
Yutao Guo1

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

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 Riskometer7 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 CHADS2 and CHA2DS2-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 CHA2DS2-VASc score has been extensively utilized in some national databases involving in up to 10 million individuals.9–12 More complicated risk scores with many clinical variables13 or the addition of biomarkers14–17 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 AF19–23 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.
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.), on predicting outcomes.

In this issue of *Thrombosis and Haemostasis*, Lip et al report on stroke risk prediction, using two common clinical rules (CHADS<sub>2</sub>, CHA<sub>2</sub>D<sub>2</sub>-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. 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.

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. 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. 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. 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