Artificial intelligence for decision support in acute stroke — current roles and potential

Andrew Bivard[1,2], *Leonid Churilov*[2] *and Mark Parsons*[1,2]✉

Abstract | The identification and treatment of patients with stroke is becoming increasingly complex as more treatment options become available and new relationships between disease features and treatment response are continually discovered. Consequently, clinicians must constantly learn new skills (such as clinical evaluations or image interpretation), stay up to date with the literature and incorporate advances into everyday practice. The use of artificial intelligence (AI) to support clinical decision making could reduce inter-rater variation in routine clinical practice and facilitate the extraction of vital information that could improve identification of patients with stroke, prediction of treatment responses and patient outcomes. Such support systems would be ideal for centres that deal with few patients with stroke or for regional hubs, and could assist informed discussions with the patients and their families. Moreover, the use of AI for image processing and interpretation in stroke could provide any clinician with an imaging assessment equivalent to that of an expert. However, any AI-based decision support system should allow for expert clinician interaction to enable identification of errors (for example, in automated image processing). In this Review, we discuss the increasing importance of imaging in stroke management before exploring the potential and pitfalls of AI-assisted treatment decision support in acute stroke.

[1]*Department of Medicine and Public Health, University of Melbourne, Melbourne, VIC, Australia.*

[2]*Melbourne Medical School, University of Melbourne, Melbourne, VIC, Australia.*

✉*e-mail: markp@unimelb.edu.au*

https://doi.org/10.1038/s41582-020-0390-y

The past decade has seen an enormous shift in how patients with acute stroke are assessed and treated, and these changes have improved patient outcomes[1]. The widespread availability and speed of multimodal CT imaging has transformed the information that clinicians are using to make treatment decisions[2]. However, more options for imaging and treatment have increased the complexity of acute stroke assessment, which makes staying up to date with advances a constant challenge for stroke clinicians. In parallel, increases in the depth and breadth of imaging analysis techniques[3] have increased the skills required for the interpretation of imaging data[4], reflected in the rapid alterations of guidelines for stroke treatment and assessment over the past 2 years[5]. Traditional randomized trials to assess new imaging tools that are being developed are unlikely to be done because assessing the impact of imaging on treatment decisions would require limiting the information available to clinicians — preventing access to potentially vital information in this way could be contentious.

We suggest that further automation of image capture and analysis with artificial intelligence (AI) will be needed to harness the full potential of modern stroke assessment methods. However, to be practice changing, AI should increase the accuracy of diagnosis, improve the speed of decision making and add precision to individual patient care. In this Review, we explore the use of AI and decision assistance in clinical practice for stroke — AI could have other roles in stroke management, but we focus on its application to image interpretation. We first summarize the importance of imaging in stroke and the clinical benefits it can provide, before considering the types of AI and how they can be applied to the development of decision support systems in the management of acute ischaemic stroke. We consider the potential of such systems, the difficulties involved and suggest how the field could move forward. We do not discuss in depth the AI techniques available, but their resulting application.

The importance of imaging in stroke

Revolutionary developments in the treatment of ischaemic stroke, including pharmacological thrombolysis and mechanical clot retrieval, have been seen in the past decade. The use of imaging for the selection of patients has had a crucial role in these developments, and clinical imaging is essential to the application of these therapies in practice. In this section, we discuss

Key points

- Imaging-based treatment guidance has been demonstrated as an effective approach in patients with a suspected stroke.
- Clinical trials in which imaging is not used for patient selection are likely to include many patients with minor stroke or stroke mimics, making treatment effects difficult to detect.
- Artificial intelligence (AI) and machine learning could provide image interpretation that equals or exceeds that of experts and could collate key features to assist clinicians with treatment decisions.
- AI could be used to generate estimations of likely patient outcomes, which would not only be useful for assisting treatment decisions but also for informing family discussions.
- AI-based decision assistance systems could be especially useful for centres without dedicated stroke specialists.
- Any decision assistance tools must be validated and applied appropriately, and clear guidelines are needed to define how useful systems are in clinical practice.

the imaging modalities used and the benefits of imaging in the clinical setting.

Imaging modalities

Modern imaging assessments for ischaemic stroke are based on CT and MRI. CT is currently the most ubiquitous imaging platform owing to its low cost relative to MRI, the limited requirements for prescan safety assessments and ease of access in most health settings globally. Before the advent of thrombectomy, most stroke guidelines included only non-contrast CT as a requirement for the assessment of patients with suspected stroke. Additional imaging was considered wasteful and time consuming, a point that is still argued by some[6]. However, a radiological diagnosis of hyperacute stroke by use of non-contrast CT alone is often a diagnosis of exclusion and, in most patients with stroke, no noticeable changes are seen with non-contrast CT for up to 6 h after symptom onset[7]. Furthermore, CT has a sensitivity of just 26% for acute detection of ischaemic tissue when compared with MRI[8].

These limitations of non-contrast CT mean that trials of treatments for stroke in which patients are only assessed with non-contrast CT are likely to include a high number of patients with stroke mimics (for example, migraine with aura, focal seizure or psychogenic symptoms) and patients in whom spontaneous reperfusion occurs; in these subsets of patients any benefit of therapy is difficult to measure owing to an otherwise good natural history. Only 17% of patients with stroke have a large-vessel occlusion, for which early treatment has the greatest effect, so a broadly selected cohort makes it challenging to identify treatment benefits. These problems were particularly apparent in the NORTEST trial[9], in which 0.4 mg/kg tenecteplase was compared with 0.9 mg/kg alteplase for acute stroke thrombolysis and no significant difference was seen between the effects of treatments. Patient selection for NORTEST was based on non-contrast CT, and 18% of enrolled patients did not have a final diagnosis of stroke on subsequent review. In other large phase III trials of thrombolysis in stroke, the rates of stroke mimics have not been reported, detailed post hoc analysis has not been conducted and/or the presence of spontaneous reperfusion has not been assessed[10]. The inability to identify the true target population for thrombolysis (that is, those who could benefit from reperfusion therapy) by the use of non-contrast CT alone means that a large proportion of patients who are enrolled in such trials have a good natural history, making it impossible to show a treatment effect[11].

Modern imaging techniques, including CT angiography, magnetic resonance angiography, CT perfusion and magnetic resonance perfusion can positively identify a stroke by detecting a blocked cerebral blood vessel (angiography) or by identifying ischaemic tissue (perfusion imaging). For this reason, assessment with CT angiography is a core requirement when considering use of thrombectomy because this approach provides positive identification of a large-vessel occlusion (the treatment target) and confirmation of good access to the thrombus. Importantly, however, up to 70% of all ischaemic strokes involve an occlusion that is challenging to identify with CT angiography because it is too distal, but patients with such strokes could still be eligable for and benifit from treatment with thrombolysis.

The relative clinical values of each of the modern image acquisition techniques is a matter of debate because trials of thrombolysis versus placebo have involved only non-contrast CT to rule out a primary haemorrhage, and patients with suspected stroke — but no positive confirmation of ischaemia — could have been enrolled and treated. Nevertheless, looking beyond clinical utility to the development of AI and decision support, we must consider that the measures generated by additional imaging modalities could be important in a model even if they are not widely considered to be clinically important.

Clinical benefits of imaging

In the past few years, several trials have been conducted to assess whether modern imaging in acute stroke can help to improve selection of patients who are likely to respond to treatment and improve patient outcomes, including fewer adverse events. These trials have demonstrated that imaging-based selection of patients for treatment with thrombectomy enables successful identification of patients who are likely to respond to thrombectomy treatment beyond the traditional treatment time window of 4.5 h (the DAWN[12] and DEFUSE 3 (REF.[13]) trials) and successful thrombolysis in patients with an unknown time of stroke onset (the WAKE-UP[14] and EXTEND[15] trials). These four trials relied on modern imaging with CT or MRI to enable positive identification of a stroke and identify patients with an imaging profile that favoured reperfusion therapy. The results of these trials have major implications for the use of imaging in everyday clinical practice. In particular, imaging-based selection could justify the transfer of patients from primary stroke centres to a thrombectomy-capable site in cases when this would not previously have been done because the transfer could not be completed in time for treatment within 6 h of stroke onset.

Clinical studies have demonstrated the particular benefits of CT perfusion in the management of acute stroke. Currently, CT perfusion is generally not included in the standard of care for ischaemic stroke, yet the technique can characterize tissue viability by measuring haemodynamic changes in cerebral tissue. Perfusion imaging (CT and MRI) provides measures of cerebral blood flow, cerebral blood volume, mean transit time and tissue transit (expressed as time to peak, time to maximum, or delay). Processing of these images enables identification of a threshold to distinguish ischaemic tissue (FIG. 1). In this way, the volume and locations of irreversibly damaged tissue (the ischaemic core) and potentially salvageable ischaemic tissue (the penumbra) can be estimated. This approach has been validated against the use of MRI in patients who were treated either successfully or unsuccessfully with thrombolysis. Some variability exists in the thresholds between patients and software and so these tissue classifications are said to be estimates rather than exact measures. However, these thresholds have been reproduced multiple times so seem to be robust[16–18].

The benefit of using CT perfusion to identify ischaemic tissue and salvageable tissue prior to reperfusion therapy is highlighted by comparing trials of tenecteplase. In the ATTEST trial of tenecteplase[19], patients were enrolled on the basis of standard clinical criteria and the overall results were negative. In another trial of tenecteplase conducted in Australia, CT perfusion was used to enrol patients with positively identified salvageable tissue and the overall outcome was a positive treatment effect[20]. CT perfusion data had been collected in the ATTEST trial as a biomarker of treatment response, and a post hoc pooling analysis showed that only 34% of patients in the ATTEST trial had salvageable tissue identifiable with CT perfusion, meaning the majority of patients were unlikely to benefit from the treatment because they did not have a treatment target, such as penumbral tissue or a vessel occlusion. The pooled data revealed a substantial clinical benefit of tenecteplase for patients with salvageable tissue (that is, a reperfusion target)[21].

Analysis has shown that the ischaemic core volume estimated with CT perfusion is related to patient outcomes, treatment responses and rates of adverse events[22]. For this reason, CT perfusion, and specifically the ischaemic core volume, were used in the extended time window trials to select or exclude patients for enrolment[17–20]. In light of these positive trials, CT perfusion is being widely implemented along with CT angiography for the identification of patients who are eligible for thrombectomy and thrombolysis. With the increasing uptake of CT perfusion, applications have become available to automate the processing of perfusion data and distribute the results on email or picture archiving and communication systems. On this basis, we see automated CT perfusion as one of the first steps towards decision support in stroke since the volumes of the ischaemic core and penumbra are strongly related to patient outcomes and treatment responses[23].

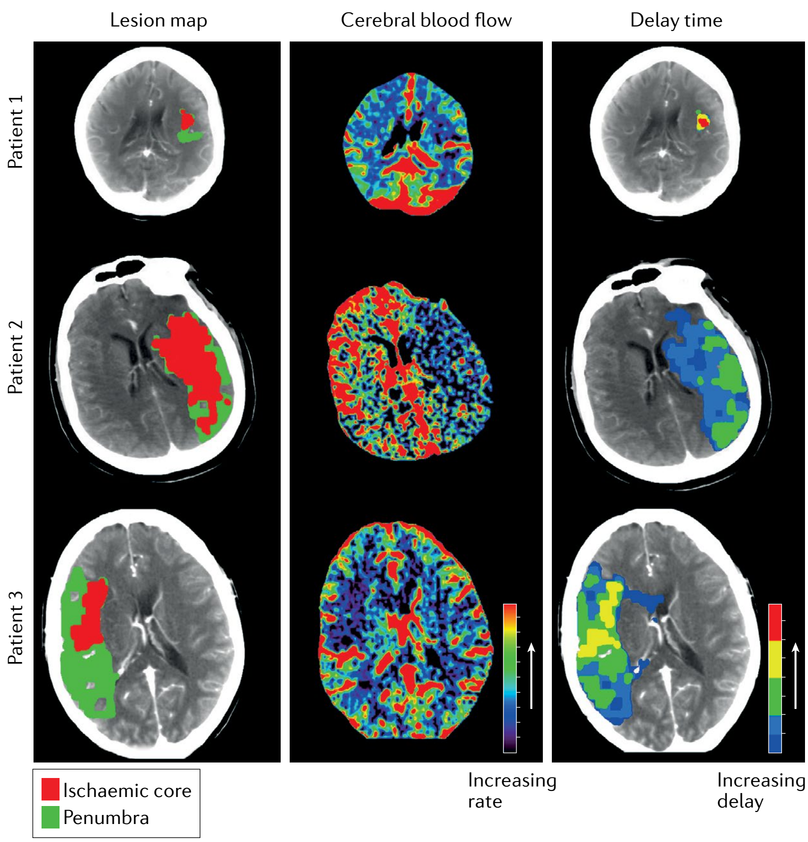

Fig. 1 | **Perfusion imaging outputs from validated automated post-processing software.** Each column illustrates a different CT perfusion measure. The lesion map displays the ischaemic core and penumbra. Cerebral blood flow represents the speed and volume of contrast agent arrival to the voxel. Contrast delay time is the time taken for the contrast agent to arrive at each individual voxel; prolonged (>3s) delays indicate ischaemia. The cerebral blood flow and contrast delay maps are used to generate the lesion map. Patient 1 has no penumbra and a very small lesion, so thrombolysis is likely to be futile because no treatment target exists. Patient 2 has a very large infarct core (>70 ml) and thrombolysis is therefore likely to be futile. Patient 3 has a large penumbra and small ischaemic core, making them an ideal candidate for thrombolysis treatment.

The potential of AI

Improvements in stroke therapy options and available imaging assessments have provided clear evidence that treatment responses in stroke are highly heterogeneous and are affected by clinical factors, such as the patient's age and the severity of stroke, and by factors that can be assessed with imaging, such as how much salvageable brain tissue (ischaemic penumbra)[3] is present before treatment. If brain tissue is irreversibly injured (infarcted), restoration of perfusion with thrombectomy or intravenous thrombolysis might not improve clinical outcomes[24]. Such variability in the response to reperfusion treatment for ischaemic stroke is well documented, and modern imaging can be used to identify subgroups of patients with particular treatment responses to intravenous thrombolysis. As such, there is enormous potential for an AI-based decision support system to be accurate, fast and widely available if it is reliable and related to individual patient outcomes.

Importantly, the data used to generate a prediction for decision support must be relevant to the underlying disease process. The routine use of CT angiography and CT perfusion to positively identify features of ischaemic stroke, such as a vessel occlusion or ischaemic brain tissue, provides such data. As these assessment methods improve and provide more granularity for the identification of patient-specific features, the accuracy of outcome prediction models will also improve, resulting in outcome predictions that can inform clinical decisions for an individual patient. However, currently, the means of validation and identification of measures that are meaningful for clinicians need to be further explored.

Artificial intelligence — the basics

The field of AI is broadly defined as the design, evaluation and use of non-human 'intelligent agents' — systems that perceive their environments and can mimic human cognitive tasks such as learning, problem solving[25] and acting to maximize their chances of success. Since its inception as an academic discipline >70 years ago, the field of AI has undergone several major transformative waves that can be characterized by shifts in emphasis between broad classes of problems (for example, planning, knowledge representation, formal reasoning, classification and prediction).

Machine learning is a modern branch of AI that involves the development of algorithms to identify patterns or groups in data in an automated process. In contrast to earlier incarnations of AI in which the emphasis was on intelligence through understanding and explaining of data, machine learning algorithms do not necessarily involve a semantic understanding of the input data, often resulting in a 'black box' between the problem and the proposed solution and a limited ability to explain the solution process[26,27]. Nevertheless, machine learning algorithms are often demonstrably more successful than other AI approaches at solving whole classes of problems.

Machine learning is still developing, and a lack of consistency remains in its definition and application. Emphasis on a particular feature is often used to define the whole machine learning domain — for example, the reliance on data rather than statistical hypotheses, or the automated behaviour that requires minimal human intervention. The proposed definitions of machine learning include: "a class of computational algorithms; a modelling strategy to let the data speak for themselves, to the extent possible, which makes it an attractive option for characterizing and predicting complex biological phenomena that do not have a priori models"[28]; and "a method of data analysis that automates analytical model building … based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention"[26,29].

A framework has been proposed[27] that outlines three requirements for any machine learning algorithm: the algorithm, including the inputs, must be represented in a formal language that a computer can handle; an evaluation function, such as an accuracy and error rate or a cost to utility ratio, that can distinguish good instances of the algorithm from bad ones; and a mathematics-based method to search the algorithm instances and identify the highest scoring ones for optimization. The choice of this optimization technique is key to the efficiency of the algorithm.

Supervised learning

In supervised learning algorithms, the underlying process provides inputs and associated outputs, and the task of the machine learning algorithm is to learn the best mapping function from the inputs to the output. This type of algorithm is particularly suited to addressing classification problems (in which the output is categorical) and regression problems (in which the output is measured on a difference or a ratio scale). Typical examples of machine learning algorithms for supervised learning include various regression models, classification and regression trees, random forests, naive Bayes trees, support vector machines and various artificial neural networks, such as feedforward, recurrent and convolutional neural networks. For example, a support vector machine has been used to detect the hyperdense middle cerebral artery dot sign on CT[22], and random forests have been used for automated quantification of cerebral oedema in patients with a malignant hemispheric infarct[30].

Unsupervised learning

In unsupervised learning algorithms, no outputs are specified and so no learning feedback is provided, and the goal is to identify the underlying structure or inherent grouping patterns in the input features. Unsupervised learning is used to address clustering problems, association rule problems and dimensionality reduction problems. Typical examples of machine learning algorithms for unsupervised learning include k-means clustering, principal component analysis, Boltzmann machine learning and different varieties of self-organizing neural networks (for example, Kohonen's self-organized maps). For example, convolutional neural networks have been applied to lesion segmentation of acute ischaemic stroke with diffusion-weighted imaging[31].

Deep learning

Deep learning is the latest paradigm in machine learning, and is characterized by machine learning algorithms with complex structures (for example, layers, hidden layers, feature maps and layer pools) combined with automated input feature selection that the algorithm derives directly from data. By contrast, standard machine learning relies on human involvement in feature selection to ensure representation of the system being modelled. The deep learning paradigm is therefore better suited to applications in which the input is presented in a relatively raw form with little prior expert knowledge.

Decision support in stroke care

Treatment decision making in acute stroke is complex and time pressured[32,33]. Given the number of variables involved in the decision-making process and their heterogeneous relationships with patient outcomes, the development of comprehensive predictive models based on individual pathophysiology is critical for guiding clinical practice and providing patients and their families with realistic expectations. Aspects of these predictive models that can be automated, and for which we can trust algorithms, should be automated; for example, calculation of the volumes of the ischaemic core or penumbra from CT perfusion images could be automated. Such automated processes would provide decision support by automating aspects of the decision process and classifying patient data, but would not generate individual patient management decisions.

To ensure that decision support systems have a meaningful impact, the requirements for these systems should reflect the nature of the decision-making process faced by stroke clinicians. This process includes making rapid decisions on the basis of a potentially overwhelming number of input features, including patient-specific features (for example, age, sex, clinical severity, comorbidities, ischaemic core volume, collateral status, clot location and burden, extent of hypodensity and accessibility of vessels for endovascular treatment) and features related to the systems and processes of care (such as logistical, financial or human resource constraints that limit treatment choices). Two types of decision support model are particularly suitable for such tasks[34]. The first is automation of processes for which use is “frequent and routine, with, in general, no need to prepare the model for each use”[12]. In these models — known as decision automation — tolerance for any type of error is usually very low because decisions based on the results of the model are reinforced by the model and are rarely corrected. The second type are models for routine decision support, which are “used to assist, but not replace, people making routine, repeated decisions”[12] and with which a user can override the automated process. Below, we provide illustrative examples of both decision support contexts (FIG. 2).

Decision automation

Given the large body of imaging data available and the large volume of raw information available during assessment of a patient with acute stroke, even experienced stroke clinicians rarely feel comfortable interpreting all imaging data under the time pressure required for acute treatment. For new clinicians, these assessments are challenging to learn and apply after only a few years of training. In this context, decision support has a role in reducing disagreement between clinicians and standardizing imaging assessment of patients with stroke. Ideally, trained models for imaging analysis and interpretation would be fully automated and would provide a summary for a clinical end user, who would then appreciate the clinical relevance of the information. Fully automated decision making is unlikely owing to the subtleties in clinical medicine that require an expert to apply their experience. However, automation of processes with decision automation can facilitate clinical decision making by addressing relatively straightforward questions, such as whether a CT scan contains hypodensities or early ischaemic changes that are indicative of a stroke. Examples of such automation of processes are given below.

Hypodensity detection. Non-contrast CT has been part of routine stroke care since the initial trials that validated the use of intravenous thrombolysis. Non-contrast CT is primarily used to identify haemorrhagic strokes but cannot reliably distinguish an ischaemic stroke from common stroke mimics. The early changes seen with non-contrast CT that indicate an ischaemic stroke are parenchymal hypodensity and focal swelling[35]. Parenchymal hypodensity is often seen as a loss of definition between grey and white matter, reflecting accumulation of water (oedema) in intracellular and extracellular areas of the ischaemic region[29]. This hypodensity on non-contrast CT is the most common sign

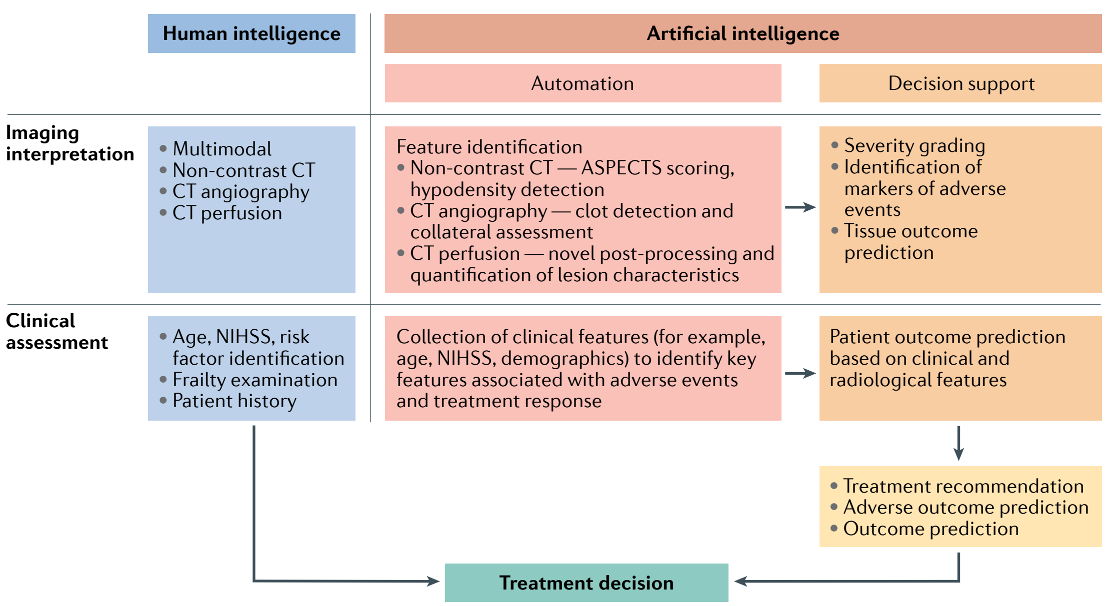

Fig. 2 | **Treatment decision making by humans alone and in combination with artificial intelligence.** The human intelligence column indicates what happens in current routine care, whereas the artificial intelligence column illustrates how automation of processes and decision support can add to human intelligence. Artificial intelligence is split into two; automation indicates the imaging assessments and clinical data collection that can provide outputs that feed into decision support. Combining treatment recommendations and outcome predictions from the decision support system with human assessments leads to the final treatment decision. ASPECTS, Alberta Stroke Program Early CT Score; NIHSS, National Institutes of Health Stroke Scale.

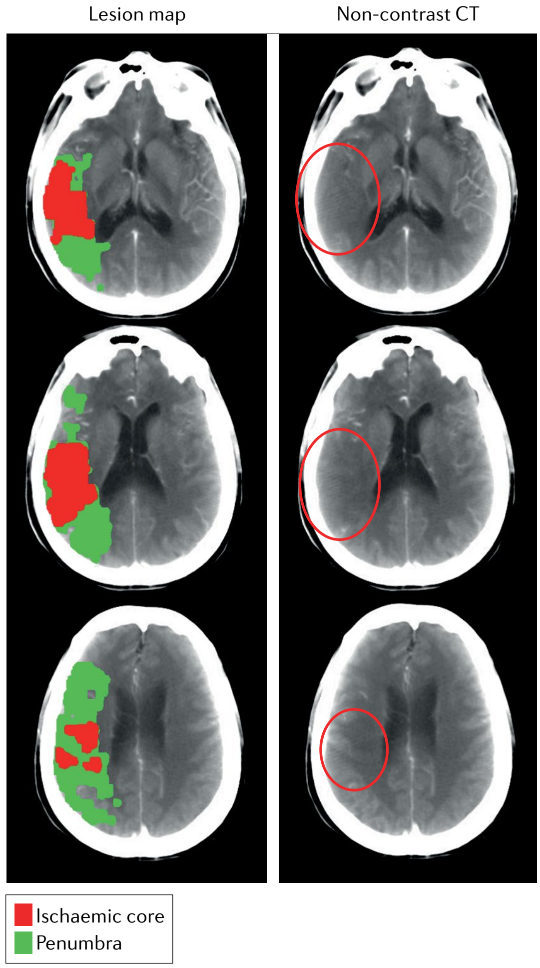

Fig. 3 | **Irreversible injury in ischaemic stroke.** Three CT scans (at different levels) from one patient who presented 4.5 h after onset of stroke-like symptoms. The lesion map shows the ischaemic region. The non-contrast CT images on the right illustrate an established hypodensity (circled) that corresponds to the area of ischaemia. Assessing the value of reperfusion therapy in this scenario is challenging but thrombolysis is associated with a high risk of haemorrhage.

of ischaemic stroke and probably indicates irreversible injury (FIG. 3).

The problem with assessing hypodensity is that, in most patients, very subtle or no changes are seen with non-contrast CT for up to 6 h after stroke onset. Infarctions and tissue hypodensity become more obvious after 12–24 h, which is when reperfusion therapy is less likely to be beneficial. Several attempts have been made to automate the measurement of hypodensity with machine learning techniques, and the performance has been comparable to that of expert manual segmentation[36,37]. This result is excellent given that the output would predominantly be used by junior staff to provide them with the equivalent of an expert non-contrast CT assessment.

ASPECTS assessment. The Alberta Stroke Program Early CT Score (ASPECTS) is a systematic method of assessing early ischaemic change on non-contrast CT that was proposed to improve the accuracy of acute assessment and has since been used in clinical trials[38]. ASPECTS is a ten-point quantitative topographic CT scan score determined from the evaluation of two standardized regions of the middle cerebral artery territory — the basal ganglia level and the supra-ganglionic level. The score is calculated by subtracting one point from ten for any evidence of early ischaemic change in each of the defined regions. However, the inter-observer variability for early changes on non-contrast CT is high (Cohen's κ coefficient 0.48–0.67)[39] and outcome prediction with ASPECTS is modest at best (sensitivity 0.78 and specificity 0.96)[38], and so incorporation of such a scoring system into a decision assistance algorithm is a challenge. The reliability of ASPECTS does improve with guidance from other information, such as the lesion location detected with CT perfusion or knowledge of the symptomatic hemisphere, indicating that integration of ASPECTS with clinical and CT perfusion data into one system could provide exponential gains in the reliability and clinical relevance of the assessments.

Automated ASPECTS scoring has been developed, and performance of this e-ASPECTS system with 132 scans was non-inferior when compared with expert readers[22,40]. The fact that ASPECTS only takes into account the large middle cerebral artery region and not other locations where strokes commonly occur, such as the anterior cerebral artery or posterior cerebral artery, means that the automated ASPECTS measurement has little role in automated clinical decision support, and the overall usefulness of such a limited score to a clinician is questionable. Importantly, however, only one side-by-side comparison of automated ASPECTS software packages has been conducted and demonstrated no equivalence between the software[41]. These comparisons are vital because the underlying algorithms differ between applications and are likely to produce different results, especially given the difficulty of scoring hypodensity on non-contrast CT.

Defining the aim of automation. The ultimate aim of using machine learning for automation of processes that contribute to decision making is an important consideration. Junior clinicians are most likely to want fast results that are equally as reliable as an expert opinion, so in this context algorithms do not need to improve on the accuracy of experts but reproduce it. However, cut-offs need to be defined to indicate when a measure can be considered clinically useful, otherwise any application can be considered useful simply because it uses AI. An alternative aim for machine learning in acute stroke assessments could be to derive new measures from existing data that would alter clinical practice altogether.

For example, when assessing the clinical utility of a tool such as automated detection of a clot on CT angiography, a clinician should expect the model to generate an area under the curve (AUC) of at least 0.9 and to detect all symptomatic occlusions in any vessel (not just M1 occlusions, which are easily detected). Algorithms should also be validated with random datasets, such as externally acquired data that include proximal or distal M2 occlusions, anterior circulation, posterior circulation and basilar artery occlusions, and suboptimal scans (for example, scans with motion artefacts). Without this depth of data and robustness of a model, the real clinical accuracy will not equal, let alone surpass, that of expert stroke physicians and neuroradiologists.

Routine clinical decision support

A fully automated decision process for the management of stroke is unlikely to ever be acceptable because decisions ultimately rely on clinical judgement in the face of uncertainty rather than simply on a set number of features that need to be considered. For example, clinicians often acquire information that is not part of a standard script, such as assessments of patient frailty, that a machine model is unlikely to appreciate but alters decision making. Indeed, application of an automated process based on limited information, such as the presence or absence of hypodensity on non-contrast CT, is unlikely to make a substantive change to a yes–no treatment decision as many additional factors influence the decision, but such variables will influence the ultimate clinical outcome of a patient. As a result, information gained from automated processes could alter the clinician's prediction about the response to treatment and/or the risk of an adverse event, such as a haemorrhage, without changing the treatment decision, meaning the information is used to support the decision rather than dictate it. For example, patients with an ischaemic core larger than 70 ml seen on CT perfusion rarely have a good outcome and are highly likely to die within the first month after stroke onset (FIG. 4), and so automated measurement of ischaemic core volume provides valuable assistance for treatment decisions.

An ideal clinical decision support tool would combine broad clinical and imaging information into a model that generates accurate estimations of likely outcomes under different treatment scenarios. Furthermore, information used by a decision support system should be malleable and modifiable in light of new information, re-processing of automated imaging data by an expert and updated clinical information. Indeed, images often include erroneous information as a result of external issues, such as patient motion or acquisition anomalies (FIG. 5). These errors are often detected by trained reviewers (FIG. 6) but in routine hyperacute practice they are often overlooked and treatment decisions are made on the basis of erroneous data. Treatment decision assistance could identify potential errors that could then be fixed or could enable erroneous data to be excluded.

Initial assessment. Clinical diagnosis of stroke requires a physical examination that ideally includes a standardized stroke severity assessment with the National Institutes of Health Stroke Scale (NIHSS), although this scale is not universally applied. A routine assessment also includes an attempt to rule out stroke mimics, such as a migraine or psychogenic presentations. These are standard classification tasks, but the number of classification options is enormous. The clinical definition of a stroke is very broad and routine assessments vary widely between centres and the experience of the clinicians[42], adding to the complexity of designing a tool for routine clinical use. Furthermore, an ever-increasing number of imaging variables are being related to patient outcomes; for example, very low cerebral blood volume and extremely delayed contrast have been associated with an increased risk of haemorrhage[43,44], different cerebral blood flow thresholds can predict the size of the ischaemic core[43], clot burden could be a marker of the response to thrombolysis treatment[44], and quantification of white matter disease can be used to predict haemorrhage[42]. The volume of this information is extraordinary, and the ways in which each of these individual variables affect responses to reperfusion therapy in individual patients are not at all clear. Given this complexity, it is not feasible that a decision automation system could output a clear and valid yes–no treatment recommendation that would be acceptable in today's clinical space. However, these variables can be used to guide treatment or alter clinical

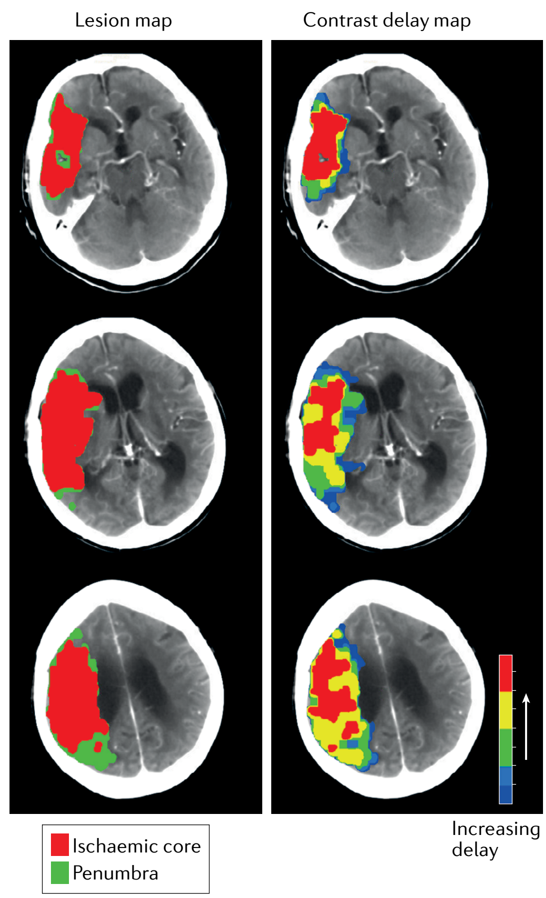

Fig. 4 | **Identification of patients who are unlikely to benefit from therapy.** Three CT scans (at different levels) from one patient who presented to hospital within 4.5 h of stroke onset but with a very large ischaemic core, shown in the lesion map. The contrast delay map indicates extensive delayed arrival of contrast agent (indicating ischaemia) throughout the ischaemic hemisphere. The scans indicate that the majority of the acute perfusion lesion is infarcted and little tissue remains to salvage. Little evidence suggests that therapy is beneficial in this scenario, raising ethical questions about therapy. An artificial intelligence-powered outcome prediction model in this situation would at least inform family discussions.

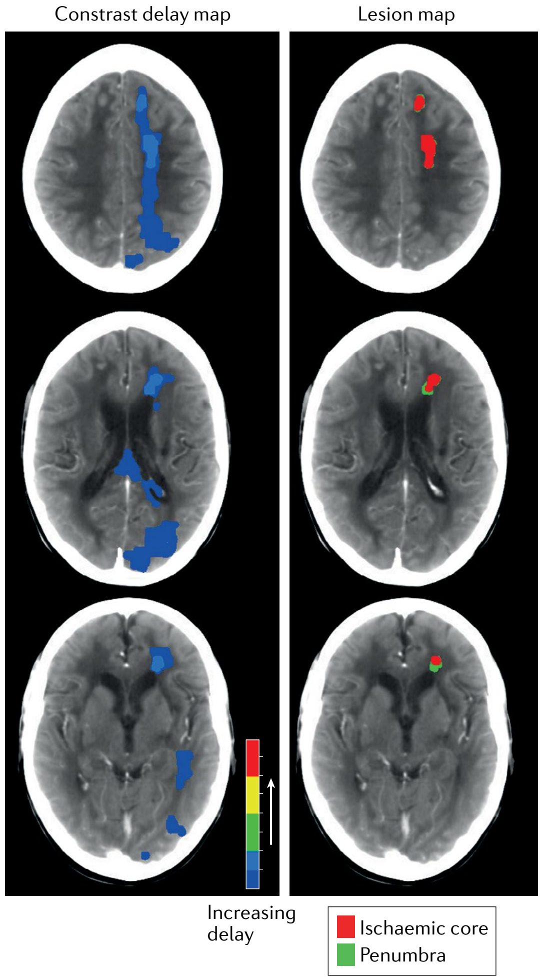

Fig. 5 | **Clinical implications of poor image acquisition.** Acute CT perfusion scans (at different levels) of a patient who presented with clinical symptoms that suggested a stroke, processed with automated software. The contrast delay map shows delayed arrival of the contrast agent, indicating ischaemia, but the delay is largely below the ischaemic threshold, and so the lesion map is likely to be inaccurate. The widespread delay can result from poor image acquisition (for example, as a result of patient motion, slow contrast injection or incorrect contrast injection) or chronic disease. Decision support systems will need to take into account such factors that can influence imaging results.

practice to enable collection of information that is considered important but would not routinely be collected in practice owing to inter-clinician variation. Decision support can reduce this inter-clinician variation and therefore guide treatment decision making to optimize patient selection for reperfusion therapy.

Treatment selection. When a patient presents with an acute neurological deficit, the priority for the frontline clinician is to correctly diagnose whether they have a stroke and, if so, whether it is a primary haemorrhagic stroke or an ischaemic stroke. In the case of an ischaemic stroke, the next step is to determine whether the patient is eligible for treatment with thrombolysis or thrombectomy and to estimate whether treatment will improve the outcome of the patient and to what extent. The goal of such an assessment is to estimate the potential benefit (or harm) of reperfusion treatment for ischaemic stroke. In this context, treatment decision assistance seems to be a beneficial development as it would aid in the identification of contraindications for therapy and would automatically incorporate clinical and imaging assessments that can be summarized for the clinician.

Global variations between health services and health cultures need to be taken into account in the development of any decision support system so that local clinical and non-clinical variations can be incorporated; any decision support system is unlikely to be 'one size fits all'. For example, the application of thrombectomy and intravenous thrombosis varies regionally for reasons that do not relate to the interpretation of clinical or imaging data but are influenced by factors such as a lack of stroke physicians or of imaging and endovascular facilities. Conversely, high rates of treatment might relate to various incentives (such as substantial payments) in some regions. In this way, the structures of health systems and economic environments have critical roles in the delivery of health care and so would need to be incorporated into any decision support system. These factors also define the role of decision support; for example, if no thrombectomy services are available in a given region, the system could only be used to identify candidates for thrombolysis with limited contraindications. Alternatively, if travel times required for thrombectomy are long, such as in regional Australia, the decision to transfer a patient must take into account factors such as how the use of resources for a transfer would limit resources for transfer or retrieval of other patients. In this context, clinicians may only decide to transfer patients for whom the likelihood of clinical benefit is large and the risk of adverse events is minimal.

Similarly, predefined treatment benefit to cost ratio thresholds differ between countries and health systems on the basis of affordability (and thus also the ability to accurately predict individual patient responses to therapy), so the treatment recommendation supported by a decision assistance system is likely to differ according to the health service in which the system is implemented. Other factors include local variations in protocols and guidelines, disagreements about what type of imaging should be used to assess patients (the accuracy of decision support systems could differ with different imaging approaches owing to variation in imaging features included), and which patients should not be offered treatment because of regional variations in treatment criteria.

Acceptable accuracy. Assessment with multimodal imaging substantially improves prognostication in stroke, enabling clinicians to make more informed treatment decisions and to provide clarity in discussions with patients and their family. However, an element of expertise is required for interpretation and application of results from multimodal imaging of an individual patient, so such information is rarely (if ever) used to its full potential. For example, CT enables clinicians to positively identify ischaemia and to more accurately predict whether an individual will have an excellent clinical outcome after thrombolysis[45]. This accuracy of prediction is usually expressed as the AUC, which is a performance measure for classification problems that indicates how well the classification model can distinguish between classes. The AUC can vary between 0 and 1, and higher values indicate better classification — for example, an AUC >0.9 would indicate accurate prediction of clinical outcomes. However, the pretreatment volumes of the ischaemic core and the penumbra are directly related to individual patient responses to reperfusion therapy[46] and so improving the outcome prediction with CT perfusion beyond an AUC of 0.9 is challenging, and evidence suggests that AI does not substantially improve on such results[47].

A decision support model that combines data from CT perfusion with clinical data could take into account the age of the patient, their premorbid status and the penumbral volume to estimate that, for example, saving 50 ml of penumbral volume with effective reperfusion from thrombectomy would lengthen the patient's healthy life by 2 years (compared with no treatment)[48]. This kind of information is currently not available to clinicians in the acute stroke setting.

To achieve accurate predictions with machine learning algorithms, features presented to the algorithm need to provide a reasonable representation of the underlying clinical problem. For example, data from the successful MR CLEAN trial of thrombectomy have been used to compare machine learning techniques (such as random forests, support vector machine, neural network, logistic regression models) for the prediction of outcomes for trial participants. All models performed well in predicting 3-month functional outcomes (AUC 0.77–0.79)[49], but perfusion imaging data were not included in the predictive dataset, leaving potential for the AUC to increase substantially. Perfusion imaging data that provided a measure of ischaemia in patients with ischaemic stroke would have improved the outcome. Contrary to the common idea that if all data are included in a machine learning model then the model will get better, meaningful data are required for optimal outputs.

Machine learning could also help to make predictions about adverse events, thereby supporting the clinician in assessing the risks of treatment. Adverse events, such as symptomatic brain haemorrhage, generally result in severe disability or death. Furthermore, the occurrence of such events can affect future decision making of the clinicians involved. Many factors have been associated with adverse events or poor outcomes[48,50–52], yet application of this knowledge in individual cases is challenging for clinicians. Furthermore, whether the probability of a haemorrhage would change the treatment decision also depends on the probability and likely extent of the benefit from therapy if haemorrhage occurs. This situation is ideally suited to the use of decision support, as a support system would provide the clinician with the information they need to weigh up the probability of good and poor outcomes for an individual and to discuss treatment with the patient and their family. Machine learning has been used in one study to predict the risk of symptomatic intracerebral haemorrhage in 116 patients with acute ischaemic stroke who were treated with intravenous thrombolysis[53]. Of those patients, 16 developed symptomatic intracerebral haemorrhage, and the model achieved an AUC of 0.744. However, this AUC is still low for implementation in routine practice, where the AUC would ideally be >0.9.

Real-world testing

To develop a clinical decision support system, we need to choose the outcome that is to be predicted; for example, we need to decide whether we want a system that only positively identifies patients with stroke, or a system

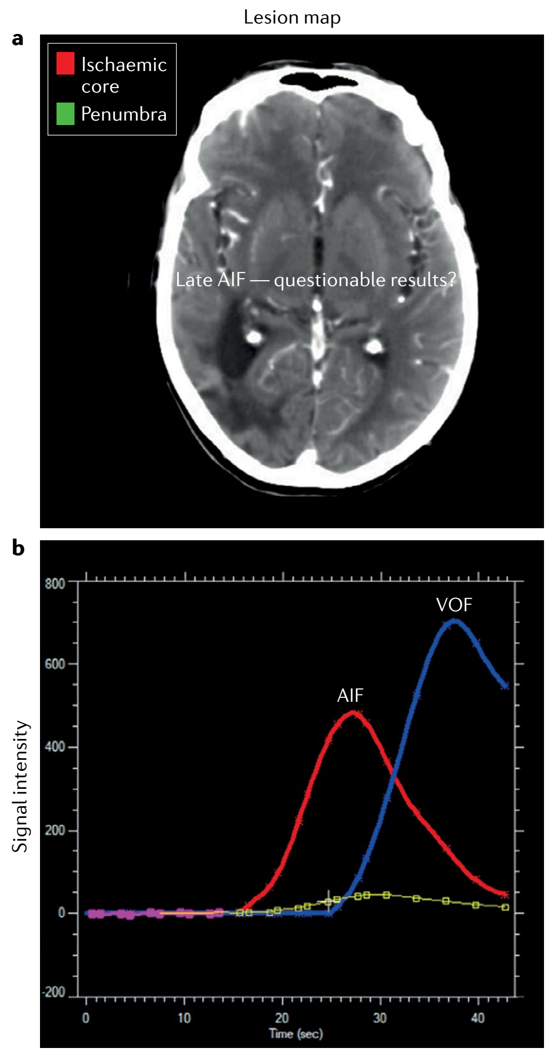

Fig. 6 | **A poorly acquired acute CT perfusion scan.** **a** | No lesion is present in the lesion map — a red ischaemic core and green penumbra would be expected if there were a lesion. However, examination of the data on which these images are based in part **b** shows incomplete data. **b** | The arterial input function (AIF) and venous outflow function (VOF) from which the lesion map in part **a** was calculated. Many of the data are missing. The end of the VOF has been cut off and long tails for both the AIF and VOF would be expected but are missing. The resulting lesion map could, therefore, be inaccurate. This scenario is dangerous because a lesion can easily be miscalculated. Clinicians review these data to assess quality, and a decision assistance tool could be used for quality assurance.

that identifies patients who will respond to treatment. Before testing an AI application in the real world, further discussion is needed about the level of evidence and supporting data that will be acceptable for a system to be deployed and relied on in the clinic.

In a news release in February 2018, the FDA[54] defined AI algorithms as "a type of clinical decision support software that can assist providers in identifying the most appropriate treatment plan for a patient's disease or condition" and emphasized that these algorithms "should not be used as a replacement of a full patient evaluation or solely relied upon to make or confirm a diagnosis". These statements emphasize the intended use of AI models to support clinical experts in routine clinical decision making rather than to replace decision making with automated machine learning. The question is how accurate the system needs to be — clinicians often make mistakes, but we would probably not accept mistakes from an automated, machine-based system. The field of AI is immature and further advances are required, but expectations also need to be tempered. As an example, in one study an AI algorithm was used to analyse 112,000 chest x-ray images to detect pneumonia and the AUC was 0.76; this AUC is low and so the algorithm is unlikely to be used clinically, yet its performance was better than that of a team of four radiologists[55]. In another study, an AI algorithm for diagnosing trauma in 37,000 head CT scans achieved an AUC of 0.73, which is again low, but the results of the algorithm were generated 150-fold faster than radiologist reports[56], and so the system could be used to accelerate radiologist reporting. Furthermore, the AUC metric might not be the ideal measure of accuracy despite its common usage[57,58], adding further complexity.

Besides the nuance of acceptable measurements and outcomes, another important aspect is the level and type of validation that will be clinically acceptable for algorithms to be used to guide treatment decision making. Not everything requires a clinical trial, but testing new applications in the real world in patients who have a range of conditions (from stroke mimics to severe stroke) and with data that are not always optimal (for example, incorrect clinical information or severe motion on imaging) followed by re-testing in multiple environments (for example, different hospitals and countries) seems the best way forward where clinical trials are not feasible.

Conclusions

In the context of stroke management, AI is ideally suited to reducing inter-rater disagreement, improving standardization of assessments, enabling quick assessments of meaningful imaging data, supporting clinical decision making through identification of key treatment-relevant variables and synthesizing these variables to predict treatment responses. There is also considerable scope for the implementation of more advanced imaging analysis to provide clinicians with contextual information, such as the potential 'life expectancy' of the penumbra (which would be particularly helpful when long-distance inter-hospital transfers are required), and to analyse every voxel in an image rather than specific elements such as the penumbra, ischaemic core or hypodense region.

The need for decision support tools provides a strong motivation for researchers to pool data and standardize assessments so that clinicians can be provided with a comprehensive and validated output. Perhaps the greatest value of AI for stroke clinicians is its ability to amalgamate, prioritize and summarize a large volume of clinical and imaging characteristics of one individual and to compare these with fitted models that have undergone robust evaluation and optimization with large cohorts of data (>5,000 patients) to assist routine clinical decision making. This application of AI does not remove the need for clinician or radiologist assessment of imaging but is ideally suited to decision support.

Published online: 24 August 2020

1. Goyal, M. et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet* **387**, 1723–1731 (2016).
An excellent meta-analysis of thrombectomy trials that demonstrates a very large treatment effect.
2. Potter, C. A. et al. CT for treatment selection in acute ischemic stroke: a code stroke primer. *Radiographics* **39**, 6 (2019).
3. Sanelli, P. C. et al. Imaging and treatment of patients with acute stroke: an evidence-based review. *Am. J. Neuroradiol.* **35**, 1045–1051 (2014).
4. Bacharach, R., Niazi, M. & Ermak, D. Pitfalls of CT perfusion imaging in acute ischemic stroke: a case series. *Neurology* **90** (Suppl. 15), P3.206 (2018).
5. Adeoye, O. et al. Recommendations for the establishment of stroke systems of care: a 2019 update. A policy statement from the American Stroke Association. *Stroke* **50**, e187–e210 (2019).
6. Heit, J. & Wintermark, M. Perfusion computed tomography for the evaluation of acute ischemic stroke: strengths and pitfalls. *Stroke* **47**, 1153–1158 (2016).
7. Gao, J. et al. Visibility of CT early ischemic change is significantly associated with time from stroke onset to baseline scan beyond the first 3 hours of stroke onset. *J. Stroke* **19**, 340–346 (2017).
8. Fiebach, J. B. et al. CT and diffusion-weighted MR imaging in randomized order: diffusion-weighted imaging results in higher accuracy and lower interrater variability in the diagnosis of hyperacute ischemic stroke. *Stroke* **33**, 2206–2210 (2002).
9. Logallo, N. et al. Tenecteplase versus alteplase for management of acute ischaemic stroke (NOR-TEST): a phase 3, randomised, open-label, blinded endpoint trial. *Lancet Neurol.* **16**, 781–788 (2017).
10. IST-3 collaborative group. et al. The benefits and harms of intravenous thrombolysis with recombinant tissue plasminogen activator within 6 h of acute ischaemic stroke (the third international stroke trial [IST-3]): a randomised controlled trial. *Lancet* **379**, 2352–2363 (2012).
11. Bivard, A. et al. Too good to treat? Ischemic stroke patients with small computed tomography perfusion lesions may not benefit from thrombolysis. *Ann. Neurol.* **80**, 286–293 (2016).
A controversial article in which the benefit of thrombolysis for patients with a perfusion lesion <15 ml is questioned.
12. Nogueira, R. G. et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N. Engl. J. Med.* **378**, 11–21 (2018).
13. Albers, G. et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. *N. Engl. J. Med.* **378**, 11–21 (2018).
14. Thomalla, G. et al. MRI-guided thrombolysis for stroke with unknown time of onset. *N. Engl. J. Med.* **379**, 611–622 (2018).
Together with Albers et al. (2018), these papers represent medical breakthroughs as they show that the treatment time window for stroke caused by large-vessel occlusion can be extended from 6 h to 24 h with the aid of imaging-based patient selection.
15. Ma, H. Extending thrombolysis to 9 hours and wake-up stroke using perfusion imaging selection. *N. Engl. J. Med.* **380**, 1795–1803 (2019).
A clinical trial of thrombolysis that shows that the treatment time window can be extended to 9 h or that the treatment can be used for patients with an unknown time of onset but have an imaging treatment target seen on CT perfusion.
16. Mokin, M. et al. Predictive value of RAPID assessed perfusion thresholds on final infarct volume in SWIFT PRIME (Solitaire with the intention for thrombectomy as primary endovascular treatment). *Stroke* **48**, 932–938 (2017).
17. Hoving, J. W. et al. Volumetric and spatial accuracy of computed tomography perfusion estimated ischemic core volume in patients with acute ischemic stroke. *Stroke* **49**, 2368–2375 (2018).
18. Campbell, B. C. et al. Cerebral blood flow is the optimal CT perfusion parameter for assessing infarct core. *Stroke* **42**, 3435–3440 (2011).
19. Huang, X. et al. Alteplase versus tenecteplase for thrombolysis after ischaemic stroke (ATTEST): a phase 2, randomised, open-label, blinded endpoint study. *Lancet Neurol.* **14**, 368–376 (2015).
20. Parsons, M. et al. A randomized trial of tenecteplase versus alteplase for acute ischemic stroke. *N. Engl. J. Med.* **366**, 1099–1107 (2012).

21. Bivard, A. et al. Impact of computed tomography perfusion imaging on the response to tenecteplase in ischemic stroke: analysis of 2 randomized controlled trials. *Circulation* **135**, 440–448 (2017).
22. Takahashi, N. et al. An automated detection method for the MCA dot sign of acute stroke in unenhanced CT. *Radiol. Phys. Technol.* **7**, 79–88 (2014).
23. Bivard, A. et al. Perfusion computed tomography to assist decision making for stroke thrombolysis. *Brain* **138**, 1919–1931 (2015).
24. Hussein, H. M., Saleem, M. A. & Qureshi, A. I. Rates and predictors of futile recanalization in patients undergoing endovascular treatment in a multicenter clinical trial. *Neuroradiology* **60**, 557–563 (2018).
25. Russell, S. J. & Norvig, P. *Artificial Intelligence: a Modern Approach* 3rd edn (Prentice Hall, 2009).
26. SAS. Machine learning: what it is and why it matters. https://www.sas.com/en_au/insights/analytics/machine-learning.html (2020).
27. Domingos, P. A few useful things to know about machine learning. *Commun. ACM.* **55**, 78–87 (2012).
28. Brzdok, D., Krzywinski, M. & Altman, N. Machine learning: a primer. *Nat. Methods.* **14**, 1119–1120 (2017).
29. von Kummer, R. Ischemic stroke and tissue hypodensity on computed tomography. *Stroke* **30**, 1974 (1999).
30. Chen, Y. et al. Automated quantification of cerebral edema following hemispheric infarction: application of a machine-learning algorithm to evaluate CSF shifts on serial head CTs. *Neuroimage Clin.* **12**, 673–680 (2016).
31. Chen, L., Bentley, P. & Rueckert, D. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. *Neuroimage Clin.* **15**, 633–643 (2017).
32. Meretoja, A., Keshtkaran, M., Tatlisumak, T., Donnan, G. A. & Churilov, L. Endovascular therapy for ischemic stroke: save a minute–save a week. *Neurology* **88**, 2123–2127 (2017).
33. Meretoja, A. et al. Stroke thrombolysis: save a minute, save a day. *Stroke* **45**, 1053–1058 (2014).
34. Pidd, M. Why modelling and model use matter. *J. Oper. Res. Soc.* **61**, 14–24 (2010).
35. Tan, J. C. et al. Systematic comparison of perfusion CT and CT-angiography in acute stroke patients. *Ann. Neurol.* **61**, 533–543 (2007).
36. Maier, O., Schröder, C., Forkert, N. D., Martinetz, T. & Handels, H. Classifiers for ischemic stroke lesion segmentation: a comparison study. *PLoS ONE* **10**, e0145118 (2015).
37. Wilke, M., de Haan, B., Juenger, H. & Karnath, H. O. Manual, semi-automated, and automated delineation of chronic brain lesions: a comparison of methods. *Neuroimage* **56**, 2038–2046 (2011).
38. Barber, P. A., Demchuk, A. M., Zhang, J. & Buchan, A. M. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. ASPECTS study group. Alberta stroke programme early CT score. *Lancet* **355**, 1670–1674 (2000).
39. Kobkitsuksakul, C., Tritanon, O. & Suraratdecha, V. Interobserver agreement between senior radiology resident, neuroradiology fellow, and experienced neuroradiologist in the rating of Alberta stroke program early computed tomography score (ASPECTS). *Diagn. Interv. Radiol.* **24**, 104–107 (2018).
40. Nagel, S. et al. e-ASPECTS software is non-inferior to neuroradiologists in applying the ASPECT score to computed tomography scans of acute ischemic stroke patients. *Int. J. Stroke* **12**, 615–622 (2017).
41. Austein, F. et al. Automated versus manual imaging assessment of early ischemic changes in acute stroke: comparison of two software packages and expert consensus. *Eur. Radiol.* **29**, 6285–6292 (2019).
42. Castle, J. et al. Agreement regarding diagnosis of transient ischemic attack fairly low among stroke-trained neurologists. *Stroke* **41**, 1367–1370 (2010).
43. Bivard, A. et al. Ischemic core thresholds change with time to reperfusion: a case control study. *Ann. Neurol.* **82**, 995–1003 (2017).
44. Sillanpaa, N. et al. The clot burden score, the Boston acute stroke imaging scale, the cerebral blood volume ASPECTS, and two novel imaging parameters in the prediction of clinical outcome of ischemic stroke patients receiving intravenous thrombolytic therapy. *Neuroradiology* **54**, 663–672 (2012).
45. Bivard, A. et al. Validating a predictive model of acute advanced imaging biomarkers in ischemic stroke. *Stroke* **48**, 645–650 (2017).
46. Kawano, H. et al. Perfusion CT in stroke thrombolysis patients: save a ml of penumbra, save a week of disability-free life. *Brain* **140**, 684–691 (2017).
47. Liebeskind, D. Artificial intelligence in stroke care: deep learning or superficial insight? *EBioMedicine* **35**, 14–15 (2018).
48. Campbell, B. C. et al. Advanced imaging improves prediction of hemorrhage after stroke thrombolysis. *Ann. Neurol.* **73**, 510–519 (2013).
49. van Os, H. J. A. et al. Predicting outcome of endovascular treatment for acute ischemic stroke: potential value of machine learning algorithms. *Front. Neurol.* **9**, 784 (2018).
50. Bivard, A. et al. Global white matter hypoperfusion on CT predicts larger infarcts and hemorrhagic transformation after acute ischemia. *CNS Neurosci. Ther.* **22**, 238–243 (2016).
51. Campbell, B. C. et al. Regional very low cerebral blood volume predicts hemorrhagic transformation better than diffusion-weighted imaging volume and thresholded apparent diffusion coefficient in acute ischemic stroke. *Stroke* **41**, 82–88 (2010).
52. Yassi, N. et al. Prediction of poststroke hemorrhagic transformation using computed tomography perfusion. *Stroke* **44**, 3039–3043 (2013).
53. Bentley, P. et al. Prediction of stroke thrombolysis outcome using CT brain machine learning. *Neuroimage Clin.* **4**, 635–640 (2014).
54. FDA. FDA permits marketing of clinical decision support software for alerting providers of a potential stroke in patients. *FDA.gov* https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm596575.htm (2018).
55. Wang, X. et al. ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *IEEE CVPR* https://doi.org/10.1109/CVPR.2017.369 (2017).
56. Titano, J. J. et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. *Nat. Med.* **24**, 1337–1341 (2018).
57. Lobo, J. M., Jiménez-Valverde, A. & Real, R. AUC: a misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* **17**, 145–151 (2008).
An important analytical piece that shows how to interpret statistical results.
58. Saito, T. & Rehmsmeier, M. The precision–recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* **10**, e0118432 (2015).

Author contributions
All authors contributed to all aspects of the manuscript.

Competing interests
M.P. and A.B. have active research collaborations with Siemens Healthineers, Canon medical Systems and Apollo Medical Imaging. L.C. declares no competing interests.

Peer review information
Nature Reviews Neurology thanks the anonymous reviewer(s) for their contribution to the peer review of this work.