Risk factors of hyperuricemia calculated by random forest machine learning

Yuhan Gao

Tianjin Medical University General Hospital

Shichong Jia

Tianjin Medical University General Hospital

Chao Huang

University of Hull

Zhaowei Meng (✉ zmeng@tmu.edu.cn)

Tianjin Medical University General Hospital <https://orcid.org/0000-0002-4478-878X>

Mei Yu

Tianjin University

Tianyi Xu

Tianjin University

Ming Liu

Tianjin Medical University General Hospital

Jinhong Sun

Tianjin Medical University General Hospital

Qiyu Jia

Tianjin Medical University General Hospital

Qing Zhang

Tianjin Medical University General Hospital

Ying Gao

Tianjin Medical University General Hospital

Kun Song

Tianjin Medical University General Hospital

Xing Wang

Tianjin Medical University General Hospital

Yaguang Fan

Tianjin Medical University General Hospital

Research article

Keywords: Hyperuricemia (HUA), random forest (RF), prediction model, risk factors

Posted Date: February 27th, 2020

DOI: <https://doi.org/10.21203/rs.2.24675/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Objectives The present study aimed to develop a random forest (RF) based prediction model for hyperuricemia (HUA) and estimate associated risk factors.

Methods This cross-sectional study recruited 91,690 participants (52,607 males, 39,083 females). The prediction models were derived from training sets using RF learning machine. Performances of the prediction model were evaluated in validation datasets. Significant indicators were produced after comparing between true positive set and true negative set. Odds ratio was calculated by binary logistic regression models.

Results The area under the receiver-operating curve was 0.732 in males and 0.837 in females in the RF prediction models. The sensitivity, specificity and negative predictive value of the models were 0.686, 0.656 and 0.882 in males, 0.786, 0.738 and 0.978 in females, respectively. According to the feature value of each index in RF, a total of 10 explanatory variables were selected for each gender. Triglyceride, creatinine, body mass index, waist circumference, alanine transaminase, age, weight and total cholesterol were high-risk factors for HUA in both genders.

Conclusion RF demonstrated good stability and strong predictive power in predicting HUA in Chinese population. People with high risk factors should be encouraged to actively control the above factors to reduce the probability of developing HUA.

Declarations

Ethics approval and consent to participate

This study was approved by the Institutional Review Board and Ethics Committee of Tianjin Medical University General Hospital and the approval number was 2011-6-1. The study obtained the written informed consent of all participants. We confirm that all methods used in this study were performed in accordance with the relevant guidelines and regulations.

Consent for publication

Not applicable.

Availability of data and materials

The datasets during and/or analyzed during the current study available from the corresponding author on reasonable request.

Competing interests

We declare no competing interests in the paper.

Funding

This study was also supported by the following grants. China National Natural Science Foundation grant #81571709 and #81971650 (to Zhaowei Meng), #81870533 (to Jinhong Sun), #71804124 (to Ying Gao). Key Project of Tianjin Science and Technology Committee Foundation grant #16JCZDJC34300 (to Zhaowei Meng). Tianjin Science and Technology Committee Foundation grant, #12ZCZDSY20400 (to Qiyu Jia), #11ZCGYSY05700 (to Qing Zhang), #13ZCZDSY20200 (to Kun Song) and #17JCYBJC25400 (to Yaguang Fan).

Authors' contributions

YG and SJ were major contributors in writing the original draft and completed the investigation and formal analysis of the article. ZM participated in the conceptualization, supervision, review and editing of the article. MY, TX and YF provided algorithm and software support and participated in the experimental design. CH and ML participated in supervision of the article. JS, QJ, QZ, YG, KS, XW provided the data of the article. All authors read and approved the final manuscript.

Acknowledgements

This study was supported by the National Key Clinical Specialty Project (awarded to the Departments of Nuclear Medicine and Radiology).

This study was supported by Tianjin Medical University General Hospital New Century Excellent Talent Program; Young and Middle-aged Innovative Talent Training Program from Tianjin Education Committee; and Talent Fostering Program (the 131 Project) from Tianjin Education Committee, Tianjin Human Resources and Social Security Bureau (to Zhaowei Meng).

Key Messages

1. This is the first study conducted that predicted hyperuricemia by Random Forest (RF) machine learning.
2. The true positive and negative sets were analyzed to obtain feature contrast calculated by RF.
3. RF demonstrating strong predictive power can provide early warning for high-risk groups with hyperuricemia.

Introduction

Uric acid (UA) is a metabolite of purines (ATP, GTP, and nucleic acids) circulating in the blood. The excretion of UA plays an important role in removing nitrogenous wastes from the body[1]. Hyperuricemia(HUA)can be caused by overproduction or underexcretion of UA[2]. Increased production may be the consequence of high-purine diet, alcohol abuse or congenital enzyme deficiency. Reduced

excretion can be caused by genetic defects, renal disease or drugs which interfere with uric acid excretion such as diuretics and cyclosporine. Over the past 40 years, the level of serum UA (SUA) and prevalence of HUA have risen sharply globally[3]. From 2000 to 2014, prevalence of HUA in China was 13.3% (19.4% for male and 7.9% for female), among which HUA was more common in urban residents than rural, which may be closely related to not only the consumption of meat, seafood and alcohol but also work type, commuting method and exercise frequency[4, 5].

Besides, plenty of studies have suggested that elevated SUA level was associated with various diseases, including metabolic syndrome[6], dyslipidemia[7], chronic kidney disease[8] and cardiovascular events[9]. In addition, it is known that many patients with HUA develop gout eventually[10].

In order to elucidate the relationship between SUA and other variables, a previous study developed a prediction model based on cox proportional hazards regression for HUA using 10 selected variables[11]. Note that the risk factors for HUA remain controversial, in this study, we intended to establish a RF model through large sample from the perspective of data mining, with the aim of effectively predicting the prevalence of HUA, comprehensively exploring risk factors of HUA and timely guiding high-risk groups to take actions. Our cross-sectional study did not screen risk factors through prior knowledge but directly established a RF model to analyze and compare the correlations of all included factors.

Methods

Design

This cross-sectional study was carried out under a collaboration from various departments of Tianjin Medical University General Hospital in China, School of Medicine of Shanghai Jiao Tong University, College of Intelligence and Computing of Tianjin University in China and Hull York Medical School of Hull University in UK. During the period from September 2011 to April 2014, a total of 91,690 subjects participated in this mainly community-based health check program. All participants were asked to complete a questionnaire about medical history, lifestyle and habits.

Measurements and definitions

Anthropometric measurements and fasting blood tests of the participants were performed during their visits to our institutions. Height and weight were measured in centimeters and kilograms. Body mass index (BMI) was calculated by dividing weight (kg) by the square of height (m^2). Waist circumference (WC) was measured in centimeters. Systolic and diastolic blood pressures (SBP/ DBP) were measured by an automated sphygmomanometer while the subjects were in a seated position after resting for 5 minutes. Alanine aminotransferase (ALT), serum total cholesterol (TC), triglycerides (TG), blood urea (BUN), SUA, serum creatinine (Cr) and fasting plasma glucose (FPG) were measured by an auto-analyzer (Hitachi Model 7600 analyzer, Hitachi, Tokyo, Japan).

The criteria for diagnosis of HUA was SUA>420 μ mol/L in males and >360 μ mol/L in females[12].

Statistical analysis

All measurement data was represented as mean $\pm$ standard deviation. Independent-sample T test was used to compare the differences between the true positive set and true negative set. Odds ratio (OR) for HUA with 95% confidence interval (CI) was calculated by the binary logistic regression. The analysis was accomplished by using the statistical package for social sciences (SPSS version 22.0, Chicago, IL, USA).

RF model

In this study, the original data was cleaned before data mining to eliminate noise data, vacant data and inconsistent data, as well as some redundant data. Process of data cleaning was divided into four steps: deleting items containing vacant values, correcting items containing illegal values, normalization processing and data transformation (Appendix 1). The study cohort was divided randomly into training and validation datasets, with 80% of subjects assigned to training, and the remaining 20% utilized for validation. A RF model was trained for predicting HUA using 35 baseline clinical variables including continuous variables and categorical variables. RF is an algorithm that integrates multiple decision trees through the idea of Ensemble Learning that determines a consensus prediction for each observation by averaging the results of many individual recursive partitioning tree models[13, 14]. Each of the individual trees are fitted to a randomly selected subset of the observations, and utilize a random subset of the available predictors at each node as candidates for splitting[13]. Finally, the RF model integrates all categories and outputs the result based on a majority vote. The main advantages of RF are as follows. First, RF can handle input samples with high dimensional features and does not need dimension reduction. Second, RF can assess the importance of each feature in classification. Third, RF can balance the error to a certain extent for an unbalanced data set. Fourth, the models trained by RF can maintain accuracy even if there are many missing values. Last but not the least, any interaction or correlation between variables does not adversely affect the RF classification since it is capable of representing high order interactions[15]. The analysis was accomplished by using Python.

Evaluation Criteria

The discriminatory power of models was analyzed by ROC curve. ROC curves were constructed by plotting true positive versus the false positive fraction. Sensitivity (the probability of a positive test given the individual has the disease), specificity (SPC, the probability of a negative test given the individual does not have the disease), positive predictive value (PPV, the probability of having the disease given a positive test), and the negative predictive value (NPV, the probability of not having the disease given a negative test) were calculated for each cutoff score[16].

Results

Study characters

From 91,690 subjects participated, the training dataset was determined as 73,351 subjects, of which 42,085 was males (57.4%). The validation dataset consisted of 18,339 subjects, of which 10,522 was males (57.4%). All baseline variables in Table 1 were included in the RF model. The training and validation data sets were similar in terms of baseline covariates. Samples for statistical analysis are individuals in the validation dataset correctly diagnosed by the RF model (Figure1).

Evaluation of model predictive ability

There were 2,306 (21.9%) males in the validation set with HUA, and the detection rate was 68.6%. For females, 7817 individuals were in the set and 560 (7.2%) had HUA, with 78.6% of the detection rate. Classification matrix was shown in Supplementary Table 1 (Appendix 2). Area under the curve (AUC) was 0.732 in males and 0.837 in females (Figure 2). The sensitivity and specificity of the RF models were 0.686 and 0.656 in males, 0.786 and 0.738 in females, respectively. Besides, accuracy was 0.662 for males and 0.742 for females. While as for PPV, 0.359 was for males and 0.188 for females. NPV was 0.882 in males and 0.978 in females, which suggested that the model was suitable for initial screening.

Participants with or without HUA

In this part, we analyzed variables from the true negative set and the true positive set to obtain more distinct feature contrast. Variables with the top ten values in each gender were selected. Clinical baseline characteristics classified by gender were indicated in Table 2. All parameters showed significant difference except FPG in male ($P > 0.05$). BMI, weight and WC in individuals without HUA were significantly lower than individuals with HUA. TG, Cr and ALT showed the same results in both groups, but baselines of these characteristics were much higher in male than in female. In addition, the average age (58y) of HUA in females was much older than that in males (43y). The level of UA increased significantly with the rise of age in female ($P < 0.001$); but the level of UA showed a downward trend with increase of age in male.

Risks of developing HUA with different variables

In our study, we used the binary logistic regression model to calculate the risk of developing HUA in men and women with different variables (Table 3). Crude OR had no covariate, while adjusted OR included all other 9 selected variables as the covariates.

TG, Cr, Weight, BMI, ALT, TC and WC had hazardous effects on HUA in both genders. BUN and SBP had hazardous effects on the prevalence of HUA in females. Age was negatively correlated with prevalence in males and females under 50, but positively correlated with prevalence in females over 50 (Table 3). FPG in males showed no significance, but its negative effect appeared when included other covariates. Furthermore, we calculated the risk of HUA by including other covariables in model 2. Weight of females with HUA showed a positive effect with an OR of 1.182, but the effect disappeared when other covariates were included.

Discussion

HUA is defined by the finding of an abnormally high level of UA in the blood. HUA could be associated to factors from different aspects such as laboratory variables, food intake habits (including drinking history and smoking history) and education[17, 18]. Our previous study demonstrated correlations between lifestyle choices and HUA[4]. In the current study, we developed a RF based prediction model for HUA and analyzed its associated risk factors. In fact, we were the first to predict HUA using RF.

One important concern is parameters related to lipid metabolism. In our survey, TG showed the highest weight (feature value) in the judgment of RF models in both genders (Table 2). Elevated TG will lead to the production and utilization of more free fatty acid, accelerating synthesis of purines and UA[19]. Matsuura et al. has reported that visceral obesity patients had higher lipogenesis activity[20]. Tight correlations as further proof between weight, BMI, WC and HUA were made in our model. Besides, other indicators related to TG have been proved to be closely related to HUA: Chen et al. showed that the hyper triglyceridemic waist phenotype was strongly associated with HUA[21]. Xu et al. have shown that TG and non-high-density lipoprotein cholesterol have stronger relationship with HUA than other lipid indices[22]. All of these implied an important link between lipid metabolism and HUA.

FPG ranked tenth in males. Previous studies of relationship between UA levels and diabetes have yielded inconsistent findings, including positive, negative and no significant relationship[23–25]. Our current study demonstrated an inverse association between male FPG and HUA prevalence in diabetes individuals (OR = 0.446, CI = 0.311–0.639), and a positive association in individuals with normal glucose tolerance, while the latter disappeared when other covariates were included (OR = 1.073, CI = 0.897–1.284). Similarly, a bell-shaped relation between FPG and SUA levels has been shown in several studies[26]. The possible mechanism for a positive relationship between glucose and UA may be related to the dual biological properties of UA. UA usually has an antioxidative effect; however, it becomes a strong oxidant in the environment of metabolic syndrome[27]. Inflammation and oxidative stress induced by metabolic syndrome and HUA may predispose individuals to a higher risk for diabetes[24]. Biological mechanism underlying the relation between higher FPG and SUA levels is thought to be due to the uricosuric effect of glycosuria [28]. Glycosuria occurs when glucose in the renal tubules exceeds its maximum absorption capacity, which inhibits the reabsorption of UA at the same place. What needs to be emphasized here is that it is glycosuria, rather than FPG, that leads to increased UA excretion. Further studies, especially of UA in the normal glucose tolerance group, are urgently needed.

ALT ranked fifth in males and sixth in females. ALT is closely related to intrahepatic fat deposition and has been widely considered as a marker of nonalcoholic fatty liver (NAFLD) in some epidemiological studies[29, 30]. Many clinical studies have shown that HUA and NAFLD have similar metabolic disorders, including insulin resistance, dyslipidemia and visceral obesity[6, 31, 32]. Therefore, there may be a positive correlation between elevated SUA and elevated ALT. Another plausible explanation for the link

between HUA and ALT elevation is oxidative stress[33]. The production of UA is accompanied by the production of reactive oxygen species. In patients with NAFLD, increased SUA levels may alter endogenous antioxidant defenses of liver fat peroxidation, thereby promoting the progression of liver injury and leading to elevated ALT[29, 34].

The peak age of prevalence in females was completely opposite to that in males (Fig. 3). However, the curvilinear distribution of HUA prevalence between the sexes indicated that both sexes may be affected by hormonal factors with only a difference in degree. As one of the mechanisms underlying the gender discrepancy in the prevalence of HUA, the estrogen's uricosuric effect has been widely recognized[35]. Multivariate analysis found a negative and a positive association between HUA prevalence and age in the female groups under and over 50 years old (OR = 0.894, CI = 0.852–0.938; OR = 1.046, CI = 1.007–1.086), respectively. While the question of whether androgen could independently affect SUA levels, as in the case of estrogen, remains controversial. The urate transporter 1, a specific urate transporter, expresses higher in male mice than in female mice, which are positively affected by testosterone(T)[36], and it also has been reported that androgen played a certain role in promoting the catabolism of nucleotide[37], both of which suggest that T levels are positively correlated with UA levels. Rosen et al. showed no difference in the serum T levels between asymptomatic HUA and normouricemic group[38]. However, a few studies found a negative association between T levels and SUA, which are consistent with us[36, 39]. It has been reported that serum total T and free T concentrations fall by 0.8% and 2% per year in middle-aged men, which may provide an explanation for the high prevalence in older men[40]. Insulin resistance, obesity, alcohol intake may also be associated with higher prevalence in HUA with age [41–43]. The phenomenon that the prevalence of male HUA decreases during the third to the seventh decades then shoots up needs to be further studied. Besides, higher values among young male may be a secondary consequence of other pathological change[44].

SBP was only shown in female ranked tenth. Previous studies have shown that SUA helped to maintain blood pressure through both acute renal vasoconstriction (via stimulation of the renin angiotensin system) and chronic renal microvascular and interstitial disease (by inducing salt-sensitivity via activation MAP kinase, PDGF, and COX-2 systems)[45, 46]. When renal microvascular disease continues to progress (a lesion resembling arteriosclerosis), and sufficient narrowing of the arteriolar lumen occurs, a component of the hypertension becomes salt-driven, renal-dependent, and independent of UA levels[47], which may also explain the no-significance association between SUA and hypertension in the older female group[48] and males. In addition, Cr of male and female HUA patients in our study was significantly higher than that of non-HUA patients. It has been reported that gout patients had lower Cr clearance and fractional UA excretion[49]. Obermayr et al. found that UA levels of 7-8.9 mg/dl nearly doubled the risk for incident kidney disease (OR = 1.74, CI: 1.45–2.09), while UA levels > 9.0 mg/dl got a tripled risk (OR = 3.12, CI: 2.29–4.25)[50].

The study also has some limitations. Firstly, the dataset was based on a cross-sectional, single-center study, which may have selection bias and lack of representativeness, and such a study cannot provide causality information. Secondly, all these ostensible subjects might have some diseases they do not

know by themselves, which could influence SUA. Thirdly, data loss occurs when continuous variables are converted to categorical variables (labeled individual SUA levels with having HUA or not).

Conclusions:

In conclusion, we developed a RF based prediction model for HUA in general Chinese population by a cross-sectional dataset. The model demonstrated good stability and strong predictive power, which could be used to identify high-risk groups of HUA in the early stage and provide early warning and intervention.

References

1. Pillinger MH, Rosenthal P, Abeles AM: **Hyperuricemia and gout: new insights into pathogenesis and treatment.** *Bull NYU Hosp Jt Dis* 2007, **65**(3):215-221.
2. Ghei M, Mihailescu M, Levinson D: **Pathogenesis of hyperuricemia: recent advances.** *Curr Rheumatol Rep* 2002, **4**(3):270-274.
3. Yang J, Liu Z, Zhang C, Zhao Y, Sun S, Wang S, Zhao Y, Zhang Y, Li J, Lu F: **The prevalence of hyperuricemia and its correlates in an inland Chinese adult population, urban and rural of Jinan.** *Rheumatol Int* 2013, **33**(6):1511-1517.
4. Liu L, Lou SS, Xu K, Meng ZW, Zhang Q, Song K: **Relationship between lifestyle choices and hyperuricemia in Chinese men and women.** *Clin Rheumatol* 2013, **32**(2):233-239.
5. Liu R, Han C, Wu D, Xia X, Gu J, Guan H, Shan Z, Teng W: **Prevalence of Hyperuricemia and Gout in Mainland China from 2000 to 2014: A Systematic Review and Meta-Analysis.** *Biomed Res Int* 2015, **2015**:762820.
6. Khichar S, Choudhary S, Singh VB, Tater P, Arvinda RV, Ujjawal V: **Serum uric acid level as a determinant of the metabolic syndrome: A case control study.** *Diabetes Metab Syndr* 2017, **11**(1):19-23.
7. Zhang Y, Wei F, Chen C, Cai C, Zhang K, Sun N, Tian J, Shi W, Zhang M, Zang Y *et al*: **Higher triglyceride level predicts hyperuricemia: A prospective study of 6-year follow-up.** *J Clin Lipidol* 2018, **12**(1):185-192.
8. Li L, Yang C, Zhao Y, Zeng X, Liu F, Fu P: **Is hyperuricemia an independent risk factor for new-onset chronic kidney disease?: A systematic review and meta-analysis based on observational cohort studies.** *BMC Nephrol* 2014, **15**:122.
9. Li M, Hu X, Fan Y, Li K, Zhang X, Hou W, Tang Z: **Hyperuricemia and the risk for coronary heart disease morbidity and mortality a systematic review and dose-response meta-analysis.** *Sci Rep* 2016, **6**:19520.
10. Bardin T, Richette P: **Definition of hyperuricemia and gouty conditions.** *Curr Opin Rheumatol* 2014, **26**(2):186-191.

11. Cao J, Wang C, Zhang G, Ji X, Liu Y, Sun X, Yuan Z, Jiang Z, Xue F: **Incidence and Simple Prediction Model of Hyperuricemia for Urban Han Chinese Adults: A Prospective Cohort Study.** *Int J Environ Res Public Health* 2017, **14**(1).
12. Zhang Q, Lou SS, Meng ZW, Ren XJ: **Gender and age impacts on the correlations between hyperuricemia and metabolic syndrome in Chinese.** *Clin Rheumatol* 2011, **30**(6):777-787.
13. Gurm HS, Kooiman J, LaLonde T, Grines C, Share D, Seth M: **A Random Forest Based Risk Model for Reliable and Accurate Prediction of Receipt of Transfusion in Patients Undergoing Percutaneous Coronary Intervention.** *PLoS One* 2014, **9**(5).
14. Genuer R, Poggi JM, Tuleau-Malot C: **Variable selection using random forests.** *Pattern Recogn Lett* 2010, **31**(14):2225-2236.
15. Sugimoto K, Murata H, Hirasawa H, Aihara M, Mayama C, Asaoka R: **Cross-sectional study: Does combining optical coherence tomography measurements using the 'Random Forest' decision tree classifier improve the prediction of the presence of perimetric deterioration in glaucoma suspects?** *BMJ Open* 2013, **3**(10):e003114.
16. Liu J, Tang ZH, Zeng F, Li Z, Zhou L: **Artificial neural network models for prediction of cardiovascular autonomic dysfunction in general Chinese population.** *BMC Med Inform Decis Mak* 2013, **13**:80.
17. Liu H, Zhang XM, Wang YL, Liu BC: **Prevalence of hyperuricemia among Chinese adults: a national cross-sectional survey using multistage, stratified sampling.** *J Nephrol* 2014, **27**(6):653-658.
18. Poletto J, Harima HA, Ferreira SR, Gimeno SG: **Hyperuricemia and associated factors: a cross-sectional study of Japanese-Brazilians.** *Cad Saude Publica* 2011, **27**(2):369-378.
19. Lima WG, Martins-Santos MES, Chaves VE: **Uric acid as a modulator of glucose and lipid metabolism.** *Biochimie* 2015, **116**:17-23.
20. Matsuura F, Yamashita S, Nakamura T, Nishida M, Nozaki S, Funahashi T, Matsuzawa Y: **Effect of visceral fat accumulation on uric acid metabolism in male obese subjects: visceral fat obesity is linked more closely to overproduction of uric acid than subcutaneous fat obesity.** *Metabolism* 1998, **47**(8):929-933.
21. Chen S, Guo X, Dong S, Yu S, Chen Y, Zhang N, Sun Y: **Association between the hypertriglyceridemic waist phenotype and hyperuricemia: a cross-sectional study.** *Clin Rheumatol* 2017, **36**(5):1111-1119.
22. Xu J, Peng H, Ma Q, Zhou X, Xu W, Huang L, Hu J, Zhang Y: **Associations of non-high density lipoprotein cholesterol and traditional blood lipid profiles with hyperuricemia among middle-aged and elderly Chinese people: a community-based cross-sectional study.** *Lipids Health Dis* 2014, **13**:117.
23. Bandaru P, Shankar A: **Association between Serum Uric Acid Levels and Diabetes Mellitus.** *Int J Endocrinol* 2011.
24. Chien KL, Chen MF, Hsu HC, Chang WT, Su TC, Lee YT, Hu FB: **Plasma uric acid and the risk of type 2 diabetes in a Chinese community.** *Clin Chem* 2008, **54**(2):310-316.
25. Taniguchi Y, Hayashi T, Tsumura K, Endo G, Fujii S, Okada K: **Serum uric acid and the risk for hypertension and Type 2 diabetes in Japanese men: The Osaka Health Survey.** *J Hypertens* 2001,

26. Choi HK, Ford ES: **Haemoglobin A1c, fasting glucose, serum C-peptide and insulin resistance in relation to serum uric acid levels—the Third National Health and Nutrition Examination Survey.** *Rheumatology (Oxford)* 2008, **47**(5):713-717.
27. Hayden MR, Tyagi SC: **Uric acid: A new look at an old risk marker for cardiovascular disease, metabolic syndrome, and type 2 diabetes mellitus: The urate redox shuttle.** *Nutr Metab (Lond)* 2004, **1**(1):10.
28. Chen J, Qiu SH, Guo HJ, Li W, Sun ZL: **Increased urinary glucose excretion is associated with a reduced risk of hyperuricaemia.** *Diabetic Med* 2019, **36**(7):902-907.
29. Chen S, Guo X, Yu S, Sun G, Yang H, Li Z, Sun Y: **Association between Serum Uric Acid and Elevated Alanine Aminotransferase in the General Population.** *Int J Environ Res Public Health* 2016, **13**(9).
30. Jeong SK, Nam HS, Rhee JA, Shin JH, Kim JM, Cho KH: **Metabolic syndrome and ALT: a community study in adult Koreans.** *Int J Obes Relat Metab Disord* 2004, **28**(8):1033-1038.
31. Kuwabara M, Borghi C, Cicero AFG, Hisatome I, Niwa K, Ohno M, Johnson RJ, Lanaspas MA: **Elevated serum uric acid increases risks for developing high LDL cholesterol and hypertriglyceridemia: A five-year cohort study in Japan.** *Int J Cardiol* 2018, **261**:183-188.
32. Kim Y, Kang J, Kim GT: **Prevalence of hyperuricemia and its associated factors in the general Korean population: an analysis of a population-based nationally representative sample.** *Clin Rheumatol* 2018, **37**(9):2529-2538.
33. Gong S, Song J, Wang L, Zhang S, Wang Y: **Hyperuricemia and risk of nonalcoholic fatty liver disease: a systematic review and meta-analysis.** *Eur J Gastroenterol Hepatol* 2016, **28**(2):132-138.
34. Lanaspas MA, Sanchez-Lozada LG, Choi YJ, Cicerchi C, Kanbay M, Roncal-Jimenez CA, Ishimoto T, Li N, Marek G, Duranay M *et al*: **Uric acid induces hepatic steatosis by generation of mitochondrial oxidative stress: potential role in fructose-dependent and -independent fatty liver.** *J Biol Chem* 2012, **287**(48):40732-40744.
35. Yahyaoui R, Esteva I, Haro-Mora JJ, Almaraz MC, Morcillo S, Rojo-Martinez G, Martinez J, Gomez-Zumaquero JM, Gonzalez I, Hernando V *et al*: **Effect of long-term administration of cross-sex hormone therapy on serum and urinary uric acid in transsexual persons.** *J Clin Endocrinol Metab* 2008, **93**(6):2230-2233.
36. Park JW, Lee JH, Cho HJ, Ha YJ, Kang EH, Shin K, Byun SS, Lee EY, Song YW, Lee YJ: **Influence of androgen deprivation therapy on serum urate levels in patients with prostate cancer: A retrospective observational study.** *PLoS One* 2018, **13**(12).
37. Pizzichini M, Leoncini R, Vannoni D, Marinello E, Pagani R: **The influence of testosterone on purine nucleotide metabolism in rat liver.** *Life Sci* 1995, **57**(23):2127-2135.
38. Rosen R, Tomer Y, Carel R, Weinberger A: **Serum 17-beta-estradiol and testosterone levels in asymptomatic hyperuricaemic men.** *Clin Rheumatol* 1994, **13**(2):219-223.
39. Fukai S, Akishita M, Miyao M, Ishida K, Toba K, Ouchi Y: **Age-related changes in plasma androgen levels and their association with cardiovascular risk factors in male Japanese office workers.** *Geriatr*

Gerontol Int 2010, **10**(1):32-39.

40. Feldman HA, Longcope C, Derby CA, Johannes CB, Araujo AB, Coviello AD, Bremner WJ, McKinlay JB: **Age trends in the level of serum testosterone and other hormones in middle-aged men: Longitudinal results from the Massachusetts Male Aging Study.** *J Clin Endocr Metab* 2002, **87**(2):589-598.
41. Han TS, Lan L, Qu RG, Xu Q, Jiang RY, Na LX, Sun CH: **Temporal Relationship Between Hyperuricemia and Insulin Resistance and Its Impact on Future Risk of Hypertension.** *Hypertension* 2017, **70**(4):703-+.
42. Krishnan E, Pandya BJ, Chung L, Hariri A, Dabbous O: **Hyperuricemia in Young Adults and Risk of Insulin Resistance, Prediabetes, and Diabetes: A 15-Year Follow-up Study.** *Am J Epidemiol* 2012, **176**(2):108-116.
43. Villegas R, Xiang YB, Cai Q, Fazio S, Linton M, Li H, Elasy T, Zheng W, Shu XO: **Prevalence and determinants of hyperuricemia in middle-aged, urban Chinese men.** *Metab Syndr Relat Disord* 2010, **8**(3):263-270.
44. Yamanaka H: **Gout and hyperuricemia in young people.** *Curr Opin Rheumatol* 2011, **23**(2):156-160.
45. Mazzali M, Hughes J, Kim YG, Jefferson JA, Kang DH, Gordon KL, Lan HY, Kivlighn S, Johnson RJ: **Elevated uric acid increases blood pressure in the rat by a novel crystal-independent mechanism.** *Hypertension* 2001, **38**(5):1101-1106.
46. Watanabe S, Kang DH, Feng LL, Nakagawa T, Kanellis J, Lan H, Mazzali M, Johnson RJ: **Uric acid, hominoid evolution, and the pathogenesis of salt-sensitivity.** *Hypertension* 2002, **40**(3):355-360.
47. Johnson RJ, Feig DI, Herrera-Acosta J, Kang DH: **Resurrection of uric acid as a causal risk factor in essential hypertension.** *Hypertension* 2005, **45**(1):18-20.
48. Forman JP, Choi H, Curhan GC: **Plasma uric acid level and risk for incident hypertension among men.** *J Am Soc Nephrol* 2007, **18**(1):287-292.
49. Ma LL, Wei L, Chen HY, Zhang ZJ, Yu Q, Ji ZF, Jiang LD: **Influence of urate-lowering therapies on renal handling of uric acid.** *Clin Rheumatol* 2016, **35**(1):133-141.
50. Obermayr RP, Temml C, Gutjahr G, Knechtelsdorfer M, Oberbauer R, Klauser-Braun R: **Elevated Uric Acid Increases the Risk for Kidney Disease.** *J Am Soc Nephrol* 2008, **19**(12):2407-2413.

Tables

Table 1 Characteristics of patients in the training and the validation datasets

Characteristic	Training	Validation	P-value
Gender(male%)	42,085[57.4%]	10,522[57.4%]	1.000
Age(years)	45.621±14.710	45.587±14.747	0.781
Height(cm)	167.063±8.425	167.076±8.433	0.851
Weight(kg)	69.514±13.287	69.520±13.313	0.957
BMI(kg/cm ²)	24.782±3.649	24.782±3.668	0.985
WC (cm)	84.085±11.078	84.060±11.114	0.765
SBP(mmHg)	123.360±17.971	123.340±17.969	0.859
DBP(mmHg)	77.800±11.619	77.760±11.574	0.650
History of Smoking (non-case%)	54395[74.2%]	13649[74.4%]	0.713
History of Drinking(non-case%)	44357[60.5%]	11083[60.4%]	0.796
ALT(U/L)	24.780±21.169	24.710±19.535	0.704
TBIL(μmol/L)	12.827±5.750	12.756±5.716	0.133
BUN(mmol/L)	4.787±1.292	4.809±1.318	0.038
Cr(μmol/L)	70.920±15.382	71.190±17.718	0.060
TC(mmol/L)	5.128±1.000	5.139±1.002	0.202
TG(mmol/L)	1.539±1.295	1.554±1.324	0.165
FPG(mmol/l)	5.157±1.148	5.155±1.164	0.902
P-LCR(%)	24.737±6.870	24.756±6.915	0.746
MCV (fL)	89.423±4.436	89.442±4.482	0.612
MCH(pg)	29.888±1.823	29.895±1.825	0.659
MCHC(g/L)	334.180±11.250	334.200±11.207	0.863
MPV(fl)	9.899±0.913	9.897±0.931	0.801
Lymphocyte(×10 ⁹ /L)	1.966±0.543	1.966±0.558	0.974
Lymphocyte percentage(%)	34.955±7.394	34.912±7.425	0.480
WBC(×10 ⁹ /L)	5.711±1.395	5.719±1.536	0.491
Neutrophil(×10 ⁹ /L)	3.240±1.055	3.243±1.056	0.776
Neutrophil percentage(%)	56.149±7.845	56.160±7.916	0.864
RBC(×10 ¹² /L)	4.740±0.462	4.739±0.463	0.736
RDW-CV(%)	12.802±0.841	12.803±0.844	0.942
HCT(%)	42.319±3.898	42.315±3.904	0.902
PLT(×10 ⁹ /L)	221.690±53.536	221.220±53.055	0.293
PDW(fl)	12.506±1.881	12.511±1.885	0.748
Hb(g/L)	141.500±14.450	141.490±14.513	0.973

UWBC(/ μ L)	103.561 $\pm$ 2022.434	116.568 $\pm$ 2,147.257	0.442
URBC(/ μ L)	34.642 $\pm$ 1162.431	38.184 $\pm$ 1,217.594	0.715

BMI = body mass index, WC = waist circumference, SBP = systolic blood pressure, DBP = diastolic blood pressure, ALT = alanine aminotransferase, TBIL = total bilirubin, BUN = blood urea nitrogen, Cr = creatinine, TC = total cholesterol, TG = triglyceride, FPG = fasting plasma glucose, P-LCR = platelet-large cell ratio, MCV = mean corpuscular volume, MCH = mean corpuscular hemoglobin, MCHC = mean corpuscular hemoglobin concentration, MPV = mean platelet volume, WBC = white blood cell, RBC = red blood cell, RDW-CV = red cell distribution width-coefficient of variation, HCT = hematocrit, PLT = platelet, PDW = platelet distribution width, Hb = hemoglobin, UWBC = urine white blood cell, URBC = urine red blood cell.

Table 2 The baseline clinical characteristics in different genders

Variables	Individuals without HUA	Individuals with HUA	P value	Weight (Feature value)
Males				
TG	1.309 $\pm$ 0.941	2.639 $\pm$ 1.899	<0.001	0.074
Cr	75.980 $\pm$ 9.814	86.470 $\pm$ 23.216	<0.001	0.063
Weight	71.541 $\pm$ 9.355	84.388 $\pm$ 11.564	<0.001	0.049
BMI	24.304 $\pm$ 2.986	28.096 $\pm$ 3.199	<0.001	0.046
ALT	23.210 $\pm$ 17.406	40.840 $\pm$ 26.696	<0.001	0.045
Age	47.510 $\pm$ 14.795	42.830 $\pm$ 13.754	<0.001	0.040
TC	4.934 $\pm$ 0.898	5.469 $\pm$ 1.008	<0.001	0.036
WC	85.240 $\pm$ 8.705	94.680 $\pm$ 8.156	<0.001	0.036
Height	171.585 $\pm$ 5.968	173.171 $\pm$ 6.286	<0.001	0.031
FPG	5.315 $\pm$ 1.492	5.256 $\pm$ 0.956	0.134	0.031
Females				
TG	0.922 $\pm$ 0.486	2.097 $\pm$ 1.375	<0.001	0.090
Cr	57.620 $\pm$ 8.104	70.500 $\pm$ 19.131	<0.001	0.078
BMI	22.053 $\pm$ 2.606	27.520 $\pm$ 3.680	<0.001	0.056
WC	73.660 $\pm$ 7.527	88.340 $\pm$ 9.677	<0.001	0.050
BUN	4.167 $\pm$ 1.086	5.414 $\pm$ 1.485	<0.001	0.045
ALT	15.850 $\pm$ 9.775	26.270 $\pm$ 15.261	<0.001	0.044
Age	40.254 $\pm$ 12.575	58.058 $\pm$ 16.060	<0.001	0.042
Weight	56.965 $\pm$ 7.204	68.945 $\pm$ 10.555	<0.001	0.041
TC	4.876 $\pm$ 0.913	5.827 $\pm$ 1.082	<0.001	0.033
SBP	113.620 $\pm$ 14.367	136.990 $\pm$ 19.695	<0.001	0.032

TG = triglyceride, Cr = creatinine, BMI = body mass index, ALT = alanine aminotransferase, TC = total cholesterol, WC = waist circumference, FPG = fasting plasma glucose, BUN = blood urea, SBP = systolic blood pressure.

Feature value was rounded to three decimal places.

Table 3 The likelihood of developing HUA in different variables

Sex/Variables	Model 1	Model 2		
	Crude OR(CI)	P	Adjusted OR(CI)*	P
Males				
TG	2.994(2.780-3.225)	<0.001	3.388(3.015-3.807)	<0.001
Cr	1.083(1.077-1.090)	<0.001	1.187(1.172-1.201)	<0.001
Weight	1.132(1.124-1.141)	<0.001	1.349 (1.183-1.539)	<0.001
BMI	1.499(1.463-1.536)	<0.001	0.626(0.424-0.925)	0.019
ALT	1.051(1.047-1.055)	<0.001	1.035(1.029-1.041)	<0.001
Age	0.977(0.973-0.981)	<0.001	0.938(0.930-0.947)	<0.001
TC	1.798 (1.692-1.911)	<0.001	1.649(1.476-1.842)	<0.001
WC	1.137(1.128-1.146)	<0.001	1.094(1.068-1.119)	<0.001
Height	1.044(1.035-1.054)	<0.001	0.782(0.693-0.883)	<0.001
FPG	0.968(0.928-1.010)	0.135	0.512(0.462-0.568)	<0.001
NGT	1.542(1.408-1.698)	<0.001	1.073(0.897-1.284)	0.442
diabetes	0.828(0.699-0.982)	0.030	0.446(0.311-0.639)	<0.001
Females				
TG	6.514(5.568-7.621)	<0.001	5.708(4.199-7.758)	<0.001
Cr	1.118(1.106-1.131)	<0.001	1.239(1.205-1.274)	<0.001
BMI	1.808(1.728-1.891)	<0.001	1.857(1.600-2.155)	<0.001
WC	1.231(1.212-1.251)	<0.001	1.101(1.058-1.146)	<0.001
BUN	2.181(2.012-2.365)	<0.001	2.356(1.981-2.802)	<0.001
ALT	1.059(1.051-1.067)	<0.001	1.059(1.043-1.075)	<0.001
Age	1.090(1.082-1.099)	<0.001	0.991(0.973-1.010)	0.359
<50y	1.004(0.979-1.029)	0.770	0.894(0.852-0.938)	<0.001
≥50y	1.099(1.083-1.116)	<0.001	1.046(1.007-1.086)	0.020
Weight	1.182(1.165-1.198)	<0.001	0.981(0.936-1.028)	0.427
TC	2.458(2.226-2.713)	<0.001	2.001(1.590-2.519)	<0.001
SBP	1.075(1.069-1.082)	<0.001	1.050(1.037-1.064)	<0.001

OR = odds ratio, CI = confidence interval, TG = triglyceride, Cr = creatinine, BMI = body mass index, ALT = alanine aminotransferase, WC = waist circumference, TC = total cholesterol, FPG = fasting plasma glucose, NGT=normal glucose tolerance, BUN =blood urea, SBP =systolic blood pressure.

*Adjusted with other 9 selected covariates.

Figures

Figure 1

Figure showed the sample sizes of classifier models at different stages. 91,690 people were initially included in the experiment and were randomly divided into 80% as training set and 20% as verification

set. The samples for statistical analysis were individuals correctly diagnosed by RF model in the verification data set.

Figure 2

ROC curve showed the discriminatory power of RF for HUA by taking all risk factors into consideration Area under the curve was 0.732 (A) in male and 0.837 in female (B).

Figure 3

Figure showed the proportions of HUA in female and male with age. Age distribution of HUA varied greatly in gender. The peak age of prevalence in females was obviously different with that in males. The prevalence rate of women increased rapidly after menopause, while that of men was higher in the young and middle-aged years, then decreased, and then increased after 70y. The general population showed an increasing trend with age.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryAppendix.docx](#)