Research and Applications

Stroke risk prediction using machine learning: a prospective cohort study of 0.5 million Chinese adults

Matthew Chun,1,2 Robert Clarke,1,* Benjamin J. Cairns,1,3,* David Clifton,2,4 Derrick Bennett,1 Yiping Chen,1,3 Yu Guo,5 Pei Pei,5 Jun Lv,6 Canqing Yu,6 Ling Yang,1 Liming Li,6 Zhengming Chen,3,* and Tingting Zhu2,* on behalf of the China Kadoorie Biobank Collaborative Group†

1Clinical Trial Service Unit and Epidemiological Studies, Nuffield Department of Population Health, University of Oxford, Oxford, UK, 2Department of Engineering Science, University of Oxford, Oxford, UK, 3Medical Research Council, Population Health Research Unit, University of Oxford, Oxford, UK, 4Oxford-Suzhou Centre for Advanced Research, Suzhou, China, 5Chinese Academy of Medical Sciences, Beijing, China, and 6Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Sciences Center, Beijing, China

*Joint senior authors.
†Members of the China Kadoorie Biobank Collaborative Group are provided in the online supplement.

Corresponding Author: Robert Clarke, FRCP, Nuffield Department of Population Health, University of Oxford, Big Data Institute, Old Road Campus, Oxford OX 7LF, UK (robert.clarke@ndph.ox.ac.uk)

Received 20 November 2020; Revised 22 February 2021; Editorial Decision 28 March 2021; Accepted 1 April 2021

ABSTRACT

Objective: To compare Cox models, machine learning (ML), and ensemble models combining both approaches, for prediction of stroke risk in a prospective study of Chinese adults.

Materials and Methods: We evaluated models for stroke risk at varying intervals of follow-up (<9 years, 0–3 years, 3–6 years, 6–9 years) in 503,842 adults without prior history of stroke recruited from 10 areas in China in 2004–2008. Inputs included sociodemographic factors, diet, medical history, physical activity, and physical measurements. We compared discrimination and calibration of Cox regression, logistic regression, support vector machines, random forest, gradient boosted trees (GBT), and multilayer perceptrons, benchmarking performance against the 2017 Framingham Stroke Risk Profile. We then developed an ensemble approach to identify individuals at high risk of stroke (>10% predicted 9-yr stroke risk) by selectively applying either a GBT or Cox model based on individual-level characteristics.

Results: For 9-yr stroke risk prediction, GBT provided the best discrimination (AUROC: 0.833 in men, 0.836 in women) and calibration, with consistent results in each interval of follow-up. The ensemble approach yielded incrementally higher accuracy (men: 76%, women: 80%), specificity (men: 76%, women: 81%), and positive predictive value (men: 26%, women: 24%) compared to any of the single-model approaches.

Discussion and Conclusion: Among several approaches, an ensemble model combining both GBT and Cox models achieved the best performance for identifying individuals at high risk of stroke in a contemporary study of Chinese adults. The results highlight the potential value of expanding the use of ML in clinical practice.

Key words: stroke, cardiovascular diseases, machine learning, risk assessment, China

© The Author(s) 2021. Published by Oxford University Press on behalf of the American Medical Informatics Association. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
The China Kadoorie Biobank (CKB) is a prospective cohort study population. The aims of this study were to (i) compare Cox and ML models for prediction of risk of stroke in China at varying intervals of follow-up; (ii) identify individuals for whom ML models might be superior to conventional Cox-based approaches for stroke risk prediction; and (iii) develop and evaluate an ensemble model combining both approaches to identify individuals at high risk of stroke.


downloads from https://academic.oup.com/jamia/article/28/8/1719/6272889 by Oxford University user on 03 August 2021
ual did not have a stroke) based on individual-level characteristics. Feature importance for the decision tree was analyzed using the Gini importance metric. Screening approaches to identify individuals at high risk of stroke were then compared in the test set, including a Cox-only approach, best ML-only approach, and an ensemble approach that used the trained decision tree to select a model in cases of disagreement between Cox and the best ML model.

### Statistical analysis

The Cox and ML models were assessed for risk discrimination and calibration performance in the validation set. Risk discrimination refers to the ability to correctly discriminate between individuals with and without stroke, and was evaluated using the area under the receiver operating characteristic curve (AUROC), with higher AUROCs indicating better risk discrimination. Calibration refers to the similarity between observed and predicted numbers of stroke events for each predicted risk decile, and was evaluated using chi-squared test statistics ($\chi^2$) from the Hosmer-Lemeshow test (for binary classification models) and Nam-D’Agostino test (for survival models), with lower $\chi^2$ values indicating better calibration. The 95% confidence intervals were estimated for AUROC and $\chi^2$ values using 1000 bootstrapped samples from the validation set.

Risk screening approaches were evaluated in the test set, using sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy. Agreement between approaches was assessed using Cohen’s kappa ($\kappa$). For the ensemble screening approach, the decision tree to select between Cox and the best ML model was evaluated for accuracy as well as discriminatory performance using AUROC in the test set.

Statistical analyses were performed using Python version 3.7.0 and R version 3.6.1. Cox models were implemented using the lifelines package33 version 0.21.1 with LASSO variable selection performed in R using the glmnet package34 version 3.0-2. RSF models were implemented using the ranger package35 version 0.12.1. LR, SVM, and GBT models were implemented using scikit-learn tool-kit36 version 0.19.2, and MLP models were implemented using keras37 version 2.3.1. tSNE visualizations were implemented using the Rtsne package version 0.15.38

### RESULTS

Among the included study participants, the mean (SD) age was 51.9 (10.6) years and 59% were women (Table 1). During 9 years of follow-up, a total of 43,234 individuals had a first stroke (Supplementary Figure S1). The incidence of stroke was higher in men than in women (9.5% vs 7.9%) and varied by more than 5-fold between the 10 study areas. Compared with those who had no stroke, individuals who had a first stroke were older and more likely to have prior history of CHD, diabetes, or hypertension (Table 1). Overall, men and women had similar proportions with prior history of CHD (2.5% vs 3.0%), diabetes (5.3% vs 6.0%), and use of blood pressure-lowering medication (9.9% vs 11.4%), but the prevalence of current smoking was much higher in men than in women (67.7% vs 3.2%).

**Comparisons of cox versus ML models to predict risk of stroke**

The Cox model and ML models all outperformed the 2017 Framingham Stroke Risk Profile (FSRP) with and without recalibration and refitting and achieved similar discrimination for 9-year risk of stroke, with GBT yielding marginally higher AUROCs than other models for both men and women (Table 2, Supplementary Figure S2). However, calibration performance varied substantially between models. RSF, LR, and GBT all yielded lower $\chi^2$ values than Cox.
models in both men and women, with GBT showing significant improvements in calibration performance (Table 2, Supplementary Figure S3). MLP and SVM (after isotonic regression) were observed to yield good calibration for women but were poorly calibrated for men. Nevertheless, all models were significantly better calibrated than the original FSRP, which severely underestimated stroke risk in the CKB cohort (Figure 2, Supplementary Figure S3). Calibration plots (Figure 2) indicated that calibration for all models was better in women than in men, with models underestimating risk of stroke in men compared with women. Furthermore, deviation from perfect calibration was more extreme in those at highest risk of stroke (Figure 2) and in older individuals (Supplementary Figure S4). Due to its high AUROC and improvements to calibration over the traditional Cox model, GBT was identified as the best-performing ML model and was selected for further risk screening analyses.

Evaluation of risk prediction models at varying intervals of follow-up (0–3 years, 3–6 years, 6–9 years) demonstrated comparable relative performance between models (Figure 3, Supplementary Tables S1–S3). However, binary classification models (LR, SVM, GBT, and MLP), which were retrained for each prediction task, yielded substantial improvements over survival models (FSRP, Cox, RSF) at all intervals of follow-up and for both sexes (Figures 3A and 3B). The AUROCs for all models decreased monotonically for later intervals of follow-up. For most models, calibration also declined at later intervals of follow-up, but was less sensitive to changes in times scales than discrimination performance (Figures 3C and 3D). At all intervals of follow-up, MLP and SVM (after isotonic regression) had a tendency for poor calibration in men, but were well-calibrated in women. Once again, the FSRP without recalibration and refitting, yielded the worst calibration, substantially underestimating stroke risk at all intervals of follow-up.

GBT remained among the best-performing of the ML models based on discrimination and calibration metrics. Additional analyses of the most important GBT features for later intervals of follow-up (Supplementary Table S4) indicated comparable performance in men and women with emphasis on risk factor indicators related to

### Table 1. Distribution of established risk factors for stroke in men and women by presence or absence of stroke during follow-up

| Risk factors in 2017 Framingham Stroke Risk Profilea | Men | | | Women | | |
|---|---|---|---|---|---|---|
| No Stroke (n = 185 706) | Stroke (n = 19 387) | No Stroke (n = 274 902) | Stroke (n = 23 647) | |
| Age, mean, year | 51.8 | 60.7 | 50.6 | 59.6 |
| Current smoking, % | 68.5 | 59.9 | 3.1 | 4.8 |
| Coronary heart disease, % | 2.1 | 6.4 | 2.5 | 9.4 |
| Age 65 yrs+, % | 13.8 | 39.9 | 10.7 | 34.0 |
| Diabetes at age < 65 yrs, % | 3.7 | 6.5 | 4.0 | 8.1 |
| Diabetes at age 65+ yrs % | 1.1 | 4.8 | 1.3 | 6.0 |
| BP-lowering treatment, % | 8.6 | 22.3 | 10.1 | 26.3 |
| SBP-un treated, mean, mmHg | 130 | 142 | 126 | 138 |
| SBP- treated, mean, mmHg | 148 | 153 | 150 | 155 |

Note: “No Stroke” column includes individuals who remained stroke-free until being censored, even if lost to follow-up before 9 years.

aAtrial fibrillation is a part of the Framingham Stroke Risk Profile but was not recorded in CKB.

### Table 2. Discrimination and calibration performance for prediction of 9-year risk of stroke. Comparisons included the 2017 Framingham Stroke Risk Profile (FSRP), a recalibrated and refitted FSRP, Cox, random survival forest (RSF), logistic regression (LR), support vector machine (SVM), gradient boosted tree (GBT), and multilayer perceptron (MLP) models

| Model Type | Discrimination | Calibration | | | Discrimination | Calibration |
|---|---|---|---|---|---|---|
| | AUROCs [95%CI] | | | | AUROCs [95%CI] | |
| FSRP | 0.781 [0.772-0.790] | 5541 [4996-6107] | | | 0.772 [0.764-0.780] | 19402 [17784-21019] |
| Recalibrated and refitted FSRP | 0.824 [96-185] | | | | 0.825 [89-172] | |
| Cox | 0.829 [83-166] | | | | 0.831 [36-95] | |
| RSF | 0.826 [36-90] | | | | 0.832 [34-85] | |
| LR | 0.831 [31-86] | | | | 0.832 [24] | |
| SVM | 0.830 [31-86] | | | | 0.831 [11-41] | |
| GBT | 0.826 [582-852] | | | | 0.832 [30-69] | |
| MLP | 0.831 [410-627] | | | | 0.833 [8-35] | |
age, blood pressure, physical activity, and geographic area. However, characteristics, such as retirement and number of children, became relatively more important at later intervals of follow-up (3–6 years and 6–9 years) than at 0–3 years from baseline.

Prediction of best model for risk screening

After identifying GBT as the best-performing ML model, risk predictions estimated using the Cox and GBT models were used to screen individuals at high risk of stroke in the validation set in order to train a decision tree to select which model to use for a particular individual, given disagreement about the individual’s risk classification. For the purposes of this study, individuals were classified as “high-risk” if they had >10% predicted risk of stroke in 9 years. In cases of disagreement between models, either the Cox or GBT model was identified as the better classifier if it classified an individual with stroke as “high-risk” or classified a stroke-free individual as “not high-risk”.

The t-SNE visualizations of individuals in the validation set and test set (Figure 4) indicated high levels of agreement between both the Cox and GBT models for stroke risk prediction in both men and women. Disagreements occurred in only 5% of men and women, of whom 10% of men and 12% of women would go on to experience a stroke event.

Among the 4281 men and 6441 women in the test set (after excluding stroke-free, right-censored individuals), sex-specific decision trees were used to resolve 590 disagreements (219 in men, 371 in women) and selected GBT as the better classifier 31% of the time (34% for men, 29% for women). Among individuals in the test set for whom the Cox and GBT models disagreed, the decision trees (Supplementary Figures S5–S6) yielded good discrimination for model selection (AUROC: 0.71 in men, 0.74 in women) and correctly selected the better classifier with an accuracy of 73% in men and 70% in women. The most important features for selecting the better classifier included blood pressure, age, and geographic area-related risk factor indicators for both sexes—in addition to additional features reported in the online supplement (Supplementary Table S5).

Comparison of risk screening approaches

In both men and women, the ensemble approach yielded a higher accuracy (76% in men, 80% in women) and specificity (76% in men, 81% in women) compared with either Cox-only or GBT-only approaches (Table 3). The ensemble approach also yielded the highest PPV in men (26%). The GBT-only approach yielded the highest sensitivity (80% in men, 74% in women), and the Cox-only approach did not provide the best performance using any metric. Differences in performance were incremental for all metrics with overlapping 95% confidence intervals. The confusion matrices for each screening approach are reported for both sexes in the online supplement (Supplementary Table S6). Good agreement was observed between all models with χ values ranging from 0.85 to 0.96 (Supplementary Table S7).

DISCUSSION

In this study, involving almost a 100-fold larger population than the original Framingham Study,13 we developed novel risk scores for prediction of stroke in a contemporary Chinese cohort. Previous population-based prospective studies17–19 highlighted the need for novel risk scores for use in Chinese adults and proposed new Cox-derived models based on these populations. However, the models derived in the present study were based on a substantially larger (5-fold compared to the China-PAR study17) and more contemporary population.

In contrast to previous studies of risk prediction of stroke in Chinese adults, we compared both conventional Cox model-based approaches and ML techniques for risk prediction to assess the potential of ML techniques for improved risk prediction. Consistent
with findings for cardiovascular risk prediction, we demonstrated that ML techniques improved 9-yr risk prediction of stroke over Cox models, with GBT providing the best discrimination and calibration performance. Improvements over the Cox model were particularly evident for binary classification models that predicted stroke at narrower intervals of follow-up (0–3 years, 3–6 years, and 6–9 years from baseline). This may be due to the fact that survival models, such as Cox and RSF are optimized across the overall 9-year follow-up period, while binary classification models are able to be retrained for optimal performance in each particular time interval of interest. All models substantially outperformed the 2017 Framingham Stroke Risk Profile, which greatly underestimated stroke risk in CKB.

While the discrimination improvements of ML over Cox models were marginal, such incremental improvements can translate to meaningful population health benefits. For example, a recent analysis of 100 000 UK adults reported that polygenic risk scores for CVD with improvements of just 0.012 in the C-index could help to prevent 7% more CVD events than conventional risk scores alone. Moreover, the substantial calibration improvements of ML approaches, such as GBT, over Cox models are highly relevant for clinical practice, in which decisions on initiation of drug treatment may be informed by defined risk thresholds. Contemporary clinical guidelines recommend using absolute risk predictions from Cox models to screen individuals at high risk of stroke, who are then prioritized for initiation of drug treatments. For example, the 2013 guidelines of the American College of Cardiology and the American Heart Association (ACC–AHA) recommend initiation of statin therapy for those with a ≥7.5% 10-year CVD risk as assessed by the Pooled Cohort Equations, while in the UK, the cutoff is ≥10% risk as assessed by QRISK3. In such settings, underestimation of stroke risk due to poor calibration of models could result in failure to identify high-risk individuals who would benefit from statins or other preventative drug treatments.

Using a threshold of ≥10% 9-year stroke risk, we found that an ensemble approach that combined Cox and GBT models had a higher accuracy, specificity, and PPV for stroke prediction than either the Cox-only or GBT-only approaches. However, such improvements were marginal and warrant assessment for reproducibility in external validation studies. We have provided statistical

![Figure 3. Discrimination (subplots A and B) and calibration (subplots C and D) performance in men and women, respectively, for risk prediction of stroke at various time scales (0–3 years, 3–6 years, 6–9 years after baseline). Comparisons made between the 2017 Framingham Stroke Risk Profile (FSRP), a recalibrated and refitted FSRP, Cox, random survival forest (RSF), logistic regression (LR), support vector machine (SVM), gradient boosted tree (GBT), and multilayer perceptron (MLP) models.](https://academic.oup.com/jamia/article/28/8/1719/6272889)
In contrast with typical ensemble approaches that use voting or averaging of base model outputs, a major strength of our ensemble approach is its default reliance on the Cox model. Cox model-based approaches to risk prediction are widely used in clinical practice, and their relative simplicity and interpretability have been challenges to the adoption of novel ML-based methods.\textsuperscript{42} We found that the component Cox and GBT models failed to agree on risk prediction for stroke in about 5% of individuals, and of these, our proposed ensemble approach selected GBT as the better-performing model about one-third of the time. This suggests that, in practice, our ensemble approach would override the risk classification of the Cox model for only a small proportion (1%–2%) of individuals. Meanwhile, clinicians could continue to use the output from Cox model-derived scores, without any loss of predictive performance for the vast majority of individuals at high risk of stroke. Rather than changing the existing paradigm of stroke risk prediction, our proposed ensemble approach has been designed as an incremental change to clinical practice, which could help to facilitate more widespread use and trust of ML methods for health risk prediction.\textsuperscript{43}

Other barriers to the adoption of complex high-dimensional models for application in clinical practice include the availability of certain risk factor data and the need for regular updating and recalibration of such models. However, as electronic health records (EHRs) become more detailed and widespread, they may mitigate these issues by providing detailed individual-level data and enabling automatic updating and recalibration of complex ML models to local practices.\textsuperscript{43}

This study had several limitations. First, atrial fibrillation (AF), which is commonly included in risk scores for CVD and stroke,\textsuperscript{10,13} was not recorded in the CKB and could not be included in the models. However, other population-based studies of comparable age groups in China indicated that the prevalence of AF was substantially lower in China than in the Framingham Stroke Risk Profile cohort (1.7% vs 7.1%).\textsuperscript{13,44} Hence, omission of AF is unlikely to have had a material impact on stroke risk prediction in CKB. Second, the exclusion of right-censored data is an inherent limitation of training the binary classification models presented in this study. Although few participants in CKB were lost to follow-up, and exclusion of these individuals did not lead to a reduction in model performance, care should be taken when developing similar models in other study populations.
CONCLUSIONS

Novel risk scores for stroke have been developed using data from a contemporary cohort of 0.5 million Chinese adults. Use of ML techniques improved risk prediction over traditional Cox model approaches, with GBT providing the best discrimination and calibration performance. An ensemble approach was also proposed to screen for individuals at high risk of stroke who may benefit from more intensive treatment. The ensemble approach identified high-risk individuals with marginal improvements to accuracy, specificity, and PPV over either Cox or GBT models alone. By identifying a small portion of individuals who would benefit from ML predictions, our ensemble approach provides an incremental benefit beyond current clinical practice that has potential to translate into important benefits for population health and facilitate the adoption of ML-based risk calculators in clinical practice.

FUNDING

The baseline survey was supported by the Kadoorie Charitable Foundation, Hong Kong, China and funding for long-term follow-up was provided by the UK Wellcome Trust (202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z), Chinese National Natural Science Foundation (81390540, 81390541, 81390544), and the National Key Research and Development Program of China (2016YFC0900500, 2016YFC0900501, 2016YFC0900504, 2016YFC1303904). Core funding was provided to the CTSU, University of Oxford, by the British Heart Foundation, the UK Medical Research Council, and Cancer Research UK. MC was supported by a Rhodes Scholarship. BC is supported by a Nuffield Department of Population Health Senior Research Fellowship. The University of Oxford Medical Research Council (MRC) Population Health Research Unit is funded through a strategic partnership between the MRC and the University of Oxford. The research was also supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

AUTHOR CONTRIBUTIONS

Study concept and design: MC, TZ, DC, BC, and RC. Data collection: RC, DB, YC, PP, JL, CY, LY, LL, and ZC. Data analysis and interpretation: MC, TZ, DC, BC, and RC. Drafting of the manuscript: MC, TZ, DC, BC, and RC. Critical revision of the manuscript: all authors. Final approval: all authors.

DATA AVAILABILITY

The data underlying this article are available in the article and in its online supplementary material. Additional computer code used to estimate the machine learning and other statistical risk prediction models of stroke are available at: https://github.com/ckbiobank/ckb-stroke-risk-models.

SUPPLEMENTARY MATERIAL

Supplementary material is available at Journal of the American Medical Informatics Association online.

ACKNOWLEDGMENTS

A full list of members of the China Kadoorie Biobank Collaborative Group is included in the online supplement.

CONFLICT OF INTEREST STATEMENT

None of the authors have any conflicts of interest in relation to this report.

REFERENCES

1. Feigin VL, Krishnamurthi RV, Parmar P, GBD 2013 Stroke Panel Experts Group, et al. Update on the global burden of ischemic and hemorrhagic stroke in 1990–2013: the GBD 2013 study. Neuroepidemiology 2015; 45 (3): 161–76.
2. World Stroke Organization. Statement on Stroke Care in China: Statement by the China Stroke Stroke Organization and endorsed by WSO. World Stroke Organization website. 2019. https://www.world-stroke.org/news-and-blog/news/statement-on-stroke-care-in-china-june Accessed September 20, 2020.
3. Institute for Health Metrics and Evaluation. GBD Compare. Seattle, WA: IHME, University of Washington. 2015. https://vizhub.healthdata.org/gbd-compare Accessed September 20, 2020
4. Meschia JF, Bushnell C, Boden-Albala B, Council on Hypertension, et al. Guidelines for the primary prevention of stroke: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2014; 45 (12): 3754–832.
5. Li XY, Shi ZW, Zhao D, writing group of 2019 Chinese expert consensus statement on aspirin application in primary prevention of cardiovascular disease, et al. 2019 Chinese expert consensus statement on aspirin application in primary prevention of cardiovascular disease. Chin Med J (Engl) 2020; 133 (10): 1221–3.
6. Zhao JJ, He QQ, Gong SY, et al. Status and costs of primary prevention for ischemic stroke in China. J Clin Neurosci 2013; 20 (10): 1427–32.
7. Brainin M, Feigin V, Martins S, et al. Cut stroke in half: polypill for primary prevention in stroke. Int J Stroke 2018; 13 (6): 633–47.
8. Goff DC, Jr., Lloyd-Jones DM, Bennet G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014; 129 (25, suppl 2): S49–5.
9. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ 2007; 335 (7611): 136.
10. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. BMJ 2008; 336 (7659): 1473–82.
11. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ 2017; 357: j2099.
12. Wolf PA, D’Agostino RB, Belanger AJ, et al. Probability of stroke: a risk profile from the Framingham Study. Stroke 1991; 22 (3): 312–8.
13. Dufouil C, Besser A, McLaren LA, et al. A revised Framingham Stroke Risk Profile to reflect temporal trends. Circulation 2017; 135 (12): 1145–59.
14. Muntner P, Colantonio LD, Cushman M, et al. Validation of the Atherosclerotic Cardiovascular Disease Pooled Cohort Risk Equations. JAMA 2014; 311 (14): 1406–15.
15. Collins GS, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. BMJ 2010; 340; c2442.
16. Flueckiger P, Longstreth W, Herrington D, et al. Revised Framingham Stroke Risk Score, nontraditional risk markers, and incident stroke in a multiethnic cohort. Stroke 2018; 49 (2): 363–9.
17. Xing X, Yang X, Liu F, et al. Predicting 10-year and lifetime stroke risk in Chinese population. The China-PAR project. Stroke. Stroke 2019; 50 (9): 2371–8.
18. Chien KL, Su TC, Hsu HC, et al. Constructing the prediction model for the risk of stroke in a Chinese population: report from a cohort study in Taiwan. Stroke 2010; 41 (9): 1858–64.
19. Leung JYY, Lin SL, Lee RSY, et al. Longitudinal study of adults with atrial fibrillation in Hong Kong. Hong Kong Med J 2018; 24: S8–11.
20. Weng SF, Reps J, Kai J, et al. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One 2017; 12 (4): e0174944.
21. Khosla A, Cao Y, Lin CCY, et al. An integrated machine learning approach to stroke prediction. KDD 2010; 183–192.
22. Hung CY, Chen WC, Lai PT, et al. Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database. IEEE EMBC 2017: 3110–3.
23. Letham B, Rudin C, McCormick TH, et al. Interpretable classifiers using rules and Bayesian analysis: building a better stroke prediction model. Ann Appl Stat 2015; 9: 1350–71.
24. Li X, Bian D, Yu J, et al. Using machine learning models to improve stroke risk level classification methods of China national stroke screening. BMC Med Inform Decis Mak 2019; 19 (1): 261.
25. Chen Z, Lee L, Chen J, et al. Cohort profile: the Kadoorie Study of Chronic Disease in China (KSCDC). Int J Epidemiol 2005; 34 (6): 1243–9.
26. Chen Z, Chen J, Collins R, China Kadoorie Biobank (CKB) orative group, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol 2011; 40 (6): 1652–66.
27. Pang Y, Kartsonaki C, Guo Y, et al. Diabetes, plasma glucose and incidence of pancreatic cancer: a prospective study of 0.5 million Chinese adults and a meta-analysis of 22 cohort studies. Int J Cancer 2017; 140 (8): 1781–8.
28. Chen Y, Wright N, Guo Y, China Kadoorie Biobank Collaborative Group, et al. Mortality and recurrent vascular events after first incident stroke: a 9-year community-based study of 0.5 million Chinese adults. Lancet Glob Health 2020; 8 (4): e580–90.
29. Suthaharan S. Machine Learning Models and Algorithms for Big Data Classification. Boston, MA: Springer; 2016: 237–69.
30. Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: I. development, internal validation, and assessing the incremental value of a new (bio)marker. Heart 2012; 98 (9): 683–90.
31. D’Agostino RB, Byung-Ho N. Evaluation of the performance of survival analysis models: discrimination and calibration measures Handb Stat 2004;23:1–25.
32. Demler OV, Paynter NP, Cook NR. Tests of calibration and goodness of fit in the survival setting. Stat Med 2015; 34 (10): 1659–80.
33. Davidson-Pilon C, Kalderstam J, Zvich P, et al. CamDavidsonPilon/lifelines: v0.21.1 (Version v0.21.1). Zenodo 2019. doi: 10.5281/zenodo.2652543.
34. Friedman J, Hastie T, Tibshirani R. glmnet: lasso and elastic-net regularized generalized linear models. R package version 3.0-2. 2019. http://CRAN.R-project.org/package=glmnet Accessed April 8, 2020.
35. Wright MN, Ziegler A. A fast implementation of random forests for high dimensional data in C++ and R. J Stat Soft 2017; 77 (1): 1–17.
36. Pedregosa F, Varoquaux G, Gramfort A. Scikit-learn: machine learning in Python. JMLR 2011; 12 (85): 2825–30.
37. Chollet F. Keras. GitHub. 2015. https://github.com/fchollet/keras Accessed April 8, 2020.
38. Krijthe JR. T-distributed stochastic neighbor embedding using Barnes-Hut implementation. R package version 0.15. 2015. https://github.com/jkrijthe/rtSne Accessed April 8, 2020.
39. Sun L, Pennells L, Kaptoge S, et al. Polygenic risk scores in cardiovascular risk prediction: A cohort study and modelling analyses. PLoS Med 2021; 18 (1): e1003498.
40. Stone NJ, Robinson JG, Lichtenstein AH, American College of Cardiology/American Heart Association Task Force on Practice Guidelines, et al. ACC/AHA 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2014; 63 (25, Pt B): 2889–934.
41. National Institute for Health and Care Excellence. CVD risk assessment and management: Scenario: Management of people with an estimated risk of 10% or more. National Institute for Health and Care Excellence website. 2019. https://nice.org.uk/topics/cvd-risk-assessment-management/managements/cvd-risk-10percent-or-more/ Accessed November 5, 2020.
42. Deprose WK, Bunt N, Hua N, et al. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. JAMIA 2020; 27 (4): 592–600.
43. de Vries TI, Visseren FLJ. Cardiovascular risk prediction tools made relevant for GPs and patients. Heart 2021; 107 (4): 312–40.
44. Zhou Z, Hu D. An epidemiological study on the prevalence of atrial fibrillation in the Chinese population of mainland China. J Epidemiol 2008; 18 (5): 209–16.