Peer Review File

Article information: https://dx.doi.org/10.21037/atm-22-1916

Reviewer A Comments:

Comment 1: I am curious as to why patients with a history of tumor, liver dysfunction, and renal failure are excluded, especially patients with liver disease who have a higher disease burden for CVD. If this risk stratification is used in the clinical setting, it may be worthwhile to explore if the model differs for subsets of patients with a higher disease burden of CVD.

Reply 1: It is true as the Reviewer suggested that liver cirrhosis, renal failure, and cancer are indeed CVD risk factors. Considering that the biochemical indicators of these patients are quite different from others due to the impact of disease and disease treatment, additional risk stratification models are required for these patients. In this study, the target population is mainly healthy or metabolic syndrome people. We are carrying out a prospective cohort of cardiovascular disease (ChiCTR2100042724). In the future, we will separately build the CVD risk stratification or CVD risk prediction model for these patients. We also find that the statement about abnormal liver function in the previous article is not accurate enough. We exclude patients with liver cirrhosis, and we modified it (see Page 6, line 96).

Changes in the text: Page 6, line 96

Comment 2: The CHD cohort includes self-reported diagnoses by patients. The accuracy of self-reported diagnosis has previously been raised for CVD. Given the CHD cohort was also diagnosed by symptoms, could the health record be queried to verify the diagnosis?

Reply 2: Thank you for the suggestion. Doctors would confirm the patient's self-report through the medical insurance system in practice. We added this description in the paper (see Page 7, lines 111-112).

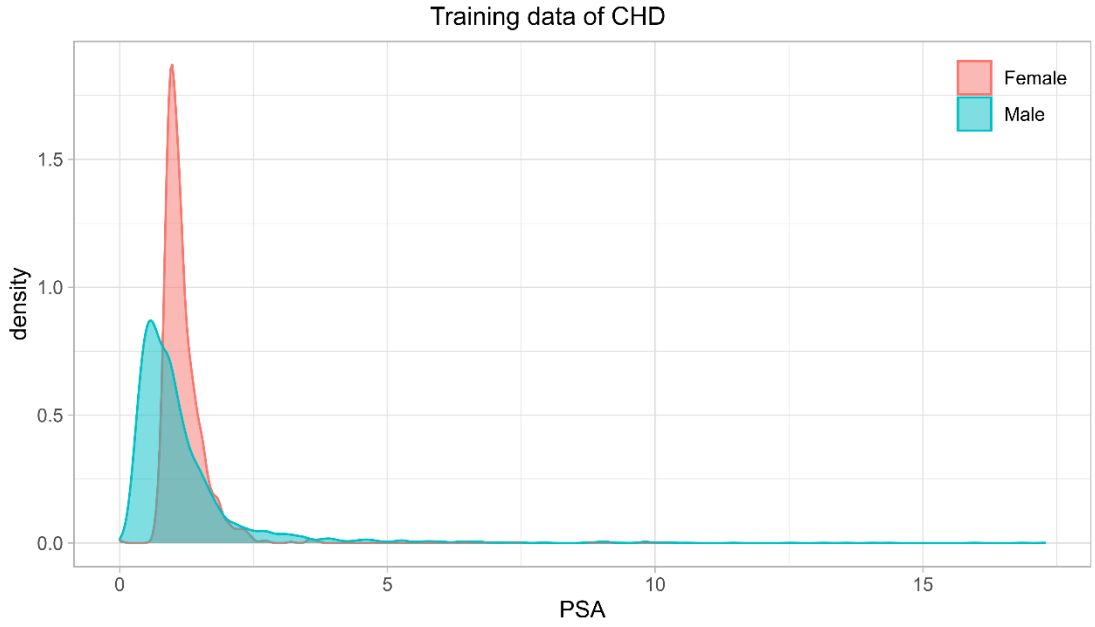
Changes in the text: Page 7, lines 111-112

Comment 3: While a sub-analysis by sex has been done, it may be worth considering having two separate models for male vs. female given nearly ¾ of the cohort male. With

PSA being one of the five crucial features in both CHD and CIS, there is reservations on whether features may differ if the cohort is stratified by sex.

Reply 3: Thank you for pointing this out. Previous classical models were often modeled separately for men and women. Considering the advantages of the machine learning model, we did not separate sex but conducted a subgroup analysis by sex. Our results showed that the contribution of hypertension to the models was different between men and women. As you suggested, we tried to build models separately for men and women. However, since both bootstrap and SHAP analysis required a lot of time, we were very sorry that we did not complete the reconstruction and analysis within three weeks. As for the PSA you mentioned, this feature was not in our final models, and we found that some women had higher PSA (see Figure below). According to previous studies, women with higher PSA may be related to hyperandrogenism (reference DOI: 10.1373/clinchem.2016.256198).

Changes in the text: N/A



Comment 4: It will be appealing to see how many participants were excluded from the study (Figure 1/flowchart).

Reply 4: As the Reviewer suggested, we added the number of people excluded due to age and specific diseases in the flowchart (see Figure 1).

Changes in the text: Figure 1

Reviewer C Comments:

Comment 1: There is no comparison based on statistics between AUCs of models in table 2. In fact, such difference seems too low, suggesting that these models could be comparable. If so, can the authors tell which model has the highest performance? And why didn't the authors utilize AUPRC in the imbalanced dataset like this?

Reply 1: Thank you for pointing these out. We added statistics between the AUCs of models (see Page 10, lines 179-180, Figure S6, S7). There was no statistically significant difference between the top three models with the highest AUC. The results of ROC curves and AUC values indicate that the complex balanced bagging classifier was the best model for the classification of CHD and the balanced bagging classifier was the best model for the classification of CIS.

As the Reviewer suggested, we also added an analysis of AUPRC (see Page 12, lines 225-227, 232-233, Figure 3B, 3D).

Changes in the text: Page 10, lines 179-180, Figure S6, S7, Page 12, lines 225-227, 232-233, Figure 3B, 3D

Comment 2: In terms of exclusion criteria, the authors excluded those with the history of tumor, liver dysfunction, or renal failure. However, I don't agree with the authors. Especially, tumor, liver fibrosis, and renal failure can be an important risk factor for stroke. This exclusion can lead to biased results that can not reflect real clinical fields.

Reply 2: It is true as the Reviewer suggested that liver cirrhosis, renal failure, and cancer are stroke risk factors. Considering that the biochemical indicators of these patients are quite different from others due to the impact of disease and disease treatment, additional risk stratification models are required for these patients. In this study, the target population is mainly healthy or metabolic syndrome people. We are carrying out a prospective cohort of cardiovascular disease (ChiCTR2100042724). In the future, we will separately build the CVD risk stratification or CVD risk prediction model for these patients.

Changes in the text: N/A

Comment 3: Detailed method of Imputation is needed. For example, what's the criteria (for example, threshold? or % of missing value of whole dataset?) when the authors decided to impute?

Reply 3: Thank you for the suggestion. We added a detailed method of imputation and a table of missing values and added new sentences read as follows “Indicators with a missing rate greater than 45% were discarded, and the remaining missing indicators were imputated (Table S1)”. (see Page 7, lines 115-116, Table S1).

Changes in the text: Page 7, lines 115-116, Table S1

Reviewer D Comments:

Comment 1: The role of temporality in the risk prediction model development is sorely lacking, which makes it very unclear if the models are actually prognostic or simply phenotyping models. There's little information provided on how the baseline characteristics were timed relative to the outcome events; rather, outcomes are simply described as dichotomous variables with respect to clinical history. It's similarly unclear how predictor feature information was collected with respect to time. When building prognostic models, it's generally important to establish a baseline, and then follow patients for a sufficient period of time to evaluate the prognostic information of those baseline data. In a similar vein, it's not clear what the risk prediction period is – 5-year? 10-year? Did patients have complete follow-up? Did some patients stop receiving medical care prior to the target risk period (i.e., right censoring)? What about death? Is left-truncation an issue (i.e., did all patients experience non-lethal events?) What data are being used to predict the "sequential" cohort? It seems like these patients would be at risk for at most 2 years (til end of 2019), and it's still not clear how timing of risk factors relative to events was handled.

Reply 1: Thank you for your rigorous consideration. We apologize for the ambiguity caused by "sequential". Sequential testing just represents the data collected after the year 2018 and is used for further evaluating the performance of the models. Our study only used cross-sectional data and we developed a stratification model rather than a predictive model. It is hard to make an accurate 10-year risk forecast in a cross-sectional study. Currently, we are carrying out a 10-year prospective cohort study in China (ChiCTR2100042724), and we will continue to improve the work of this paper and try to optimize the model.

Changes in the text: N/A

Comment 2: For risk prediction models, it's important to also demonstrate the calibration of the model, particularly in a clinical context where absolute risk is important for clinical decision making. The authors seem to focus solely on discrimination via AUC metrics, but it's not clear how well calibrated the predictions are. Figures 5A/B would be better replaced with assessments of model calibration.

Reply 2: Thank you for your advice. We added the calibration analysis. (see Page 13, lines 238-241, Figure S8). Calibration analysis showed that our models might

overestimate the risk score. This is the limitation of our models. We further explained it in the discussion section (see Page 17, lines 339-342).
Changes in the text: Page 13, lines 238-241, Figure S8; Page 17, lines 339-342

Comment 3: The issue of case imbalance feels overstated. This is generally not an issue unless you are using loss functions that fail to account for prevalence (e.g., absolute accuracy), and generally imbalance is a concern to the extent that sophisticated attention is necessary in only the most extreme of cases.
Reply 3: It is true as the Reviewer suggested that imbalanced data is a common problem, which brings challenges to class separation, and results in poor model performance. We have tried loss functions to deal with imbalanced data but the models did not perform well during the test. Traditional data processing is prone to representing the majority class and ignoring the characteristics of the minority class (reference DOI: 10.3390/ijerph17061828; 10.1016/j.jtcvs.2020.06.052). It is highly recommended to deal with class imbalance before modeling.
Changes in the text: N/A

Comment 4: It's unclear what the value of figure 5 C and D is given that age is also the strongest predictor in the model.
Reply 4: We revised this part according to the Reviewer's suggestion. Figure 5 C and D showed the relative risk of CHD and CIS patients was higher than controls in each age group rather than the predictive value of age. We have made further explanations in the text (see Page 15, lines 287-288).
Changes in the text: Page 15, lines 287-288

Comment 5: The age eligibility criterion (>30 years) may be a bit permissive given CHD/CIS is very rare in younger adults.
Reply 5: Thank you for your thoughtful comment. Although CHD/CIS is most prevalent in elderly individuals, it also affects younger adults. Driven by lifestyle changes associated with urbanization, CHD/CIS is on the rise in younger people. According to China cardiovascular health and disease report 2020 (reference DOI: 10.3969/j.issn.1000-3614.2021.06.001), the obesity rate of children is rising rapidly, which also increases the prevalence of hypertension and hyperlipidemia. The onset age

of CVD is becoming younger. Considering this trend, we have broadened the age range. Thus, early identification of individuals at high risk of developing CHD/CIS is of clinical importance.
Changes in the text: N/A

Comment 6: The VarianceThreshold criterion for categorical features seems a bit restrictive. Necessitating a 20% prevalence of the lower-prevalent conditions can lead to the removal of highly informative low-prevalence risk factors (e.g., diabetes).
Reply 6: Thank you for pointing this out. We used to apply VarianceThreshold based on different thresholds (as shown in the tables below). There are few patients with diabetes in our data. The indicator of diabetes will be retained when the threshold is 5%. Considering that continuous indicators include glycosylated hemoglobin and fasting blood glucose, we chose a strict threshold (20%).
Changes in the text: N/A

Categorical features selection of CHD dataset.

Threshold	Selected Features
5%	Sex, Smoking status, Drinking, Vascular stiffness, Hypertension, Diabetes, CCB, ARB, BLD, Urine WBC, UEM, Urine crystal, Antihypertensive drug, Hyperlipidemia
10%	Sex, Smoking status, Drinking, Vascular stiffness, Hypertension, CCB, BLD, Antihypertensive drug
15%	Sex, Smoking status, Drinking, Vascular stiffness, Hypertension, Antihypertensive drug
20%	Sex, Smoking status, Drinking, Vascular stiffness, Hypertension

ARB, Angiotensin receptor blocker; BLD, urinary occult blood; CCB, Calcium channel blocker; CHD, coronary heart disease; WBC, White blood cell; UEM, urinary erythrocyte morphology.

Categorical features selection of CIS dataset.

Threshold	Selected Features

5%	Sex, Smoking status, Drinking, Vascular stiffness, Hypertension, Diabetes, CCB, ARB, BLD, Urine WBC, UEM, Urine crystal, Antihypertensive drug
10%	Sex, Smoking status, Drinking, Vascular stiffness, Hypertension, BLD, Antihypertensive drug
15%	Sex, Smoking status, Drinking, Vascular stiffness, Hypertension, Antihypertensive drug
20%	Sex, Smoking status, Drinking, Vascular stiffness, Hypertension

ARB, Angiotensin receptor blocker; BLD, urinary occult blood; CCB, Calcium channel blocker; CIS, cerebral ischemic stroke; WBC, White blood cell; UEM, urinary erythrocyte morphology.