Deep Learning Applications for Acute Stroke Management

Isha R. Chavva, BS ,[1] Anna L. Crawford, MSc,[1] Mercy H. Mazurek, BS,[1] Matthew M. Yuen, BA,[1] Anjali M. Prabhat, BA,[1] Sam Payabvash, MD,[2] Gordon Sze, MD,[2] Guido J. Falcone, MD, ScD, MPH ,[1] Charles C. Matouk, MD,[3] Adam de Havenon, MD, MSCI,[1] Jennifer A. Kim, MD, PhD,[1] Richa Sharma, MD, MPH,[1] Steven J. Schiff, MD, PhD,[4] Matthew S. Rosen, PhD ,[5] Jayashree Kalpathy-Cramer, PhD,[5] Juan E. Iglesias Gonzalez, PhD,[5] W. Taylor Kimberly, MD, PhD,[6] and Kevin N. Sheth, MD [1]

Brain imaging is essential to the clinical care of patients with stroke, a leading cause of disability and death worldwide. Whereas advanced neuroimaging techniques offer opportunities for aiding acute stroke management, several factors, including time delays, inter-clinician variability, and lack of systemic conglomeration of clinical information, hinder their maximal utility. Recent advances in deep machine learning (DL) offer new strategies for harnessing computational medical image analysis to inform decision making in acute stroke. We examine the current state of the field for DL models in stroke triage. First, we provide a brief, clinical practice-focused primer on DL. Next, we examine real-world examples of DL applications in pixel-wise labeling, volumetric lesion segmentation, stroke detection, and prediction of tissue fate postintervention. We evaluate recent deployments of deep neural networks and their ability to automatically select relevant clinical features for acute decision making, reduce inter-rater variability, and boost reliability in rapid neuroimaging assessments, and integrate neuroimaging with electronic medical record (EMR) data in order to support clinicians in routine and triage stroke management. Ultimately, we aim to provide a framework for critically evaluating existing automated approaches, thus equipping clinicians with the ability to understand and potentially apply DL approaches in order to address challenges in clinical practice.

ANN NEUROL 2022;00:1–14

Neuroimaging techniques, such as computed tomography (CT) and magnetic resonance imaging (MRI), are used for early identification, diagnosis, treatment, and outcome prognostication of acute stroke. Neuroimaging can help physicians to differentiate stroke from stroke mimics,[1–3] assess the presence and severity of intracranial hemorrhage (ICH),[1,2] identify the location of vascular occlusions,[4,5] determine the degree of reversibility of ischemic injury,[6,7] and identify the best candidates for acute interventions like thrombectomy and thrombolysis.[8–10] Segmentation and parameterization of neuroimages can be used to quantify regions of irrevocably damaged and at-risk salvageable tissue, which can better elucidate possible treatment options depending on the time since the stroke onset.[11,12]

Despite the utility of neuroimaging in acute stroke management and the availability of several automated, neuroimaging-based platforms for stroke diagnosis and treatment, the full potential of robust, fully automated, and universal applications of neuroimaging in stroke triage

View this article online at wileyonlinelibrary.com. DOI: 10.1002/ana.26435

Received Mar 4, 2022, and in revised form May 27, 2022. Accepted for publication Jun 4, 2022.

Address correspondence to Dr Sheth, Department of Neurology, Yale School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA. E-mail: kevin.sheth@yale.edu

From the [1]Department of Neurology, Yale School of Medicine, New Haven, CT; [2]Department of Radiology, Yale School of Medicine, New Haven, CT; [3]Department of Neurosurgery, Yale School of Medicine, New Haven, CT; [4]Departments of Neurosurgery, Engineering Science and Mechanics and Physics, Penn State University, University Park, PA; [5]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA; and [6]Department of Neurology, Division of Neurocritical Care, Massachusetts General Hospital, Boston, MA

has not yet been attained. This is likely because neuroimaging analysis is a complex process and several factors hinder its utility in acute clinical decision making: (1) an inability to obtain rapid, accurate information about neuroimages without tedious manual segmentation[11,13]; (2) heterogeneity in pathology and other clinical factors which affect prognosis, such as patients' age, stroke severity, baseline status at stroke onset, and tissue response to ischemic infarction[11,14,15]; and (3) intra- and inter-clinician variability in the interpretation of scans.[14,16–18]

Contemporary medical practice is now generating more clinically relevant information than ever before, in the forms of continuously accumulating biomedical imaging data, physiological metric information, and electronic medical records (EMRs). However, the enormity and complexity of these data mean that humans cannot efficiently determine the predictive relevance of this multidimensional medical information to inform clinical decision making.

Recent advances in machine learning (ML) have enabled computational medical image analysis at a large scale and present opportunities to harness these massive medical datasets to inform medical practice across various domains.[19–21] ML is a branch of artificial intelligence (AI) wherein algorithms are trained to parse large pools of data in order to learn informative features and identify meaningful patterns without explicit instructions. ML requires structured input data, which defines relevant, task-specific features for differentiating criteria of interest. However, due to the complexity and variety of pathological manifestations in neuroimaging, manual selection of relevant features, as utilized in ML, may not be feasible for accurately characterizing the heterogeneous and evolving presentation of disease. Fortunately, the rising prevalence of large datasets and robust infrastructure has reinvigorated ML, resulting in the subfield of deep learning (DL), where feature engineering is pushed onto the model. Thus, DL models emerge as potential options for assessment of medical images without a priori information. DL algorithms use multiple computational layers ("deep") to progressively extract higher-level features from raw input (Fig 1). One subtype of deep neural networks (DNNs) suited for such image processing tasks are convolutional neural networks (CNNs), which compute spatial relations between different areas of pixels within an image. CNNs use convolutional layers, which consist of sets of filters, to pass information from each layer to subsequent layers (Fig 2). Another relevant class is that of recurrent neural networks (RNNs), which have connected nodes that form a directed graph along a temporal sequence. RNNs use their internal state, or "memory," to process inputs of variable length and take historical information into account.

DL models have been applied to facilitate surveillance and analysis of acute stroke neuroimages, offering solutions for lesion segmentation and quantification,[11,17,22–34] early stroke detection,[18,35–46] selection of candidates for therapeutic intervention,[9,47–49] and prediction of short- and long-term functional outcomes.[9,47–50] Existing automated applications for clinical settings, such as Viz.ai and RapidAI, have been developed for a variety of tasks, including identifying large vessel occlusions (LVOs), diagnosing ischemic and hemorrhagic stroke, and assessing salvageable brain tissue. However, these methods require further improvements to their accuracy and sensitivity, are often still heavily reliant on clinician review, may not integrate into other data streams available to physicians, and are generally limited by the lack of standardization and systematic comparison for validation. Regardless, the continued development of predictive models offers the possibility of routine support, or objective second opinions based on broad clinical and

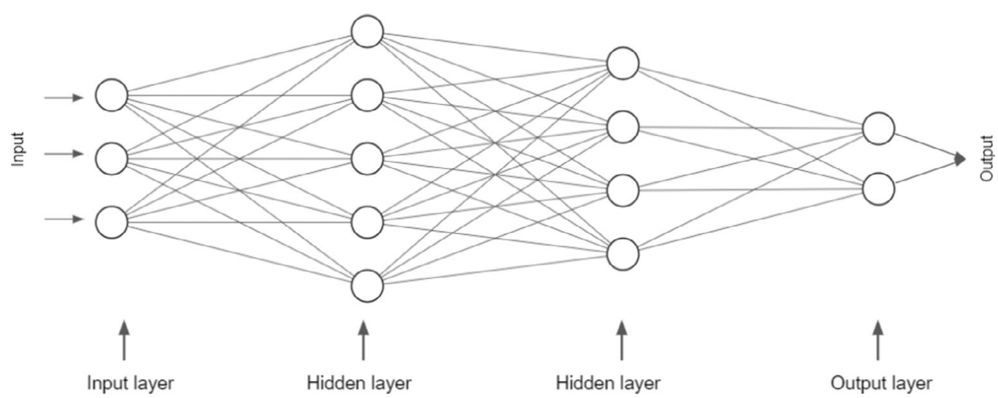

Figure 1: The basic structure of an artificial neural network (ANN) is composed of an input layer, output layer, and hidden layer(s). A series of nonlinear transformations are applied to layer inputs in order to construct internal representations of information which are then used to generate task-specific output decisions.

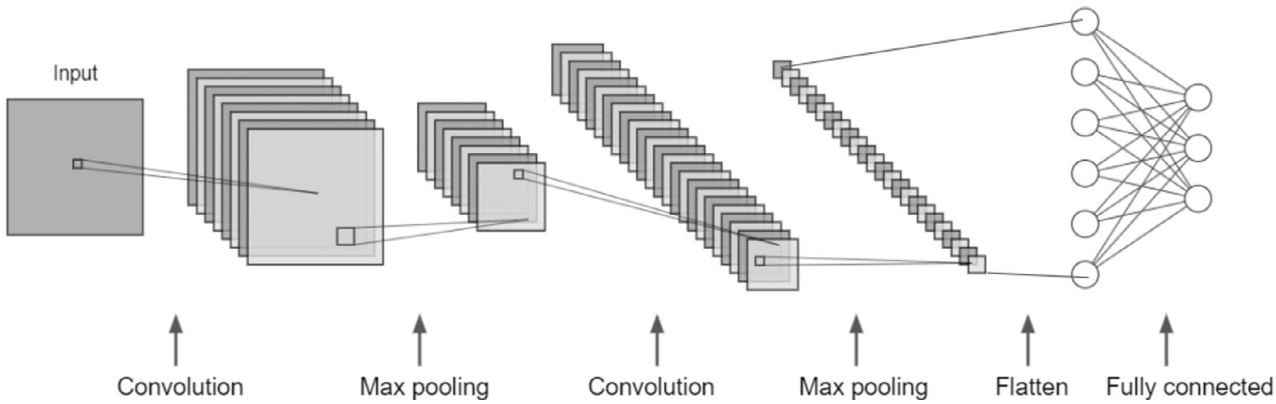

Figure 2: Convolutional neural networks (CNNs) compute using learned spatial relations between different areas of pixels within an image. First, the input image is re-dimensioned and passed to the first convolutional layer as input. Here, a filter, or kernel, is overlaid on pixels of the input image. The filter is rastered across the input image, multiplying its values with the original pixel values of the image. Results of the multiplication are summed into a single value for each receptive field, known as a feature, or activation, map. The first convolutional layer typically detects low-level features, such as curves or edges. Subsequent convolutional layers, however, are fed activation maps from preceding layers, which represent locations in the original image where certain low-level features are present. Convolving these activation maps generates activation maps representing higher-level features, such as combinations of features. Deeper convolutional layers identify more complex features; thus, at the network's end, filters may be able to activate for the high-level features of interest, such as identifying dogs or handwritten digits. Pooling layers down sample features (for example, a max pooling layer would return the maximum values in features within a region); then, the pooled feature map matrix is flattened into a single column, which is processed in the fully connected layer. The fully connected layer accepts the outputs of the preceding layer (the activation maps of complex features) and identifies which complex features most strongly correlate to a particular class that is being identified by the CNN ("tumor", "infarct", etc.). Finally, the network determines the probability that the image corresponds to that class in order to generate a task-specific output decision.

imaging information, which can accelerate critical decision making in the stroke triage workflow and improve outcomes for patients with stroke.

DNNs constitute an excellent toolkit for addressing clinical challenges. Available clinical data must inform the selection of network architectures, training strategies, and model objectives. Simply reusing generic models on custom data might not yield high performance; such models might be idiosyncratically susceptible to biases in the training data, ultimately resulting in poor reliability. The present review aims to inform clinicians about the current state of the field for DL models in acute stroke management. We examine real-world examples of DL applications for acute stroke management in the domains of pixel-wise labeling, volumetric lesion segmentation, stroke detection, and prediction of functional outcomes and tissue fate postintervention. Finally, we explore future directions for the clinical applications of DL in acute stroke management.

Stroke Detection

Methods for Stroke Detection

Rapid detection of time-sensitive pathologies, such as acute stroke, results in improved clinical outcomes.[7,8] For patients with suspected ischemic stroke, early detection with neuroimaging allows for the faster exclusion of ICH and other stroke mimics, as well as rapid segmentation and prediction of tissue fate outcome. For patients with suspected ICH, early detection allows clinicians to assess the need for urgent neurosurgical interventions.[43] However, these processes are slowed by the time required for neuroimage acquisition and clinical interpretation. Treatment decisions cannot be made until the extent of lesioned tissue and likelihood of successful reperfusion or surgical intervention is determined. Moreover, alongside lagging imaging analysis times, manual stroke detection suffers from discrepancies between raters[18] and inter-rater variability for Alberta Stroke Program Early CT score (ASPECTS) assessments that are used to assess early ischemic changes on non-contrast CT (NCCT) brain images.[51–53]

Streamlining the process of stroke detection may not only decrease time from presentation to treatment and reduce mortality related to stroke but may also increase the accuracy and reliability of neuroimaging and EMR analysis for stroke detection. Automated systems for distinguishing between stroke types and classifying patient priority as "urgent" or "routine" based on neuroimaging data and clinical data in EMRs may be helpful as supportive aides for human expert diagnosis. DL techniques are well-suited to automate stroke detection, as they can perform well despite variations in the presentation of various pathologies by incorporating multiple types of relevant clinical data into decision making.

DNNS for Acute Ischemic Stroke Detection

Several approaches have been developed for potentially incorporating computer-assisted diagnosis (CAD) into clinical practice by using DNNs to detect acute neurological illness from NCCT and CT angiography (CTA) imaging. In a broad application of DNNs for acute ischemic stroke (AIS) detection, Titano et al (2018) deployed a 3D CNN modeled on ResNet-50 architecture in a simulated clinical environment and tasked it with determining whether an image contained acute neurological illness, and, if so, whether the illness was critical or noncritical. When deployed in a prospective trial, the model was able to significantly prioritize urgent studies, which appeared earlier in the queue than routine cases in the prioritized list.[35] Another useful modality in AIS detection is 4D CTA, which permits visualization of cerebral hemodynamics. In tandem with CT perfusion (CTP) imaging, these can be used to identify candidates for endovascular therapy outside of the 6-hour time window after which potential benefit is significantly reduced and risk of fatal bleeding increased.[47,54] Mejis et al (2020) applied a 3D CNN architecture for analyzing 4D CTA for image-level detection of arterial occlusions, accomplished by deriving a normalized time-to-signal (nTTS) map of the 4D-CTA in order to capture temporal information in a downsized 3D space. The model performed well, displaying relatively high sensitivity and specificity, with three false negative results and 18 false positive results in an independent consecutive test set of 279 cases. The improved speed of both models, and the accuracy of LVO detection demonstrated by Meijs et al, suggest that DNNs can be applied in clinical environments in order to distinguish between urgent and routine clinical cases, assist clinicians in differentiating AIS from other vascular pathologies, and even suggest eligible patients for thrombectomy.[36]

To perform early detection and rapid quantification of acute ischemic lesions using magnetic resonance (MR) images, Do et al (2020) developed a classifier algorithm using a recurrent residual convolutional neural network (RRCNN) to distinguish between diffusion-weighted imaging (DWI) MRI slices belonging to low (1–6) and high (7–10) DWI-ASPECTS groups.[39] An RRCNN contains a residual unit, which allows the network to train deep architectures by incorporating skip connections, or shortcuts to surpass certain layers; in theory, the recurrent residual convolutional layers permit feature accumulation on temporal tasks, which should improve performance on segmentation tasks.[55] These CNNs benefit from the reduction in imaging parameters as a result of data preprocessing and normalization, maximizing image contrast and permitting the models to fully benefit from multi-contrast MRI[39,40] in the detection and diagnosis of AIS. Together, these results suggest that DNNs may be able to assume an ancillary role in stroke triage by providing rapid assessments of neuroimaging and assisting clinicians in acute clinical decision making.[39]

DNNS for ICH Detection

Several successful CNNs have been developed for the detection of ICH. Grewal et al (2017) developed RADnet to identify ICH from 2D NCCT slices, and obtained a positive predictive value (precision) comparable to that of radiologists.[41] In the same year, Prevedello et al (2017) developed another DL algorithm for detecting hemorrhage, mass effect, and hydrocephalus from NCCT.[42]

Mirroring the success of the CT-based triage system developed by Titano et al (2018),[35] Arbabshirani et al (2018) developed a 3D CNN architecture trained on head CT images to recategorize “routine” head CT scans as “stat” if ICH was detected.[56] When applied in real time as a radiology workflow optimization tool, the network was able to upgrade 94 out of 347 “routine” studies to “stat,” 60 of which were declared to have ICH present by an expert radiologist; 5 new ICH cases from outpatients were detected. Despite imaging-based challenges facing the models—Prevedello et al’s model under-detected several urgent findings, possibly as a result of slice thickness—these models’ performances are beginning to approach those of radiologists, providing evidence for the potential ability of DL algorithms to detect ICH in a clinically meaningful manner.

Recent studies have improved upon previously devised strategies for detecting ICH. Kuo et al (2019) developed PatchFCN (CNN trained on NCCT image “patches,” or subsections, extracted from whole images),[44] and Ojeda et al (2019) evaluated the performance of a proprietary CNN architecture developed by Aidoc (Tel Aviv, Israel), one of the first DNNs to receive US Food and Drug Administration (FDA) clearance.[46] Both models performed at a level comparable to experts, even identifying some abnormalities missed by expert radiologists. Moreover, the Aidoc proprietary CNN achieved an overall accuracy of 98% when model output was reviewed by three expert radiologists.[44,46] Whereas Aidoc’s CNN benefited from its robust invariance as a result of its large training dataset, PatchFCN was able to produce comparable results by using strong supervision on a smaller training dataset, as well as smaller subsections of images versus full images.[44] This suggests that massive amounts of training data are not necessarily required to construct accurate, reliable models for relatively simple tasks, such as ICH detection; data augmentation techniques, which increase the amount of input data without necessarily increasing the contextual or semantic data—including cropping, flipping, translation, and patch extraction—can also assist in training DNNs on relatively small datasets. Furthermore, the success of Aidoc's model in a real-world, high-volume clinical setting represents a meaningful achievement in using CAD to improve triage and routine workflow, accounting for the heterogeneity in case presentation without controlling for center-specific factors, such as scanner type and image acquisition parameters.[46]

Another useful feature in the detection of ICH is the delineation between its subtypes—intraparenchymal hemorrhage, intraventricular hemorrhage, subdural hematoma, extradural hematoma, and subarachnoid hemorrhage—of which require different treatment methods and often necessitate rapid surgical intervention.[43] Chilamkurthy et al (2018) constructed a 2D CNN for the automated detection of ICH and its subtypes, calvarial fractures, midline shift, and mass effect with NCCT. Despite additional imbalance in the training dataset, such as the under-representation of extradural hematoma and the generation of possibly ambiguous or confounding labels through natural language processing (NLP)-based label extraction, this model is one of the first to use DNNs for the identification of ICH in NCCT.[43] The winners of the 2019 Radiological Society of North America Intracranial Hemorrhage Detection Challenge, also deployed a 2D CNN with three 2D classifier pipelines—single slices with 3 windows, 3 spatially adjacent slices with one window, and a combination of the 2—from bone, subdural, and brain imaging. This sequential model was able to identify ICH and delineate subtypes. Further increasing the complexity and representational power of the 2D CNN, Ye et al (2019) and Lee et al (2019) used modified CNNs for the detection of ICH and its 5 subtypes, also using slices from NCCT.[18,45] Ye et al (2019) used a joint CNN-RNN architecture wherein the CNN extracts useful features from image slices, whereas the RNN is used to extract useful features on a subject level. Unlike the deep convolutional neural networks (DCNNs) used by Lee et al, which are limited by the single-institution data upon which they were trained, the CNN-RNN's dual-pronged structure increases its robustness, allowing it to generate diagnoses for images collected at separate locations with different scanners and imaging parameters.[18]

Together, these models suggest the potential for comparable performance marked improvement in sensitivity comparable to expert raters, dependent on more robust training datasets. However, both models struggled with detection of subarachnoid hemorrhage and extradural hematoma, the former of which is an especially challenging subtype of ICH to diagnose, and the latter of which is consistently underrepresented in training and test datasets.[18,45] Future models should seek to train on more balanced, heterogeneous datasets in order to improve rates of accurate identification of less common ICH subtypes (Table 1).

Stroke Lesion Segmentation

Methods for Lesion Segmentation

Segmentation and quantitative assessment of CT, CTA, CTP, and MR neuroimages are critical for the diagnosis, monitoring, treatment, and investigational research of stroke.[1–5] For acute ischemia, characteristics of the infarct core and penumbra can be used to predict disease progression and identify candidates who would benefit from revascularization.[6–10] For ICH, neuroimaging can identify the hematoma, quantify its volume, and characterize its location.[1,2] Precision of clinical decision making is dependent on the timely, accurate extraction of relevant information from neuroimaging.

However, quantitative analysis of brain lesions via manual segmentation of 3-dimensional images is tedious, time-consuming, and expensive. Manual segmentation is expensive in terms of time, and even simple measurement methods are affected by inter-rater variability and error.[18,51–53] Accurate segmentation is hindered by the extreme heterogeneity in lesion shape, size, location, and evolution.[11,14,15] Furthermore, the imaging modalities used to visualize lesions often suffer from low signal-to-noise ratios (SNRs) and artifacts. Thus, clinicians often adopt qualitative measures, such as visual inspection, to determine the presence or absence of any acute traumatic intracranial abnormality,[57,58] midline shift exceeding 5 mm,[57,58] or intracranial hematoma exceeding 25 cubic centimeters[57]—or less-precise measures, such as approximate lesion volume, to inform clinical decision making.[59] However, incorporation of quantitative, rather than qualitative, imaging features has been linked to significant improvements in the prediction of clinical outcomes over a variety of injury severity levels.[59]

Automated segmentation and quantification methods may assist physicians in critical decision making by acquiring more precise measurements of injury. Such strategies can take advantage of statistical regularities across patient populations in order to rapidly identify and provide information about ischemic and hemorrhagic lesions in patients with suspected stroke, ultimately decreasing the time between image acquisition and therapeutic intervention without sacrificing precision. Although automated methods are unlikely to replace physicians in interpreting images and making treatment decisions, their ability to operate on larger scales and provide time-sensitive clinical information with greater reliability and reproducibility can help inform physicians' decision making processes.

DNNS for Lesion Segmentation

Prior attempts to solve the problems of automated lesion semantic and volumetric image segmentation and quantification have been met with various degrees of success. Earlier works conceptualized lesion segmentation as an

Table 1. Overview of a Selection of DL Algorithms for Lesion Detection and Segmentation Highlighted in the Text

Problem	Author	Specific Task	Architecture	Input	Pros	Cons	Improvements in Practice
Detection	Titano et al. (2018)	Detection of acute neurological illness	CNN	2D CT slices	Average inference time of 134 ms—150 times faster than that of humans	High false-alarm rate	Able to significantly prioritize urgent studies, which appeared earlier in queue
	Mejis et al. (2020)	Detection of intracranial ACA occlusions	CNN	3D time-to-signal (TTS) representation of 4D CTA	nTTS permits higher precision with respect to arrival and acquisition times	No direct localization of the occlusion	First method for detection of intracranial anterior circulation occlusions using 4D-CTA
	Do et al. (2020)	Binary classification of ASPECTS scores	CNN-RNN	2D DWI MRI slices	Recurrent residual layers permit feature accumulation on temporal tasks	Classification of binary ASPECTS scores, rather than individual regions	May serve as an ancillary tool for assisting in treatment decisions
	Grewal et al. (2017)	Detection of ICH	CNN	2D CT slices	3D context from neighboring slices can improve prediction ability	Requires improvement in prediction precision	Performance comparable to radiologists; higher recall than 2/3 radiologists in analysis
	Arbabshirani et al. (2018)	Classification of images as “routine” or “stat”	CNN	2D CT slices	Detected cases from heterogeneous dataset with a variety of conditions	Requires improvement in specificity and sensitivity	When implemented into clinical workflow, significant benefit on time to diagnosis in outpatients
	Kuo et al. (2019)	Detection of ICH	CNN	2D CT patches	Classification comparable to experts and robust abnormality localization	Requires more training data to mitigate random effects and boost accuracy	High sensitivity and specificity; screening tool with low proportion of false positives
	Ye et al. (2019)	Detection and classification of ICH and subtypes	CNN-RNN	2D CT slices	RNN captures feature info from consecutive slices, adding context	Prevalence of ICH in dataset higher than that in real clinical setting	Classification performance generally superior to avg. of junior radiology trainees
Segmentation	Kamnitsas et al. (2017)	Segmentation of brain lesions	CNN	2D CT slices	Dual pathway architecture incorporates both local and larger contextual information	Differences in scanner type and acquisition protocols impact images	Generic, FCN structure applicable to dif. lesion seg. tasks w/o major adaptations

abnormality detection problem, and designed models to compare pathological tissue with healthy counterparts in order to identify anomalous regions.[22,23] Presently, supervised CNN-based models have moved to the forefront of biomedical image segmentation.[4,28,31,60] Deep methods permit the incorporation of novel network modules, such as the U-Net,[61] attention modules,[62] and dense connectivity,[4] to retain input specificity while improving model performance.

Although 2D models are typically less demanding in terms of memory, training time, and dataset requirements, 3D models can sometimes make learning easier for a CNN and are capable of using 3D contextual information present in volumetric data, better equipping them to perform multiscale semantic segmentation.[4] Winners of the public Ischemic Stroke Lesion Segmentation (ISLES) competition, designed to facilitate the creation of accurate, reproducible advanced data analysis techniques for localization of lesioned tissue, represent the state of the field for 3D segmentation. In 2015, Kamnitsas et al (2017) won with DeepMedic, a 3D CNN which utilized a hybrid scheme between patch training — training on "patches," or subsections, extracted from whole images — and dense training on the whole image. DeepMedic correctly segmented 34 of 36 of cases, which consisted of 3D fluid-attenuated inversion recovery (FLAIR), T2-weighted turbo spin echo (T2w TSE), T1-weighted turbo field echo/turbo spin echo (T1w TFE/TSE), and diffusion-weighted (DWI) MR images, and obtained a Dice coefficient of 0.59 ± 0.31.[17] This model utilized a fully connected conditional random field (CRF), a statistical modeling method which can incorporate context[63] for biomedical data.[17]

Recent models use fully convolutional layers incorporated into stacked encoder-decoder models. In these models, convolutional and pooling encoder layers first compress input into a latent space representation. Later, upsampling decoder layers predict segmentation output from this representation. Subsequent ISLES winners have capitalized on the efficacy of the U-Net, based on encoder-decoder, for biomedical image semantic and volumetric image segmentation.[29,30,61] To address the trade-off between localization accuracy and image context,[64] Ronneberger et al (2015) devised the U-Net, a CNN with 2 paths: a contracting path to capture context, and an expanding path to perform upsampling and increase localization. The U-Net uses an "overlap-tile" strategy, which allows the network to predict the image part by part by breaking it down into overlapping tiles. Pixels at the boundaries of these tiles are extrapolated by mirroring the neighboring tiles. This strategy is especially effective because it allows the network to segment arbitrarily large images without experiencing slowdown from high GPU memory demands.[61] Additionally, when data augmentation is performed with elastic deformations mirroring the deformations observed in pathological imaging, the U-Net grows invariant to such warping, allowing the network to reach high levels of accuracy with relatively few annotated training images.[61]

The U-Net has become a popular architecture for biomedical semantic and volumetric image segmentation; modified U-Nets have been used for segmenting and quantifying lesions in acute stroke for a variety of imaging modalities (DWI and FLAIR MRI, NCCT, CTA, and CTP), input formats (whole-head images and image slices), and dimensions (2D and 3D).[2,27,32,33,65] Cutting-edge segmentation methods combine U-Nets with both novel and common strategies, such as pyramid scene parsing networks,[34] which extract global context information through region-based context aggregation[29]; and deep residual attention modules,[66] which extract high-quality features from input images.[27] Presently, the most robust DNNs are those with features including dense connectivity to boost gradient flow in the network,[4] image sampling that balances the data distribution, and smaller convolutional kernels for greater discriminative ability[17] (Table 1).

Stroke Outcome Prognostication

Methods for Prognostication

For both ischemic and hemorrhagic stroke, prediction of the benefits of therapeutic and surgical interventions in terms of mortality and functional outcome are critical for medical decision making. For ischemic stroke, several recent clinical trials have found mechanical thrombectomy to be a safe, effective therapy within 6 hours of stroke symptom onset (MR-CLEAN,[67] ESCAPE,[68] EXTEND-IA,[69] SWIFT-PRIME,[9] REVASCAT,[70] and THRACE[71]). Furthermore, clinical trials suggest that perfusion levels of brain tissue can be used to select candidates for therapeutic intervention outside of the traditional time windows for treatment (DEFUSE[9,47] and EXTEND[48]). Safe, reliable prediction of expected tissue salvage in patients with ischemic stroke is crucial for selection of candidates who would most benefit from mechanical thrombectomy, especially those who could achieve revascularization outside of the traditional 6-hour time window for reperfusion interventions.[49] One early investigation into stroke outcome prognostication found a relationship between time to recanalization and recanalization status: for every hour until thrombolysis in cerebral infarction (TICI), the likelihood of infarction increased by 18.9% until 0/2a recanalization (no perfusion restored to less than two-thirds of perfusion restored), and by 33.2% for every hour until TICI 2b/3 recanalization (complete perfusion restored, ranging from slow to normal filling speed).[50] These findings further emphasized the need for rapid intervention for AIS.

Despite the fact that ICH is the most severe form of stroke, few trials have examined the benefits of medical treatment and surgical intervention on mortality and functional outcome in hemorrhagic stroke. The ICH score is a commonly used metric for grading the severity of presentations,[72] but along with other outcome prognostication models for ICH, it has not been proven beneficial for improving patient outcomes.[73] Furthermore, alongside the complexity and progression of ICH, most predictive models fail to account for common changes in the clinical care of patients with ICH, such as the withdrawal of care, do-not-resuscitate orders (DNRs), and end-of-life palliative care.[73,74] As ICH-related mortality approaches 50%,[75,76] the lack of applicable models for outcome prediction in ICH belie their necessity. Advanced neuroimaging analysis may offer insights into which patients are likely to benefit from therapeutic interventions and which patients are likely to experience adverse, life-threatening risks[77] or poor functional outcomes. Understanding the likelihood of various clinical outcomes allows clinicians and family members to make more informed decisions for care in acute stroke, taking into account realistic expectations for patients' post-stroke health and functionality.

DNNS for Outcome Prognostication

In ischemic stroke, DL models using tissue information from symptom onset to predict tissue fate typically use perfusion map information – time to peak (TTP), cerebral blood flow (CBF), cerebral blood volume (CBV), and time-to-maximum perfusion (T_{max}).[49,78,79] Early attempts focused on creating CNNs for ischemic lesion evolution by operating not only on a voxel-by-voxel basis, but also by incorporating data in the area surrounding a target location in order to improve prediction accuracy. However, such models were limited by the selection of specific 2D lesion slices as inputs, a process which disregards lesion heterogeneity and prevents the network from modeling interplay between lesion appearance and other clinical factors discernible on neuroimaging.[79] More robust, recent models have been trained on all voxels of interest in neuroimaging[78] and include attention gates to focus on target structures.[80] Their success likely stems from the realistic clinical variability in the training and test datasets, which allows the models to harness the interplay between various biomarkers present in acute and follow-up imaging in ischemic stroke.

Presently, there are few ML applications, and even fewer DNNs, that incorporate both neuroimaging and other clinical data for tissue fate and functional outcome prediction. For ischemic stroke, McKinley et al (2017), winners of the ISLES 2015 acute stroke outcome/penumbra estimation (SPES) task, based their predictions on a multitude of relevant clinical data: clinical parameter details, such as TICI score, modified Rankin scale (mRS), time since stroke (TSS), and time to treatment (TTT); acute imaging details, such as (antibody-drug conjugate [ADC]) maps, raw 3D and 4D perfusion data and maps; and follow-up stroke imaging. They implemented 2 joint models and random forests, which output the mean prediction of an ensemble of decision trees, to predict outcome in the event of good response to therapy, as well as the natural course of the stroke, on a voxel-by-voxel basis using compound spatial information from multimodal MRI. Their model correctly identified 20 of 20 cases and obtained a Dice coefficient of 0.82 ± 0.08.[49] Attempting the novel task of predicting mRS from neuroimaging and clinical data, 2016 winners Maier et al (2017) extracted information about the states of the ischemic lesion, surrounding tissue, and remainder of the brain to train regression random forests (RRFs) to predict mRS scores for patients after 90 days. Their model correctly predicted 90-day mRS in 19 of 19 cases with an average absolute error of 1.05 ± 0.62.[28]

In a pilot study that investigated the prediction of functional outcomes following tPA in ischemic stroke using DL, Bacchi et al (2020) developed a CNN for mRS-based classification and a general artificial neural network (ANN) for clinical data-based classification (demographic information, time from stroke onset to presentation, National Institutes of Health Stroke Scale [NIHSS] at presentation, clinical variables, such as blood pressure and blood glucose at presentation, and the medical history).[81] Their findings suggested that combined predictive DNN models were more effective at outcome prediction 24 hours after presentation than common prognostication score calculations such as THRIVE[82] and HIAT.[83] Although this study demonstrates proof of concept that DNNs incorporating both clinical and neuroimaging information can potentially be used to predict patient outcomes, further studies are required to refine models for deployment in acute clinical decision making, such as for selecting patients for clinical trials and novel interventions. These studies provide evidence for the potential for ML and DL-based algorithms as supportive tools to streamline conventional neuroimaging analysis procedures and to improve prognostication in clinical settings (Table 2).

Discussion

The aim of DNNs, and the ML applications upon which they are based, is to extract the maximum amount of predictive power from the available multidimensional clinical

Table 2. Overview of a Selection of DL Algorithms for Outcome Prognostication Highlighted in the Text

Problem	Author	Specific Task	Architecture	Input	Pros	Cons	Improvements in Practice
Prognostication	McKinley et al. (2017)	Estimation of penumbral volume	Random forest	2D DWI, T1, T2 MRI slices - ADC maps, raw 3D/4D perfusion data - Clinical data	Able predict tissue risk for new patients with respect to treatment success	Requires improvement in sensitivity and precision	Can yield more accurate pred. of tissue-at-risk than expert rater using lin. thresholded maps
	Maier et al. (2017)	Prediction of lesion and clinical outcome	Random forest	- ADC/PWI maps, raw 4D perfusion data - Clinical data	Hemispheric dif. measure to makes use of PWI maps; lesion and local features extracted	Fast to train, but can be slow to create predictions due to large number of trees	Competitive lesion seg. results in outcome prediction and as well as seg. in acute/semi-acute stroke
	Wang et al. (2019)	Prediction of mRS at two timepoints	Random forest	- Demographic characteristics - Laboratory studies - CT imaging findings	Predictors of 30-day mortality in the ICH score contributed to predicting 1-month outcome	Small sample size; did not include early hematoma growth/edema extension info	Optimal prediction based on few attributes; can be recorded easily without extra clinical loading
	Nawabi et al. (2021)	Prediction of functional outcome	Random forest	- Radiomic features extracted from consensus ROIs on 2D CT slices - ICH score	Quantitative imaging features provide a high discriminatory power in outcome prediction	Manual ROI definition still implies a certain degree of observer dependence	Performs equally rel. to multi-dimensional scoring systems; may assist clinical trials

information. The almost intractable burden of manually sorting through countless clinical variables to identify the key parametric combinations which influence the desired clinical support system illustrate the difficulty of such a task for a human. Despite the potential value of advanced neuroimaging techniques coupled with clinical information for assisting clinicians in the diagnosis, localization, and treatment of acute stroke, the contributions of neuroimaging in acute stroke management are tempered by time delays between presentation and treatment,[35,36,43] inter-clinician variability in clinical assessment of neuroimaging,[18,51–53] and the lack of systemic conglomeration of clinical information to generate robust predictions. The DL models presently surveyed offer developing solutions to many of these challenges. They enable rapid assessments of meaningful neuroimaging data,[35,36] whereas reducing inter-rater variability and boosting assessment standardization and reliability.[18,44] Moreover, they allow for automatic selection and demonstration of relevant clinical features for acute stroke decision making[18,45] and harness the massive amounts of neuroimaging data and EMRs generated in clinical settings to aid clinicians in routine and triage stroke management.[35,84]

Several options exist for types of available clinical support systems. DNNs can be used for fully automated processes; hybrid processes, which allow clinicians to manually survey and, if needed, edit outputs; and routine decision support. Some automated detection algorithms could serve as a beginning foothold for DNNs in clinical practice. On an individual level, the goal of incorporating DL algorithms into medical decision making in this manner is to make clinical care and diagnosis faster and more reliable; on an institutional level, these approaches are applied with the aims of reducing inefficiencies in the clinical workflow and addressing inequities in health care. As automated diagnostic and prognostic tools are integrated into patient care, the structure and nature of stroke management will evolve with emerging challenges and solutions. The traditional role of a physician is to accurately diagnose pathology based on limited evidence, in limited time, and determine the best treatment strategy among many options for a specific patient. Ideally, the prediction generated by a DL algorithm serves as an additional datapoint for a clinician to consider in medical decision making. However, in cases of disagreement in diagnosis, the issue emerges of how much weight a physician should assign to an algorithm's diagnosis, and, furthermore, how to characterize and address this conflict, considering that the DNN has not been trained in the same manner and according to the same standards as a human physician. Although the inputs and outputs of DNNs are generally easily understood, the complex mechanisms underlying changes to weights in hidden layers pose a significant challenge to interpretation by humans, leading to DNNs' common description as "black boxes."[85] Despite developments in increasing the transparency and comprehensibility of DNNs, opacity persists in the forms of corporate confidentiality, user-side technical illiteracy, and complexity of algorithmic representations of data.[86] In considering the role DL applications play in supporting clinical decision making, one must consider the relationship between the algorithm and clinician. Without certainty that a decision produced by a DL algorithm qualifies as "knowledge," which may be difficult to ascertain without insight into the process behind a given output, overriding decisions made by clinicians in deference to those made by algorithms lacks proper epistemic support.[86] One potential working solution may be the visualization of clinically relevant factors in AI-based decision-making. For example, the gradient-weighted class activation mapping (Grad-CAM) technique generates coarse localization maps demonstrating the level of importance of individual neurons in output generation in a network's final convolutional layer.[87] Such displays, hiding the complexity of the ML from the clinician but displaying the results in a clinically relevant form, enable a physician to accept or reject automated diagnostics based on whether the class activation maps make biological and medical sense. As DL algorithms become more explainable, and relevant standards which appeal to epistemic norms of informed consent are established, the role of a DL algorithm as a "peer" may be renegotiated; presently, however, physicians, and the human approach to clinical care, involving information gathering, integration, interpretation, and diagnosis, are unlikely to be "replaced" by DNNs.

The question persists: what are, and should be, the consequences of ignoring the algorithm's decision? The potential emergence of "defensive medicine," deferring to automated applications' decisions out of fear of being marked as acting imprudently,[86] portends the risk of devaluing physicians' training and knowledge in medical decision making. The ethical and legal responsibilities associated with artificial intelligence are muddied by the multifaceted structure of decision making, especially in situations where an algorithm generates an incorrect decision, or one that leads to an unfavorable outcome. The burden of responsibility may lie with the developers of the algorithm, the compilers of the dataset the algorithm was trained on, the physician who incorporated available evidence and made the final decision, the health care institution that introduced algorithmically generated predictions into clinical decision making, or some intractable combination of these agents. In evaluating the ethics of DL-based clinical care, a restructuring of traditional notions of responsibility may need to take place. Overall, ethical and regulatory concerns must be considered in three broad categories: data sourcing, product development, and clinical deployment.[88] Ethical data sourcing requires companies to be compliant with data protection guidelines in their and constituent patients' countries of residence, especially in cases where datasets contain sensitive data subject to unique regulatory provisions based on patient populations or data types. Ethical product development requires DL algorithm developers to explicitly combat healthcare disparities by compiling training datasets which accurately represent true rates of epidemiology and modes of prevalence in source and target populations. Finally, ethical clinical development and deployment requires physicians to play a key role in the introduction of DL algorithms in clinical practice, patients to be informed about the use of novel AI-based systems in their care, and for administrative systems to be established to systematically assess the efficacy of DL algorithms in real-time clinical settings. The latter may precipitate the need for artificial intelligence professionals in clinical settings to test, certify, and retest artificial intelligence diagnostic platforms as they are introduced and updated, as a system where clinicians alone are responsible to judge the accuracy and reliability of such algorithms would be fraught with trouble. Ultimately, for DL to have value in medical decision making, physicians' and patients' autonomy must be protected, and future deployments of DL algorithms in clinical practice must emphasize and move toward increased transparency.

Critically, the potential of DL techniques in decision support for acute stroke must be evaluated based on the level of performance is acceptable in clinical care. Many of the best-performing models in lesion segmentation and quantification, stroke detection, and stroke outcome prognostication fail to reach the gold standard area under the receiver operating characteristic curve (AUC; which conveys how well a classification model can delineate between classes) of >0.90. This suggests that the majority of DNNs require further development before being deployed in clinical settings. However, because patient outcomes after therapeutic intervention are extremely variable and often directly related to lesion size at presentation, it is difficult for DL algorithms to attain or surpass gold standard metrics, even with excellent training and test data.[77] However, the aim of most DNN applications in clinical settings is to aid, rather than to replace, clinicians. Therefore, models capable of reaching AUC values comparable to those of clinicians may improve clinical care simply by offering second opinions.

A common study limitation in DNN application for stroke management has been data heterogeneity. Variability in pathological presentation, scanner model, and neuroimaging parameters is often not represented in a balanced manner in training datasets, which reduces model generalizability to data collected at other centers. Future studies should pool data from multiple institutions, collected with various imaging acquisition techniques and clinical parameters, for more robust training datasets and algorithm standardization. Typically, academic hospitals strive be the first to publish papers on data collected from clinical trials, and thus only share data after studies are finished and publications have been written. However, this process often prevents data from being used by others for analysis for years, minimizing the potential widespread utility of novel datasets. Furthermore, stricter privacy laws, such as the EU General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), are covering more of the world's personal information.[89] Scientific societies and regulatory agencies, researchers, and developers should maximize efforts to take advantage of the opportunities big data provides for improving health care while prioritizing the agency and rights of patients to control their own personal information. This may be potentially accomplished through involving patients or patient advocacy groups in study design. Additionally, developers of DL-based models use patient-level data to develop their algorithms, raising the issues of whether patients should be compensated for their contributions, however indirect to company profits, and if this data should be openly available for research by other companies and institutions. In a review of data sharing initiatives in healthcare, Hulsen (2020) discusses the BigData@Heart platform of the Innovative Medicines Initiative (IMI), which compiled potential conditions for data sharing, including (1) only sharing health data for scientific research, (2) in de-identified form, (3) after approval from a designated review committee, (4) observing appropriate measures for data security in compliance with applicable laws and regulations.[89,90] Ultimately, the goal of data sharing is to encourage collaboration between researchers, enable independent confirmation of results, promote the testing of novel hypotheses, and reinforce the credibility of conclusions drawn from completed clinical trials, all goals which will be possible only through an emphasis on patient rights and agency. A key element in ethical, accessible clinical development of DL-based tools is the engagement of physicians in deployment. An often-neglected bottleneck in applying of DL-based diagnostic systems is the importance of physician "buy-in." Engaging physicians in automated approaches requires making clear how DL models can augment and extend their clinical practice. Early adopters should strive to emphasize how automated tools can improve practice efficiency, increase patient safety and positive outcomes, and reduce physician burnout. Furthermore, it is critical that technology is easy to learn how to use, and for physicians to feel involved in the integration process. Easy access to training, resources, and results will likely facilitate the adoption of technology into practice. Physicians and healthcare leaders alike must not only understand the benefits and limitations of AI, ML, and DL-based tools, but be equipped with the knowledge to evaluate and assess these tools as "consumers," rather than be dependent on marketing and institutional pressure to determine their applicability in their own practice. Informed, critical inquiry and constructive criticism are vital to gaining maximal potential use from DL algorithms in improving clinical care for patients. Training rooted in evidence-based medicine, Cornelius et al (2021) posit the importance of topics in study design, epidemiology, and biostatistics, must be brought to the forefront of clinical care in order to enable trainees to evaluate medical literature and understand how to benefit from DL applications in their own practice.[91] Systematic changes are required in medical curricula to equip future and present physicians to critically evaluate developing technologies, with an emphasis on high-level principles, including the terminology used in ML and DL, optimal models for different kinds of clinical and imaging datasets, types of clinical problems AI-based strategies are most useful for solving, and performance trade-offs of different kinds of models. When evaluating clinical applications of artificial intelligence, physicians should be able to determine whether the results of a study are valid, if the datasets used were representative of heterogeneity and prevalence of epidemiology in given populations of interest, what ground truth data were used to train the algorithm, if the study design and algorithm used was appropriate for the kind of question asked, what benchmarks were used to test and validate the model, and if the findings are applicable to their own patient populations. Literacy in these elements, achievable through an emphasis on research methods and statistics courses, will enable more realistic integration of DL applications into clinical workflows. Alongside technical training in ML literacy, it is paramount to emphasize thinking through the ethical issues that naturally arise when incorporating artificial intelligence into medical decision making, specifically with the aim of preserving humanity in medicine.[92] As the role of a physician co-evolves with that of technology, so does the relationship between physicians and patients. It is critical to encourage medical trainees to regularly evaluate and re-assess their roles as empathetic decision makers, caretakers, and educators throughout their changing careers, especially in tandem with understanding the role and potential for artificial intelligence in improving patient care. Further developments should be pursued in DNNs to maximize their benefits in acute stroke care and allow for faster, more accurate, and more comprehensible stroke detection, lesion segmentation, and tissue outcome prognostication for improved patient outcomes.

Author Contributions

I.R.C., A.L.C., M.H.M., M.M.Y., A.M.P., S.P., S.J.S., and K.N.S. contributed to the conception and design of the study. I.R.C. acquired and analyzed the data. I.R.C. drafted the text and prepared the figures.

Potential Conflicts of Interest

No conflicts of interests to report.

References

1. Sanelli PC, Sykes JB, Ford AL, et al. Imaging and treatment of patients with acute stroke: an evidence-based review. AJNR Am J Neuroradiol 2014;35:1045–1051.
2. Arab A, Chinda B, Medvedev G, et al. A fast and fully-automated deep-learning approach for accurate hemorrhage segmentation and volume quantification in non-contrast whole-head CT. Sci Rep 2020;10:19389.
3. von Kummer R, Meyding-Lamadé U, Forsting M, et al. Sensitivity and prognostic value of early CT in occlusion of the middle cerebral artery trunk. AJNR Am J Neuroradiol 1994;15:9–15. discussion 16–8.
4. Zhang R, Zhao L, Lou W, et al. Automatic segmentation of acute ischemic stroke from DWI using 3-D fully convolutional DenseNets. IEEE Trans Med Imaging 2018;37:2149–2160.
5. Bash S, Villablanca JP, Jahan R, et al. Intracranial vascular stenosis and occlusive disease: evaluation with CT angiography, MR angiography, and digital subtraction angiography. AJNR Am J Neuroradiol 2005;26:1012–1021.
6. Bivard A, McElduff P, Spratt N, et al. Defining the extent of irreversible brain ischemia using perfusion computed tomography. Cerebrovasc Dis 2011;31:238–245.
7. Neumann-Haefelin T, Wittsack HǊ Wenserski F, et al. Diffusion- and perfusion-weighted MRI. The DWI/PWI mismatch region in acute stroke. Stroke 1999;30:1591–1597.
8. Albers GW, Thijs VN, Wechsler L, et al. Magnetic resonance imaging profiles predict clinical response to early reperfusion: the diffusion and perfusion imaging evaluation for understanding stroke evolution (DEFUSE) study. Ann Neurol 2006;60:508–517.
9. Albers GW, Straka M, Kemp S, et al. Abstract 52: results of DEFUSE 2. Imaging Endpoints 2012;43:A52.
10. Parsons M, Spratt N, Bivard A, et al. A Randomized Trial of Tenecteplase versus Alteplase for Acute Ischemic Stroke. N Engl J Med 2012;366:1099–1107.
11. Chen L, Bentley P, Rueckert D. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. Neuroimage Clin 2017;15:633–643.
12. Wouters A, Dupont P, Norrving B, et al. Prediction of stroke onset is improved by relative fluid-attenuated inversion recovery and perfusion imaging compared to the visual diffusion-weighted imaging/fluid-attenuated inversion recovery mismatch. Stroke 2016;47:2559–2564.
13. Delcourt C, Carcel C, Zheng D, et al. Comparison of ABC methods with computerized estimates of intracerebral hemorrhage volume: the INTERACT2 study. Cerebrovasc Dis Extra 2019;9:148–154.
14. Pérez Malla CU, Valdés Hernández MC, Rachmadi MF, Komura T. Evaluation of enhanced learning techniques for segmenting ischaemic stroke lesions in brain magnetic resonance perfusion images using a convolutional neural network scheme. Front Neuroinform 2019;13:33.
15. van der Worp HB, van Gijn J. Acute Ischemic. Stroke 2007;357:572–579.
16. Toma AK, Holl E, Kitchen ND, Watkins LD. Evans' index revisited: the need for an alternative in Normal pressure hydrocephalus. Neurosurgery 2011;68:939–944.
17. Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal 2017;36:61–78.
18. Ye H, Gao F, Yin Y, et al. Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. Eur Radiol 2019;29:6191–6201.
19. Hesamian MH, Jia W, He X, Kennedy P. Deep learning techniques for medical image segmentation: achievements and challenges. J Digit Imaging 2019;32:582–596.
20. Renard F, Guedria S, Palma ND, Vuillerme N. Variability and reproducibility in deep learning for medical image segmentation. Sci Rep 2020;10:13724.
21. Isensee F, Jaeger PF, Kohl SAA, et al. nnU-net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18:203–211.
22. Prastawa M et al. A brain tumor segmentation framework based on outlier detection*1. Med Image Anal 2004;8:275–283.
23. Doyle S, Vasseur F, Dojat M, and Forbes F, Fully automatic brain tumor segmentation from multiple MR sequences using hidden Markov fields and variational EM. Procs. NCI-MICCAI BraTS, 2013:18–22.
24. Gooya A, Pohl KM, Bilello M, et al. Joint segmentation and deformable registration of brain scans guided by a tumor growth model. Med Image Comput Comput Assist Interv 2011;14:532–540.

25. Parisot S, Duffau H, Chemouny S, Paragios N. Joint Tumor Segmentation and Dense Deformable Registration of Brain MR Images. Med Image Comput Comput Assist Interv 2012;15:651–658.

26. Rao, A., C. Ledig, V. Newcombe, D. Menon and D. Rueckert, Contusion segmentation from subjects with traumatic brain injury: a random forest framework. in *2014 IEEE 11th international symposium on biomedical imaging (ISBI)*. 2014.

27. Liu L, Hu X, Zhu L, et al. ψ-Net: stacking densely convolutional LSTMs for sub-cortical brain structure. Segmentation 2020;39:2806–2817.

28. Maier O, Menze BH, von der Gablentz J, et al. ISLES 2015 - a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Med Image Anal 2017;35:250–269.

29. Winzeck S, Hakim A, McKinley R, et al. ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI. Front Neurol 2018;9:679.

30. Song TJ, *Generative model-based ischemic stroke lesion segmentation*. 2019.

31. Liu X, Niethammer M, Kwitt R, et al. Low-rank to the rescue – atlas-based analyses in the presence of pathologies. Med Image Comput Comput Assist Interv 2014;17:97–104.

32. Huff TJ, Ludwig PE, Salazar D, Cramer JA. Fully automated intracranial ventricle segmentation on CT with 2D regional convolutional neural network to estimate ventricular volume. Int J Comput Assist Radiol Surg 2019;14:1923–1932.

33. Kim Y-C, Lee JE, Yu I, et al. Evaluation of diffusion lesion volume measurements in acute ischemic stroke using encoder-decoder convolutional network. Stroke 2019;50:1444–1451.

34. Zhao, H., J. Shi, X. Qi, X. Wang and J. Jia Pyramid scene parsing network. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

35. Titano JJ, Badgeley M, Schefflein J, et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. Nat Med 2018;24:1337–1341.

36. Meijs M, Meijer FJA, Prokop M, et al. Image-level detection of arterial occlusions in 4D-CTA of acute stroke patients using deep learning. Med Image Anal 2020;66:101810.

37. Meijs, M., A. Patel, S. van de Leemput, B. van Ginneken, M. Prokop and R. Manniesing Fast, robust and accurate segmentation of the complete cerebral vasculature in 4D-CTA using deep learning. In Annual meeting of the Radiological Society of North America. 2018.

38. Meijs M, Patel A, van de Leemput SC, et al. Robust segmentation of the full cerebral vasculature in 4D CT of suspected stroke patients. Sci Rep 2017;7:15622.

39. Do L-N, Baek BH, Kim SK, et al. Automatic assessment of ASPECTS using diffusion-weighted imaging in acute ischemic stroke using recurrent residual convolutional neural network. Diagnostics (Basel, Switzerland) 2020;10:803.

40. Liu L, Chen S, Zhang F, et al. Deep convolutional neural network for automatically segmenting acute ischemic stroke lesion in multi-modality MRI. Neural Comput Appl 2020;32:6545–6558.

41. Grewal, M., M. M. Srivastava, P. Kumar and S. Varadarajan Radnet: radiologist level accuracy using deep learning for hemorrhage detection in ct scans. in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. 2018. IEEE.

42. Prevedello LM, Erdal BS, Ryu JL, et al. Automated critical test findings identification and online notification system using artificial intelligence in imaging. Radiology 2017;285:923–931.

43. Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. Lancet 2018;392:2388–2396.

44. Kuo W, Häne C, Mukherjee P, et al. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. Biol Sci 2019;116:22737–22745.

45. Lee H, Yune S, Mansouri M, et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. Nat Biomed Eng 2019;3:173–182.

46. 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. *SPIE Medical Imaging*. Vol 10949. SPIE, San Diego, CA, United States, 2019.

47. Albers GW, Marks MP, Kemp S, et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion. Imaging 2018;378: 708–718.

48. Campbell BC, Mitchell PJ, Yan B, et al. A multicenter, randomized, controlled study to investigate EXtending the time for thrombolysis in emergency neurological deficits with intra-arterial therapy (EXTEND-IA). Int J Stroke 2014;9:126–132.

49. McKinley R, Häni L, Gralla J, et al. Fully automated stroke tissue estimation using random forest classifiers (FASTER). J Cereb Blood Flow Metab 2017;37:2728–2741.

50. Kemmling A, Flottmann F, Forkert ND, et al. Multivariate dynamic prediction of ischemic infarction and tissue salvage as a function of time and degree of recanalization. J Cereb Blood Flow Metab 2015; 35:1397–1405.

51. Austein F, Wodarg F, Jürgensen N, et al. Automated versus manual imaging assessment of early ischemic changes in acute stroke: comparison of two software packages and expert consensus. Eur Radiol 2019;29:6285–6292.

52. Wilson AT, Dey S, Evans JW, et al. Minds treating brains: understanding the interpretation of non-contrast CT ASPECTS in acute ischemic stroke. Expert Rev Cardiovasc Therapy 2018;16:143–153.

53. Farzin B, Fahed R, Guilbert F, et al. Early CT changes in patients admitted for thrombectomy: intrarater and interrater agreement. Neurology 2016;87:249–256.

54. Ma H et al. Thrombolysis guided by perfusion imaging up to 9 hours after onset of. Stroke 2019;380:1795–1803.

55. Alom, M.Z., Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha, Vijayan K. Asari, Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. 2018.

56. Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ, et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. npj Digital Medicine 2018;1:9.

57. Marshall LF, Marshall SB, Klauber MR, et al. The diagnosis of head injury requires a classification based on computed axial tomography. J Neurotrauma 1992;9:S287–S292.

58. Maas AI, Hukkelhoven CW, Marshall LF, Steyerberg EW. Prediction of outcome in traumatic brain injury with computed tomographic characteristics: a comparison between the computed tomographic classification and combinations of computed tomographic predictors. Neurosurgery 2005;57:1173–1182. discussion 1173-82.

59. Yuh EL, Cooper SR, Ferguson AR, Manley GT. Quantitative CT improves outcome prediction in acute traumatic brain injury. J Neurotrauma 2012;29:735–746.

60. Krizhevsky, A., I. Sutskever, and G.E. Hinton. ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems. NeurIPS Proceedings 2012.

61. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention – MICCAI 2015*. Cham: Springer International Publishing, 2015.

62. Jin Q, Meng Z, Sun C, et al. RA-UNet: a hybrid deep attention-aware network to extract liver and tumor in CT scans. Front Bioeng Biotechnol 2020;8:605132.

63. Krähenbühl, P. and V. Koltun, Efficient inference in fully connected CRFs with Gaussian edge potentials 2011.

64. Cireşan D and Meier U. Multi-column deep neural networks for offline handwritten Chinese character classification. in *2015 international joint conference on neural networks (IJCNN)*. 2015. IEEE.

65. Wang Z, Zou N, Shen D, Ji S. Non-local U-nets for biomedical image segmentation. Proceedings of the AAAI Conference on Artificial Intelligence 2020;34:6315–6322.

66. Chen L-C, Papandreou G, Kokkinos I., Murphy K, Yuille AL, *Deeplab*: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs 2017. 40: p. 834–848.

67. Berkhemer OA, Fransen PS, Beumer D, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. N Engl J Med 2015; 372:11–20.

68. Goyal M, Demchuk AM, Menon BK, et al. Randomized assessment of rapid endovascular treatment of ischemic stroke. N Engl J Med 2015;372:1019–1030.

69. Campbell BCV, Mitchell PJ, Kleinig TJ, et al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. N Engl J Med 2015;372:1009–1018.

70. Jovin TG, Chamorro A, Cobo E, et al. Thrombectomy within 8 hours after symptom onset in ischemic. Stroke 2015;372:2296–2306.

71. Bracard S, Ducrocq X, Mas JL, et al. Mechanical thrombectomy after intravenous alteplase versus alteplase alone after stroke (THRACE): a randomised controlled trial. Lancet Neurol 2016;15:1138–1147.

72. Hemphill JC, Bonovich DC, Besmertis L, et al. The ICH Score. Stroke 2001;32:891–897.

73. Wang H-L, Hsu WY, Lee MH, et al. Automatic Machine-Learning-Based Outcome Prediction in Patients With Primary Intracerebral Hemorrhage. Front Neurol 2019;10:910.

74. Holloway RG, Arnold RM, Creutzfeldt CJ, et al. Palliative and end-of-life care in stroke: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2014;45:1887–1916.

75. Jakubovic R, Aviv RI. Intracerebral hemorrhage: toward physiological imaging of hemorrhage risk in acute and chronic bleeding. Front Neurol 2012;3:86.

76. Nawabi J, Kniep H, Elsayed S, et al. Imaging-based outcome prediction of acute intracerebral hemorrhage. Transl Stroke Res 2021;12: 958–967.

77. Bivard A, Churilov L, Parsons M. Artificial intelligence for decision support in acute stroke - current roles and potential. Nat Rev Neurol 2020;16:575–585.

78. Nielsen A, Hansen MB, Tietze A, Mouridsen K. Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning. Stroke 2018;49:1394–1401.

79. Stier N, Vincent N, Liebeskind D, Scalzo F. Deep learning of tissue fate features in acute ischemic stroke. Proceedings (IEEE Int Conf Bioinformatics Biomed) 2015;2015:1316–1321.

80. Yu Y, Xie Y, Thamm T, et al. Use of deep learning to predict final ischemic stroke lesions from initial magnetic resonance imaging. JAMA Netw Open 2020;3:e200772.

81. Bacchi S, Zerner T, Oakden-Rayner L, et al. Deep learning in the prediction of Ischaemic stroke thrombolysis functional outcomes: a pilot study. Acad Radiol 2020;27:e19–e23.

82. Flint AC, Faigeles BS, Cullen SP, et al. Thrive score predicts ischemic stroke outcomes and thrombolytic hemorrhage risk in vista. Stroke 2013;44:3365–3369.

83. Hallevi H, Barreto AD, Liebeskind DS, et al. Identifying patients at high risk for poor outcome after intra-arterial therapy for acute ischemic stroke. Stroke 2009;40:1780–1785.

84. Chen-Ying H, Chen W-C, Lai P-T, et al. Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database. Annu Int Conf IEEE Eng Med Biol Soc 2017;2017:3110–3113.

85. Kriegeskorte N. Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. Annu Rev Vis Sci 2015;1:417–446.

86. Grote T, Berens P. On the ethics of algorithmic decision-making in healthcare. J Med Ethics 2020;46:205–211.

87. Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis 2020;128:336–359.

88. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. PLoS Med 2018;15:e1002689.

89. Hulsen T. Sharing is caring—data sharing initiatives in healthcare. Int J Environ Res Public Health 2020;17:3046.

90. Kalkman S, Mostert M, Udo-Beauvisage N, et al. Responsible data sharing in a big data-driven translational research platform: lessons learned. BMC Med Inform Decis Mak 2019;19:283.

91. James CA, Wheelock KM, Woolliscroft JO. Machine learning: the next paradigm shift in medical education. Acad Med 2021;96: 954–957.

92. Pucchio A, Eisenhauer EA, Moraes FY. Medical students need artificial intelligence and machine learning training. Nat Biotechnol 2021; 39:388–389.