Machine Learning and the Conundrum of Stroke Risk Prediction

Yaacoub Chahine [ID],[1] Matthew J Magoon [ID],[2] Bahetihazi Maidu [ID],[3] Juan C del Álamo [ID],[3,4,5] Patrick M Boyle [ID][2,4,5] and Nazem Akoum [ID][1,2]

1. Division of Cardiology, University of Washington, Seattle, WA, US; 2. Department of Bioengineering, University of Washington, Seattle, WA, US; 3. Department of Mechanical Engineering, University of Washington, Seattle, WA, US; 4. Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA, US; 5. Center for Cardiovascular Biology, University of Washington, Seattle, WA, US

Abstract

Stroke is a leading cause of death worldwide. With escalating healthcare costs, early non-invasive stroke risk stratification is vital. The current paradigm of stroke risk assessment and mitigation is focused on clinical risk factors and comorbidities. Standard algorithms predict risk using regression-based statistical associations, which, while useful and easy to use, have moderate predictive accuracy. This review summarises recent efforts to deploy machine learning (ML) to predict stroke risk and enrich the understanding of the mechanisms underlying stroke. The surveyed body of literature includes studies comparing ML algorithms with conventional statistical models for predicting cardiovascular disease and, in particular, different stroke subtypes. Another avenue of research explored is ML as a means of enriching multiscale computational modelling, which holds great promise for revealing thrombogenesis mechanisms. Overall, ML offers a new approach to stroke risk stratification that accounts for subtle physiologic variants between patients, potentially leading to more reliable and personalised predictions than standard regression-based statistical associations.

Stroke is a leading cause of death and permanent disability worldwide.[1] Ischaemic stroke is the most common stroke variety, comprising more than 80% of strokes in the US.[2] One mechanism of ischaemic stroke is atherosclerosis in the extracranial and intracranial arteries, with plaque rupture leading to thrombosis. The second major category is embolic stroke, in which thrombi form in the heart or the arterial/venous beds and then embolise to occlude downstream arteries, typically in the intracranial domain.[3] Most embolic strokes are associated with AF. Understanding stroke mechanisms and developing effective risk stratification strategies are crucial for primary and secondary prevention. Control of modifiable risk factors can prevent approximately half of all strokes in high-risk individuals.[4]

Machine learning (ML) is being increasingly implemented in various disciplines and is emerging as a powerful tool in healthcare. ML offers algorithms capable of modelling complex and hidden relationships between multiple clinical and physiological variables and desired outcomes. A common application of ML in healthcare is precision medicine, where various options are available to treat a particular condition, and tools are developed to predict the treatment protocol most likely to succeed based on the patient's characteristics.[5] Another valuable implementation of ML in

healthcare is risk stratification. Prediction of which patients are at risk for a particular disease or associated outcome(s), especially those with greater mortality and morbidity, would enable the use of early interventions and reduce strain on the healthcare system. Most cardiovascular disease (CVD) risk assessment models currently in widespread use are based on conventional statistical methods that incorporate small numbers of predictors quantified by human observers (e.g. left ventricular ejection fraction on transthoracic echocardiography). Such models thus oversimplify complex relationships, including large numbers of risk factors. The emerging field of ML in healthcare, using computer algorithms that learn from a dataset without explicit programming, has the potential to address these limitations and outperform current clinical prediction models. ML models can objectively exploit full datasets (e.g. automated analysis of whole echocardiography clips) and, with the help of multiscale computational modelling, offer patient-specific insights. The ML method used varies depending on the type of data. Classification and regression models (e.g. support vector machine [SVM], random forest [RF], gradient-boosted tree [GBT]) are most commonly used in clinical research. These methods search among the available predictor variables to find the features best linked to the outcome. Deep learning models based on neural networks, especially

convolutional neural networks (CNNs), can be used to extract features from ECGs and can be used in image recognition algorithms to find data patterns in image pixels.[6]

This review discusses risk stratification tools in stroke prediction, exploring new avenues for incorporating ML-based risk assessment (*Figure 1*).

Cardiovascular Disease Prediction

The term CVD refers to a wide variety of commonly linked pathologies, including coronary artery disease, stroke, peripheral artery disease, deep vein thrombosis and pulmonary embolism. CVD prediction is currently based on risk scores that use patients' clinical characteristics and comorbidities to approximate the likelihood of future CVD, thus facilitating clinical decision-making. Models widely used in the US include the Framingham Risk Score (FRS) and the American College of Cardiology/ American Heart Association (ACC/AHA) guidelines.[7,8] Based on the risk profile calculated by these algorithms, the physician may choose to start a patient on statin therapy, for example, in addition to recommending adequate lifestyle modifications.

Machine Learning Models Compared with Traditional Risk Scores

The 2013 ACC/AHA Pooled Cohort Equations Risk Calculator is based on traditional risk factors to predict cardiovascular events over 10 years. It can overestimate or underestimate the overall risk of cardiovascular events in certain groups.[9] Kakadiaris et al. used the same risk variables collected in the Multi-Ethnic Study of Atherosclerosis cohort to train an SVM ML algorithm.[10] The resulting predictive tool outperformed the ACC/ AHA calculator, recommending less drug therapy and missing fewer cardiovascular events. The 2017 Framingham Stroke Risk Profile was compared with ML techniques for stroke risk prediction (random survival forest [RSF], SVM, GBT and multilayer perceptron) in a Chinese cohort of 503,842 adults.[11] These ML models improved risk prediction over the conventional Cox model-based approach, with GBT providing the best discrimination and calibration performance. However, the best means of identifying individuals at high risk of stroke was the use of an ensemble model combining the GBT and Cox methods (accuracy of 76% in men and 80% in women). Weng et al. compared four ML algorithms with an established evaluation rubric (ACC guidelines) to predict the first cardiovascular event over 10 years in a prospective cohort of 378,256 patients from UK family practices free from CVD at the study's outset.[12] All ML-based approaches tested were better at identifying individuals at risk of CVD than the established algorithm, with neural networks providing the best risk prediction (area under the receiver operating characteristic curve [AUC] 0.764 versus 0.728 for neural networks versus the conventional method, respectively).

Hung et al. compared deep neural networks (DNNs) with other ML techniques in predicting the 5-year stroke occurrence for a large population-based electronic medical claims database of around 800,000 patients.[13] DNNs and GBTs performed similarly high in predicting stroke risk compared with logistic regression and SVM; however, DNNs performed better using fewer patient data when compared with GBT. Four analysis methods were processed and compared in a Chinese hypertension population to assess stroke risk: logistic regression, stepwise logistic regression, extreme gradient boosting (XGBoost) and RF. The best model performance was observed in a random under-sampling applied RF model.[14] Wu and Fang found that synthetic minority over-sampling applied to regularised logistic regression outperformed other tested models in an older Chinese population for predicting the risk of stroke.[15]

## Machine Learning Models Can Analyse More Complex Features

In addition to patient characteristics and demographics, ML can use blood biomarker levels or imaging data as a basis for predicting CVD. Ambale-Venkatesh et al. found that imaging, ECG and serum biomarkers were among the top predictors of CVD as opposed to traditional cardiovascular risk factors and identified RSF as the most predictive cardiovascular risk (including stroke) model out of nine tested models in an initially asymptomatic multiracial population.[16] In another study, an ML-based, 27-protein prognostic cardiovascular model was trained and validated to predict the 4-year likelihood of MI, stroke, heart failure or death. The proteomic model performed better than the clinical model in predicting CVD (AUC 0.73 versus 0.64).[17] Preliminary results from an ML-based model for predicting white matter lesions as a surrogate for CVD from retinal microvasculature changes (arteriovenous nicking, opacity, focal arteriolar narrowing) are also compelling.[18] Four ML classifiers were tested (RF, artificial neural network [ANN], SVM and linear regression), and both the RF and ANN achieved optimal classification accuracy. Thus, retinal imaging is proposed as a screening tool at the primary care level. Sridhar et al. developed a rapid and inexpensive triage tool based on a DNN to predict major adverse cardiovascular outcomes from ECG data in patients with COVID-19.[19] This approach yielded comparable results to traditional statistical models incorporating extensive clinical data with the advantage of not requiring clinical expertise to gather medical history. *Figure 2* shows the different data types that can be used in stroke risk prediction.

Using a US cohort of 3,435,224 patients from prospective medical databases, Lip et al. showed that the ML approach accounting for the complex and dynamic relationships between variables (including age and incident comorbidities) provided the highest discriminant validity values for stroke prediction (concordance index=0.866).[20] This approach may offer a better alternative to the static or single time point evaluations by the traditional risk scores. Although ML models seem to outperform conventional CVD risk scores, and although the existing approach to CVD risk assessment may benefit from an overhaul, additional studies are needed to validate these ML-based models in different cohorts, independent from the ones used in their inception. For example, the ACC/ AHA Risk Calculator, one of the most common traditional risk scores, was derived from independent centres and diverse populations, and has been validated multiple times.[8]

Cardioembolic Stroke

The current clinical paradigm of stroke associated with AF is that arrhythmia, in the presence of cardiovascular risk factors, leads to blood stasis and hypercoagulability (components of Virchow's triad) with subsequent thrombus formation and systemic embolisation.[21] AF is associated with a fivefold increase in stroke risk, and stroke prevention is a cornerstone of AF management.[22] Fibrotic remodelling in the atrium plays a central role in the pathogenesis of the arrhythmic substrate for AF.[23] It is also increasingly recognised as a significant factor in left atrium appendage (LAA) thrombosis and stroke in AF. Histologically quantified LAA fibrosis burden is higher in patients with AF than in those in sinus rhythm, and in patients with LAA thrombus than in those without.[24,25] The mechanism by which atrial fibrosis contributes to thrombogenesis is not completely understood. A decrease in left atrial strain was shown in the presence of extensive late gadolinium enhancement MRI (LGE-MRI) fibrosis.[26] Patients with AF had marked alterations in the blood fluid dynamics parameters, particularly a higher mean blood residence time and a lower mean kinetic energy compared with patients without AF.[27–30] Whether fibrosis is at the centre of the interaction between overall atrial

Figure 1: The Conundrum of Stroke Risk Prediction
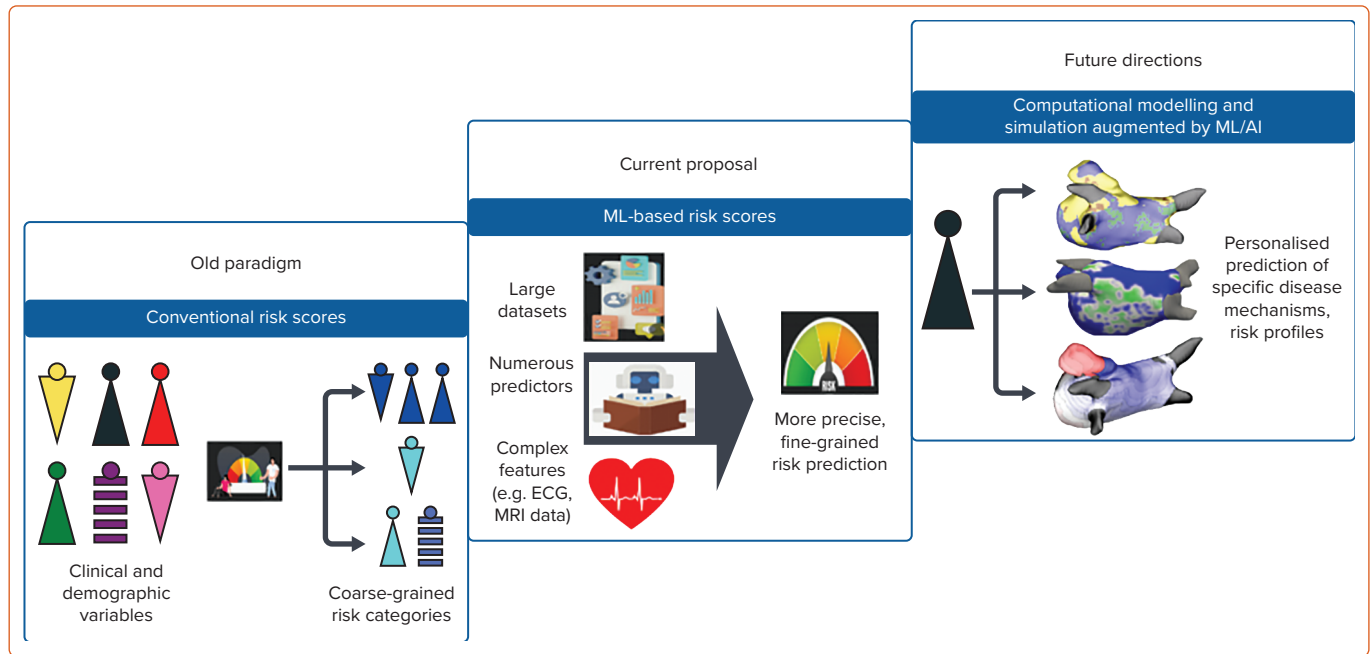
Conventional stroke risk scores based on patient demographic and comorbidity data have poor predictive accuracy. Machine learning (ML)-based risk scores can provide precise stroke risk prediction with a higher degree of granularity. Computational modelling and simulation augmented by ML and/or artificial intelligence (AI) have the potential to predict the consequences of subtle inter-individual differences in physiology and/or anatomy. This offers a promising avenue for even more accurate personalised prediction of stroke risk.

chamber and LAA remodelling, blood fluid dynamics, AF substrate and thrombogenesis is the subject of ongoing studies.
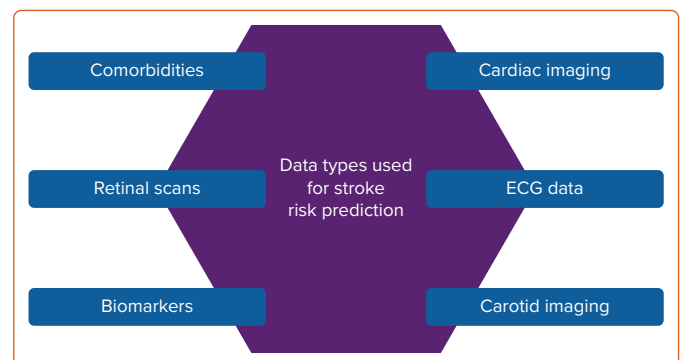
Clinical Stroke Risk Assessment in Atrial Fibrillation

Prior work aiming to characterise ischaemic stroke risk in AF patients has focused on clinical scores, such as $CHADS_2$, CHA_2DS_2-VASc and ATRIA. $CHADS_2$ was limited by its difficulty in accurately evaluating low-risk groups.[31] To improve the predictive performance in this subset of patients, the CHA_2DS_2-VASc score was implemented. The ATRIA score outperforms CHA_2DS_2-VASc in some cases, but both scores remain imperfect predictors of stroke with modest predictive accuracy.[32,33]

Early efforts to develop ML algorithms for predicting stroke risk in AF patients have shown some promise, and have achieved an AUC as high as 0.892 in one cohort analysis.[34] Whereas $CHADS_2$ and CHA_2DS_2-VASc use 6–7 features to stratify stroke risk, an attention-based DNN model identified up to 48 features that influenced stroke risk using the Korean National Health Insurance data, and provided better prediction of ischaemic stroke in AF patients compared with CHA_2DS_2-VASc scores (AUC 0.727 versus 0.651, respectively).[35] ML was used in a German cohort to predict ischaemic strokes in AF patients started on non-vitamin K oral antagonists. The Sub-population Optimisation and Modelling Solutions (SOMS) tool identified age, male sex, ischaemic heart disease, urinary tract infection, and dementia as predictors of ischaemic stroke in this population.[36]

Arrhythmia Burden and Stroke

Arrhythmia burden is a significant factor in predicting stroke risk, given that patients with persistent or permanent AF have a higher stroke propensity than patients with paroxysmal AF.[37] Catheter ablation to suppress AF has been linked to a reduction in stroke risk in observational studies and in a recent randomised trial of persistent and long-standing AF.[38,39] Early rhythm control of patients within 1 year of AF diagnosis has also been associated with stroke reduction.[40] Machine-learned signatures of AF burden from continuous remote monitoring data of cardiac

Figure 2: Different Data Types Can be Used in Stroke Risk Prediction

Traditional risk scores mostly rely on comorbidities and patient clinical characteristics to predict stroke risk, however, machine learning-based approaches can combine complex features to yield a more accurate risk assessment..

implantable electronic devices were superior to CHA_2DS_2-VASc in predicting stroke, and ensemble models such as RF incorporating CHA_2DS_2-VASc may be useful extensions to CHA_2DS_2-VASc alone.[41]

Anticoagulation reduces stroke risk in patients with AF and risk factors for thromboembolism. Unfortunately, AF is often unrecognised and untreated because it is frequently asymptomatic. A deep learning model can identify patients at high risk of new-onset AF from resting 12-lead ECGs. In patients with no history of AF who develop an AF-related stroke, nearly two-thirds would have been identified pre-stroke as having a high risk for AF via deep learning.[42] Distinguishing cardioembolic from non-cardioembolic stroke also has important clinical implications because anticoagulation is generally indicated for patients with cardioembolic stroke. ML-based identification of this subtype of stroke has been demonstrated not only to be feasible using electronic health records, but also helpful in identifying patients who would benefit from oral anticoagulation.[43] Wearable devices are currently used to detect AF. In a recent study, a novel

Table 1: Select Studies Comparing Machine Learning Algorithms with Conventional Methods for Cardiovascular Disease and Stroke Risk Prediction

Outcome	Best Performing Model Developed	Input	Population	Diagnostic Accuracy	Conventional Model Used for Comparison	Reference
CV event over 13 years	Support Vector Machine	Same risk factors as ACC/AHA risk calculator	Multi-Ethnic Study of Atherosclerosis cohort (6,459 participants)	Sensitivity 0.86, specificity 0.95 and AUC 0.92	ACC/AHA risk calculator	Kakadiaris et al. 2018[10]
Stroke	Ensemble model combining Gradient Boosted trees and Cox models	Sociodemographic factors, diet, medical history, physical activity and physical measurements	503,842 Chinese adults	Accuracy: ♂ 76%, ♀ 80%; sensitivity: ♂ 76%, ♀ 67%; specificity: ♂ 76%, ♀ 81%; PPV: ♂ 26%, ♀ 24%; NPV: ♂ 97%, ♀ 97%	2017 Framingham Stroke Risk Profile	Chun et al. 2021[11]
CV event over 10 years	Neural networks	ACC/AHA risk factors and 22 additional variables	378,256 patients from the UK	Sensitivity 67.5%, specificity 70.7%, PPV 18.4%, NPV 95.7%, AUC 0.76	ACC/AHA risk calculator	Weng et al. 2017[12]
CV event over 12 years	Random Forest	Imaging, ECG and serum biomarkers (735 variables)	Multi-Ethnic Study of Atherosclerosis cohort (6,814 participants)	C-index 0.81, Brier score 0.083	AHA/ACC ASCVD score, Framingham score	Ambale-Venkatesh et al. 2017[16]
CV event over 5 years	AutoPrognosis algorithm	473 clinical and lab variables	423,604 UK Biobank participants	Sensitivity 69.9%, PPV 2.6%, AUC 0.774	Framingham score	Alaa et al. 2019[102]
Stroke in AF patients	Logistic regression	Diversified list of comorbidities/demographic/temporal exposure variables	6,457,412 patients from two health plans	AUC 0.892	CHADS$_2$, CHA$_2$DS$_2$-VASc, multimorbid index	Lip et al. 2022[34]
Stroke in AF patients	Attention-based deep neural network	Demographics, health examination, and medical history information	Korean National Health Insurance data (150,989 AF patients)	AUC 0.727	CHA$_2$DS$_2$-VASc score	Jung et al. 2022[35]
Stroke in AF patients	Random Forest	AF burden recorded on cardiac implantable electronic devices	Veterans Health Administration serviced AF patients (9,836 participants)	Sensitivity 52%, specificity 63%, AUC 0.662	CHA$_2$DS$_2$-VASc score	Han et al. 2019[41]

♂ = male; ♀ = female; ACC = American College of Cardiology; AHA = American Heart Association; ASCVD = atherosclerotic cardiovascular disease; AUC = area under the receiver operating characteristic curve; CV = cardiovascular; NPV = negative predictive value; PPV = positive predictive value.

photoplethysmography software algorithm had a positive predictive value of 98.2% for concurrent AF. However, detection of AF during periods of active motion remains a challenge, and wearing the devices at night may maximise the sensitivity.[44] ML-based models may facilitate accurate real-time prediction of AF onset. Guo et al. developed a photoplethysmography-based ML model capable of predicting AF onset in advance.[45] Digital and mobile health technologies can also increase the general awareness of stroke risk factors and prevention.[46] The use of real-time detection of AF onset by wearable devices and the guiding of anticoagulation in patients with AF to reduce stroke risk is currently being investigated. *Table 1* summarises results from various studies comparing ML algorithms with conventional methods in predicting CVD or stroke.

Embolic Stroke of Undetermined Source

Even after an extensive diagnostic workup, a large proportion of ischaemic strokes are still classified as cryptogenic, including embolic stroke of undetermined source (ESUS). This leaves patients without a treatment tailored to the specific pathophysiology.[47] Aetiologies of ischaemic stroke in patients with ESUS include cardiac sources (undetected AF, atrial cardiomyopathy), non-stenotic large artery atherosclerosis, paradoxical emboli related to a patent foramen ovale, and hypercoagulable states. Cardiac LGE-MRI studies have shown comparable levels of atrial fibrosis in ESUS and AF patients.[48] ESUS patients with high atrial fibrosis have a high risk of new-onset AF or recurrent stroke.[49] Recent computational modelling studies showed that the proarrhythmic substrate properties of fibrosis are similar in ESUS and AF patients, prompting the hypothesis that ESUS patients with

fibrotic atria are spared from arrhythmia due to a lack of triggers.[50] The subgroup of patients with a cardioembolic aetiology of ESUS could potentially benefit from oral anticoagulation, but correctly identifying this subgroup of patients is challenging. ML may provide novel insights by applying rules extracted from labelled cases (strokes with known aetiology) to unlabelled cases (ESUS). ML models based on patient demographics, comorbidities, laboratory variables and echocardiography can distinguish cardioembolic from non-cardioembolic strokes with relatively high accuracy (AUC ranging from 0.82 to 0.85) and it is estimated that 40–44% of cases of ESUS are cardioembolic in origin.[51,52] Undiagnosed paroxysmal AF is often suspected in patients with ESUS and warrants prolonged ambulatory cardiac rhythm monitoring following the stroke episode. In a recent study, artificial intelligence (AI)-enabled ECG was implemented to estimate the risk of paroxysmal AF in ESUS patients; this tool can potentially be used to guide prolonged rhythm monitoring in these patients, however, AF probability by AI analysis of ECGs was not associated with ESUS.[53] The fact that AI-enabled ECGs cannot currently predict ESUS reliably highlights the heterogeneous nature of ESUS, and emphasises that it is often caused by mechanisms other than silent AF.

Finally, ML can help categorise ESUS patients into different aetiological groups. A hierarchical k-means clustering algorithm identified four subgroups in a cohort of 800 ESUS patients based on baseline data: arterial disease, atrial cardiopathy, patent foramen ovale, and left ventricular disease. More than half of the patients were assigned to the arterial disease cluster.[54] The hypothesis that incorporation of quantitative

measurements from advanced carotid and cardiac, including fibrosis, imaging into ML algorithms, improves the classification of stroke aetiology and potentially leads to even more optimal treatment and prevention strategies is currently being tested.

Large Artery Atherosclerosis

Large artery atherosclerosis is a major cause of stroke. Plaque rupture or ulceration often results in the formation of a thrombus that may embolise and occlude a vessel lumen in the brain, obstructing the blood flow and causing a stroke. One way of predicting stroke is via risk assessment using conventional factors responsible for atherosclerosis growth.[55] However, this ignores a potentially fruitful avenue for personalisation: each patient's atherosclerotic plaque morphology, which can be assessed by imaging.[56] Integration of carotid imaging phenotypes with conventional risk factors has shown improved risk stratification compared with either method alone.[57] However, such approaches still rely on traditional prediction models and have not yet fully leveraged the potential of ML.

Araki et al. described stroke risk prediction by integrating assessment of the near and far walls of the carotid artery using grayscale morphology of the atherosclerotic plaque and showed that this method is superior to the traditional analysis of the far wall of the carotid artery alone.[58] Johri et al. measured plaque characteristics using different imaging modalities, and two ML-based algorithms (RF and RSF) were superior to conventional statistical methods in predicting risk of cardiovascular events and coronary artery disease.[59] Plaque rupture in atherosclerotic carotid arteries is correlated with high mechanical stresses. Guvenir et al. extracted multicomponent properties of atherosclerotic carotid arteries using ultrasound, MRI, finite element modelling, and ML-based Bayesian optimisation.[60] This framework has great potential to be advanced for patient-specific *in vivo* applications. Wu et al. established an explainable ML model based on XGBoost to predict the presence of carotid plaques in asymptomatic individuals.[61] It identified high-risk patients who could benefit from a carotid ultrasound, creating a framework to assist in large-scale stroke screening. Saba et al. developed a deep learning algorithm that can classify carotid plaque into symptomatic (causing one or more transient ischaemic attacks or strokes referable to the appropriate internal carotid artery distribution) versus asymptomatic based on carotid ultrasound characteristics of the plaque.[62]

Carotid ultrasound imaging can provide valuable data that can help predict CVD and stroke. A low-cost ML-based integrated model taking into account both conventional risk factors and carotid ultrasound image-based phenotypes (CUSIP), using RF as a classifier, showed an improvement of 18% in predicting cardiovascular/stroke risk compared with an ML model incorporating only conventional risk factors.[63] The same team later developed an office-based ML cardiovascular risk stratification tool (AtheroEdge Composite Risk Score 2.0) based on clinical risk factors and CUSIP, which provided better cardiovascular risk estimates than the FRS and the WHO risk score.[64] Granularity in expressing CVD risk prediction is crucial for personalised medicine. Multiclass ML algorithms can predict several risk categories for a particular disease. Three kinds of multiclass ML cardiovascular risk assessment calculators (based on SVM, RF and XGBoost) showed superior performance compared with conventional CVD risk calculators (FRS, Systematic Coronary Risk Evaluation Score and Atherosclerotic Cardiovascular Disease calculators), with RF performing best.[65] Carotid ultrasound imaging, hypertension, use of medications, and smoking had the highest influence on risk prediction. Finally, this team created

Figure 3: Computational Modelling and Simulations for Risk Stratification and Treatment Strategies for Patients with Stroke, AF or both
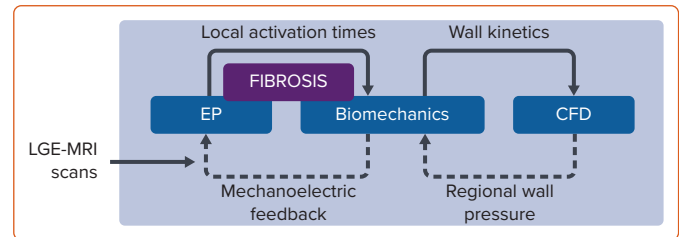
Electrophysiological and computational fluid dynamics (CFD) simulations enable specific assessment of patient-specific consequences of fibrotic remodelling and enable prediction of thrombus formation. EP = electrophysiology; LGE = late gadolinium enhancement.

the first ML algorithm for multilabel cardiovascular event prediction, using both CUSIP and conventional risk factors.[66]

Machine Learning as a Means of Improving or Complementing Computational Simulations of AF

Multiscale, biophysically detailed electrophysiology simulations serve as a complement to conventional experiments and have emerged as a useful approach for deriving insights into arrhythmia mechanisms, with many avenues for patient-specific personalisation.[67] Simulations have corroborated AF dynamics observed via intracardiac mapping, which could ultimately lead to improved understanding of AF-related strokes.[68,69] ML techniques are being used in this research area, both to improve model reconstruction and as a means of augmenting patient-specific simulation results.

Atrial electrophysiology simulations are typically conducted in models reconstructed from clinical imaging of patient hearts, with regions of fibrosis delineated by LGE-MRI.[70] Cell- and tissue-scale bioelectrical properties based on experimental and clinical measurements from human tissue are combined with atlas-based approximations of myocardial fibre orientations in the atrial volume.[71,72] Recent developments have exploited ML to improve this process, such as the use of physics-informed neural networks (PINNs) to estimate each individual's unique fibre orientation by solving an inverse problem based on intracardiac mapping data.[73]

In parallel research, results from AF simulations are being used alongside other multimodal inputs to ML-based classifiers. Roney et al. reconstructed models from MRI scans in a cohort of 100 AF patients who underwent ablation, then used 10-fold cross-validation to test different ML approaches as a means of predicting long-term recurrence with various inputs.[74] Using patient history alone, the optimal classifier (RF) achieved an AUC of 0.61 ± 0.14; when measurements from patient LGE-MRI scans were added, this improved to 0.66 ± 0.17 (K-nearest neighbours with principal component analysis [PCA]). Further augmentation with outcomes from simulations of ablation remarkably boosted the AUC to 0.85 ± 0.09 (SVM with PCA). Working in a separate cohort of 32 paroxysmal AF patients, Shade et al. trained a quadratic discriminant analysis classifier to predict post-ablation recurrence with a similar degree of accuracy (AUC 0.82) from imaging features and simulation outcomes alone.[75] Overall, these exciting new findings point towards ML as a powerful means to achieve translationally relevant insights from computational simulations in the context of AF. The hope is that such electrophysiological simulations could ultimately be combined with biomechanics, fluid dynamics and coagulation modelling to better understand each AF patient's overall stroke risk profile (*Figure 3*).[71]

Table 2: Comparison of the Strengths and Limitations of Conventional Regression-based Models and Machine Learning Algorithms

	Conventional Statistical Methods	Machine Learning Models
Number of predictors	Can take in only a limited set of predictor variables	Can handle numerous predictors and large datasets, such as electronic health records or omics data
Computation time	Results are computed instantaneously	Generally require longer computational time
Interpretability and presentation of results	Estimate easy-to-interpret parameters such as OR, RR and HR. Risk models are usually converted into point-based scores that can be easily calculated (e.g. FRS)	ML parameters may be more difficult to interpret by clinicians, especially in the case of deep learning algorithms such as neural networks. In addition, presentation of the results may be more complicated
Causality	Elucidate associations between predictor variables and outcomes that may be on the causal pathway	Causal interpretation is limited, but ML algorithms can be useful in gaining insight into subclinical disease markers
Data types	Can use patient characteristics, but variables need to be derived from imaging and ECG data	Can use different types of data including patient characteristics and comorbidities, raw imaging and ECG data
Assumptions	Are based on theory and assumptions limiting their flexibility	Lack of assumptions increases flexibility of use

FRS = Framingham Risk Score; ML = machine learning.

Use of Machine Learning to Better Understand Blood Flow Dynamics and Thrombogenesis

Studies using cardiac MRI and echocardiography show that flow imaging can map intracardiac stasis and predict mural thrombosis and cerebral microembolism.[28,29,76–79] However, these studies have been largely limited to the left ventricle and sinus rhythm, given that flow imaging in the fibrillating atrium is challenging. Patient-specific computational fluid dynamics (CFD) analysis has been shown to outperform traditional functional imaging for coronary ischaemia diagnosis and could be a promising tool to assess atrial thrombosis risk.[27,30,80–88] However, simulating the flow inside the cardiac chambers requires significant computational power and detailed patient-specific knowledge of inflow/outflow boundary conditions that is often hard to attain (e.g. flow profiles in all pulmonary veins). Although these demands are accessible in research environments, they are a significant barrier to adoption in clinical settings. In the past two decades, there have been significant advances in ML algorithms applied to fluid mechanics.[89] ML algorithms, in particular neural networks, have emerged as computationally efficient tools to establish correlations between vascular or chamber anatomy and global flow metrics, including thrombogenic potential.[90,91] Neural networks can also be used to declutter or enhance flow imaging data. Two particularly interesting applications are the unwrapping of phase velocity maps and the recovery of wall shear stresses from low-resolution or sparse image data.[92–94] Finally, neural networks are widely used in cardiovascular wall segmentation, a crucial step in patient-specific CFD simulations.[95,96]

Despite their high accuracy and short computation time, many current ML frameworks for flow prediction are trained using large amounts of labelled data from costly, high-resolution CFD simulations, which can be challenging. Training simulations can be performed in patient-specific anatomies, but an alternative approach that saves computational resources is to use simplified geometries and test fully trained ML frameworks on different patients. Thus, this approach is not always fully patient specific. PINNs can alleviate these limitations by prescribing constraints from underlying physical models, including the partial differential equations of fluid motion and their boundary conditions. A paradigmatic example of an application was the prediction of blood flow velocity fields and pressure fields inside patient-specific 3D aneurysm anatomies from contrast agent concentration fields.[97] Extensions of this framework are rapidly emerging: two notable examples are the non-invasive measurement of thrombus mechanical properties and of arterial blood pressure from imaging data.[98,99] Although the limits of application of these novel ML techniques are still being investigated, it is worth noting

that they have the potential to be truly patient specific when provided with sufficient information from a single patient.

Moving Beyond Conventional Methods

Multiple analytical challenges may not be well handled by regression-based models (*Table 2*). Regression-based models handle a limited set of risk predictors and fail to provide accurate risk assessment and event prediction when a large number of risk factors is used.[6] ML is an alternative approach to traditional statistical inference that offers high predictive power for identifying disease conditions and selecting treatment choices without a deep understanding of underlying mechanisms. Thanks to recent advances in computational hardware and numerical algorithms, ML analysis of large-scale databases of thousands of patients and rich data sources is now feasible. These features make ML a promising new technology to guide prevention and patient management.

Generally, ML methods described in this review outperformed traditional statistical methods in terms of predictive accuracy. The lack of underlying models in ML can be useful in gaining insight into subclinical disease markers that are not captured without assuming causality.[16] However, this can also make ML models difficult to interpret and the factors driving model predictions tend to be opaque. For instance, a predictor variable may be selected not because it causes the outcome but simply because it is indirectly associated with a cryptic variable that causes the outcome, which raises practical and ethical concerns.[100] The explainability and interpretability of ML algorithms is a subject of active research, and the clinical adoption of increasingly transparent models is highly anticipated.[101] Last, ML models differ vastly in computation time: some may be as fast as conventional methods, especially models based on amendments to regression methods and classification trees, while some ensemble approaches can take longer, especially during training and validation.

Although ML seems promising in the stroke prediction realm, there is a need for cautious optimism given that several criteria are essential for the transition of such algorithms into routine clinical practice. Some ML algorithms have been described as a 'black box', when there is little insight into how the model is basing its prediction or which variables are contributing most to the predicted risk. In addition, most of these models 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 in predicting stroke risk in patients with progressive multiple risk factors and comorbidities.

Future Directions

ML models have been increasingly implemented in a variety of healthcare applications and have demonstrated superior predictive value compared with many traditional models for predicting risk of stroke or overall CVD. However, there are no ML-based predictive models approved by the US Food and Drug Administration or similar regulatory bodies for CVD prevention.

For ML to be more widely incorporated in stroke risk prediction, the level of evidence needs to be raised to the same level as that of the studies used to validate current widely used risk calculators. Use of large electronic health records may help achieve this. The use of increasingly large datasets to train, validate and test ML algorithms is expected to lead to better understanding of their theoretical maximum potential. Then, fair comparisons between ML and statistical methods can be fully established and reviewed. Further research is necessary to determine the feasibility of applying ML models in clinical settings and to determine whether this could improve clinical care and patient outcomes. ❑

Clinical Perspective

- The current paradigm of stroke risk assessment and mitigation is centred around clinical risk factors and comorbidities.
- Although functional and easy to use, these models have poor predictive accuracy.
- Machine learning techniques have emerged as a powerful approach to solving data analysis challenges that are poorly addressed by typical regression approaches.
- Multiscale, multiphysics computational modelling, supported by machine learning, holds tremendous promise in providing a mechanistic understanding of thrombogenesis.

1. GBD 2019 Diseases and Injuries Collaborator. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 2020;396:1204–22. https://doi.org/10.1016/S0140-6736(20)30925-9; PMID: 33069326.
2. Virani SS, Alonso A, Aparicio HJ, et al. Heart disease and stroke statistics – 2021 update. *Circulation* 2021;143:e254–743. https://doi.org/10.1161/CIR.0000000000000950; PMID: 33501848.
3. Adams HP, Bendixen BH, Kapelle LJ, et al. Classification of subtype of acute ischemic stroke definitions for use in a multicenter clinical trial. *Stroke* 1993;24:35–41. https://doi.org/10.1161/01.STR.24.1.35; PMID: 7678184.
4. Brainin M, Feigin V, Martins S, et al. Cut stroke in half: polypill for primary prevention in stroke. *Int J Stroke* 2018;13:633–47. https://doi.org/10.1177/1747493018761190; PMID: 29461155.
5. Lee SI, Celik S, Logsdon BA, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. *Nat Commun* 2018;9:42. https://doi.org/10.1038/s41467-017-02465-5; PMID: 29298978.
6. Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. *Eur Heart J* 2017;38:1805–14. https://doi.org/10.1093/eurheartj/ehw302; PMID: 27436868.
7. D'Agostino RB, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008;117:743–53. https://doi.org/10.1161/circulationaha.107.699579; PMID: 18212285.
8. Goff DC, Lloyd-Jones DM, Bennett 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(Suppl 2):S49–73. https://doi.org/10.1161/01.cir.0000437741.48606.98; PMID: 24222018.
9. DeFilippis AP, Young R, Carrubba CJ, et al. An analysis of calibration and discrimination among multiple cardiovascular risk scores in a modern multiethnic cohort. *Ann Intern Med* 2015;162:266–75. https://doi.org/10.7326/M14-1281; PMID: 25686167.
10. Kakadiaris IA, Vrigkas M, Yen AA, et al. Machine learning outperforms ACC/AHA CVD risk calculator in MESA. *J Am Heart Assoc* 2018;7:e009476. https://doi.org/10.1161/JAHA.118.009476; PMID: 30571498.
11. Chun M, Clarke R, Cairns BJ, et al. Stroke risk prediction using machine learning: a prospective cohort study of 0.5 million Chinese adults. *J Am Med Inform Assoc* 2021;28:1719–27. https://doi.org/10.1093/JAMIA/OCAB068; PMID: 33969418.
12. Weng SF, Reps J, Kai J, et al. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One* 2017;12:e0174944. https://doi.org/10.1371/journal.pone.0174944; PMID: 28376093.
13. 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. *Annu Int Conf IEEE Eng Med Biol Soc* 2017;2017:3110–3. https://doi.org/10.1109/embc.2017.8037515; PMID: 29060556.
14. Huang X, Cao T, Chen L, et al. Novel insights on establishing machine learning-based stroke prediction models among hypertensive adults. *Front Cardiovasc Med* 2022;9:901240. https://doi.org/10.3389/fcvm.2022.901240; PMID: 35600480.
15. Wu Y, Fang Y. Stroke prediction with machine learning methods among older Chinese. *Int J Environ Res Public Health* 2020;17:1828. https://doi.org/10.3390/ijerph17061828; PMID: 32178250.
16. Ambale-Venkatesh B, Yang X, Wu CO, et al. Cardiovascular event prediction by machine learning: the Multi-Ethnic Study of Atherosclerosis. *Circ Res* 2017;121:1092–101. https://doi.org/10.1161/circresaha.117.311312; PMID: 28794054.
17. Williams SA, Ostroff R, Hinterberg MA, et al. A proteomic surrogate for cardiovascular outcomes that is sensitive to multiple mechanisms of change in risk. *Sci Transl Med* 2022;14:eabj9625. https://doi.org/10.1126/scitranslmed.abj9625; PMID: 35385337.
18. Bhuiyan A, Kanti Roy P, Bhuiyan T, et al. A novel retinal vascular feature and machine learning-based brain white matter lesion prediction model. *medRxiv* 2021. https://doi.org/10.1101/2021.09.27.21264168; preprint.
19. Sridhar AR, Chen AZ-H, Mayfield JJ, et al. Identifying risk of adverse outcomes in COVID-19 patients via artificial intelligence-powered analysis of 12-lead intake electrocardiogram. *Cardiovasc Digit Health J* 2022;3:62–74. https://doi.org/10.1016/j.cvdhj.2021.12.003; PMID: 35005676.
20. Lip GYH, Genaidy A, Tran G, et al. Improving stroke risk prediction in the general population: a comparative assessment of common clinical rules, a new multimorbid index, and machine-learning-based algorithms. *Thromb Haemost* 2022;122:142–50. https://doi.org/10.1055/a-1467-2993; PMID: 33765685.
21. Watson T, Shantsila E, Lip GY. Mechanisms of thrombogenesis in atrial fibrillation: Virchow's triad revisited. *Lancet* 2009;373:155–66. https://doi.org/10.1016/S0140-6736(09)60040-4; PMID: 19135613.
22. Wang TJ, Larson MG, Levy D, et al. Temporal relations of atrial fibrillation and congestive heart failure and their joint influence on mortality: the Framingham Heart Study. *Circulation* 2003;107:2920–5. https://doi.org/10.1161/01.cir.0000072767.89944.6E; PMID: 12771006.
23. Nattel S, Heijman J, Zhou L, Dobrev D. Molecular basis of atrial fibrillation pathophysiology and therapy: a translational perspective. *Circ Res* 2020;127:51–72. https://doi.org/10.1161/circresaha.120.316363; PMID: 32717172.
24. Miyauchi S, Tokuyama T, Uotani Y, et al. Association between left atrial appendage fibrosis and thrombus formation: a histological approach. *J Cardiovasc Electrophysiol* 2022;33:677–87. https://doi.org/10.1111/jce.15384; PMID: 35066945.
25. Chahine Y, Akoum N. Fibrosis: a nexus between atrial fibrillation and left atrial appendage thrombosis. *J Cardiovasc Electrophysiol* 2022;33:688–9. https://doi.org/10.1111/jce.15382; PMID: 35064708.
26. Kuppahally SS, Akoum N, Burgon NS, et al. Left atrial strain and strain rate in patients with paroxysmal and persistent atrial fibrillation: relationship to left atrial structural remodeling detected by delayed-enhancement MRI. *Circ Cardiovasc Imaging* 2010;3:231–9. https://doi.org/10.1161/circimaging.109.865683; PMID: 20133512.
27. García-Villalba M, Rossini L, Gonzalo A, et al. Demonstration of patient-specific simulations to assess left atrial appendage thrombogenesis risk. *Front Physiol* 2021;12:596596. https://doi.org/10.3389/fphys.2021.596596; PMID: 33716763.
28. Markl M, Lee DC, Ng J, et al. Left atrial 4D flow MRI: stasis and velocity mapping in patients with atrial fibrillation. *Invest Radiol* 2016;51:147–54. https://doi.org/10.1097/rli.0000000000000219; PMID: 26488375.
29. Costello BT, Voskoboinik A, Qadri AM, et al. Measuring atrial stasis during sinus rhythm in patients with paroxysmal atrial fibrillation using 4 dimensional flow imaging: 4D flow imaging of atrial stasis. *Int J Cardiol* 2020;315:45–50. https://doi.org/10.1016/j.ijcard.2020.05.012; PMID: 32439367.
30. Koizumi R, Funamoto K, Hayase T, et al. Numerical analysis of hemodynamic changes in the left atrium due to atrial fibrillation. *J Biomech* 2015;48:472–8. https://doi.org/10.1016/j.jbiomech.2014.12.025; PMID: 25547024.
31. Lip GYH, Nieuwlaat R, Pisters R, et al. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the Euro Heart Survey on atrial fibrillation. *Chest* 2010;137:263–72. https://doi.org/10.1378/chest.09-1584; PMID: 19762550.
32. Van Den Ham HA, Klungel OH, Singer DE, et al. Comparative performance of ATRIA, CHADS2, and CHA2DS2-VASc risk scores predicting stroke in patients with atrial fibrillation: results from a national primary care database. *J Am Coll Cardiol* 2015;66:1851–9. https://doi.org/10.1016/J.jacc.2015.08.033; PMID: 26493655.
33. Aspberg S, Chang Y, Atterman A, et al. Comparison of the ATRIA, CHADS2, and CHA2DS2-VASc stroke risk scores in predicting ischaemic stroke in a large Swedish cohort of patients with atrial fibrillation. *Eur Heart J* 2016;37:3203–10. https://doi.org/10.1093/eurheartj/ehw077; PMID: 26941204.
34. Lip GYH, Tran G, Genaidy A, et al. 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* 2022;8:548–56. https://doi.org/10.1093/ehjqcco/qcab037; PMID: 33999139.
35. Jung S, Song MK, Lee E, et al. Predicting ischemic stroke in patients with atrial fibrillation using machine learning. *Front Biosci* (*Landmark Ed*) 2022;27:80. https://doi.org/10.31083/J.fbl2703080; PMID: 35345312.
36. Kostev K, Wu T, Wang Y, et al. Predicting the risk of ischemic stroke in patients treated with novel oral anticoagulants: a machine learning approach. *Neuroepidemiology* 2021;55:387–92. https://doi.org/10.1159/000517512; PMID: 34350851.
37. Steinberg BA, Hellkamp AS, Lokhnygina Y, et al. Higher risk of death and stroke in patients with persistent vs. paroxysmal atrial fibrillation: results from the ROCKET-AF trial. *Eur Heart J* 2015;36:288–96. https://doi.org/10.1093/eurheartj/ehu359; PMID: 25209598.
38. Bunch TJ, May HT, Bair TL, et al. Atrial fibrillation ablation patients have long-term stroke rates similar to patients without atrial fibrillation regardless of CHADS2 score. *Heart Rhythm* 2013;10:1272–7. https://doi.org/10.1016/j.hrthm.2013.07.002; PMID: 23835257.
39. Wu G, Huang H, Cai L, et al. Long-term observation of catheter ablation vs. pharmacotherapy in the management of persistent and long-standing persistent atrial fibrillation (CAPA study). *Europace* 2021;23:731–9. https://doi.org/10.1093/europace/euaa356; PMID: 33367669.
40. Kirchhof P, Camm AJ, Goette A, et al. Early rhythm-control therapy in patients with atrial fibrillation. *N Engl J Med* 2020;383:1305–16. https://doi.org/10.1056/NEJMoa2019422;

PMID: 32865375.

41. Han L, Askari M, Altman RB, et al. Atrial fibrillation burden signature and near-term prediction of stroke: a machine learning analysis. *Circ Cardiovasc Qual Outcomes* 2019;12:e005595. https://doi.org/10.1161/circoutcomes.118.005595; PMID: 31610712.

42. Raghunath S, Pfeifer JM, Ulloa-Cerna AE, et al. Deep neural networks can predict new-onset atrial fibrillation from the 12-lead ECG and help identify those at risk of atrial fibrillation-related stroke. *Circulation* 2021;143:1287–98. https://doi.org/10.1161/circulationaha.120.047829; PMID: 33588584.

43. Guan W, Ko D, Khurshid S, et al. Automated electronic phenotyping of cardioembolic stroke. *Stroke* 2021;52:181–9. https://doi.org/10.1161/strokeaha.120.030663; PMID: 33297865.

44. Lubitz SA, Faranesh AZ, Selvaggi C, et al. Detection of atrial fibrillation in a large population using wearable devices: the Fitbit heart study. *Circulation* 2022;146:1415–24. https://doi.org/10.1161/circulationaha.122.060291; PMID: 36148649.

45. Guo Y, Wang H, Zhang H, et al. Photoplethysmography-based machine learning approaches for atrial fibrillation prediction: a report from the Huawei Heart Study. *JACC Asia* 2021;1:399–408. https://doi.org/10.1016/j.jacasi.2021.09.004; PMID: 36341222.

46. Guo Y. 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. *Thromb Haemost* 2022;122:5–7. https://doi.org/10.1055/a-1508-7980; PMID: 33984864.

47. Hart RG, Diener HC, Coutts SB, et al. Embolic strokes of undetermined source: the case for a new clinical construct. *Lancet Neurol* 2014;13:429–38. https://doi.org/10.1016/s1474-4422(13)70310-7; PMID: 24646875.

48. Tandon K, Tirschwell D, Longstreth WT, et al. Embolic stroke of undetermined source correlates with atrial fibrosis without atrial fibrillation. *Neurology* 2019;93:e381–7. https://doi.org/10.1212/wnl.0000000000007827; PMID: 31239359.

49. Kühnlein P, Mahnkopf C, Majersik JJ, et al. Atrial fibrosis in embolic stroke of undetermined source: a multicenter study. *Eur J Neurol* 2021;28:3634–9. https://doi.org/10.1111/ene.15022; PMID: 34252263.

50. Bifulco SF, Scott GD, Sarairah S, et al. Computational modeling identifies embolic stroke of undetermined source patients with potential arrhythmic substrate. *eLife* 2021;10:e64213. https://doi.org/10.7554/eLife.64213; PMID: 33942719.

51. Kamel H, Navi BB, Parikh NS, et al. Machine learning prediction of stroke mechanism in embolic strokes of undetermined source. *Stroke* 2020;51:e203–10. https://doi.org/10.1161/STROKEAHA.120.029305; PMID: 32781943.

52. Martin MC, Sichtermann T, Schürmann K, et al. Classification of patients with embolic stroke of undetermined source into cardioembolic and non-cardioembolic profile subgroups. *Eur J Neurol* 2022;29:2275–82. https://doi.org/10.1111/ene.15356; PMID: 35420727.

53. Rabinstein AA, Yost MD, Faust L, et al. Artificial intelligence-enabled ECG to identify silent atrial fibrillation in embolic stroke of unknown source. *J Stroke Cerebrovasc Dis* 2021;30:105998. https://doi.org/10.1016/j.jstrokecerebrovasdis.2021.105998; PMID: 34303963.

54. Ntaios G, Weng SF, Perlepe K, et al. Data-driven machine-learning analysis of potential embolic sources in embolic stroke of undetermined source. *Eur J Neurol* 2021;28:192–201. https://doi.org/10.1111/ene.14524; PMID: 32918305.

55. Thompson JB, Blaha M, Resar JR, et al. Strategies to reverse atherosclerosis: an imaging perspective. *Curr Treat Options Cardiovasc Med* 2008;10:283–93. https://doi.org/10.1007/s11936-008-0049-2; PMID: 18647584.

56. Jamthikar A, Gupta D, Khanna NN, et al. A special report on changing trends in preventive stroke/cardiovascular risk assessment via B-mode ultrasonography. *Curr Atheroscler Rep* 2019;21:25. https://doi.org/10.1007/s11883-019-0788-4; PMID: 31041615.

57. Khanna NN, Jamthikar AD, Gupta D, et al. Performance evaluation of 10-year ultrasound image-based stroke/cardiovascular (CV) risk calculator by comparing against ten conventional CV risk calculators: a diabetic study. *Comput Biol Med* 2019;105:125–43. https://doi.org/10.1016/j.compbiomed.2019.01.002; PMID: 30641308.

58. Araki T, Jain PK, Suri HS, et al. Stroke risk stratification and its validation using ultrasonic echolucent carotid wall plaque morphology: a machine learning paradigm. *Comput Biol Med* 2017;80:77–96. https://doi.org/10.1016/J.compbiomed.2016.11.011; PMID: 27915126.

59. Johri AM, Mantella LE, Jamthikar AD, et al. Role of artificial intelligence in cardiovascular risk prediction and outcomes: comparison of machine-learning and conventional statistical approaches for the analysis of carotid ultrasound features

and intra-plaque neovascularization. *Int J Cardiovasc Imaging* 2021;37:3145–56. https://doi.org/10.1007/S10554-021-02294-0; PMID: 34050838.

60. Guvenir S, Torun HM, Hansen HHG, et al. Multicomponent material property characterization of atherosclerotic human carotid arteries through a Bayesian optimization based inverse finite element approach. *J Mech Behav Biomed Mater* 2022;126:104996. https://doi.org/10.1016/J.jmbbm.2021.104996; PMID: 34864574.

61. Wu D, Cui G, Huang X, et al. An accurate and explainable ensemble learning method for carotid plaque prediction in an asymptomatic population. *Comput Methods Programs Biomed* 2022;221:106842. https://doi.org/10.1016/J.cmpb.2022.106842; PMID: 35569238.

62. Saba L, Sanagala SS, Gupta SK, et al. Ultrasound-based internal carotid artery plaque characterization using deep learning paradigm on a supercomputer: a cardiovascular disease/stroke risk assessment system. *Int J Cardiovasc Imaging* 2021;37:1511–28. https://doi.org/10.1007/S10554-020-02124-9; PMID: 33423132.

63. Jamthikar A, Gupta D, Khanna NN, et al. A low-cost machine learning-based cardiovascular/stroke risk assessment system: integration of conventional factors with image phenotypes. *Cardiovasc Diagn Ther* 2019;9:420–30. https://doi.org/10.21037/cdt.2019.09.03; PMID: 31737514.

64. Jamthikar AD, Gupta D, Johri AM, et al. Low-cost office-based cardiovascular risk stratification using machine learning and focused carotid ultrasound in an Asian-Indian cohort. *J Med Syst* 2020;44:208. https://doi.org/10.1007/s10916-020-01675-7; PMID: 33175247.

65. Jamthikar AD, Gupta D, Mantella LE, et al. Multiclass machine learning vs. conventional calculators for stroke/CVD risk assessment using carotid plaque predictors with coronary angiography scores as gold standard: a 500 participants study. *Int J Cardiovasc Imaging* 2021;37:1171–87. https://doi.org/10.1007/S10554-020-02099-7; PMID: 33184741.

66. Jamthikar A, Gupta D, Johri AM, et al. A machine learning framework for risk prediction of multi-label cardiovascular events based on focused carotid plaque B-mode ultrasound: a Canadian study. *Comput Biol Med* 2022;140:105102. https://doi.org/10.1016/J.COMPBIOMED.2021.105102; PMID: 34973521.

67. Bifulco SF, Akoum N, Boyle PM. Translational applications of computational modelling for patients with cardiac arrhythmias. *Heart* 2020;107:456–61. https://doi.org/10.1136/heartjnl-2020-316854; PMID: 33303478.

68. Cochet H, Dubois R, Yamashita S, et al. Relationship between fibrosis detected on late gadolinium-enhanced cardiac magnetic resonance and re-entrant activity assessed with electrocardiographic imaging in human persistent atrial fibrillation. *JACC Clin Electrophysiol* 2018;4:17–29. https://doi.org/10.1016/j.jacep.2017.07.019; PMID: 29479568.

69. Zahid S, Cochet H, Boyle PM, et al. Patient-derived models link re-entrant driver localization in atrial fibrillation to fibrosis spatial pattern. *Cardiovasc Res* 2016;110:443–54. https://doi.org/10.1093/cvr/cvw073; PMID: 27056895.

70. Gharaviri A, Bidar E, Potse M, et al. Epicardial fibrosis explains increased endo-epicardial dissociation and epicardial breakthroughs in human atrial fibrillation. *Front Physiol* 2020;11:68. https://doi.org/10.3389/fphys.2020.00068; PMID: 32153419.

71. Boyle PM, Del Álamo JC, Akoum N. Fibrosis, atrial fibrillation and stroke: clinical updates and emerging mechanistic models. *Heart* 2021;107:99–105. https://doi.org/10.1136/heartjnl-2020-317455; PMID: 33097562.

72. Roney CH, Pashaei A, Meo M, et al. Universal atrial coordinates applied to visualisation, registration and construction of patient specific meshes. *Med Image Anal* 2019;55:65–75. https://doi.org/10.1016/j.media.2019.04.004; PMID: 31026761.

73. Herrera CR, Grandits T, Plank G, et al. Physics-informed neural networks to learn cardiac fiber orientation from multiple electroanatomical maps. *Eng Comput* 2022;38:3957–73. https://doi.org/10.1007/s00366-022-01709-3.

74. Roney CH, Sim I, Yu J, et al. Predicting atrial fibrillation recurrence by combining population data and virtual cohorts of patient-specific left atrial models. *Circ Arrhythm Electrophysiol* 2022;15:e010253. https://doi.org/10.1161/circep.121.010253; PMID: 35089057.

75. Shade JK, Ali RL, Basile D, et al. Preprocedure application of machine learning and mechanistic simulations predicts likelihood of paroxysmal atrial fibrillation recurrence following pulmonary vein isolation. *Circ Arrhythm Electrophysiol* 2020;13:e008213. https://doi.org/10.1161/circep.119.008213; PMID: 32536204.

76. Martinez-Legazpi P, Rossini L, Pérez del Villar C, et al. Stasis

mapping using ultrasound: a prospective study in acute myocardial infarction. *JACC Cardiovasc Imaging* 2018;11:514–5. https://doi.org/10.1016/j.jcmg.2017.06.012; PMID: 28917683.

77. Garg P, Van Der Geest RJ, Swoboda PP, et al. Left ventricular thrombus formation in myocardial infarction is associated with altered left ventricular blood flow energetics. *Eur Heart J Cardiovasc Imaging* 2019;20:108–17. https://doi.org/10.1093/ehjci/jey121; PMID: 30137274.

78. Delgado-Montero A, Martinez-Legazpi P, Desco MM, et al. Blood stasis imaging predicts cerebral microembolism during acute myocardial infarction. *J Am Soc Echocardiogr* 2020;33:389–98. https://doi.org/10.1016/j.echo.2019.09.020; PMID: 31813676.

79. Demirkiran A, Hassell MECJ, Garg P, et al. Left ventricular four-dimensional blood flow distribution, energetics, and vorticity in chronic myocardial infarction patients with/without left ventricular thrombus. *Eur J Radiol* 2022;150:110233. https://doi.org/10.1016/j.ejrad.2022.110233; PMID: 35278980.

80. Driessen RS, Danad I, Stuijfzand WJ, et al. Comparison of coronary computed tomography angiography, fractional flow reserve, and perfusion imaging for ischemia diagnosis. *J Am Coll Cardiol* 2019;73:161–73. https://doi.org/10.1016/j.jacc.2018.10.056; PMID: 30654888.

81. Otani T, Al-Issa A, Pourmorteza A, et al. A computational framework for personalized blood flow analysis in the human left atrium. *Ann Biomed Eng* 2016;44:3284–94. https://doi.org/10.1007/S10439-016-1590-X; PMID: 26968855.

82. Bosi GM, Cook A, Rai R, et al. Computational fluid dynamic analysis of the left atrial appendage to predict thrombosis risk. *Front Cardiovasc Med* 2018;5:34. https://doi.org/10.3389/FCVM.2018.00034; PMID: 29670888.

83. García-Isla G, Olivares AL, Silva E, et al. Sensitivity analysis of geometrical parameters to study haemodynamics and thrombus formation in the left atrial appendage. *Int J Numer Method Biomed Eng* 2018;34:e3100. https://doi.org/10.1002/cnm.3100; PMID: 29737037.

84. Masci A, Barone L, Dedè L, et al. The impact of left atrium appendage morphology on stroke risk assessment in atrial fibrillation: a computational fluid dynamics study. *Front Physiol* 2018;9:1938. https://doi.org/10.3389/fphys.2018.01938; PMID: 30723422.

85. Masci A, Alessandrini M, Forti D, et al. A proof of concept for computational fluid dynamic analysis of the left atrium in atrial fibrillation on a patient-specific basis. *J Biomech Eng* 2020;142:011002. https://doi.org/10.1115/1.4044583; PMID: 31513697.

86. Feng L, Gao H, Griffith B, et al. Analysis of a coupled fluid-structure interaction model of the left atrium and mitral valve. *Int J Numer Method Biomed Eng* 2019;35:e3254. https://doi.org/10.1002/cnm.3254; PMID: 31454470.

87. Sanatkhani S, Nedios S, Menon PG, et al. Subject-specific calculation of left atrial appendage blood-borne particle residence time distribution in atrial fibrillation. *Front Physiol* 2021;12:633135. https://doi.org/10.3389/fphys.2021.633135; PMID: 34045972.

88. Gonzalo A, García-Villalba M, Rossini L, et al. Non-Newtonian blood rheology impacts left atrial stasis in patient-specific simulations. *Int J Numer Method Biomed Eng* 2022;38:e3597. https://doi.org/10.1002/cnm.3597; PMID: 35344280.

89. Brunton SL, Noack BR, Koumoutsakos P. Machine learning for fluid mechanics. *Annu Rev Fluid Mech* 2020;52:477–508. https://doi.org/10.1146/annurev-fluid-010719-060214.

90. Morales Ferez X, Mill J, Juhl KA, et al. Deep learning framework for real-time estimation of in-silico thrombotic risk indices in the left atrial appendage. *Front Physiol* 2021;12:694945. https://doi.org/10.3389/fphys.2021.694945; PMID: 34262482.

91. Itu L, Rapaka S, Passerini T, et al. A machine-learning approach for computation of fractional flow reserve from coronary computed tomography. *J Appl Physiol (1985)* 2016;121:42–52. https://doi.org/10.1152/japplphysiol.00752.2015; PMID: 27079692.

92. Nahas H, Au JS, Ishii T, et al. A deep learning approach to resolve aliasing artifacts in ultrasound color flow imaging. *IEEE Trans Ultrason Ferroelectr Freq Control* 2020;67:2615–28. https://doi.org/10.1109/TUFFC.2020.3001523; PMID: 32746180.

93. Arzani A, Wang JX, D'Souza RM. Uncovering near-wall blood flow from sparse data with physics-informed neural networks. *Phys Fluids* 2021;33:071905. https://doi.org/10.1063/5.0055600.

94. Ferdian E, Dubowitz DJ, Mauger CA, et al. WSSNet: aortic wall shear stress estimation using deep learning on 4D flow MRI. *Front Cardiovasc Med* 2022;8:769927. https://doi.org/10.3389/fcvm.2021.769927; PMID: 35141290.

95. Avendi MR, Kheradvar A, Jafarkhani H. A combined deep-learning and deformable-model approach to fully automatic

segmentation of the left ventricle in cardiac MRI. *Med Image Anal* 2016;30:108–19. https://doi.org/10.1016/J. MEDIA.2016.01.005; PMID: 26917105.

96. Bernard O, Lalande A, Zotti C, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE Trans Med Imaging* 2018;37:2514–25. https://doi.org/10.1109/tmi.2018.2837502; PMID: 29994302.

97. Raissi M, Yazdani A, Karniadakis GE. Hidden fluid mechanics: learning velocity and pressure fields from flow visualizations. *Science* 2020;367:1026–30. https://doi.org/10.1126/science.aaw4741; PMID: 32001523.

98. Yin M, Zheng X, Humphrey JD, Karniadakis GE. Non-invasive inference of thrombus material properties with physics-informed neural networks. *Comput Methods Appl Mech Eng* 2021;375:113603. https://doi.org/10.1016/j.cma.2020.113603; PMID: 33414569.

99. Kissas G, Yang Y, Hwuang E, et al. Machine learning in cardiovascular flows modeling: predicting arterial blood pressure from non-invasive 4D flow MRI data using physics-informed neural networks. *Comput Methods Appl Mech Eng* 2020;358:112623. https://doi.org/10.1016/J.CMA.2019.112623.

100. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. *J Glob Health* 2019;9:010318. https://doi.org/10.7189/jogh.09.020318; PMID: 31788229.

101. Guidotti R, Monreale A, Ruggieri S, et al. A survey of methods for explaining black box models. *ACM Comput Surv* 2019;51:1–42. https://doi.org/10.1145/3236009.

102. Alaa AM, Bolton T, Angelantonio E Di, et al. Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK Biobank participants. *PLoS One* 2019;14:e0213653. https://doi.org/10.1371/journal.pone.0213653; PMID: 31091238.