A New Paradigm of "Real-Time" Stroke Risk Prediction and Integrated Care Management in the Digital Health Era: Innovations Using Machine Learning and Artificial Intelligence Approaches

Yutao Guo[1]

[1] Department of Pulmonary Vessel and Thrombotic Disease, Medical School of Chinese PLA, Chinese PLA General Hospital, Beijing, China

Thromb Haemost 2022;122:5–7.

Address for correspondence Yutao Guo, Department of Pulmonary Vessel and Thrombotic Disease, Medical School of Chinese PLA, Chinese PLA General Hospital, Beijing, China (e-mail: dor_guoyt@hotmail.com).

Stroke remains the second leading cause of death and disability worldwide, calling for effective risk assessment and prevention approaches that are needed to reduce the increasing stroke burden.[1] The established major stroke risk factors are smoking, overweight/obesity, diet, dyslipidemia, diabetes mellitus, hypertension, renal disease, coronary heart disease, congestive cardiac failure, valvular heart disease, atrial fibrillation (AF), and vascular disease.[2,3] An increasing cluster of multiple cardiovascular risk factors contributes to even greater risks for ischemic stroke, especially in the elderly population.[4]

The more common and validated stroke risk factors have been used to formulate clinical risk scores as risk stratification tools using traditional statistical models. For example, the Framingham 10-Year Risk Score,[5] MyRisk_Stroke Calculator,[6] and the Stroke Riskometer[7] have been developed in population-based cohort studies, ranging from 3,000 to 17,805 persons, while the QStroke score was derived from 3.5 million primary care population aged 25 to 84 years.[8] There is a great heterogeneity of reported performance in predicting 10-year stroke risk of these models, due to different risk profiles of derived cohorts and the limitation of traditional statistical models.

Given the multiple cardiovascular risk factors incorporated into these clinical risk scores, they are likely to be increasingly more complex to be handled in everyday clinical practice. For example, there are 18 variables in QStroke score, 15 questions with a total 138 points in MyRisk_Stroke Calculator, and 21 variables in Stroke Riskometer.[6–8] Other substantially simpler clinical scores, such as $CHADS_2$ and CHA_2DS_2-VASc scores, are commonly utilized for stroke risk stratification in patients with AF and are easy enough to calculate mentally in busy wards or clinics. Indeed, the CHA_2DS_2-VASc score has been extensively utilized in some national databases involving in up to 10 million individuals.[4,9–12] More complicated risk scores with many clinical variables[13] or the addition of biomarkers[14–17] do not necessarily mean improved prediction in the real world.

Nonetheless, many of these scores are based on the impact of a risk factor determined at baseline, and outcomes ascertained many years later. Given that stroke risk is strongly determined by aging and incident comorbidities, there is uncertainty for predicting stroke risk among patients with progressive multiple risk factors and comorbidities. Some attempts to address the dynamic nature of risk have been published.[10,15,18] Also, there are some methodological limitations about traditional clinical risk prediction models, including unidentified clinical risk factors, unmeasured confounding, information bias, potential for bias due to missing data, as well as limitation of variables input with traditional statistical analysis itself. These would impact on the diagnostic accuracy of any risk stratification tool.

What are the possible options? With the surge in artificial intelligence (AI) technology and machine learning (ML)-based algorithms for predictive analytics, the development of risk predictive models can move from traditional clinical risk tools to a new era of smart technologies and digital health (►**Fig. 1**).

In many cases, risk models based on ML-based algorithms have outperformed clinical risk factor assessment tools, in some scenarios, including AF,[19–23] with their powerful ability of dealing with far more multivariate variables, compared with traditional statistical models (including logistic regression). However, ML-based algorithms seemly did not demonstrate significant advantages over traditional clinical risk models in other clinical settings.[24,25]

Although there is preliminary promise of AI technology and ML-based algorithms in risk prediction, there are many knowledge gaps. For example, the influencing factors on the predictive ability of ML-based algorithms remain unclear.
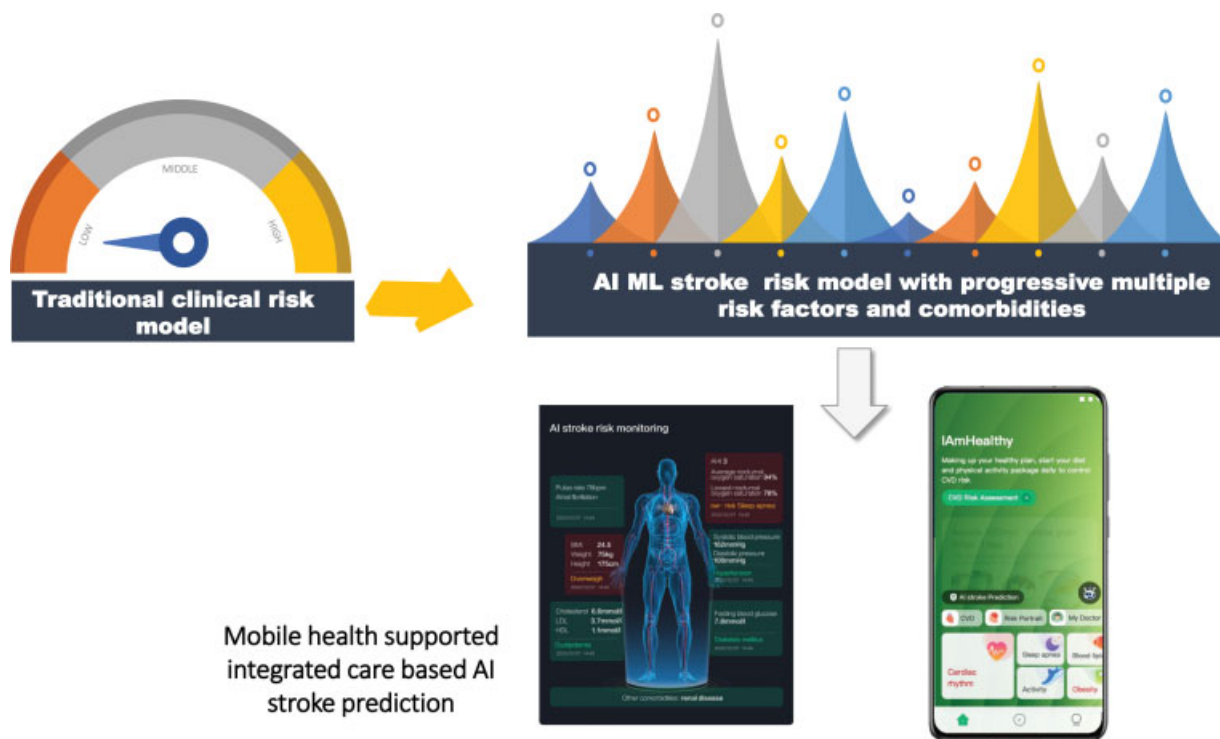
Fig. 1 The new landscape of stroke prevention with AI ML algorithms in the digital health era, incorporating innovations using machine learning and artificial intelligence approaches. AI, artificial intelligence; ML, machine learning.

Which variables would be suitable for ML-based algorithms —the extent to which they would facilitate to predict the risk —is possibly dependent on the factors that are used to train the AI model. Moreover, the impact of different ML-learning approaches is unclear, including supervised (decision tree analysis, neural networks, extreme gradient boosting, or XGBoost) and unsupervised ML algorithms (K-means clustering, hierarchical clustering, etc.),[26] on predicting outcomes.

In this issue of *Thrombosis and Haemostasis,* Lip et al report on stroke risk prediction, using two common clinical rules (CHADS$_2$, CHA$_2$DS$_2$-VASc scores), a clinical multimorbid index and a ML approach accounting for the complex relationships among variables, using a prospective U.S. cohort of 3,435,224 patients from medical databases.[27] This is a first large-scale investigation, with respect to the progressive risk factors for stroke, the difference between traditional statistical methods and ML-based algorithms in predicting stroke risk, together with the comparison of different AI ML approaches. The authors found that a clinical multimorbid index had higher discriminant validity values than common clinical rules, perhaps unsurprisingly given that more clinical variables were used. The synergistic concomitant effects of multiple stroke risk factors would contribute to "real-world" and "real-time" stroke risk assessments, which changes over time with aging and incident comorbidities. Hence, the primary preventive or management strategy may focus not only on "one" major disease, but also on multiple risk factors to reduce the individual stroke risk. This is a sound argument for a more integrated or holistic care approach to characterization and managing chronic cardiovascular conditions, including AF.[28,29]

Indeed, the article by Lip et al also found that the ML-based algorithms yielded the highest discriminant validity values for the gradient boosting/neural network logistic regression formulations, with no marked significant differences among the ML approaches. Hence, the ML-based algorithms would be a better alternative method other than "static" or "one-off" evaluations by the traditional logistic regression model.[30] Such an AI model could incorporate "dynamic" changing risk factors to improve the risk prediction ability.

Beyond the improved risk prediction, the practicality of using ML or AI models in everyday clinical practice should be considered, balancing the complexity of collecting more variables and simple clinical factors in the logistic model. In addition, how ML-based algorithm approaches could be transferred to effectively deliver stroke primary preventive strategies requires further evaluation.

Innovative technologies, including AI, smartwear, and mobile health technologies, make it possible to increase general awareness about stroke and its risk factors as well as to improve stroke prevention.[7] The application of AI technology facilitates stroke risk assessment and monitoring its progress over time. It is expected that AI ML-based algorithms, combined with other smart technologies, would improve holistic primary stroke prevention, through individual risk profile-oriented recommendations based on AI stroke risk monitoring, incorporating educational programs and self-management. The era of structured mHealth approaches to deliver integrated care has shown significant improvements in clinical outcomes (especially hospitalization), with good long-term adherence and persistence.[31,32] Such innovations using ML and AI approaches offer a new paradigm of "real-time" stroke

risk prediction and integrated care management in the digital health era (►**Fig. 1**).

Conflict of Interest
None declared.

Reference

1 Krishnamurthi RV, Ikeda T, Feigin VL. Global, regional and country-specific burden of ischaemic stroke, intracerebral haemorrhage and subarachnoid haemorrhage: a systematic analysis of the Global Burden of Disease Study 2017. Neuroepidemiology 2020;54(02):171–179

2 Bos MJ, Koudstaal PJ, Hofman A, Ikram MA. Modifiable etiological factors and the burden of stroke from the Rotterdam study: a population-based cohort study. PLoS Med 2014;11(04):e1001634

3 O'Donnell MJ, Chin SL, Rangarajan S, et al; INTERSTROKE investigators. Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. Lancet 2016;388(10046):761–775

4 Guo Y, Wang H, Tian Y, Wang Y, Lip GY. Multiple risk factors and ischaemic stroke in the elderly Asian population with and without atrial fibrillation. An analysis of 425,600 Chinese individuals without prior stroke. Thromb Haemost 2016;115(01):184–192

5 Wolf PA, D'Agostino RB, Belanger AJ, Kannel WB. Probability of stroke: a risk profile from the Framingham Study. Stroke 1991;22(03):312–318

6 Nobel L, Mayo NE, Hanley J, Nadeau L, Daskalopoulou SS. MyRisk_Stroke Calculator: a personalized stroke risk assessment tool for the general population. J Clin Neurol 2014;10(01):1–9

7 Parmar P, Krishnamurthi R, Ikram MA, et al; Stroke RiskometerTM Collaboration Writing Group. The Stroke Riskometer(TM) App: validation of a data collection tool and stroke risk predictor. Int J Stroke 2015;10(02):231–244

8 Hippisley-Cox J, Coupland C, Brindle P. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. BMJ 2013;346:f2573

9 Lee HL, Kim JT, Lee JS, et al. CHA2DS2-VASc score in acute ischemic stroke with atrial fibrillation: results from the Clinical Research Collaboration for Stroke in Korea. Sci Rep 2021;11(01):793

10 Yoon M, Yang PS, Jang E, et al. Dynamic changes of CHA2DS2-VASc score and the risk of ischaemic stroke in Asian patients with atrial fibrillation: a nationwide cohort study. Thromb Haemost 2018;118(07):1296–1304

11 Chao TF, Lip GY, Liu CJ, et al. Validation of a modified CHA2DS2-VASc score for stroke risk stratification in Asian patients with atrial fibrillation: a nationwide cohort study. Stroke 2016;47(10):2462–2469

12 Karlsson LO, Nilsson S, Bång M, Nilsson L, Charitakis E, Janzon M. A clinical decision support tool for improving adherence to guidelines on anticoagulant therapy in patients with atrial fibrillation at risk of stroke: a cluster-randomized trial in a Swedish primary care setting (the CDS-AF study). PLoS Med 2018;15(03):e1002528

13 Proietti M, Rivera-Caravaca JM, Esteve-Pastor MA, Marín F, Lip GYH. Stroke and thromboembolism in warfarin-treated patients with atrial fibrillation: comparing the CHA2DS2-VASc and GARFIELD-AF risk scores. Thromb Haemost 2021;121(08):1107–1114

14 Camelo-Castillo A, Rivera-Caravaca JM, Marín F, Vicente V, Lip GYH, Roldán V. Predicting adverse events beyond stroke and bleeding with the ABC-Stroke and ABC-bleeding scores in patients with atrial fibrillation: the Murcia AF project. Thromb Haemost 2020;120(08):1200–1207

15 Chao TF, Lip GYH, Lin YJ, et al. Incident risk factors and major bleeding in patients with atrial fibrillation treated with oral anticoagulants: a comparison of baseline, follow-up and Delta HAS-BLED scores with an approach focused on modifiable bleeding risk factors. Thromb Haemost 2018;118(04):768–777

16 Rivera-Caravaca JM, Roldán V, Esteve-Pastor MA, et al. Long-term stroke risk prediction in patients with atrial fibrillation: comparison of the ABC-Stroke and CHA2DS2-VASc scores. J Am Heart Assoc 2017;6(07):e006490

17 Esteve-Pastor MA, Roldán V, Rivera-Caravaca JM, Ramírez-Macías I, Lip GYH, Marín F. The use of biomarkers in clinical management guidelines: a critical appraisal. Thromb Haemost 2019;119(12):1901–1919

18 Chao TF, Liao JN, Tuan TC, et al. Incident co-morbidities in patients with atrial fibrillation initially with a CHA2DS2-VASc score of 0 (males) or 1 (females): implications for reassessment of stroke risk in initially 'low-risk' patients. Thromb Haemost 2019;119(07):1162–1170

19 Mortazavi BJ, Bucholz EM, Desai NR, et al. Comparison of machine learning methods with national cardiovascular data registry models for prediction of risk of bleeding after percutaneous coronary intervention. JAMA Netw Open 2019;2(07):e196835

20 Hill NR, Ayoubkhani D, McEwan P, et al. Predicting atrial fibrillation in primary care using machine learning. PLoS One 2019;14(11):e0224582

21 Lip GYH, Tran G, Genaidy A, Marroquin P, Estes C, Landsheftl J. Improving dynamic stroke risk prediction in non-anticoagulated patients with and without atrial fibrillation: comparing common clinical risk scores and machine learning algorithms. Eur Heart J Qual Care Clin Outcomes 2021. Doi: 10.1093/ehjqcco/qcab037

22 Lip GYH, Genaidy A, Tran G, Marroquin P, Estes C. Incident atrial fibrillation and its risk prediction in patients developing COVID-19: a machine learning based algorithm approach. Eur J Intern Med 2021;91:53–58

23 Lip GYH, Tran G, Genaidy A, Marroquin P. Revisiting the dynamic risk profile of cardiovascular/non-cardiovascular multimorbidity in incident atrial fibrillation patients and five cardiovascular/non-cardiovascular outcomes: a machine-learning approach. J Arrhythmia 2021;22;37(04):931–941

24 Tiwari P, Colborn KL, Smith DE, Xing F, Ghosh D, Rosenberg MA. Assessment of a machine learning model applied to harmonized electronic health record data for the prediction of incident atrial fibrillation. JAMA Netw Open 2020;3(01):e1919396

25 Khera R, Haimovich J, Hurley NC, et al. Use of machine learning models to predict death after acute myocardial infarction. JAMA Cardiol 2021;6(06):633–641

26 Kilic A. Artificial intelligence and machine learning in cardiovascular health care. Ann Thorac Surg 2020;109(05):1323–1329

27 Lip GY, Genaidy A, Tran G, Marroquin P, Estes C, Sloop S. Improving stroke risk prediction in the general population: common clinical rules, a new multimorbid index and machine learning based algorithms. Thromb Haemost 2021. Doi: 10.1055/a-1467-2993

28 Potpara TS, Lip GYH, Blomstrom-Lundqvist C, et al. The 4S-AF Scheme (Stroke Risk; Symptoms; Severity of Burden; Substrate): a novel approach to in-depth characterization (rather than classification) of atrial fibrillation. Thromb Haemost 2021;121(03):270–278

29 Yoon M, Yang PS, Jang E, et al. Improved population-based clinical outcomes of patients with atrial fibrillation by compliance with the simple ABC (Atrial Fibrillation Better Care) pathway for integrated care management: a nationwide cohort study. Thromb Haemost 2019;119(10):1695–1703

30 Jamthikar A, Gupta D, Saba L, et al. Cardiovascular/stroke risk predictive calculators: a comparison between statistical and machine learning models. Cardiovasc Diagn Ther 2020;10(04):919–938

31 Guo Y, Lane DA, Wang L, et al. mAF-App II Trial Investigators. Mobile health technology to improve care for patients with atrial fibrillation. J Am Coll Cardiol 2020;75(13):1523–1534

32 Guo Y, Guo J, Shi X, et al; mAF-App II Trial investigators. Mobile health technology-supported atrial fibrillation screening and integrated care: a report from the mAFA-II trial Long-term Extension Cohort. Eur J Intern Med 2020;82:105–111