Journal of Clinical Neuroscience 99 (2022) 217–223

Contents lists available at ScienceDirect

Journal of Clinical Neuroscience

journal homepage: www.journals.elsevier.com/journal-of-clinical-neuroscience

Charting the potential of brain computed tomography deep learning systems

Quinlan D. Buchlak [a,b,c,*], Michael R. Milne [a], Jarrel Seah [a,d], Andrew Johnson [a], Gihan Samarasinghe [a], Ben Hachey [a], Nazanin Esmaili [b,e], Aengus Tran [a], Jean-Christophe Leveque [f], Farrokh Farrokhi [f], Tony Goldschlager [c], Simon Edelstein [a,g,h], Peter Brotchie [a,i]

[a] *Annalise.ai, Sydney, NSW, Australia*
[b] *School of Medicine, University of Notre Dame Australia, Sydney, NSW, Australia*
[c] *Department of Neurosurgery, Monash Health, Melbourne, VIC, Australia*
[d] *Department of Radiology, Alfred Health, Melbourne, VIC, Australia*
[e] *Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW, Australia*
[f] *Neuroscience Institute, Virginia Mason Medical Center, Seattle, WA, United States*
[g] *Department of Radiology, Monash Health, Melbourne, VIC, Australia*
[h] *I-MED Radiology Network, Brisbane, QLD, Australia*
[i] *Department of Radiology, St Vincent's Health Australia, Melbourne, VIC, Australia*

A R T I C L E I N F O

Keywords:
Brain computed tomography
Machine learning
Deep learning
Patient safety
Clinical decision making

A B S T R A C T

Brain computed tomography (CTB) scans are widely used to evaluate intracranial pathology. The implementation and adoption of CTB has led to clinical improvements. However, interpretation errors occur and may have substantial morbidity and mortality implications for patients. Deep learning has shown promise for facilitating improved diagnostic accuracy and triage. This research charts the potential of deep learning applied to the analysis of CTB scans. It draws on the experience of practicing clinicians and technologists involved in development and implementation of deep learning-based clinical decision support systems. We consider the past, present and future of the CTB, along with limitations of existing systems as well as untapped beneficial use cases. Implementing deep learning CTB interpretation systems and effectively navigating development and implementation risks can deliver many benefits to clinicians and patients, ultimately improving efficiency and safety in healthcare.

1. Introduction

1.1. The past and present of brain computed tomography

Prior to the invention of computed tomography (CT) it was not possible to directly image the brain. Invasive techniques such as cerebral angiography and pneumoencephalography were used to infer the presence of intracranial lesions by the shift in the position of intracranial vessels or displacement of the ventricles. The first CT brain (CTB) scan was performed in 1971 and represented a major advancement in the field of medical imaging. Godfrey Hounsfield and Allan Cormack were awarded the Nobel Prize for Physiology and Medicine in 1979 for the invention of the CT scanner.

The spatial and contrast resolution of CT allowed for the imaging of a wide range of intracranial pathologies including tumours, intracranial haemorrhage (ICH) and hydrocephalus. This technology led to significant reductions in the number of pneumoencephalograms and cerebral angiograms performed [1,2] and the associated complications and morbidities of these invasive imaging techniques. Improvements were particularly notable in the emergency department (ED) setting, with 3D visualisation of ICH offering considerable benefit over conventional methods at the time [3]. In the 50 years since its adoption, advancements in CT technology have been substantial. Initially, 30 minutes were required for a single non-contrast CT brain (NCCTB) scan, during which the patient's head was immobilised to prevent artefacts [2] and available postprocessing techniques to create non-axial projections were

* Corresponding author.
E-mail address: quinlan.buchlak1@my.nd.edu.au (Q.D. Buchlak).

https://doi.org/10.1016/j.jocn.2022.03.014
Received 20 October 2021; Accepted 8 March 2022
Available online 12 March 2022

limited and time-consuming. A whole brain can now be scanned in a few seconds, due to the introduction of multiple rows of detectors and helical scanning techniques, which allow for improved spatial resolution and thinner scan slices and enable isotropic and multiplanar visualisation. The wide availability and short acquisition time of CTB scans make them a common first-line imaging modality for patients with suspected intracranial pathology, with over 15 million CTB studies conducted in 2016 in the United States alone [4]. They are particularly useful in emergency settings (e.g., trauma, stroke, etc.) [5] where time is brain, and when rapid diagnosis is important for the provision of appropriate care to patients who might otherwise suffer significant morbidity and mortality [6]. Image reconstruction has also improved through the use of iterative reconstruction algorithms, which allow for lower dose scans with less beam hardening artefact [1]. More recently, deep learning applications have led to reduced artifact and improved image quality in low dose CT reconstruction [3].

2. Recent advancements in machine learning

Artificial intelligence (AI) encompasses a wide range of functions performed by computer systems, which attempt to reproduce human-level cognitive capabilities using learning algorithms applied to data. Currently the most successful algorithms for visual perception are deep neural networks, which are composed of many layers of highly interconnected processing elements and share basic structural similarities with biological neural networks in the human brain [7]. These networks are non-linear differentiable functions that contain many trainable free parameters and are capable of modelling complex nonlinear relationships by representing progressively higher-level abstractions of data [8]. Current neural nets only achieve human-level cognitive performance in narrow application domains, but nonetheless have demonstrated substantial utility. Several developments have accelerated the impact of these algorithms including computational power improvements, increased availability of digital imaging datasets to train and test models [9], advancements in algorithm design and the democratisation of access to high performing algorithms [10].

Convolutional neural networks (CNNs) learn through the use of many filters that are capable of finding features in images and they excel at visual perception applications because the filters scan the entire image allowing them to learn repeated patterns in different parts of an image, enabling them to detect patterns even when translated or shifted. CNNs have demonstrated high performance in medical image analysis [11], alongside recent developments in transformer architectures [12]. The number of applied machine learning studies in medicine is growing rapidly [13,14]. In this article, we review machine learning applications to CTB and consider potentially beneficial use cases from the perspective of clinicians and technologists who develop and deploy machine learning systems to clinical practice.

3. The current state of machine learning applied to CTB data

Despite the widespread use of CTB scans [15] the detection of pathology is still associated with a considerable level of diagnostic error and misinterpretation. Patterns of interpretive and perceptual error have been reported, particularly in the setting of extra-axial masses, thrombosis, and infarct detection [16,17]. In the ED, emergency physician interpretation of CTB scans is a substantial source of diagnostic error, with suboptimal accuracy and high misinterpretation rates [18,19]. Modern machine learning systems have the potential to improve detection of pathology, diagnostic accuracy and the clinical workflow.

Numerous machine learning systems have been developed to improve the accuracy and timeliness of CTB diagnostics and reporting (Table 1). ICH detection, in particular, has been the subject of attention, with machine learning yielding diagnostic accuracies similar to or exceeding those of expert radiologists [20-23]. Jnawali et al. developed a deep learning system to classify ICH on NCCTB scans using a large dataset consisting of 40,367 cases. Their ensemble achieved an area under the receiver operating characteristic curve (AUC) of 0.87 [24]. Irene et al. applied a combination of deep learning classification and regression methods to automate the detection, segmentation and volume approximation of ICH, with a system sensitivity and specificity of 0.97 and 0.96, respectively [25]. Kuo et al. trained a deep learning algorithm to detect and segment ICH using a dataset consisting of 4,396 NCCTB scans. Their model performed exceedingly well, achieving an AUC of 0.99 and demonstrating a higher performance on the test set than two of the four radiologists involved in the study [21]. Ojeda et al. evaluated the performance of the commercially available Aidoc system to triage ICH patients using a test set of 7,112 NCCTB scans. The performance of their model was compared to ground truth labels by expert neuroradiologists and it performed well, demonstrating a specificity of 0.99, a sensitivity of 0.95, and an accuracy of 0.98 [26]. Multiple similar systems have been used to facilitate haematoma segmentation on CT [27,28].

Machine learning has been applied to facilitate ischaemic stroke diagnosis and treatment. Kniep et al. developed a machine learning model to detect early ischaemic stroke changes using NCCTB scans [29]. Qui et al. developed a model to detect and segment stroke on NCCTB, automating the process of lesion volume quantification [30]. Beecy et al. developed a deep learning model to detect acute ischemic stroke on NCCTB. Algorithm AUC for voxel accuracy was 0.97. Accuracy, sensitivity, and specificity were 0.92, 0.93, and 0.92, respectively. AUC for the diagnosis of infarction at an image level was 0.91. Corresponding diagnostic accuracy, sensitivity, and specificity were 0.88, 0.65, and 0.91, respectively [31]. Dourado et al. developed a deep learning model to detect and classify ischaemic and haemorrhagic stroke. Authors trained and evaluated 144 model ensembles using only 50 scans. Top performing models achieved perfect classification accuracy, F1-score, recall and precision [32]. Gonzalez et al. developed a model to classify clot histology and chemical composition, with the potential to guide more targeted stroke therapy [33]. Chen et al. developed a system to segment cerebrospinal fluid on CT and automate the quantification of cerebral oedema after stroke [34]. Several models have been developed to automate the Alberta stroke programme early CT score (ASPECTS) process. Such models have demonstrated non-inferior performance in detecting early ischaemic stroke changes when compared to radiologists [35] and model outputs appear to agree well with neuroradiologist consensus [36].

Machine learning has been applied to CT image data to detect such varied diagnoses as dementia, intracranial aneurysms, and metastases. Gao et al. applied deep learning to CTB data to facilitate early diagnosis of Alzheimer's disease and to detect intracranial lesions, with a mean accuracy of 0.88 and individual accuracies for Alzheimer's disease, lesion and normal brain of 0.85, 0.80 and 0.95, respectively [40]. Dai et al. applied deep learning to a dataset consisting of 311 CT angiography (CTA) scans to facilitate detection of cerebrovascular aneurysms. Their model demonstrated an overall sensitivity of 0.92 while the sensitivity for detecting aneurysms larger than 3 mm was 0.97 [39]. Shi et al. also applied deep learning to detect intracranial aneurysms on CTA using a trained model validated on four internal and three external cohorts [49]. Takao et al. applied deep learning to detect brain metastases on both contrast enhanced and NCCTB scans, with their highest performing model achieving a sensitivity of 0.89. The model that was trained on both contrast-enhanced and non-contrast CTs significantly outperformed the model trained on just non-contrast studies [51] demonstrating the value of varied data sets for training.

Deep learning systems can facilitate patient triage and improve clinical efficiency. O'Neill et al. developed a model to detect ICH on NCCTB and showed that using this system as a triage tool reduced image interpretation turnaround times [52]. A similar system reduced time to diagnosis of outpatient ICH by 96% [53]. Arbabshirani et al. developed a model capable of detecting ICH on CT and prioritising those studies to

Table 1
Characteristics of studies applying deep learning to the CT brain scan. ASPECT, Alberta Stroke Program Early Computed Tomography. CNN, convolutional neural network. CTA, computed tomography angiography. ICH, intracranial haemorrhage. LSTM, Long-Short Term Memory. NCCTB, non-contrast computed tomography of the brain. NN, neural network. PPV, positive predictive value. RMSE, root mean square error.

Reference	N patients	Data type	Algorithm / architecture	N clinical findings/ outputs	Output	Performance
Beecy et al. (2018) [31]	114	NCCTB	3D multiscale, fully convolutional NN	1	Acute ischemic stroke	Voxel accuracy AUC 0.97. Diagnostic accuracy 0.92, sensitivity 0.93, and specificity 0.92. Diagnosis of infarction AUC at an image level 0.91. Corresponding diagnostic accuracy 0.88, sensitivity 0.65, and specificity 0.91.
Bermudez et al. (2019) [37]	1,313	NCCTB	3D CNN	1	Patient age	Models predicted patient age with a median absolute error of 9.99 years.
Chang et al. (2018) [22]	11,021	NCCTB	Hybrid 3D/2D mask region of interest based CNN	1	ICH	Test set: accuracy 0.97, AUC 0.98, sensitivity 0.95, specificity 0.97, PPV 0.83, and NPV 0.99. Dice scores were 0.93, 0.86, and 0.77.
Chilamkurthy et al. (2018) [38]	313,318	NCCTB	Ensemble of ResNet18 3D CNNs	9	ICH, intraparenchymal haemorrhage, intraventricular haemorrhage, subdural haemorrhage, extradural haemorrhage, subarachnoid haemorrhage, calvarial fractures, midline shift, mass effect	AUCs: intraparenchymal haemorrhage 0.90, 0.95; intraventricular haemorrhage 0.96, 0.93; subdural haemorrhage 0.92, 0.95; extradural haemorrhage 0.93, 0.97; subarachnoid haemorrhage 0.90, 0.96; calvarial fractures 0.92, 0.96; midline shift 0.93, 0.97; mass effect 0.86, 0.92.
Dai et al. (2020) [39]	311	CTA	ResNet-50 region-based CNN	1	Cerebrovascular aneurysms	Sensitivity 0.92.
Dourado et al. (2019) [32]	50	NCCTB	3D CNN	2	ICH, ischaemic stroke	Accuracy, F1-score, recall and precision 1.00.
Farzaneh et al. (2020) [27]	110	NCCTB	U-Net and random forest classifier	1	Subdural haematoma	Recall 0.79, precision 0.76, and Dice score 0.75.
Gao et al. (2019) [40]	285	NCCTB	2D (MatConvNet) and 3D CNNs	2	Alzheimer's disease, lesion (e.g., tumour)	Mean accuracy 0.88. Accuracies for Alzheimer's disease, lesion and normal: 0.85, 0.80 and 0.95.
Grewal et al. (2018) [23]	329	NCCTB	DenseNet and bidirectional LSTM	1	ICH	Accuracy 0.82.
Harms et al. (2019) [41]	24	NCCTB	Cycle-GAN and CNN	1	Image correction	Superior image quality compared to the scatter correction method, reducing noise and artifact severity.
Irene et al. (2020) [25]	27	NCCTB	Dynamic Graph CNN	1	ICH - segmentation and approximation of blood volume	Sensitivity 0.98, specificity 0.96.
Jain et al. (2019) [42]	213	NCCTB	2D and 3D U-Net	3	Acute intracranial lesions, midline shift, cistern volume	Median volume differences were 0.07 mL for acute intracranial lesions and −0.01 mL for cistern segmentation. Correlations 0.91 for volume of intracranial lesions, 0.94 for volume of cisterns, and 0.93 with expert assessments. Median error for midline shift computation was −0.22 mm.
Jnawali et al. (2018) [24]	40,367	NCCTB	3D CNN	1	ICH	AUC 0.87.
Ker et al. (2019) [43]	399	NCCTB	3D CNN	1	ICH	F1 scores ranged from 0.92 to 0.95.
Klimont et al. (2020) [44]	131	NCCTB	U-Net and ResNet18	1	Angiography generation	Dice coefficients 0.64 and 0.67.
Kuo et al. (2019) [21]	4,396	NCCTB	Patch-based fully CNN	1	ICH	AUC 0.99.
Murata et al. (2021) [45]	236	SPECT	Autoencoder and U-Net	1	Image correction	Mean errors < 1%.
Nag et al. (2019) [28]	48	CTB	Autoencoder and Chan-Vese model	1	ICH	Sensitivity 0.71, PPV 0.73, Dice score 0.70.
Nagel et al. (2017) [35]	132	NCCTB	e-ASPECTS software	1	ASPECT score	Sensitivity 0.44, specificity 0.93, accuracy 0.87. Noninferior to neuroradiologists.
Ojeda et al. (2019) [26]	7,112	NCCTB	Aidoc CNN	1	ICH	Specificity 0.99, sensitivity 0.95, accuracy 0.98.
Park et al. (2018) [46]	65	NCCTB	U-Net	1	Image correction	Image noise was lower than the ground truth.
Poirot et al. (2019) [47]	182	Dual-energy CT	2D CNN ResNet	1	Image generation	Significantly lower RMSE. Spearman's rank correlation coefficients for relative pixel intensities and correspondence for true and deep learning images were 0.71 and 0.62.
Remedios et al. (2019) [48]	45	NCCTB	3D CNN	1	ICH	Lesion mask Dice similarity 0.64; correlation of segmented hematoma volumes vs manual 0.87.
Shi et al. (2020) [49]	3,029	CTA	3D CNN	1	Cerebrovascular aneurysms	Specificity 0.99, lesion-level sensitivity 0.96 with a Dice ratio of 0.75.
	1,539			10	Image type	

(*continued on next page*)

Table 1 (*continued*)

Reference	N patients	Data type	Algorithm / architecture	N clinical findings/ outputs	Output	Performance
Sugimori (2018) [50]		Contrast & NCCTB	AlexNet and GoogLeNet			10-class accuracy 0.72, 5-class accuracy 0.86.
Sundaram (2019) [36]	58	NCCTB	Brainomix software	1	ASPECT score	Agreement with neuroradiologist consensus (k 0.84).
Takao et al. (2021) [51]	116	Contrast & NCCTB	Single-shot detector models	1	Brain metastases	Sensitivity 0.89, PPV 0.44.

reduce time to diagnosis. This predictive model was implemented prospectively in a real-world ED setting for three-months and demonstrated an AUC of 0.85, reprioritizing 94 of 347 NCCTB studies to the top of the routine worklist, 60 of which were confirmed to contain ICH by radiologists. Median interpretation time in prioritized studies containing ICH was 96% lower (19 min) than routine studies (512 min), demonstrating a clinical benefit from using machine learning for triage [53].

Even in situations in which diagnosis detection may be performed by human operators, machine learning can play a role through the automation of ancillary data, including intracranial pathology measurements, estimation of patient age, and improvement in image processing. Jain et al. developed a deep learning system to automate and improve the quantification of feature measurements in acute brain injury patients, including the measurement of midline shift, cistern volume and acute intracranial lesion volume. The system correlated strongly with ground truth measurements by experts (0.91–0.94) and error was low [42]. Bermudez et al. developed a model to conduct imaging-based age prediction using deep convolutional features and traditional structural features. Models were able to predict patient age with a median absolute error of 9.99 years [37]. Park et al. applied deep learning to improve NCCTB image resolution, converting low resolution (thick slice) images to high resolution (thin slice) images. The model was capable of accurately producing high resolution images from low resolution inputs, reducing blur and image noise. Murata et al. applied deep learning to facilitate attenuation correction in single-photon emission computed tomography, potentially removing the need for intermediate CT imaging [45]. Klimont et al. applied deep learning to segment cerebral arteries on NCCTB scans and generate angiography images without requiring the administration of contrast with test dataset and cross-validation Dice scores of 0.638 and 0.673 [44]. Poirot et al. applied deep learning to improve dual-energy CT (DECT) scan processing. Their model was trained to generate NCCTB images from DECT images and then these generated NCCTB scan images were compared to true scans and those generated using a physics-based algorithm. Images generated using deep learning were significantly more similar to the true images than those generated using the physics-based algorithm [47]. Harms et al. applied generative adversarial network methods to facilitate cone-beam CT image correction. Their method resulted in superior image quality compared to scatter correction and other machine learning methods, with reductions in image noise and artifact severity [41]. Sugimori tested different image slice sample sizes and deep learning architectures on the problem of classifying the body region (brain, neck, chest, abdomen, pelvis) of non-contrast and contrast-enhanced CT images and demonstrated that model accuracy varied substantially depending on image dataset size, algorithm applied and the number of output classes [50].

Most machine learning models developed to facilitate the interpretation of CTB scans have demonstrated promising results, but they are generally narrow in focus, encompassing only a small number of clinical findings. In the largest study in this domain to date, Chilamkurthy et al. [38] developed an algorithm to identify nine clinical findings on CTB, including haemorrhage subtypes, calvarial fractures, midline shift and mass effect. Algorithms were trained and tested using datasets consisting of 292,223 and 21,095 CTB scans, respectively. External validation was conducted using 491 scans. This system achieved a level of sensitivity that was non-inferior to the consensus of a panel of radiologists [54]. AUCs were high, illustrating that deep learning algorithms can accurately identify CTB abnormalities requiring urgent attention. This project was a milestone in the field and highlighted the potential of applying deep learning to facilitate pathology detection on CTB scans.

Most extant CTB machine learning systems remain limited, however. Models are rarely clinically comprehensive and most have not been externally validated [55]. While demonstrating high levels of classification performance, they may result in limited clinical utility, poor uptake, and low impact. A model capable of identifying a wide range of clinically critical findings simultaneously (e.g., multiple haemorrhage types, ischaemic stroke, fractures, mass effect, herniation, tumours, etc.) on CTB is expected to be much more useful to clinicians as a diagnostic and triage tool than a model that is only capable of identifying a few clinical findings. There is an opportunity to develop low bias, comprehensive CTB deep learning systems that are more clinically useful than those currently in existence.

Deep learning is a data hungry technology. A deep learning project may quickly become hobbled by a lack of data and it is often prohibitive or impractical to transfer protected health information (PHI) between institutions. To ameliorate this issue, Remedios et al. applied a distributed approach to training deep learning models to detect ICH without the need to transfer PHI. They trained three models on three different institutional datasets and combined these models into a classification ensemble. The multisite ensemble achieved higher segmentation performance than any of the three individual models alone [48]. Although there have been advancements in the availability of imaging datasets to facilitate the development of machine learning models, the range of publicly available datasets remains limited and they are mostly inadequate for the development of high performing clinically comprehensive models, as these public datasets often lack adequate pathology diversity, case volumes and diagnostic labelling. These deficiencies can lead to the development of narrow, brittle, and often overfitted, models that demonstrate deceptively strong performance on training data but perform poorly when externally validated [56]. These types of algorithms may not perform well on images acquired in different healthcare centres, on different scanners, or in different populations. Rare intracranial findings also pose a challenge due to the paucity of available training data in most datasets, and varying data quality can hamper training. Accurate models and effective validation activities generally require large samples of well-labelled cases to achieve high levels of performance and credible validation assessments. The development of datasets that meet adequate quality standards often requires substantial investment and deep medical expertise to carry out ontology tree development and case labelling activities.

4. Potential future benefits of CTB machine learning systems

A survey of current CTB machine learning systems suggests that use cases have been diverse with substantial opportunity for improvement. The creation of models that can augment the diagnostic accuracy of radiologists is only the first step in the successful integration of functional machine learning tools into clinical practice. CTB machine learning systems show substantial untapped promise for the improvement of service delivery in the developing world, treatment planning, patient safety, organisational efficiency, clinician capability development and quality control. Machine learning for advanced image enhancement and noise reduction may reduce contrast dose and radiation exposure for patients. Systems leveraging state-of-the-art natural language processing models may go beyond just providing a list of likely clinical findings to automatically and quickly generating accurate reports that mention salient negatives and offer useful relevant clinical conclusions. Quantitative imaging that automatically provides estimates of lesion volumes may facilitate triage, surgical planning and research. Automation and rapid delivery of objectively measured volumes and contours of certain pathologies (e.g., subfalcine shift, intracranial haemorrhagic volumes and diffuse axonal injury burden) is likely to facilitate short- and long-term outcome prediction as well as prospective tracking of traumatic pathologies over time. It is not unreasonable to envision that CTB machine learning systems will become integral to all components of medical care including diagnosis, treatment and follow up.

Machine learning systems have the potential to improve clinical safety in hospitals overnight and during times of fatigue or stress thereby decreasing error or missed findings. For junior radiologists or on-call trainees with less supervision in busy teaching hospitals, these systems may feel like a reassuring senior colleague looking over their shoulder, encouraging a systematic approach and reinforcing diagnostic confidence. By highlighting "check areas", they may assist in detecting pathologies that might otherwise have been missed or mitigate the occurrence of "sentinel cases" in high–risk environments. Shift changeovers and overnight shifts are time periods associated with higher risk for medical errors [57] and machine learning systems could assist in smoothing these periods of increased risk. These improvements in safety and efficiency may be especially useful in publicly funded systems in which resources are often limited.

Current machine learning systems almost universally focus on diagnosing and delineating regions of interest for CTB pathologies at the time of scan acquisition. Acutely, they may be applied to aid radiographers in alerting radiologists sooner to the presence of urgent findings or those requiring contrast administration. However, systems capable of projecting forward in time and visualising pathology changes over time without the need for re-scanning may also be useful to clinicians. These systems may better facilitate triage and lead to more effective identification of patients who may benefit from intervention. They may facilitate long-term clinical planning, predicting and visualising the speed of cerebral atrophy in dementia, for example, or visualising the likely therapeutic effects of clot retrieval or neurosurgery.

Clinical teams may use machine learning systems to meet organisational metrics more effectively. EDs are required to meet targets for time-to-treatment or disposition. A frequent substantial bottleneck in the ED treatment process is the time for a radiology report to be returned to the treating clinician. At present, a CTB scan in some public teaching hospitals may take hours to be reported. In our experience, some trauma patients may experience hours of discomfort lying flat in the ED on cervical spine precautions while they wait for their CT scans to be reported. An automated report delivered by a well-validated machine learning system has the potential to markedly improve ED efficiency, reduce wait times and discomfort for patients and speed up the treatment process.

Accurate machine learning systems may improve the efficiency, efficacy and scalability of teaching and performance evaluation processes. Currently training often requires a consultant physician and the registrar to review many hundreds of cases together, which is an inefficient use of scarce resources. A machine learning system may offer a limitless expert coach for trainees, which may free up consultants to focus on clinical priorities. Such systems could even be used for continuing medical education, with senior clinicians provided with a set of test cases and the ability to compare their performance with that of the machine learning system. This method may comprise an efficient form of feedback to facilitate clinical development, capability improvement and quality assurance. If needed, machine learning systems could offer remedial assistance in a non-confronting way for radiologists who may have fallen into suboptimal reporting patterns, and play the role of a "check buddy" for radiologists who are identified as requiring assistance. Clinicians managing anxiety disorders [58], in particular, may benefit from these diagnostic support systems. These systems have the potential to decrease the likelihood of long-term burnout amongst radiologists, who, like most medical specialists, represent a major societal investment and are difficult to replace. Robust quality control and service auditing processes are increasingly becoming contractual requisites for hospital radiology service providers. As the evidence mounts that machine learning decision support software devices can demonstrate robust performance augmentation effects, these systems may become contractual requirements [59].

Machine learning systems may be used to assist clinicians in parts of the developing world that experience a shortage of trained radiologists. CT scanners may exist in some hospitals, but there may not be enough people trained to report the scans. Accurate machine learning systems would help clinicians in these resource-constrained contexts with accurate diagnosis and timely management and thereby facilitate improvements to patient safety and service quality.

In our experience, some radiologists perceive machine learning systems initially as a challenge to their profession, authority and careers. We have seen this apprehension drive competitive and defensive attitudes in some clinicians. We have observed encouraging signs, however, within a large practice currently using machine learning systems to facilitate radiology interpretation and diagnosis. These signs include subjectively improved concentration levels and a broad acceptance (even amongst early critics) that these systems do offer useful assistance and that their effects are likely to benefit patients [60]. In addition, we have observed renewed enthusiasm amongst radiologists for their clinical work. Some clinicians have commented that the machine learning system removes some of the drudgery of routine reporting and allows focus on nuanced interpretation of more challenging clinical findings and unusual cases.

The risks associated with machine learning systems have been well described [61,62] and systems should be developed, implemented and evaluated rigorously and correctly to mitigate these risks [62-64]. Clinical safety is a core concern and it is imperative therefore to involve clinicians and maintain a clear understanding of the quantity and impacts of false positives and false negatives. Mature development of machine learning systems will require sufficient data quantity and quality, validated expert annotations, minimised overfitting, prevention of data leakage, maximised interpretability, effective validation and prospective evaluation. Mature implementation necessitates an evaluation of effective clinical workflow integration and change management. With appropriate risk management and mature development and implementation, applied CTB machine learning systems have the potential to drive substantial clinical benefits and meaningful improvements in healthcare [60].

5. Conclusion

We are on the cusp of the implementation epoch. Developments in applied CTB machine learning technologies afford confidence in clinical finding detection and have the potential to improve healthcare efficiency and patient safety. More sophisticated and comprehensive CTB machine learning systems promise to facilitate more accurate diagnosis, higher quality clinical decision making, more timely treatment, error rate reduction, clinical audits, and education and may fundamentally drive better outcomes for patients. In the context of growing healthcare costs, strained healthcare systems across the globe, ageing populations, ever-increasing patient volumes, climate change and the COVID-19 pandemic, high-quality machine learning systems may offer some respite in the form of system-wide quality and efficiency improvements. Future well-validated, low-bias, high-quality, clinically comprehensive

CTB machine learning systems have the potential to improve timely delivery of quality care to patients at scale.

This study was funded by annalise.ai. QB, MM, JS, AJ, GS, BH, AT, SE, and PB were employed by or seconded to annalise.ai. Other authors have no competing interests.

References

[1] Rubin GD. Computed tomography: revolutionizing the practice of medicine for 40 years. Radiology 2014;273(2S):S45–74.
[2] Wesolowski JR, Lev MH. CT: history, technology, and clinical aspects. Semin Ultrasound CT MR 2005;26(6):376–9.
[3] Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. Med Phys 2017;44(10):e360–75.
[4] Mettler FA, Mahesh M, Bhargavan-Chatfield M, Chambers CE, Elee JG, Frush DP, et al. Patient exposure from radiologic and nuclear medicine procedures in the United States: procedure volume and effective dose for the period 2006–2016. Radiology 2020;295(2):418–27.
[5] Coles JP. Imaging after brain injury. Br J Anaesth 2007;99(1):49–60.
[6] Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, et al. 2018 Guidelines for the Early Management of Patients with Acute Ischemic Stroke: A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association. Stroke 2018;49(3).
[7] Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science 2015;349(6245):255–60.
[8] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44.
[9] Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018;18(8):500–10.
[10] Flanders AE, Prevedello LM, Shih G, Halabi SS, Kalpathy-Cramer J, Ball R, et al. Construction of a Machine Learning Dataset through Collaboration: The RSNA 2019 Brain CT Hemorrhage Challenge. Radiology: Artificial Intelligence 2020;2(4):e209002.
[11] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE; 2015. https://doi.org/10.1109/cvpr.2015.7298594.
[12] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. ArXiv [CsCV] 2020.
[13] Buchlak QD, Esmaili N, Leveque J-C, Farrokhi F, Bennett C, Piccardi M, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. Neurosurg Rev 2020;43(5):1235–53.
[14] Buchlak QD, Esmaili N, Leveque J-C, Bennett C, Farrokhi F, Piccardi M. Machine learning applications to neuroimaging for glioma detection and classification: An artificial intelligence augmented systematic review. J Clin Neurosci 2021;89:177–98.
[15] Klang E, Barash Y, Soffer S, Bechler S, Resheff YS, Granot T, et al. Promoting head CT exams in the emergency department triage using a machine learning model. Neuroradiology 2020;62(2):153–60.
[16] Donald JJ, Barnard SA. Common patterns in 558 diagnostic radiology errors. J Med Imaging Radiat Oncol 2012;56:173–8.
[17] Jordan MJ, Lightfoote JB, Jordan JE. Quality outcomes of reinterpretation of brain CT imaging studies by subspecialty experts in neuroradiology. J Natl Med Assoc 2006;98:1326–8.
[18] Boyle A, Staniciu D, Lewis S, Hugman A, Bauza-Rodriguez B, Kirby D, et al. Can middle grade and consultant emergency physicians accurately interpret computed tomography scans performed for head trauma? Cross-sectional study. Emerg Med J 2009;26(8):583–5.
[19] Arendts G, Manovel A, Chai A. Cranial CT interpretation by senior emergency department staff. Australas Radiol 2003;47:368–74.
[20] Lee H, Yune S, Mansouri M, Kim M, Tajmir SH, Guerrier CE, et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. Nat Biomed Eng 2019;3(3):173–82.
[21] Kuo W, Häne C, Mukherjee P, Malik J, Yuh EL. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. Proc Natl Acad Sci 2019;116(45):22737–45.
[22] Chang PD, Kuoy E, Grinband J, Weinberg BD, Thompson M, Homo R, et al. Hybrid 3D/2D convolutional neural network for hemorrhage evaluation on Head CT. AJNR Am J Neuroradiol 2018;39(9):1609–16.
[23] Grewal M, Srivastava MM, Kumar P, Varadarajan S. RADnet: Radiologist level accuracy using deep learning for hemorrhage detection in CT scans. vol. 2018-April, IEEE Computer Society; 2018, p. 281–4.
[24] Jnawali K, Arbabshirani MR, Rao N, Patel AA. Deep 3D convolution neural network for CT brain hemorrhage classification. In: Mori K. PN, editor. vol. 10575, SPIE; 2018. https://doi.org/10.1117/12.2293725.
[25] Irene K, Ma'sum MA, Yunus RE, Jatmiko W. Segmentation and approximation of blood volume in intracranial hemorrhage patients based on computed tomography scan images using deep learning method. 2020 International Workshop on Big Data and Information Security (IWBIS), IEEE; 2020. https://doi.org/10.1109/iwbis50925.2020.9255593.
[26] Ojeda P, Zawaideh M, Mossa-Basha M, Haynor D. The utility of deep learning: Evaluation of a convolutional neural network for detection of intracranial bleeds on non-contrast head computed tomography studies. vol. 10949. SPIE; 2019.
[27] Farzaneh N, Williamson CA, Jiang C, Srinivasan A, Bapuraj JR, Gryak J, et al. Automated segmentation and severity analysis of subdural hematoma for patients with traumatic brain injuries. Diagnostics (Basel) 2020;10(10):773.
[28] Nag MK, Chatterjee S, Sadhu AK, Chatterjee J, Ghosh N. Computer-assisted delineation of hematoma from CT volume using autoencoder and Chan Vese model. Int J Comput Assist Radiol Surg 2019;14(2):259–69.
[29] Kniep HC, Sporns PB, Broocks G, Kemmling A, Nawabi J, Rusche T, et al. Posterior circulation stroke: machine learning-based detection of early ischemic changes in acute non-contrast CT scans. J Neurol 2020;267(9):2632–41.
[30] Qiu Wu, Kuang H, Teleg E, Ospel JM, Sohn SI, Almekhlafi M, et al. Machine learning for detecting early infarction in acute stroke with non-contrast-enhanced CT. Radiology 2020;294(3):638–44.
[31] Beecy AN, Chang Qi, Anchouche K, Baskaran L, Elmore K, Kolli K, et al. A novel deep learning approach for automated diagnosis of acute ischemic infarction on computed tomography. JACC Cardiovasc Imaging 2018;11(11):1723–5.
[32] Dourado Jr CMJM, da Silva SPP, da Nóbrega RVM, da S. Barros AC, Filho PPR, de Albuquerque VHC. Deep learning IoT system for online stroke detection in skull computed tomography images. Comput Netw 2019;152:25–39.
[33] Velasco Gonzalez A, Buerke B, Görlich D, Fobker M, Rusche T, Sauerland C, et al. Clot analog attenuation in non-contrast CT predicts histology: An experimental study using machine learning. Transl Stroke Res 2020;11(5):940–9.
[34] Chen Y, Dhar R, Heitsch L, Ford A, Fernandez-Cadenas I, Carrera C, 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 2016;12:673–80.
[35] Nagel S, Sinha D, Day D, Reith W, Chapot R, Papanagiotou P, 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 2017;12(6):615–22.
[36] Sundaram VK, Goldstein J, Wheelwright D, Aggarwal A, Pawha PS, Doshi A, et al. Automated ASPECTS in Acute Ischemic Stroke: A Comparative Analysis with CT Perfusion. AJNR Am J Neuroradiol 2019. https://doi.org/10.3174/ajnr.A6303.
[37] Bermudez C, Plassard AJ, Chaganti S, Huo Y, Aboud KS, Cutting LE, et al. Anatomical context improves deep learning on the brain age estimation task. Magn Reson Imaging 2019;62:70–7.
[38] Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. Lancet 2018;392(10162):2388–96.
[39] Dai X, Huang L, Qian Y, Xia S, Chong W, Liu J, et al. Deep learning for automated cerebral aneurysm detection on computed tomography images. Int J Comput Assist Radiol Surg 2020;15(4):715–23.
[40] Gao XW, Hui R, Tian Z. Classification of CT brain images based on deep learning networks. Comput Methods Programs Biomed 2017;138:49–56.
[41] Harms J, Lei Y, Wang T, Zhang R, Zhou J, Tang X, et al. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. Med Phys 2019;46(9):3998–4009.
[42] Jain S, Vyvere TV, Terzopoulos V, Sima DM, Roura E, Maas A, et al. Automatic quantification of computed tomography features in acute traumatic brain injury. J Neurotrauma 2019;36(11):1794–803.
[43] Ker J, Singh SP, Bai Y, Rao J, Lim T, Wang L. Image thresholding improves 3-dimensional convolutional neural network diagnosis of different acute brain hemorrhages on computed tomography scans. Sensors 2019;19(9):2167.
[44] Klimont M, Oronowicz-Jaśkowiak A, Flieger M, Rzeszutek J, Juszkat R, Jończyk-Potoczna K, et al. Deep learning for cerebral angiography segmentation from non-contrast computed tomography. PLoS ONE 2020;15(7):e0237092.
[45] Murata T, Yokota H, Yamato R, Horikoshi T, Tsuneda M, Kurosawa R, et al. Development of attenuation correction methods using deep learning in brain-perfusion single-photon emission computed tomography. Med Phys 2021;48(8):4177–90.
[46] Park J, Hwang D, Kim KY, Kang SK, Kim YK, Lee JS. Computed tomography super-resolution using deep convolutional neural network. Phys Med Biol 2018;63(14):145011.
[47] Poirot MG, Bergmans RHJ, Thomson BR, Jolink FC, Moum SJ, Gonzalez RG, et al. Physics-informed deep learning for dual-energy computed tomography image processing. Sci Rep 2019;9(1).
[48] Remedios S, Roy S, Blaber J, Bermudez C, Nath V, Patel MB, et al. Distributed deep learning for robust multi-site segmentation of CT imaging after traumatic brain injury. Proc SPIE 2019;10949. https://doi.org/10.1117/12.2511997.
[49] Shi Z, Miao C, Schoepf UJ, Savage RH, Dargis DM, Pan C, et al. A clinically applicable deep-learning model for detecting intracranial aneurysm in computed tomography angiography images. Nat Commun 2020;11(1). https://doi.org/10.1038/s41467-020-19527-w.
[50] Sugimori H. Classification of computed tomography images in different slice positions using deep learning. J Healthc Eng 2018;2018:1–9.
[51] Takao H, Amemiya S, Kato S, Yamashita H, Sakamoto N, Abe O. Deep-learning single-shot detector for automatic detection of brain metastases with the combined use of contrast-enhanced and non-enhanced computed tomography images. Eur J Radiol 2021;144:110015.
[52] O'Neill TJ, Xi Y, Stehel E, Browning T, Ng YS, Baker C, et al. Active reprioritization of the reading worklist using artificial intelligence has a beneficial effect on the turnaround time for interpretation of head CT with intracranial hemorrhage. Radiol Artif Intell 2021;3(2):e200024.

[53] Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ, Suever JD, Geise BD, Patel AA, et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. NPJ Digital Med 2018;1:1–7.
[54] Prevedello LM, Erdal BS, Ryu JL, Little KJ, Demirer M, Qian S, et al. Automated critical test findings identification and online notification system using artificial intelligence in imaging. Radiology 2017;285(3):923–31.
[55] Yeo M, Tahayori B, Kok HK, Maingard J, Kutaiba N, Russell J, et al. Review of deep learning algorithms for the automatic detection of intracranial hemorrhages on computed tomography head imaging. J Neurointerv Surg 2021;13(4):369–78.
[56] Zhu G, Jiang B, Tong L, Xie Y, Zaharchuk G, Wintermark M. Applications of deep learning to neuro-imaging techniques. Front Neurol 2019;10:869.
[57] Olson EJ, Drage LA, Auger RR. Sleep deprivation, physician performance, and patient safety. Chest 2009;136(5):1389–96.
[58] Pougnet R, Pougnet L. Anxiety disorders and mood disorders in hospital doctors: a literature review. Med Pr 2021;72:163–71.
[59] Seah JCY, Tang CHM, Buchlak QD, Holt XG, Wardman JB, Aimoldin A, et al. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. Lancet Digit Health 2021;3(8):e496–506.
[60] Jones CM, Danaher L, Milne MR, Tang C, Seah J, Oakden-Rayner L, et al. Assessment of the effect of a comprehensive chest radiograph deep learning model on radiologist reports and patient outcomes: a real-world observational study. BMJ Open 2021;11(12):e052902.
[61] Yu K-H, Kohane IS. Framing the challenges of artificial intelligence in medicine. BMJ Qual Saf 2019;28(3):238–41.
[62] Seneviratne MG, Shah NH, Chu L. Bridging the implementation gap of machine learning in healthcare. BMJ Innov 2020;6(2):45–7.
[63] Buchlak QD, Esmaili N, Leveque J-C, Bennett C, Piccardi M, Farrokhi F. Ethical thinking machines in surgery and the requirement for clinical leadership. Am J Surg 2020;220(5):1372–4.
[64] Seah J, Tang C, Buchlak QD, Milne MR, Holt X, Ahmad H, et al. Do comprehensive deep learning algorithms suffer from hidden stratification? A retrospective study on pneumothorax detection in chest radiography. BMJ Open 2021;11(12): e053024.