ahead of print]. [www.chinagp.net]

Zhang S J, Yongzhu L C, Luosang D B, et al. Development and analysis of a hypertension risk prediction model for tibetan older adults in the tibet region: based on data from the national health services survey in Tibet[J]. Chinese General Practice, 2026. [Epub ahead of print].

hypertension risk identification tool for Tibetan older adults in the Tibet Autonomous Region. **Objective** Based on data from the National Health Services Survey, to develop a hypertension risk prediction model for this population, and to provide a potential auxiliary tool for community-based preliminary screening and targeted health management of hypertension risk. **Methods** A cross-sectional study design was used. A total of 2 140 Tibetan older adults aged ≥ 60 years from the Tibet region were included. Variables were first screened by univariate analysis, and then a multistage feature selection strategy (LASSO regression combined with multivariate Logistic regression) was used to determine the final predictive variables. Based on the selected variables, five machine learning models were constructed and compared for their diagnostic value in predicting hypertension among Tibetan older adults: LASSO regression, RandomForest, XGBoost, LightGBM, and NaiveBayes. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, and specificity. Clinical utility was further assessed using calibration curves, decision curve analysis, and clinical impact curves. **Results** Among the 2 140 Tibetan older adults included, 926 had hypertension, yielding a prevalence of 43.27%. The participants were randomly divided into a training set (n = 1 499) and a test set (n = 641) at a 7:3 ratio. In the training set, 649 participants had hypertension (prevalence 43.30%), and in the test set, 277 participants had hypertension (prevalence 43.21%); the difference in prevalence between the two sets was not statistically significant ($\chi^2 = 0.001$, $P = 0.972$). Multivariate Logistic regression identified 8 predictive factors significantly associated with hypertension status: different cities, BMI category, age, insomnia symptoms within the past 2 weeks, number of other chronic diseases, daily tooth brushing frequency, social medical insurance, and employment status. Model comparison showed that NaiveBayes performed best in the test set, with an AUC of 0.640, accuracy of 59.8%, sensitivity of 54.9%, and specificity of 63.5%. A nomogram was constructed based on the final 8 factors. **Conclusion** Based on cross-sectional data, this study developed a hypertension risk prediction model for Tibetan older adults incorporating 8 factors. The NaiveBayes model showed limited discriminative ability in the test set, with an AUC of 0.640, accuracy of 59.8%, sensitivity of 54.9%, and specificity of 63.5%. The model and nomogram may provide a reference for community-based hypertension screening in this population, but their diagnostic performance and practical utility require further validation in prospective studies.

【Key words】 Hypertension; Machine learning; Predictive learning models; Tibetan nationality; Aged; Tibet

Hypertension is one of the major global public health problems and is also the leading risk factor for cardiovascular and cerebrovascular diseases[1]. The prevalence of hypertension in China has remained persistently high, and the disease burden is particularly heavy[2]. The average altitude of the Tibet Autonomous Region exceeds 4 000 m[3], and the plateau environment characterized by low pressure and hypoxia has a significant impact on hypertension prevalence. The *Chinese Guidelines for the Prevention and Treatment of Hypertension (2024 Revised Edition)*[2] (hereinafter referred to as the Guidelines) points out that, in the Qinghai–Tibet Plateau region of China, the prevalence of hypertension among residents in high-altitude areas of Qinghai Province (> 2 000 m) is 36.3%, significantly higher than that in low-altitude areas (< 1 000 m) (19.6%); among residents in Tibetan areas with an altitude above 3 000 m, the prevalence of hypertension increases by 2% for every 100 m increase in altitude. In addition, factors such as the lifestyle, dietary structure, and possible genetic background of residents in Tibet further increase the risk of hypertension. Existing studies have shown that the comorbidity rate of the "three highs" (hypertension, hyperglycemia, and hyperlipidemia) among residents in Tibet reaches 18.48%, with hypertension being one of the major comorbid factors[4]; the crude prevalence of hypertension among residents aged 35–75 years in this region is 46.5%, while the awareness rate, treatment rate, and control rate are only 45.2%, 30.8%, and 3.0%, respectively[5]. The situation of regional hypertension prevention and control is therefore extremely severe. However, Tibet is vast and sparsely populated, with dispersed settlements; traditional screening models are costly and inefficient, making routine blood-pressure monitoring difficult to achieve. Moreover, the Guidelines note that the popularization rate of home blood-pressure monitoring in western China still needs to be improved[2], resulting in delayed identification of high-risk groups and making it difficult to comprehensively and deeply advance early intervention and systematic management. Therefore, in view of the epidemiological characteristics of hypertension and the prevention-and-control situation in Tibet, constructing a simple and efficient hypertension risk prediction model applicable to high-risk populations in this region is an important prerequisite for early identification, risk stratification, and precise intervention for high-risk groups, as well as for reducing disease burden and improving prevention-and-control effectiveness[6]. It has important clinical and public health value.

At present, most studies on hypertension risk prediction models are based on general populations or specific disease cohorts, while models specifically targeting Tibetan older adults in plateau areas are relatively rare[7]. National Health Services Survey data cover multidimensional information such as demographic characteristics, health status, behavioral habits, and health-service utilization, providing a data basis for exploring risk factors related to hypertension in this population[8]. Using these data, this study aimed to apply five machine-learning methods to attempt to construct a hypertension risk identification model suitable for Tibetan older adults in Tibet, in order to provide a simple auxiliary assessment tool for primary-level hypertension prevention and control in this region.

1 Participants and Methods

1.1 Study Participants

This was a cross-sectional study. The data came from the 2023 National Health Services Survey Tibet expanded-site project. The inclusion criteria were: (1) age $\geq$ 60 years; (2) Tibetan ethnicity; and (3) complete survey data. Individuals with missing information on key variables were excluded. A total of 2 140 survey participants were ultimately included in this study; the study flow is shown in Figure 1.

1.2 Data Collection and Variable Definitions

The data for this study came from the 2023 National Health Services Survey Tibet expanded-site project, with the survey conducted from October 2023 to December 2023. The outcome variable was hypertension status at the time of the survey, based on the survey questionnaire

"Do you have diagnosed hypertension?" This question was used for determination, and a definite diagnostic certificate for hypertension was required. According to the study objectives and literature review, this study collected and included a total of 21 initial predictor variables in the following 4 categories: (1) sociodemographic characteristics: age, sex, educational level, marital status, employment status, different prefectures/cities, and urban–rural residence. (2) Socioeconomic indicators: type of social medical insurance, category of total annual household expenditure, category of total annual household income, type of drinking water, type of toilet, whether the household was receiving minimum living security, and whether the participant took part in endowment insurance. (3) Health behaviors and lifestyle: BMI category, smoking status, drinking frequency, weekly frequency of physical exercise, and daily frequency of tooth brushing. (4) Health-status-related variables: insomnia symptoms within the past two weeks and number of other chronic diseases.

Figure 1. Study flowchart

Establishment of a risk prediction model for hypertension among Tibetan older adults in Tibet

↓

The 2023 National Health Services Survey Tibet expanded-site project investigated a total of 11,104 participants

→ 8,964 participants were excluded:
- Age <60 years ($n = 8\,946$)
- Non-Tibetan older adults ($n = 14$)
- Abnormal BMI ($n = 4$)

↓

A total of 2,140 study participants were included

↓

Training group ($n = 1\,499$)   Test group ($n = 641$)

↓

Feature screening: univariate analysis, LASSO regression, Logistic regression

↓

Age; BMI category; daily frequency of tooth brushing; insomnia symptoms within the past two weeks; number of other chronic diseases; different prefectures/cities; social medical insurance; employment status

↓

Construction of the nomogram

Figure 1 Study flowchart

1.3 Data preprocessing and feature selection

R 4.5.1 software was used for analysis. For a small number of missing values, numerical variables were imputed with the mean, and categorical variables were imputed with the mode. All data were randomly divided at a 7:3 ratio into a training set ($n = 1\,499$) and a test set ($n = 641$). After grouping, the distribution of hypertension prevalence was balanced between the groups (training set 43.30%, test set 43.21%). The feature-selection process was conducted entirely in the training set to prevent information leakage. First, univariate analysis ($P < 0.05$) was performed to preliminarily screen variables; then LASSO regression was used for further variable reduction, with `lambda.min` selected by 10-fold cross-validation; finally, the variables screened by LASSO regression were entered into multivariable Logistic regression, with an inclusion criterion of $P < 0.05$, to determine the predictors ultimately included in the model.

1.4 Model construction and evaluation

Based on the final selected features, 5 prediction models were constructed in the training set: LASSO regression, RandomForest, XGBoost, LightGBM, and NaiveBayes. Ten-fold cross-validation was used to optimize model hyperparameters. Model performance was evaluated on the independent test set. The main evaluation indicators included the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, F1 score, and Kappa value. Meanwhile, calibration curves were plotted to evaluate the agreement between predicted probability and actual risk; decision curve analysis (DCA) was applied to assess the clinical net benefit of the models at different risk thresholds; and clinical impact curves were plotted. To enhance model interpretability, SHAP analysis was performed for the best model. The optimal model was selected by comprehensively considering the test-set AUC value and the stability of the model performance indicators between the training and test sets.

1.5 Statistical analysis

Measurement data conforming to a normal distribution were expressed as ($\bar{x} \pm s$), and count data were described using frequencies (percentages). Comparisons among models were mainly based on performance indicators calculated on the test set. R packages such as pROC, rmda, rms, and fastshap were used for the corresponding analyses. A difference was considered statistically significant at $P < 0.05$.

2 Results

2.1 Basic characteristics of the study participants

A total of 2,140 Tibetan older adults were included, among whom 926 had hypertension, yielding an overall hypertension prevalence of 43.27%. Participants were randomly divided at a 7:3 ratio into a training set ($n = 1\,499$) and a test set ($n = 641$). In the training set, there were 649 patients with hypertension, with a prevalence of 43.30%; in the test set, there were 277 patients with hypertension, with a prevalence of 43.21%. Comparison of hypertension prevalence between the two groups showed no statistically significant difference ($\chi^2 = 0.001$, $P = 0.972$), indicating that the data split was balanced.

2.2 Feature-selection results

Univariate analysis in the training set screened out a total of 14 variables associated with hypertension prevalence ($P < 0.05$). After further reduction by LASSO regression ($\lambda_{\text{min}} = 0.004$), 12 variables were retained; the LASSO regression results are shown in Figure 2. Multivariable Logistic regression analysis ultimately identified 8 independent associated factors: different prefectures/cities, BMI category, age, insomnia symptoms within the past two weeks, number of other chronic diseases, daily frequency of tooth brushing, social medical insurance, and employment status (the variable-screening results are shown in Table 1, and the variable assignment and results of Logistic regression are shown in Tables 2 and 3). The variance inflation factor (VIF) test values were all less than 5, indicating that there was no severe multicollinearity among the variables.

2.3 Comparison of model performance

The performance indicators of the 5 models for diagnosing hypertension among Tibetan older adults in the training and test sets are shown in Table 4. A comprehensive comparison of the performance stability of the models on the test set, measured by changes in key indicators from the training set to the test set, found that although the performance of the NaiveBayes model on the test set decreased slightly compared with that on the training set, the magnitude of decline was relatively small. Taking both generalization ability and model performance into account, it was determined to be the best model. On the test set, this model had an accuracy of 59.8%, sensitivity of 54.9%, specificity of 63.5%, and F1 score of 0.541. The ROC curve (Figure 3) and calibration curve (Figure 4) showed that the NaiveBayes model had relatively limited discrimination and calibration.

Note: A is the coefficient path plot, and B is the cross-validation plot.

Figure 2 LASSO regression coefficient path plot and cross-validation plot

Decision curve analysis (Figure 5) showed that, within a certain range of risk thresholds (0.1–0.6), application of this model may yield clinical net benefit. Considering the intrinsic structure of the NaiveBayes model (the independence assumption based on conditional probability) and its optimal diagnostic performance in this study, no additional interpretability analysis was performed; instead, the focus was on validating the model' s diagnostic capability.

Table 1 Variable screening table

No.	Stage	Variables
1	Univariate analysis	Different cities/prefectures, BMI classification, age, education level, marital status, insomnia symptoms in the past two weeks, number of other chronic diseases, daily toothbrushing frequency, social medical insurance, employment status, drinking-water type, toilet type, annual household expenditure classification, weekly physical exercise frequency
2	LASSO regression	Different cities/prefectures, BMI classification, age, insomnia symptoms in the past two weeks, number of other chronic diseases, daily toothbrushing frequency, social medical insurance, employment status, drinking-water type, toilet type, annual household expenditure classification, weekly physical exercise frequency

No.	Stage	Variables
3	Logistic regression	Different cities/prefectures, BMI classification, age, insomnia symptoms in the past two weeks, number of other chronic diseases, daily toothbrushing frequency, social medical insurance, employment status

Table 2 Variable assignment table

Variable	Assignment
Different cities/prefectures	Lhasa = 0, Shigatse = 1, Shannan = 2, Nyingchi = 3, Qamdo = 4, Nagqu = 5, Ngari = 6
Age	60–98
BMI classification	<18.5 kg/m^2 = 0, 18.5-23.9 kg/m^2 = 1, 24.0-27.9 kg/m^2 = 2, $ 28.0kg/m^{2}$ = 3
Insomnia symptoms in the past two weeks	None = 0, several days = 1, more than half the time = 2, almost every day = 3
Number of other chronic diseases	None = 0, 1 disease = 1, 2 diseases = 2, 3 or more diseases = 3
Daily toothbrushing frequency	Does not brush = 0, less than once = 1, once = 2, twice = 3
Employment status	Employed (including flexible employment) = 0, retired = 1, unemployed = 2, no work = 3
Social medical insurance	Not insured = 0, resident medical insurance = 1, employee medical insurance = 2

Table 3 Multivariate Logistic regression analysis of factors associated with hypertension in Tibetan older adults in the Tibet region

Variable	Coefficient	Standard error	Wald χ^2 value	OR (95%CI)	P value
Different cities/prefectures	0.119	0.033	12.609	1.126 (1.055–1.202)	<0.001
Age	0.031	0.008	15.037	1.032 (1.016–1.048)	<0.001
BMI classification	0.489	0.073	45.064	1.630 (1.413–1.880)	<0.001
Insomnia symptoms in the past two weeks	0.200	0.069	8.280	1.221 (1.066–1.399)	0.004
Number of other chronic diseases	0.148	0.074	4.016	1.160 (1.003–1.341)	0.045
Daily toothbrushing frequency	−0.143	0.061	5.610	0.867 (0.770–0.976)	0.018
Employment status	0.134	0.059	5.170	1.143 (1.019–1.283)	0.023
Social medical insurance	−0.704	0.257	7.497	0.494 (0.299–0.819)	0.006

2.4 Nomogram construction

Based on the 8 factors identified by multivariate Logistic regression, a nomogram for predicting hypertension risk was constructed (Figure 6). The C-index of the nomogram model in internal validation of the training set was 0.676.

3 Discussion

Based on cross-sectional data, this study used five machine-learning methods—LASSO regression, RandomForest, XGBoost, LightGBM, and NaiveBayes—

to construct a risk identification model for hypertension among Tibetan older adults in Tibet. Comparison of model performance showed that RandomForest performed best in the training set (AUC = 0.906), but its AUC decreased to 0.633 in the test set, suggesting obvious overfitting. The AUCs of XGBoost and LightGBM in the test set were 0.659 and 0.654, respectively, indicating moderate performance. The NaiveBayes model showed the smallest decline in performance between the training set (AUC = 0.693) and the test set (AUC = 0.640), and the test

Figure 3. ROC curves of the five models for diagnosing hypertension in older adults in the training and test sets

- **A.** Training-set ROC curve
  - y-axis: Sensitivity
  - x-axis: 1− specificity
  - LASSO (AUC = 0.676)
  - RandomForest (AUC = 0.906)
  - XGBoost (AUC = 0.789)
  - LightGBM (AUC = 0.77)
  - NaiveBayes (AUC = 0.693)
- **B.** Test-set ROC curve
  - y-axis: Sensitivity
  - x-axis: 1− specificity
  - LASSO (AUC = 0.623)
  - RandomForest (AUC = 0.633)
  - XGBoost (AUC = 0.659)
  - LightGBM (AUC = 0.654)
  - NaiveBayes (AUC = 0.64)

Note: A is the training-set ROC curve, and B is the test-set ROC curve.

Table 4. Diagnostic performance metrics of the five prediction models for hypertension in Tibetan older adults in the training and test sets

Model	Dataset	AUC	Accuracy	Sensitivity	Specificity	Positive predictive value	F1 score	Kappa value
LASSO regression	Training set	0.676	0.630	0.428	0.785	0.603	0.501	0.221
RandomForest	Training set	0.906	0.841	0.695	0.953	0.919	0.791	0.667
XGBoost	Training set	0.789	0.720	0.564	0.840	0.729	0.636	0.415
LightGBM	Training set	0.770	0.691	0.518	0.824	0.691	0.592	0.352
NaiveBayes	Training set	0.693	0.644	0.598	0.680	0.588	0.593	0.277
LASSO regression	Test set	0.623	0.594	0.440	0.712	0.537	0.484	0.155
RandomForest	Test set	0.633	0.616	0.458	0.736	0.570	0.508	0.199
XGBoost	Test set	0.659	0.638	0.502	0.742	0.597	0.545	0.248
LightGBM	Test set	0.654	0.616	0.451	0.742	0.571	0.504	0.198
NaiveBayes	Test set	0.640	0.598	0.549	0.635	0.533	0.541	0.183

Figure 4. Calibration curves of the best model in the training and test sets

- **NaiveBayes–training-set calibration curve**
  - y-axis: Observed probability
  - x-axis: Predicted probability
  - Original
  - Bias-corrected
  - Ideal
- **NaiveBayes–test-set calibration curve**
  - y-axis: Observed probability
  - x-axis: Predicted probability

- Original
- Bias-corrected
- Ideal

Figure 5 Decision curve analysis (DCA) plots of the best model in the training and test sets

- Panel A: training set; Panel B: test set.
- Axes: net benefit; risk threshold.
- Legend: NaiveBayes; treat all; treat none.

Figure 6 Nomogram for predicting hypertension risk in Tibetan older adults

Visible nomogram variables: Points; different prefectures/cities; BMI classification; age; insomnia symptoms within two weeks; number of other chronic diseases; times brushing teeth per day; social medical insurance; employment status; total points; probability of hypertension risk.

The accuracy in the set (59.8%) and sensitivity (54.9%) were relatively the highest among the five models. Considering both generalization ability and diagnostic performance, it was determined to be the best model in this study. Multivariable logistic regression analysis identified a total of eight factors independently associated with hypertension prevalence, including: different prefectures/cities, BMI classification, age, insomnia symptoms within two weeks, number of other chronic diseases, times brushing teeth per day, social medical insurance, and employment status.

3.1 Risk factor analysis

This study found that different prefectures/cities were an important influencing factor for hypertension among Tibetan older adults in Tibet ($OR = 1.126,\ 95\%CI = 1.055 \sim 1.202$). Compared with Lhasa, older adults living in higher-altitude prefectures/cities (such as Nagqu and Ngari) had a higher risk of hypertension, which may comprehensively reflect the influence of regional factors such as the natural conditions of the place of residence (e.g., higher altitude and a relatively harsher natural environment), level of economic development, accessibility of medical resources, and dietary culture on disease onset[9]. For every 1-year increase in age, the risk of hypertension increased by 3.2% ($OR = 1.032,\ 95\%CI = 1.016 \sim 1.048$), consistent with the general pattern of reduced vascular compliance caused by aging[5]. For each one-level increase in BMI classification, the risk of hypertension increased by 63.0% ($OR = 1.630,\ 95\%CI = 1.413 \sim 1.880$), suggesting that weight management is a key focus for prevention and control[5].

Insomnia symptoms within two weeks were positively associated with hypertension risk ($OR = 1.221,\ 95\%CI = 1.066 \sim 1.399$), consistent with previous findings[10]; insomnia may affect blood pressure regulation by activating the sympathetic nervous system. The greater the number of other chronic diseases, the higher the risk of hypertension ($OR = 1.160,\ 95\%CI = 1.003 \sim 1.341$), suggesting that multiple chronic diseases (such as diabetes, dyslipidemia, heart disease, etc.) may interact through shared pathophysiological mechanisms (such as inflammatory responses and oxidative stress), jointly increasing the risk of hypertension[11]. The number of times brushing teeth per day was negatively associated with hypertension risk ($OR = 0.867,\ 95\%CI = 0.770 \sim 0.976$). Poor oral hygiene may increase cardiovascular risk through systemic inflammatory responses, suggesting that maintaining good oral hygiene has a positive effect on reducing the risk of cardiovascular and metabolic diseases[12]. In terms of employment status, compared with employed individuals, retired, unemployed, or jobless individuals had a 14.3% increased risk of hypertension ($OR = 1.143,\ 95\%CI = 1.019 \sim 1.283$), similar to findings from foreign studies[13], suggesting that social participation and physical activity brought by occupation may play a protective role in maintaining blood pressure health among older adults. Among types of social medical insurance, those with employee medical insurance had a lower risk than those with resident medical insurance ($OR = 0.494,\ 95\%CI = 0.299 \sim 0.819$). As of the end of 2022, the coverage rate of basic medical insurance enrollment across Tibet remained stable at above 95%[14]; in this study, the proportions of employee medical insurance, resident medical insurance, and no insurance were 5.2%, 94.2%, and 0.6%, respectively, and more than half of those with employee medical insurance resided in Lhasa. Differences in medical insurance type may reflect differences among individuals in the level of social security, accessibility of medical resources, economic status, and health literacy; those with employee medical insurance generally have higher medication adherence and more complete health management services[15].

3.2 Clinical application value of the model

The NaiveBayes model and nomogram tool constructed in this study provide an initial screening method for hypertension risk based on eight easily obtainable variables. In primary health services in Tibet, community workers can, without using a sphygmomanometer, quickly estimate an individual' s probability of hypertension by asking about age, prefecture/city, height and weight, insomnia status, history of chronic disease, toothbrushing habits, medical insurance, and employment status,

thereby giving priority to precise blood-pressure measurement and intervention for high-risk individuals. Decision-curve analysis showed that applying this model within the risk-threshold range of 0.1–0.6 can yield clinical net benefit.

3.3 Advantages and Limitations

This study focused on a high-risk population with regional particularities, used a standardized machine-learning workflow and multiple evaluation methods, and provided a visualization tool. At the same time, it also has limitations. First, this study used a cross-sectional design and therefore cannot establish causal relationships; the associations identified are only contemporaneous associations. Second, the model was constructed on the basis of self-reported data and may be affected by recall bias and other factors. Third, the study did not include potentially important biomarkers such as blood lipids, hemoglobin, low-density lipoprotein cholesterol, and genetic factors

16

. Fourth, and most importantly, all performance evaluations were internal validations; the model' s performance in other populations not involved in model building or at future time points—namely, its external validity—remains unknown, and its true predictive value still needs to be confirmed by prospective studies.

4 Conclusion

In summary, based on cross-sectional data from older Tibetan adults in Tibet, this study preliminarily explored and established a model for identifying the risk of prevalent hypertension. The NaiveBayes model and its corresponding nomogram tool provide a simple reference method for community screening of hypertension in this population. It should be emphasized that this model is only an auxiliary tool for screening disease risk and cannot replace clinical diagnosis. In the future, rigorous external validation of this model in independent prospective cohorts is needed, and its effectiveness and cost-effectiveness in real-world community settings should be evaluated to determine its ultimate practical value.

Author contributions: Zhang Shijie proposed the research concept and design, was responsible for data cleaning and statistical analysis, and drafted and revised the manuscript; Yongzhu Lacuo, Luosang Danba, Cidan Zhuoga, and Zhaxi Cuomu were responsible for data collection and organization and participated in data analysis; Yangzong was responsible for guidance on the research protocol and for manuscript review and revision; Zhaxi Dawa was responsible for the overall study design and supervision, and for approval of the final version. All authors confirmed the final version of the manuscript and agreed to take full responsibility for the research work.

The authors declare no conflicts of interest.

References

1

Brauer M, Roth G A, Aravkin A Y, et al. Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021

J

. Lancet, 2024, 403(10440): 2162–2203. DOI: 10.1016/S0140-6736(24)00933-4.

2

Revision Committee of the Chinese Guidelines for Hypertension Prevention and Treatment, Hypertension Alliance (China), Hypertension Branch of the China International Exchange and Promotive Association for Medical and Health Care, et al. Chinese guidelines for the prevention and treatment of hypertension (2024 revised edition)

J

. Chinese Journal of Hypertension: Chinese and English, 2024, 32(7): 603–700.

3

People' s Government of the Tibet Autonomous Region. Overview of Tibet and related government-affairs information

EB/OL

. (2025-01-01)

$2025 - 12 - 12$

. http://www.xizang.gov.cn/rsxz/.

4

Yu Yue, Jinmei Quzhen, Bai Guoxia, et al. Current status and influencing factors of comorbidity of hypertension, dyslipidemia, and diabetes among residents of Tibet

J

. Chinese Journal of Prevention and Control of Chronic Diseases, 2025, 33(6): 442–446.

5

Yu Y, Peng J R, Zhaxi Z, et al. Epidemiological features of hypertension in a high-altitude population in Tibet, China: a cross-sectional study

J

. BMC Public Heal, 2025, 25(1): 1729. DOI: 10.1186/s12889-025-22792-3.

6

Zhou Yaxi, Xiong Hai, Zhong Huaichang, et al. Analysis of the prevalence and influencing factors of simple systolic hypertension among Tibetan populations in Tibet

J

. Journal of Sichuan University: Medical Science Edition, 2024, 55(2): 360-366. DOI: 10.12182/20240360501.

7

Tang J H, Yang R, Li H, et al. Derivation and internal validation of prediction models for pulmonary hypertension risk assessment in a cohort inhabiting Tibet, China

J

. ELife, 2024, 13: RP98169. DOI: 10.7554/elife.98169.

8

National Bureau of Statistics. National Seventh Health Service Statistical Survey System

EB/OL

. (2023-12-12)

2025 − 12 − 12

. https://www.stats.gov.cn/fw/bmdcxmsp/bmzd/202312/t20231212_1945458.html.

9

Yan Jiaojiao, Zhang Yufei, Wan Yang, et al. Analysis of the epidemiological status and risk factors of hypertension in the ultra-high-altitude area of Nagqu, Tibet

J

. Modern Preventive Medicine, 2023, 50(20): 3790-3794. DOI: 10.20043/j.cnki.MPM.202304482.

10

Wang Y L, Jiang G G, Hou N N, et al. Effects and differences of sleep duration on the risk of new-onset chronic disease conditions in middle-aged and elderly populations

J

. Eur J Intern Med, 2023, 107: 73-80. DOI: 10.1016/j.ejim.2022.11.005.

11

Xiang Z H, Li H D. Study on the risk of comorbid hypertension in multiple chronic diseases

J

. Arch Gerontol Geriatr Plus, 2025, 2(2): 100135. DOI: 10.1016/j.aggp.2025.100135.

12

Park H J, Kim N H, Shin S J, et al. Association of tooth brushing behavior with risks of major chronic health outcomes: a scoping review

J

. BMC Oral Heal, 2025, 25(1): 943. DOI: 10.1186/s12903-025-06332-4.

13

Maurya P, Chattopadhyay A, Sharma P. Unravelling the occupational factor and hypertension puzzle among Indian older adults

J

. Arch Gerontol Geriatr Plus, 2025, 2(3): 100172. DOI: 10.1016/j.aggp.2025.100172.

14

Tibet Autonomous Region Healthcare Security Bureau. The regional healthcare-security work conference was held in Lhasa

EB/OL

. (2023-02-18)

$2025-12-12$

. http://ylbzj.xizang.gov.cn/xwzx/yhyw/202302/t20230218_342713.html.

15

Lan L, Hai P C, Luo J W, et al. Medical behaviours and medication adherence of older hypertensive patients in different medical insurance programs in Beijing, China: a cross-sectional study

J

. BMC Geriatr, 2023, 23(1): 878. DOI: 10.1186/s12877-023-04476-y.

16

Liu W F, Yang C Z, Lei F, et al. Major lipids and lipoprotein levels and risk of blood pressure elevation: a Mendelian Randomisation study

J

. eBioMedicine, 2024, 100: 104964. DOI: 10.1016/j.ebiom.2023.104964.

(Received: 2025-11-10; Revised: 2026-05-15)

(Editor: Zhao Yuecui)

Note: Figure translations are in progress. See original paper for figures.