Computational Approaches for Acute Traumatic Brain Injury Image Recognition

Emily Lin and Esther L. Yuh*

Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, CA, United States

In recent years, there have been major advances in deep learning algorithms for image recognition in traumatic brain injury (TBI). Interest in this area has increased due to the potential for greater objectivity, reduced interpretation times and, ultimately, higher accuracy. Triage algorithms that can re-order radiological reading queues have been developed, using classification to prioritize exams with suspected critical findings. Localization models move a step further to capture more granular information such as the location and, in some cases, size and subtype, of intracranial hematomas that could aid in neurosurgical management decisions. In addition to the potential to improve the clinical management of TBI patients, the use of algorithms for the interpretation of medical images may play a transformative role in enabling the integration of medical images into precision medicine. Acute TBI is one practical example that can illustrate the application of deep learning to medical imaging. This review provides an overview of computational approaches that have been proposed for the detection and characterization of acute TBI imaging abnormalities, including intracranial hemorrhage, skull fractures, intracranial mass effect, and stroke.

Keywords: traumatic brain injury, deep learning, artificial intelligence, precision medicine, evidence-based medicine, image recognition

INTRODUCTION

Acute traumatic brain injury (TBI), defined as sudden physical trauma that results in damage to the brain, is diagnosed through clinical assessment, with considerable reliance on structural neuroimaging studies such as computed tomography (CT) and occasionally, magnetic resonance imaging (MRI) (1). CT is the primary imaging modality worldwide for diagnosis of acute TBI due to its high sensitivity for acute intracranial hemorrhage and skull and facial fractures, rapid scan times, and lack of absolute contraindications (2, 3). While MRI is more sensitive to certain pathoanatomic features of acute TBI such as traumatic axonal injury and small cortical contusions, it is not routinely performed on acute TBI patients, but may be used as a follow-up study in patients with persistent or unexplained neurological deficits (4, 5). The rapid and accurate interpretation of these structural imaging studies in acute TBI is critical; imaging informs immediate clinical decisions such as hospitalization, intensive care unit admissions, and neurosurgical intervention, and also bears prognostic information. Structural imaging studies are interpreted by radiologists, but there has also been increasing interest in the development of image recognition algorithms that can aid, augment, and streamline the image interpretation process.

The use of algorithms for the detection of acute TBI pathoanatomic features has several key advantages. Automated algorithms could reduce the time to diagnosis and treatment and

help improve patient outcomes. They also have the potential to extract useful information from medical imaging studies in a quantitative and objective manner, which can greatly facilitate the integration of diagnostic medical imaging with advances in precision and evidence-based medicine. Radiological interpretations by human observers, in contrast, are subjective and vary based on differences in experience and judgment. In addition to this, algorithms are not affected by fatigue or "satisfaction of search," a very common problem in which detection of one radiographic abnormality causes a human observer to overlook an additional critical, but unexpected, abnormality (6). In order for algorithmic approaches to be of clinical relevance, however, they must demonstrate very high accuracy levels on large datasets that are representative of the population.

Early attempts to develop algorithms to interpret head CT exams used rule-based or traditional machine learning approaches. For the former, researchers would define a set of heuristics for the detection of pathoanatomic features (7–10). Despite a considerable amount of painstaking manual tuning through trial and error and promising preliminary results, however, it was not possible to curate a set of rules that could distinguish abnormal from normal images with accuracies in the range of those needed for clinical utility. Machine learning approaches, consisting of models that detected lesions by learning directly on a set of training examples, were also used. These approaches were attractive because they reduced the need for painstaking trial and error and pre-programming. Studies reported a variety of machine learning strategies for head CT interpretation, including random forest (11, 12), support vector machine (13), and decision tree (14–16). However, the rigid analytic forms of traditional machine learning models still made it difficult to accurately model complex imaging data. Ultimately, neither rule-based nor machine learning strategies were sufficient to achieve the strong performance levels needed for clinical use.

Just over 10 years ago, advances in computer hardware enabled a 100-fold acceleration of matrix computations that are fundamental to neural networks, a class of machine learning algorithms (17). Reduced data storage costs also made it easier to amass large datasets. These developments made it feasible, for the first time, to perform massive numbers of matrix manipulation tasks on enormous datasets in a reasonable amount of time, and led to a resurgence of interest in deep neural networks, or "deep learning." Unlike traditional machine learning models, which are fixed in analytic form, deep learning models are flexible, have potential for essentially limitless complexity, and can theoretically model any arbitrary mathematical function (18). In real-world applications, the required quantity of training data, which roughly scales with the depth of a deep learning model, does impose a practical limit on the depth and complexity of a deep learning model. Nevertheless, the increased representation power, elimination of the need for manual tuning, and more efficient large-scale learning have resulted in profound increases in accuracy in image recognition tasks. As a result, deep learning has rapidly dominated the computer vision field in the last decade, with advances also diffusing into the medical field. Examples of early demonstrations of deep learning applied to

medical imaging include detection of pneumonia on chest x-ray (19) and tuberculosis (20) and lymph node metastases (21) and non-small cell lung cancer (22) on histopathology slides.

Deep learning models have also been successfully developed for the analysis of neuroimaging studies in acute TBI. Numerous triage algorithms have been developed for the automated classification of imaging abnormalities, with the goal of future implementation into the radiological workflow to decrease time to diagnosis (23–26). Simultaneously, localization algorithms were also being developed and trained to automatically segment abnormality boundaries to extract granular information such as lesion size, subtype, number, and location (27–29). This review will provide an overview of computational approaches for the detection of TBI imaging abnormalities, including intracranial hemorrhage, fractures, mass effect, and stroke.
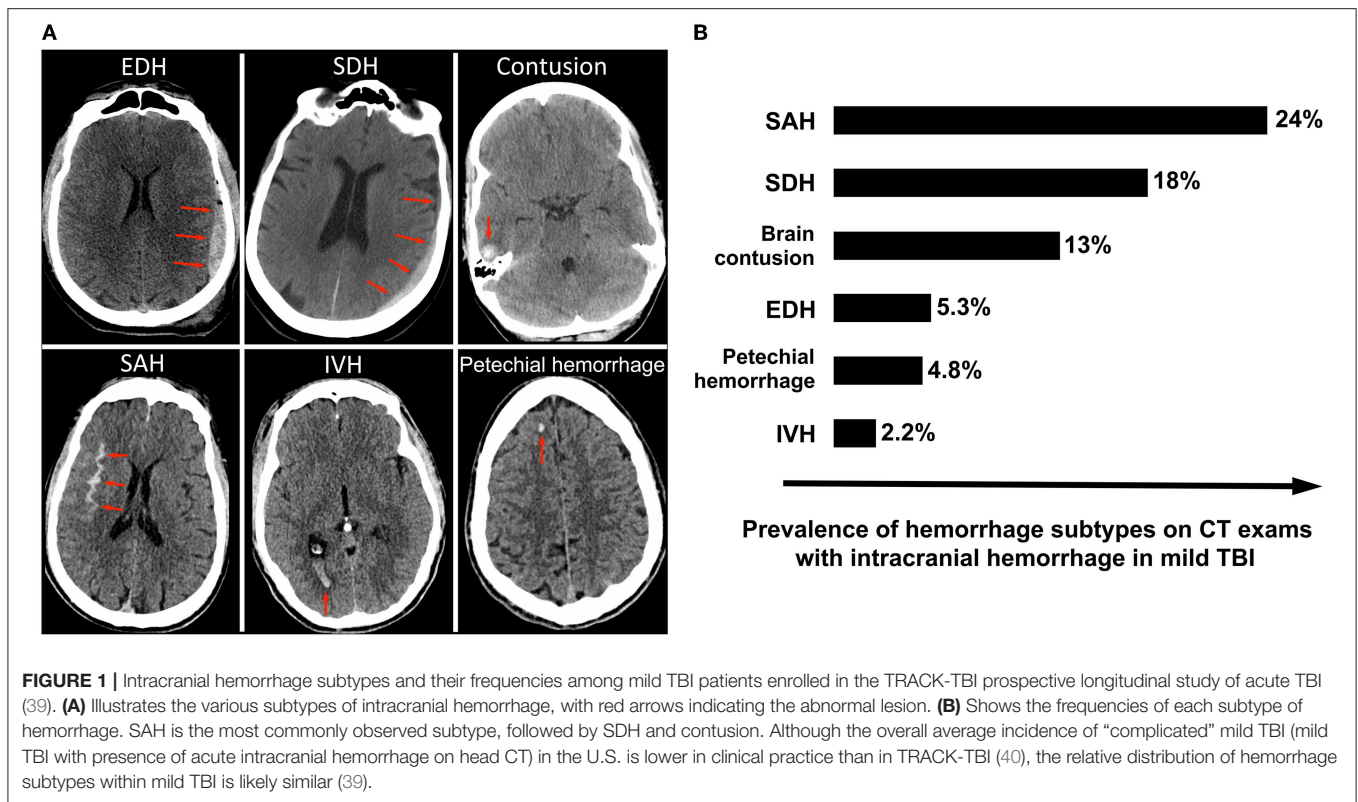
METHODS

A targeted literature review was performed using Google Scholar to identify publications related to computational approaches for image recognition in acute TBI. Search queries included combinations of the following keywords: "Traumatic brain injury," "Deep learning," "Machine learning," "Automated detection," "Head computed tomography," "Magnetic Resonance Imaging," "Intracranial hemorrhage," "Skull fracture," "Intracranial mass effect," "Edema," and "Stroke" without any restrictions on publication year. Articles that were most relevant to the review topic were selected, with an effort to identify and include all key articles with major contributions to the field.

TRAUMATIC BRAIN INJURY

Traumatic brain injuries can be classified either as penetrating, in which a foreign body traverses the skull and enters the intracranial space, or non-penetrating (also known as closed head injury or blunt TBI). TBI is divided into three clinical severity categories: mild, moderate and severe. The primary criterion that determines clinical severity category is the Glasgow Coma Scale (GCS) score. Patients presenting with GCS scores from 3 to 8, 9 to 12, and 13 to 15 are classified as having severe, moderate, and mild TBI respectively, with mild TBI accounting for ∼90% of acute TBI patients (30). In some grading schemes, mild TBI classification also requires that the patient exhibit no focal neurological deficit on initial examination (31, 32). Although clinical severity has long been based primarily on the GCS score, this coarse classification belies considerable heterogeneity in the underlying neuropathology and clinical outcomes of patients with identical GCS scores (33, 34). The current consensus is that development of more granular TBI classification schemes is a critical need in order to develop effective therapies for acute TBI (34).

Computed Tomography (CT)

With rare exception, head CT is the imaging modality used to evaluate patients with suspected acute TBI. A small proportion of patients in the Emergency Department (ED) are judged

FIGURE 1 | Intracranial hemorrhage subtypes and their frequencies among mild TBI patients enrolled in the TRACK-TBI prospective longitudinal study of acute TBI (39). **(A)** Illustrates the various subtypes of intracranial hemorrhage, with red arrows indicating the abnormal lesion. **(B)** Shows the frequencies of each subtype of hemorrhage. SAH is the most commonly observed subtype, followed by SDH and contusion. Although the overall average incidence of "complicated" mild TBI (mild TBI with presence of acute intracranial hemorrhage on head CT) in the U.S. is lower in clinical practice than in TRACK-TBI (40), the relative distribution of hemorrhage subtypes within mild TBI is likely similar (39).

to have sustained very mild head injuries based on a benign mechanism of injury, presenting GCS score of 15, and no loss of consciousness or posttraumatic amnesia, and may be discharged from the ED without head imaging. For all others, non-contrast head CT is performed (32). Over 20 million head CT scans are performed annually in the U.S. Head CT is the near universal choice for initial imaging in acute TBI in infants, children and adults, due to its very high sensitivity for acute intracranial hemorrhage and skull and facial fractures, widespread 24-h availability (35), and lack of any absolute contraindication. In addition, it's extremely short acquisition time, which allows it to perform whole-head imaging in as little as 0.3 s using modern multidetector-row CT scanners (36), is essential in many cases in which patients are unable to remain still due to altered mental status, pain, or young age.

Approximately 9% of head CT scans on acute TBI patients demonstrate acute intracranial hemorrhage (Korley, 2016). GCS score is a significant predictor of intracranial hemorrhage, identified on ∼9% of CT scans in mild TBI (37), 56% in moderate TBI, and 81% in severe TBI (38). Subtypes of acute intracranial hemorrhage include epidural hematoma (EDH), subdural hematoma (SDH), contusion, subarachnoid hemorrhage (SAH), intraventricular hemorrhage (IVH), and petechial hemorrhage (**Figure 1A**). The most common subtypes are SAH, SDH and brain contusion (**Figure 1B**). Although there is a correlation between the initial GCS score and the presence of intracranial hemorrhage, all subtypes of intracranial hemorrhage are observed at all clinical severity levels. For example, SAH is

the most common subtype of intracranial hemorrhage across all severity levels, although it is observed in 24.4% of mild TBI patients and 43% of moderate and severe TBI patients (38, 39). Skull fractures are also seen across all GCS scores, although are much more common in severe TBI [∼47% (41)] compared to mild TBI [∼3% (42)]. While midline shift and other types of brain herniation are also present in all severity categories, they are much more common in moderate and severe TBI, affecting ∼60% of these patients (43) compared to 3% of mild TBI patients (44). Lastly, while ischemic stroke does occasionally occur in acute TBI patients, it is rare and occurs in ∼2.5% of moderate and severe TBI (45). While a brief discussion of deep learning applications for stroke is included in this review, it is usually treated as a distinct disorder.

Brain MRI

Brain MRI, including basic structural (anatomic) brain MRI, is not currently recommended as the initial diagnostic imaging modality for acute TBI in adults (46, 47), children or infants (48). Brain MRI protocols for evaluation of acute TBI generally include T1-weighted, T2-weighted, T2-weighted Fluid Attenuated Inversion Recovery (FLAIR) sequences, diffusion-weighted imaging (DWI), and either susceptibility-weighted imaging (SWI) or T2*-weighted gradient echo (GRE) (49). MRI is highly sensitive to certain traumatic intracranial findings such as traumatic/diffuse axonal injury, small cortical contusions, and small extra-axial collections (50, 51), and is occasionally performed in hospitalized TBI patients as a problem solving tool

in cases when the level of consciousness is persistently impaired and not accounted for by findings on initial head CT. Brain MRI is also used to evaluate persistently symptomatic patients in the subacute or chronic stages after TBI, in medico-legal cases, and in cases of suspected abusive head trauma to assess for evidence of brain injuries of different ages (52).

The use of more advanced MRI techniques such as diffusion tensor imaging (DTI) and functional magnetic resonance imaging (fMRI) in acute TBI are highly promising techniques for more detailed and nuanced characterization of damage to the brain in TBI (49, 53). DTI in particular has been shown in numerous studies to demonstrate significant group differences in physical characteristics of the white matter tracts between acute TBI and control patients (54–56) that are also correlated with patient outcome. While most studies to date of DTI for acute or post-acute TBI have demonstrated group differences between TBI and control patients, there is not yet a consensus on how these techniques should be used clinically for diagnosis of individual acute TBI patients. Unlike CT and "basic" structural brain MRI, DTI, and fMRI data are not directly evaluated through human visual interpretation, but are inherently quantitative studies that undergo extensive post-processing using traditional statistical and machine learning techniques (57), and are not currently recommended for clinical use (46, 47). Interpretation of these studies is largely based on what the data show relative to "statistical significance" threshold values, rather than on subjective human visual inspection. Although machine learning, including deep learning, will likely have an increasing role in the analysis of DTI and fMRI data as these techniques continue to mature, these topics are beyond the scope of the current focused review.

OVERVIEW OF ALGORITHMS FOR COMPUTER ANALYSIS OF MEDICAL IMAGING

Rule-Based Algorithms

The term "computer-aided diagnosis" (CAD) has existed in the scientific literature for ∼50 years (58). For nearly all of that time, researchers have proposed heuristic, or rule-based, approaches for the analysis of neuroimaging studies in acute TBI. Rule-based approaches for image analysis use a set of "hand-crafted" rules created by a human observer based on morphology, brightness, and other characteristics of an image or imaging feature that can be perceived by the human visual system. Examples of such rule-based approaches, e.g., for acute intracranial hemorrhage detection on head computed tomography (CT), include algorithms based on top-hat transformation and left-right asymmetry (7); thresholding and connectivity (8); thresholding, linear and circular Hough transformations, and cluster analysis (9), and fuzzy c-means clustering (FCM) and region-based active contour (10). Though these and other rule-based approaches demonstrated the feasibility of using CAD to extract useful information from head CT images, they generally fell far short of accuracy levels that would be needed to make these approaches clinically useful.
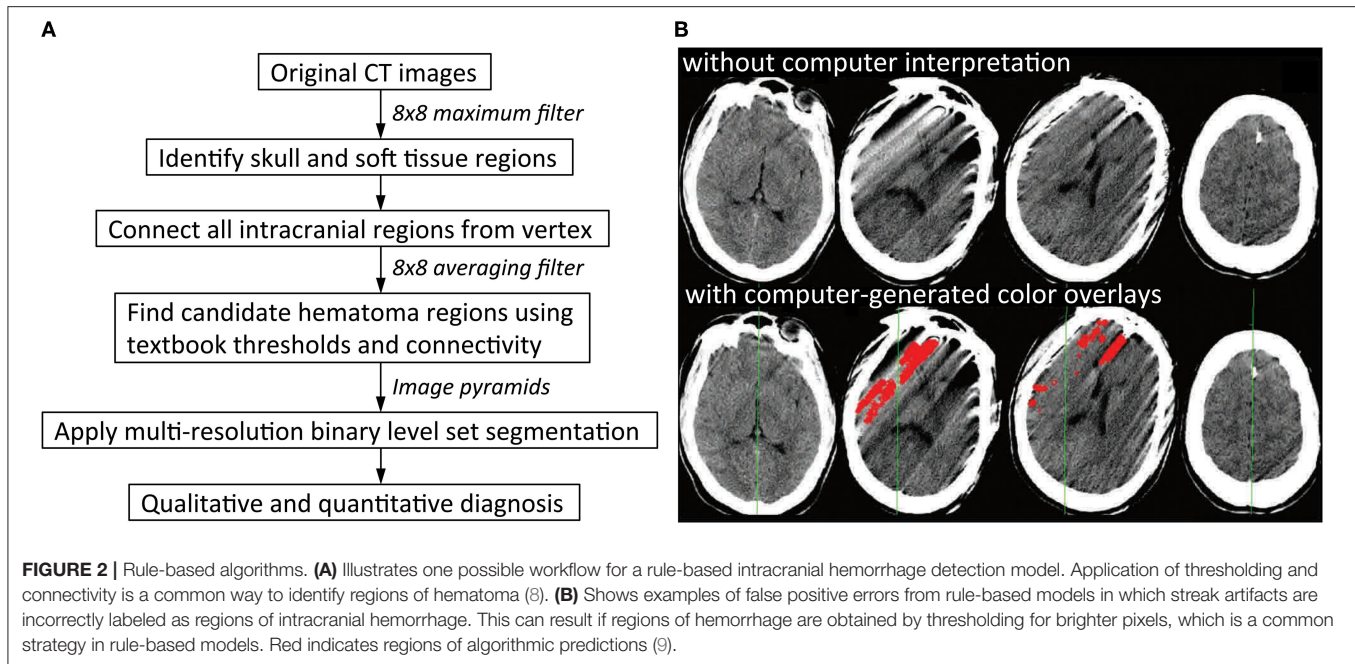
Reported performances either remained far below the threshold needed for clinical utility or were tested on CT exams that were not randomly selected.

Rule-based strategies are limited because it is challenging to enumerate all the rules that would fully describe the diversity in appearance of normal anatomy as well as pathological imaging findings seen in clinical practice. This problem is compounded by a large number of potential technical artifacts that are easily recognized by human observers, but have highly variable imaging appearances that are difficult to codify reliably. For instance, thresholding for brighter areas is a common rule-based strategy to identity acute intracranial hemorrhage, because acute hemorrhage is characterized by CT densities from 50 to 100 Hounsfield units and present as hyperdense, or "bright," regions on head CT images (**Figure 2A**). However, **Figure 2B** shows an example in which a rule-based algorithm employing thresholding, cluster analysis, and morphological analysis incorrectly labels streak artifacts as acute ICH due to its brightness.

Traditional Machine Learning Algorithms

Machine learning, a subset of artificial intelligence, seeks to minimize the reliance on heuristics and manual pre-programming and make predictions based on learned experience instead. Classical machine learning algorithms are broadly divided into supervised and unsupervised models. In supervised learning, the model receives as input both the data and the desired output, and "learns" from training data such that its predictions match the desired output as closely as possible. The use of labeled training data (with labeling performed by human experts and/or other computer algorithms) often results in stronger performances, making supervised learning the paradigm of choice for most practical applications including medical imaging. In contrast, unsupervised learning models receive input data with no labels and learn by identifying the underlying structural patterns or groupings in the data itself. Depending on the algorithm choice, model training is an iterative process in which the algorithm periodically updates its own parameters to better approximate the desired function and minimize the error in its output predictions. Numerous machine learning algorithms, including logistic regression (12), random forest (11, 12, 59), Bayesian decision theory (60), k-means clustering (61), and support vector machines (13, 62), have been used for the detection and segmentation of acute TBI imaging features. Each of these is briefly described below.

Logistic regression produces a binary output based on independent predictors (63). The random forest algorithm is an ensemble learning strategy that leverages multiple decision trees to make a classification prediction (64). Bayesian decision theory quantifies the tradeoffs between multiple choices in a probabilistic manner (65, 66). K-means clustering divides data into k groupings, where each data point is associated with the most similar group (67). The support vector machine maps input data onto a higher-dimensional grid and identifies the hyperplane that best differentiates each class, and can be more effective on complex and higher-dimensional data (68). Even though machine learning techniques such as these have considerable

FIGURE 2 | Rule-based algorithms. **(A)** Illustrates one possible workflow for a rule-based intracranial hemorrhage detection model. Application of thresholding and connectivity is a common way to identify regions of hematoma (8). **(B)** Shows examples of false positive errors from rule-based models in which streak artifacts are incorrectly labeled as regions of intracranial hemorrhage. This can result if regions of hemorrhage are obtained by thresholding for brighter pixels, which is a common strategy in rule-based models. Red indicates regions of algorithmic predictions (9).

overlap with statistical modeling, they often differ in objectives. The goal in machine learning is usually to make accurate predictions, while the goal in statistics is often to uncover significant relationships between variables that allows broad interpretation of the data (69).

While traditional machine learning algorithms are more adaptive than rule-based methods, there remain limitations. The use of traditional machine learning for medical imaging tasks still generally requires explicit feature extraction, usually as one or more early steps in the algorithm. Such features include but are not limited to voxel CT density thresholding (12), properties of surrounding voxels (11), and shape features (13), with predictions often made based on these extracted lesion predictors rather than the input image itself. This means that traditional machine learning models are rarely applied end-to-end on raw input image pixels, which could also make it more difficult to scale the algorithm to large complex datasets. Lastly, they are limited in the functions they are able to model. Logistic regression, for example, is optimal when the data can be linearly separated. However, medical imaging data is much more complex and often cannot be modeled in a linear fashion. As a result, many of these traditional machine learning techniques are more frequently used and better suited for applications such as TBI outcome prediction (70, 71).
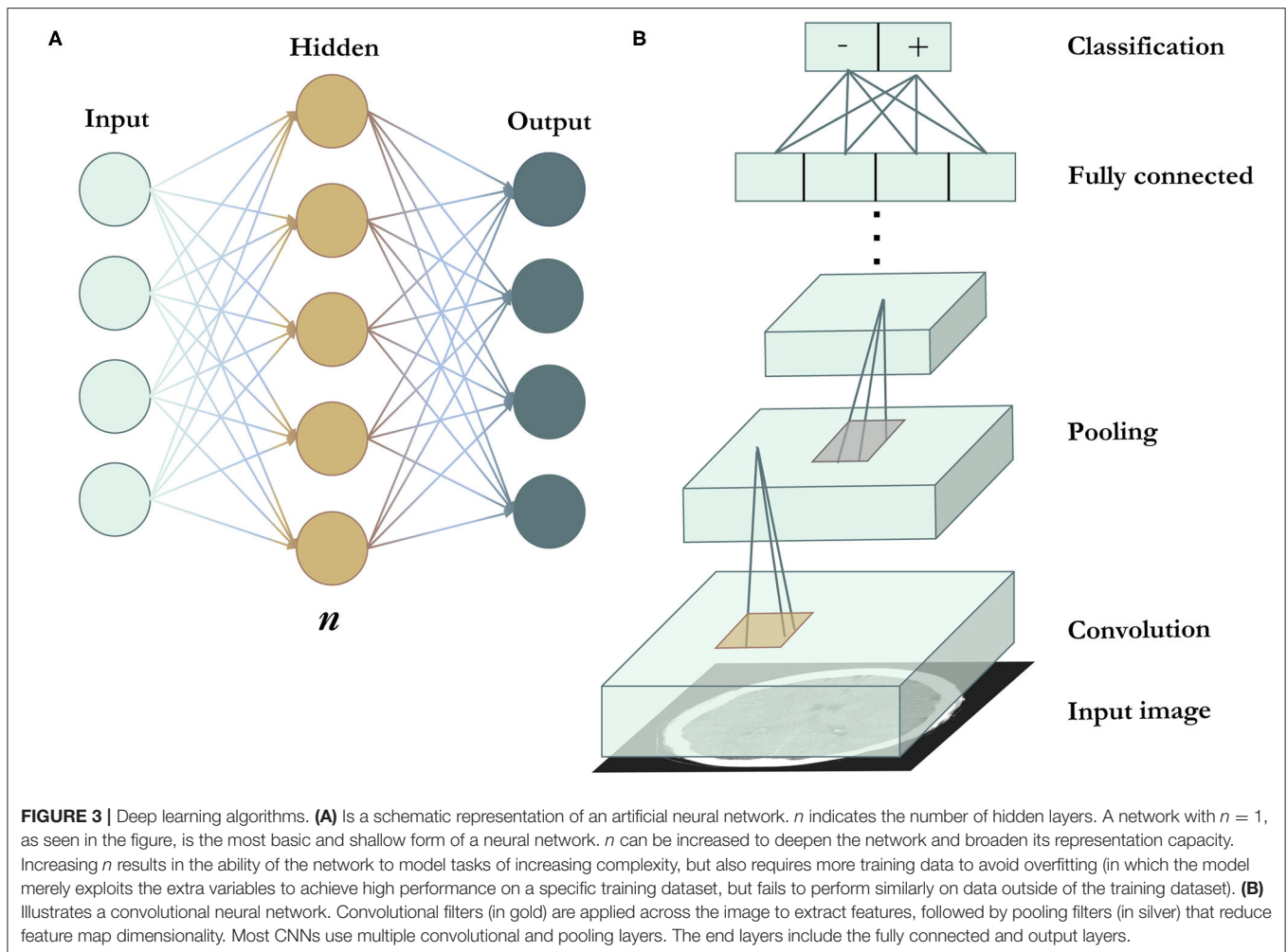
Deep Learning Algorithms

The feasibility of a large-scale data-driven approach for image classification is exemplified by deep learning (72). Deep learning circumvents many of the limitations of rule-based and traditional machine learning algorithms, including heavy reliance on painstaking handcrafted rules and feature selection. It also allows "learning" of image patterns directly from large volumes of data, making it a scalable method to more efficiently generalize to external datasets. This has led to substantial boosts in accuracy

of machine learning algorithms for image recognition tasks, and a paradigm shift in the field of computer vision. Deep learning underpins all current state-of-the-art algorithms for widely-benchmarked computer vision tasks such as image classification and object detection (73–77). While deep learning originated from the concept of the artificial neural network (ANN), it has since progressed to much deeper and more advanced networks including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and, more recently, transformers. This section will provide an overview of the major deep learning algorithms and how they have been applied to medical imaging.

The earliest artificial neural networks (ANNs) were inspired by the transmission of information between the neurons in the human brain (78–81). The ANN includes an input layer, an output layer, and at least one "hidden" layer, with each layer designed as an array of neuron-like nodes (**Figure 3A**) (82, 83). The input layer accepts data, and the output layer produces a prediction. Between the input and output layers are the hidden layers, each of which accepts the output from the previous layer's nodes and performs forward computations to produce its own output. Although ANNs were highly novel at the time, they have limited representation power. This can be solved in part by increasing the numbers of layers and deepening the neural network, which increases the power of the network to "learn" subtle differences in complex patterns, because deeper networks have larger numbers of values that can be adjusted in order to model complex data. However, the simple network architecture still limited the performance of basic neural networks on more complex tasks such as visual recognition.

This problem was addressed by the development of the convolutional neural network (CNN). In 2012, a CNN called AlexNet was the winner of the ImageNet Large Scale Visual Recognition Competition (ILSVRC) (74). As one of the earliest CNNs, AlexNet used convolutional layers to detect local features,

FIGURE 3 | Deep learning algorithms. **(A)** Is a schematic representation of an artificial neural network. *n* indicates the number of hidden layers. A network with $n = 1$, as seen in the figure, is the most basic and shallow form of a neural network. *n* can be increased to deepen the network and broaden its representation capacity. Increasing *n* results in the ability of the network to model tasks of increasing complexity, but also requires more training data to avoid overfitting (in which the model merely exploits the extra variables to achieve high performance on a specific training dataset, but fails to perform similarly on data outside of the training dataset). **(B)** Illustrates a convolutional neural network. Convolutional filters (in gold) are applied across the image to extract features, followed by pooling filters (in silver) that reduce feature map dimensionality. Most CNNs use multiple convolutional and pooling layers. The end layers include the fully connected and output layers.
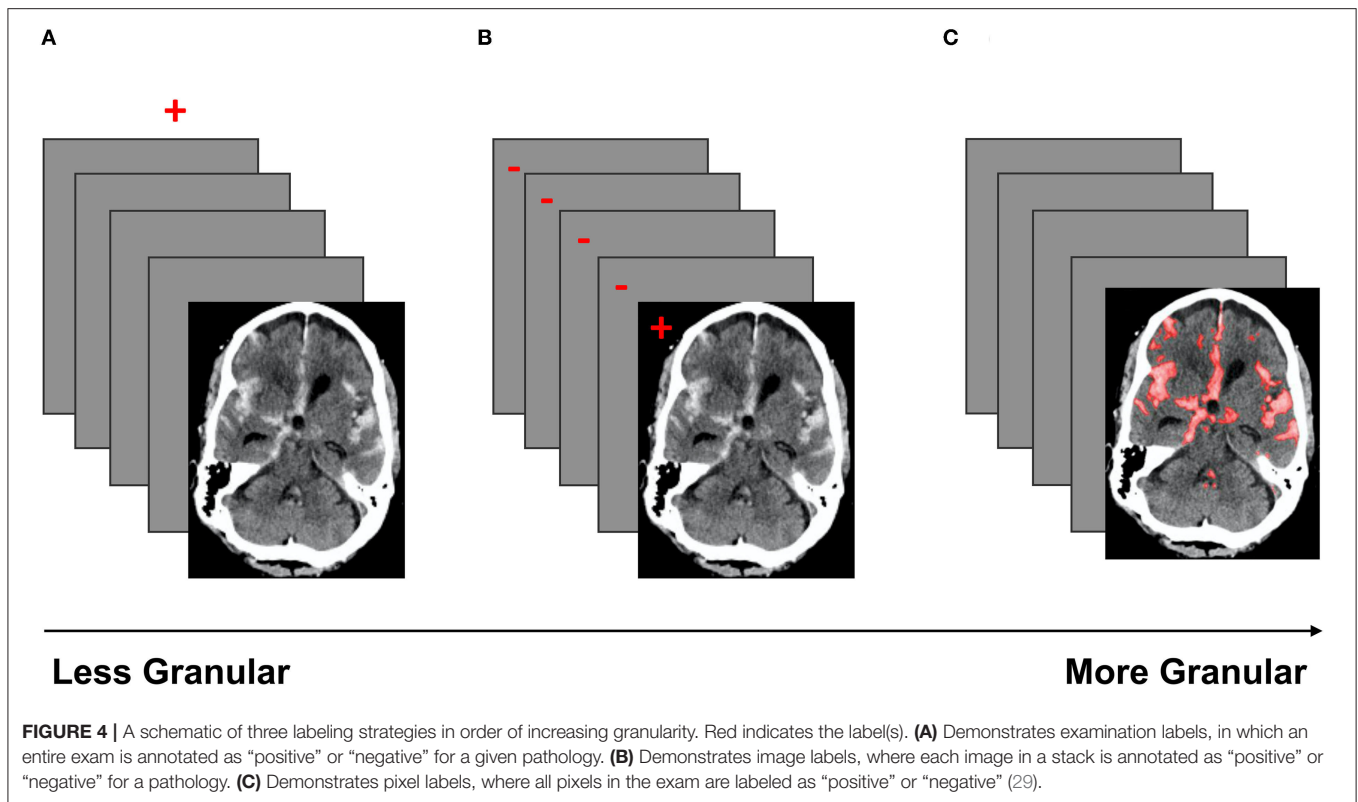
with early layers detecting simple features such as edges, corners, and colors, and higher layers detecting higher-level features such as objects (84). Filters that perform the convolution operation are applied across the input image, producing feature maps that preserve spatial invariance (**Figure 3B**). CNNs also apply pooling layers to reduce the dimensionality of feature maps and increase computational efficiency while retaining important image features. The CNN architecture continued to improve significantly in the subsequent years, with increased network depths and added features such as multi-scale filters (85), factorized convolutions (86), and residual connections (87). Coupled with greater computational efficiency, the CNN has since become the standard network architecture of choice for most visual tasks and has been applied to medical image analysis, including the detection (23, 25, 26) and localization of intracranial hemorrhage (27–29).

Aside from the CNN, other deep learning models have been applied to sequential data such as language, audio, and video tasks. Algorithms such as RNNs (88) and long short-term memory (LSTM) (89) augment a model with sequential memory to model temporal relationships, but often struggle to

retain information when encountered with longer sequences. Transformers were recently proposed to address this limitation; they have shown great success in modeling sequential data by capturing long-range pairwise relationships and have since replaced RNNs in most practical applications (90–92). Because RNNs are best suited for sequential data, they are more commonly used for the analysis of electronic medical records and clinical reports (93, 94) rather than medical images. However, there has been recent interest in the use of transformers in medical imaging (95–97). This area is still relatively new and unexplored, but may be a promising direction in coming years.

Overall, deep learning models have benefited tremendously from recent breakthroughs in computational resources (98–100), data volume (101–103), and algorithmic advancements (74, 87, 90, 104, 105). The large number of free parameters associated with hidden layers has enabled a vastly more flexible algorithm that can model complex data and excel over rule-based and traditional machine learning approaches for image recognition tasks. Coupled with greater data availability, deep learning has facilitated the development of acute TBI imaging algorithms that are more accurate than previously seen.
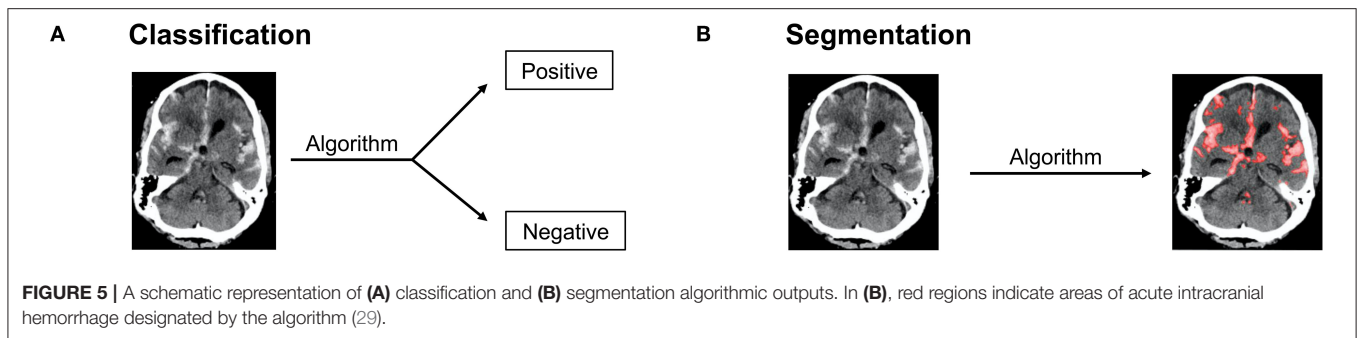
FIGURE 4 | A schematic of three labeling strategies in order of increasing granularity. Red indicates the label(s). **(A)** Demonstrates examination labels, in which an entire exam is annotated as "positive" or "negative" for a given pathology. **(B)** Demonstrates image labels, where each image in a stack is annotated as "positive" or "negative" for a pathology. **(C)** Demonstrates pixel labels, where all pixels in the exam are labeled as "positive" or "negative" (29).

DATA LABELING STRATEGIES

Deep learning algorithms learn by gleaning patterns from large volumes of data, but they require high quality labeled data to do so. The granularity of the labels determines the maximum granularity, or level of detail, of the model's output prediction, and also affects the final accuracy of the model for a given task. Therefore, the labeling strategy partly determines the clinical applications the model may be useful for. Because deep learning models' accuracy is critically dependent on the quality and type of data they learn from, it may be useful to discuss and compare common data labeling strategies. We discuss below several common data labeling strategies, in order of increasing labeling granularity.

The coarsest data labeling strategy in medical imaging is examination-level labeling, which refers to categorization of each imaging exam based on the presence or absence of a specific type of imaging finding (**Figure 4A**). For example, a head CT exam typically contains a stack of 25–70 images, depending on the size of the head and the specific imaging protocol. If a CT exam has at least one instance of intracranial hemorrhage on at least one image, the entire exam would be annotated as "positive" for intracranial hemorrhage. The use of examination-level labels has several advantages, the primary being that these labels can generally be obtained with less time and cost, making it possible to amass larger datasets. In addition, examination-level labels can potentially be extracted from pre-existing clinical radiological reports by human reviewers without

specialized training in radiology, or through the use of automated natural language processing algorithms. Examination-level labels represent the coarsest level of labeling and generally do not provide information on the location or size of abnormal findings.

Image-level annotations are an example of labels with an intermediate level of granularity. In this strategy, each image of an exam is annotated for the presence or absence of a specific type of imaging finding (**Figure 4B**). Unlike examination-level labels, image-level annotations provide some localization information, identifying the image(s) that contain instances of the finding of interest. Image-level labels provide significantly more information than examination-level labels, and can be used to improve the accuracy of exam-level predictions, as well as to provide coarse estimates of the locations and sizes of instances of the imaging finding of interest. However, image-level labels require significantly more cost and effort to obtain. Unlike examination labels, which are already present in clinical reports, image labels do not exist in the clinical setting and must be manually annotated by highly trained specialists.

Pixel-level annotation is the most granular labeling approach, and consists of the designation of each pixel in each image of the exam as positive or negative for the finding of interest. Pixel-level labels provide the densest information content (**Figure 4C**), and can be used to further improve the accuracy of exam-level predictions, as well as provide fine-grained localization information. Algorithms trained on high-quality pixel-level annotations have the potential to produce the strongest results because they are trained on the data with the

FIGURE 5 | A schematic representation of **(A)** classification and **(B)** segmentation algorithmic outputs. In **(B)**, red regions indicate areas of acute intracranial hemorrhage designated by the algorithm (29).

densest information content. The primary disadvantage of this strategy is that pixel labels are costly and time-consuming to obtain, requiring manual delineation ("segmentation") of the boundaries of each lesion or finding of interest. This is not a task that radiologists are accustomed to performing in a clinical setting, and is highly labor-intensive even with the use of optimized manual segmentation tools. Unlike examination-labeled datasets, the expansion of an existing pixel-labeled dataset may be prohibitively costly due to the labeling requirement. Regardless, algorithms trained on pixel labeled data can create stunning delineations of the boundaries of lesions, producing detailed location and size information that can be useful for clinical management decisions.

Bounding-box labels constitute a labeling approach that is intermediate between image-level and pixel-level labels. For example, Mask R-CNN (87) first detects an object and indicates its location by placing a bounding box that encloses all or part of the object, and subsequently performs a pixel-wise segmentation within the bounding box. However, hemorrhage is fluid and can take on nearly limitless morphologies, and is not well-suited to this approach. Mask R-CNN was designed primarily for the detection of discrete objects, and may be better suited to medical imaging applications that require identification of discrete pathological findings such as mass lesions.

These approaches illustrate several major strategies that have been used to label training data for deep learning algorithms for TBI image analysis. These concepts will appear again in later discussions of specific algorithms.

INTRACRANIAL HEMORRHAGE DETECTION

Acute intracranial hemorrhage, which refers to acute bleeding within the confines of the cranial vault, is a key neuroimaging finding that determines the disposition of acute TBI patients from the ED. It is the only neuroimaging finding accepted by the U.S. FDA as a prognostic marker in acute TBI. Because neurosurgical interventions such as intracranial pressure monitor placement and craniotomy, including decompressive hemicraniectomy, must be performed emergently, algorithms that can reduce the time to diagnosis through the rapid detection and localization of acute intracranial hemorrhage could improve

patient outcomes (106). A variety of deep learning approaches have been explored for intracranial hemorrhage detection. Some algorithms focus strictly on the classification of CT exams as positive or negative for intracranial hemorrhage, because this differentiation is a key determinant of immediate management steps (**Figure 5A**). A number of models that also attempt to predict different hemorrhagic subtypes have been reported. To further increase clinical usefulness and interpretability, algorithms may also perform segmentation, or detailed delineation of the hemorrhage boundaries (**Figure 5B**). Segmentation has the great advantage of allowing clinicians the opportunity to visually appreciate the algorithm's findings and to have an understanding of the extent/volume and location of the hemorrhage. Both segmentation and determination of the subtype of intracranial hemorrhage (intraventricular, intraparenchymal, subarachnoid, epidural, subdural, and petechial hemorrhage) are also critical for guiding immediate management decisions.

Table 1 summarizes prior algorithmic approaches for the detection and segmentation of acute intracranial hemorrhage in TBI. Although there is no formal distinction in the literature, we group the algorithms into two broad classes for illustrative purposes: triage algorithms and localization algorithms. Hereafter, approaches to image analysis in acute TBI will be described within the context of these two broad clinical objectives. Dataset sizes are described either in terms of number of CT exams or number of CT images. It is important to note that results from different papers cannot be compared head-to-head due to the use of different datasets and labeling strategies.

The publications in **Table 1** are related to head CT, the standard neuroimaging study in the clinical management of acute TBI. In current clinical practice, brain MRI is only occasionally performed secondarily, as a problem-solving tool when neurological deficits persist throughout the hospital stay and/or are not adequately explained by head CT findings. While deep learning has been applied to brain MRI for other pathologies and disorders (121–123), there has been little development of such algorithms for MRI image recognition in acute TBI due to the relatively uncommon use of brain MRI in this clinical setting. MRI's current limited role in acute TBI also limits the quantity of available training data. Despite this, a small number of MRI hemorrhage detection algorithms have been proposed and are briefly discussed below for completeness (124–127).

TABLE 1 | Deep learning approaches for intracranial hemorrhage detection and segmentation on head CT.

Publication	Model description	Training label type	Output type	Dataset size	Model performance
Phong et al. (107)	Three models used: LeNet, GoogLeNet, Inception-ResNet	Examination labels	Hemorrhage detection	Train: 1,360 images Test: 340 images	LeNet accuracy: 0.997 GoogLeNet accuracy: 0.982 Inception-ResNet accuracy: 0.992
Prevedello et al. (23)	CNN	Examination labels	Hemorrhage, mass effect, hydrocephalus detection	Development: 246 exams Test: 130 exams	Sensitivity: 0.90 Specificity: 0.85 AUC: 0.91
Patel and Manniesing (108)	Convolutional neural network (CNN)	Image labels	Hemorrhage detection	Development: 150 exams Validation: 20 exams Test: 20 exams	Sensitivity: 0.87 Specificity: 0.97 Accuracy: 0.95
Majumdar et al. (109)	CNN	Pixel labels for hemorrhage subtypes	Hemorrhage detection	Development: 60 exams Validation: 5 exams Test: 69 exams	Sensitivity: 0.81 Specificity: 0.98
Grewal et al. (110)	Recurrent Attention DenseNet (RADnet)	Pixel labels for presence of hemorrhage	Hemorrhage detection	Development: 185 exams Validation: 67 exams Test: 77 exams	Accuracy: 0.818 Recall: 0.886 Precision: 0.813
Jnawali et al. (111)	3D-CNN	Examination labels	Hemorrhage detection	Development: 34,848 exams Validation: 2,000 exams Test: 3,509 exams	AUC: 0.87
Titano et al. (24)	3D-CNN modeled after ResNet-50	Examination labels	Urgent or non-urgent classification	Development: 37,236 exams Test: 180 images	AUC: 0.73 Accuracy: 0.56
Arbabshirani et al. (25)	CNN	Examination labels	"Routine" vs. "stat" classification	Development: 37,074 exams Test: 9,499 exams	AUC: 0.846
Chilamkurthy et al. (26)	CNN	Image labels	Detection of hemorrhage and subtypes	Development: 290,066 exams Validation: 21,095 exams Test: 491 exams	Overall hemorrhage AUC: 0.94 Intraparenchymal AUC: 0.95 Intraventricular AUC: 0.93 SDH AUC: 0.95 Extradural AUC: 0.97 SAH AUC: 0.96
Chang et al. (27)	Hybrid 3D/2D CNN	Pixel labels	Hemorrhage detection and segmentation	Development: 10,159 exams Test: 862 exams	Hemorrhage detection accuracy: 0.975 Hemorrhage detection AUC: 0.983 Hemorrhage detection sensitivity: 0.971 Hemorrhage detection specificity: 0.975
Ye et al. (112)	3D joint convolution and recurrent network (CNN-RNN)	Image labels	Detection of hemorrhage and subtypes	Hemorrhage development: 2,255 exams Hemorrhage validation: 282 exams Hemorrhage test: 299 cases Subtype development: 1,461 exams Subtype validation: 181 cases Subtype test: 194 cases	Hemorrhage classification: >= 0.98 across all metrics Subtype AUC: > 0.8 across all subtypes
Cho et al. (113)	Cascaded deep learning model	Pixel labels	Hemorrhage detection and segmentation	5.702 exams	Hemorrhage detection sensitivity: 0.979 Hemorrhage detection specificity: 0.988 Segmentation precision: 0.802 Segmentation recall: 0.822
Lee et al. (28)	ImageNet pretrained deep convolutional neural networks (DCNN)	Image labels	Classification of hemorrhage and subtypes; heatmap localization	Development: 704 exams Retrospective test: 200 exams Prospective test: 196 exams	Retrospective sensitivity: 0.98 Retrospective specificity: 0.95 Prospective sensitivity: 0.92 Prospective specificity: 0.95
Kuo et al. (29)	Patch-based fully convolutional neural network (PatchFCN)	Pixel labels	Classification of hemorrhage and subtypes; segmentation	Development: 4,396 exams Test: 200 exams	Hemorrhage classification AUC: 0.991 SDH AUC: 0.954 EDH AUC: 0.940 Contusion/ICH/TAI AUC: 0.934 SAH/IVH: 0.956 Dice coefficient: 0.75

(Continued)

TABLE 1 | Continued

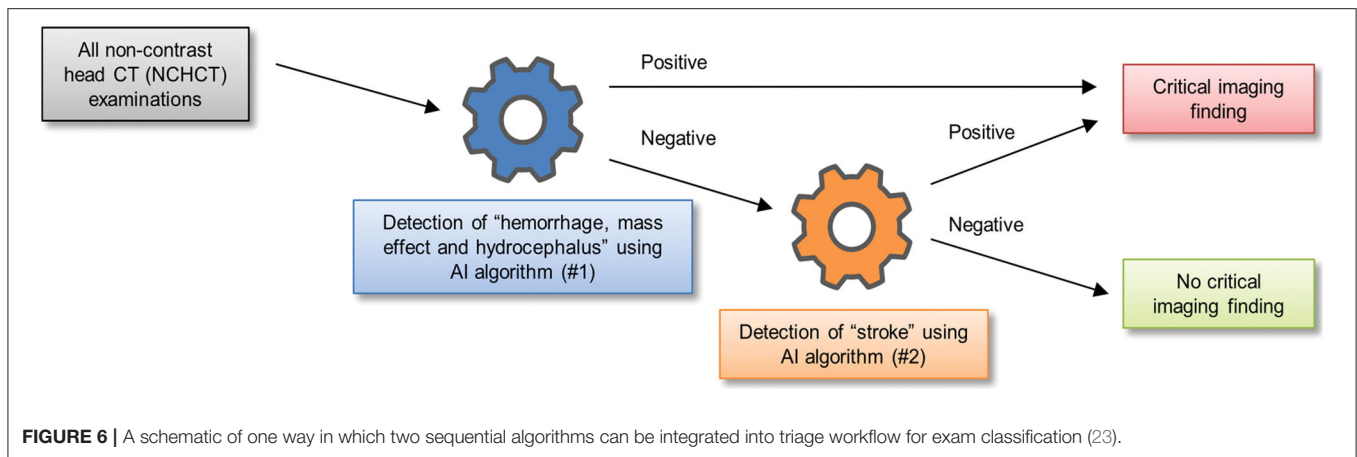| Publication	Model description	Training label type	Output type	Dataset size	Model performance
Lee et al. (114)	"Kim-Monte Carlo algorithm," an artificial neural network	Image labels with hemorrhage subtype	Hemorrhage detection and subtype classification	Training: 166 exams Validation: 84 exams	Overall AUC: 0.859 SDH/EDH accuracy: 0.534 SAH accuracy: 0.917 IPH/IVH accuracy: 0.60
Burduja et al. (115)	CNN-LSTM	Image labels with hemorrhage subtype	Hemorrhage detection and subtype classification	Training: 21,000 exams Validation: 744 exams Testing: 3,528 exams	Overall AUC: 0.9792 EDH AUC: 0.9414 IPH AUC: 0.9834 IVH AUC: 0.9866 SAH AUC: 0.9609 SDH AUC: 0.9451
Arab et al. (116)	Convolutional neural network with deep supervision (CNN-DS)	Pixel labels	Classification of hemorrhage; segmentation; volume quantification	Development: 45 exams Test: 10 exams	Dice coefficient: 0.84 Recall: 0.83
Sharrock et al. (117)	Convolutional neural networks with VNet framework	Pixel labels	Classification of hemorrhage; segmentation; volume quantification	Training: 100 exams Validation: 500 exams Testing: 12 exams	Mean dice coefficient: 0.911 Mean IVH dice coefficient: 0.911 Mean SDH dice coefficient: 0.879 Algorithm-human volume correlation: 0.98
Dhar et al. (118)	U-Net	Pixel labels	Classification of hemorrhage; segmentation; volume quantification	Training and cross-validation: 224 exams Testing: 84 exams	Dice coefficient: 0.90
Monteiro et al. (119)	CNN	Pixel labels	Classification of hemorrhage; segmentation; volume quantification	Training: 184 scans Testing: 655 exams	ICH external AUC: 0.83 IPH external AUC: 0.90 EAH external AUC: 0.80 IVH external AUC: 0.95 Overall mean dice coefficient: 0.36
Zhao et al. (120)	nnU-Net	Pixel labels	Classification of hemorrhage; segmentation; volume quantification	Training: 300 exams Validation: 80 exams	ICH dice coefficient: 0.92 IVH dice coefficient: 0.79

Triage Algorithms

Triage algorithms for analysis of clinical radiological exams are intended to expedite the interpretation of abnormal exams by bringing these to prompt attention by on-duty radiologists. One uniting feature of triage algorithms is their emphasis on exam-level classification; determination of detailed features such as location and size of abnormal findings is generally outside of the scope of algorithms in this category.

The performance requirement for a minimum viable triage algorithm is that its predictions are more accurate than random guessing. For algorithms intended for a more central role, e.g., augmenting the performance (accuracy) of radiologists through human/computer collaboration, the performance bar would likely be much higher. Still higher is the performance level that would be required for algorithms intended for use by non-radiologist clinicians for management decisions for acute TBI patients. It is important to note that these various performance bars are not defined by fixed metrics; rather, they are subjective in nature and depend on the intended "context of use" (128). Context of use is a key concept in FDA regulation of medical devices and other products, essentially providing a complete description of the intended clinical setting and manner of use of a medical product. In the case of algorithms for intracranial hemorrhage detection, context of use could include factors such as the expertise level of clinicians available to interpret acute head CT exams in a particular setting

(e.g., emergency department physicians, neurosurgeons, general radiologists, or neuroradiologists), the typical turnaround time for CT interpretation (minutes to an hour), and how much the clinical management algorithm in a particular practice (e.g., the decision to admit or discharge) relies on head CT results.

Prior work has addressed the task of intracranial hemorrhage classification using a variety of technical approaches. Phong et al. (107) evaluated three different popular neural network architectures—LeNet (129), GoogLeNet (85), and Inception-ResNet (130)—for intracranial hemorrhage classification. All three models were widely used around the time of publication and had previously achieved state-of-the-art performance on benchmarked computer vision tasks. They were trained using exam-level labels, and achieved very strong performance on the intracranial hemorrhage classification task. However, the authors manually selected positive images to include in the dataset, and it is unclear whether negative images were also included. Patel and Manniesing (108) developed a convolutional neural network for intracranial hemorrhage classification using image-level labels and also achieved strong performance. Both Majumdar et al. (109) and Grewal et al. (110) proposed intracranial hemorrhage classification models using pixel level-labeling.

These studies collectively explored a variety of labeling strategies, from examination- to pixel-level labels. These papers primarily focused on the technical side of intracranial hemorrhage detection, exploring the effectiveness of various

FIGURE 6 | A schematic of one way in which two sequential algorithms can be integrated into triage workflow for exam classification (23).

neural network architectures for hemorrhage detection and experimenting with technical hyperparameters. Each of these strategies demonstrated reasonable performances at minimum, with some even achieving accuracy metrics exceeding 0.90 (107, 108). However, a likely limitation is that each of these studies used a relatively small number of training and test cases (**Table 1**). Small training sets often lack adequate examples of the entire spectrum of imaging appearances of the pathology of interest, and the resulting models often fail to generalize to other data. In addition, performance results obtained on small test sets are often noisy.

In order to demonstrate clear clinical utility, it is important to demonstrate strong performance on larger datasets that are also representative of the intended population. Jnawali et al. (111) did this, collecting a large dataset of 40,367 head CTs to train and evaluate a three-dimensional convolutional neural network (3D-CNN) for intracranial hemorrhage classification. The model achieved an AUC of 0.87. Although these results were promising and obtained from experiments on large datasets, more work was needed to reach the performance bar needed for clinical use.

One of the earliest works to study clinical integration was a triage algorithm proposed by Prevedello et al. (23) to draw attention to critical head CT exams for expedited evaluation by a radiologist. They developed two sequential deep learning algorithms, with the first dedicated to the detection of intracranial hemorrhage, mass effect, and hydrocephalus, and the second dedicated to detection of acute infarct. Each exam was read by the two algorithms serially and marked critical if either algorithm detected abnormalities (**Figure 6**). The first algorithm was trained on cases annotated with one of six possible examination-level labels: hemorrhage, mass effect, hydrocephalus, suspected acute infarct, encephalomalacia, or non-urgent/normal. For the first algorithm, the first three labels were considered "positive" and the last three were considered "negative." During testing, the first algorithm classified cases as "positive" or "negative" for at least one of the first three imaging findings, demonstrating an AUC of 0.91. The second algorithm dedicated to acute infarct detection achieved an AUC of 0.81. Although hemorrhage, mass effect, and hydrocephalus were grouped for the purposes of algorithm evaluation, and

performance on each distinct pathology was unknown, the study was important because it was one of the earliest to report a deep learning classification pipeline that could identify acute CT exams that contained one or more of a diverse collection of abnormal findings.

Titano et al. (24) expanded upon this by proposing the integration of a deep learning system for radiological triage into a simulated clinical environment. These authors described an imaging triage system in which a 3-dimensional convolutional neural network (3D-CNN) re-ordered exams in the queue for radiological interpretation based on a much wider range of abnormal findings. The algorithm was designed to re-order exams so they would be reviewed on the basis of urgency rather than order of completion. The 3D-CNN was exposed to exams presenting with hundreds of head CT diagnoses based on the Universal Medical Language System (UMLS) concept universal identifiers, including intracranial hemorrhage, and each diagnosis was mapped to critical or non-critical categories depending on predetermined radiologist designations (**Figure 7A**). The 3D-CNN was trained on 37,236 head CT exams labeled at the exam level and produced exam-level classification outputs, demonstrating an AUC of 0.73 on 180 test images, when compared to the "gold-standard" labels of physician reviews of clinical reports. Exam re-prioritization resulted in a statistically significantly larger number of critical exams at the top of the queue, **Figures 7B,C** presents an example regarding how the algorithm identifies and reprioritizes exams in real time. This study was important because it demonstrated one of the earliest attempts to use a large dataset to learn a broad comprehensive range of pathologies. Because it involved a trial in a real-time simulated environment, it also allowed authors to quantify the speed of their algorithm in a hypothetical clinical setting and reach the conclusion that the algorithms were 150x faster than humans.

Expanding on this queue reordering study, Arbabshirani et al. (25) developed a deep convolutional neural network that was prospectively integrated into a radiological workflow for 3 months. The model's role was to re-prioritize exams from "routine" to "stat" if it detected an intracranial hemorrhage. Prior to integration, the model was trained on a large scale dataset of

FIGURE 7 | Incorporation of a reprioritization, or triage, algorithm into the radiological workflow in acute TBI (24). **(A)** Illustrates the broad range of pathological findings represented in the head CT training data, and how they were classified into non-urgent and urgent categories. **(B)** Shows the typical order in which critical (orange) and non-critical (gray) head CT exams would be interpreted by a radiologist before (left graph) and after (right graph) reprioritization by a deep learning algorithm. The gray and orange dots represent discrete CT exams, while the shaded regions are the smoothed exam frequency distributions. **(C)** Is a schematic representation of the algorithm's prioritization process.

37,084 exams with examination-level annotations and reported an AUC of 0.846 for intracranial hemorrhage classification. Following this initial development phase, the algorithm was integrated into the clinical workflow for 3 months. Over the course of that time period, the algorithm processed exams with an accuracy of 0.84. Exams prioritized as "stat" showed significantly reduced time to interpretation compared to "routine" exams, from a median time of 512–19 min. The Titano et al. and Arbabshirani et al. studies were two early papers that studied the impact of integration of deep learning triage algorithms into clinical settings.

Rather than perform broad critical and non-critical classifications of CT exams, Chilamkurthy et al. (26) focused instead on developing deep learning models that could identify specific head CT abnormalities, including intracranial hemorrhage subtypes (intraparenchymal, intraventricular, subdural, extradural, and subarachnoid), calvarial fractures,

midline shift, and intracranial mass effect (**Figure 8**). The authors collected large datasets (290,055 CT exams) for training. The training and validation exams were labeled at the image level. The exams in the test set, referred to as "CQ500," were each labeled by three highly experienced radiologists for the presence of intracranial hemorrhage and its subtypes, midline shifts, and skull fractures. Although exams were labeled at the image level, the model produced predictions at the examination level. The decision to leverage the more granular image training label likely helped achieve better exam-level predictions at test time.

This was the first study to evaluate model performance on individual pathologies, to use very large datasets (>100,000 CT exams), and to validate models on datasets representative of the patient population. Overall, the model performances across all pathologies were good, with the algorithm achieving an overall AUC of 0.94 for intracranial hemorrhage on CQ500. In
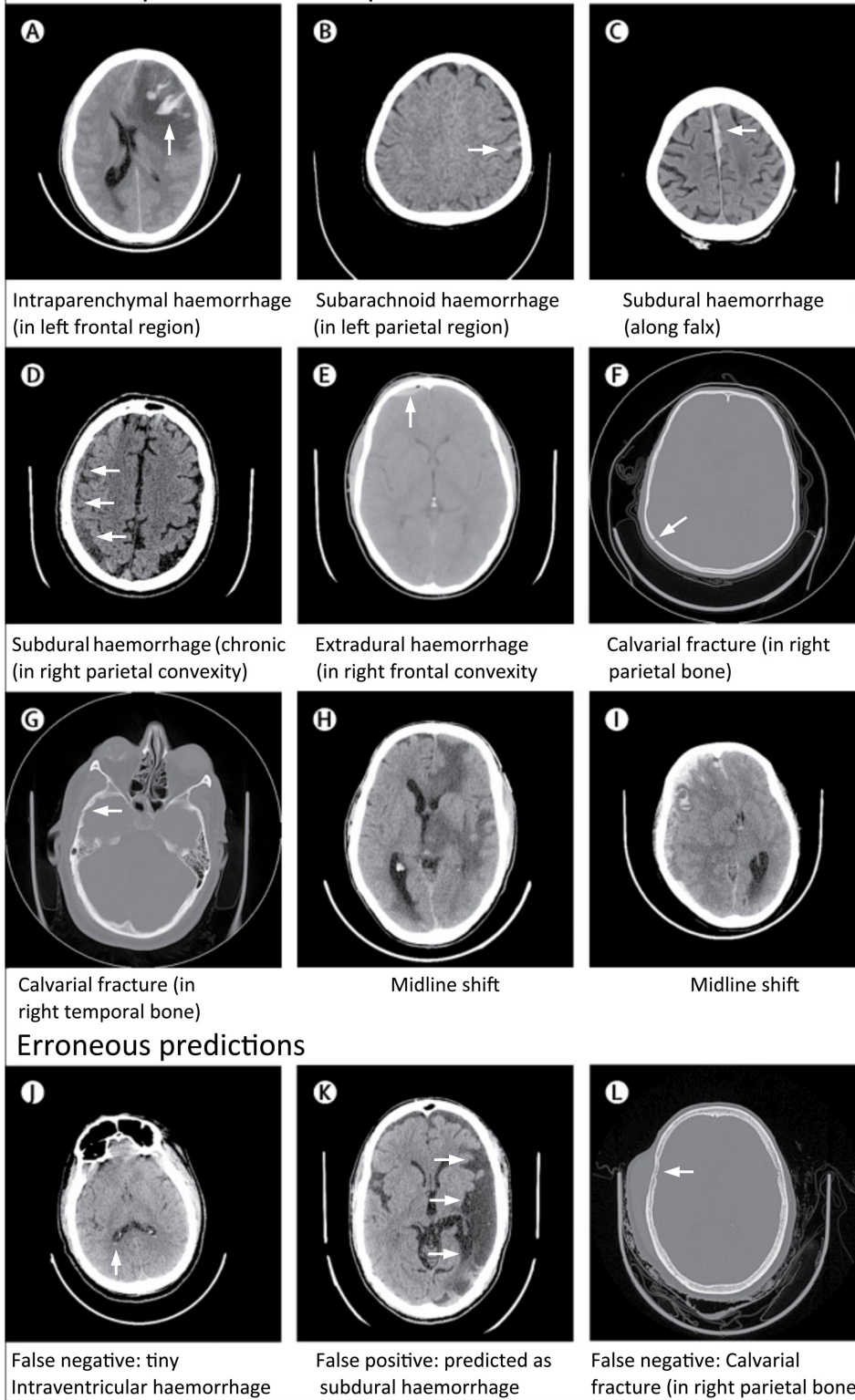
FIGURE 8 | Examples of accurate and erroneous predictions of abnormalities on head CT in acute TBI patients by a deep learning algorithm (26). Although individual images are shown, the model classifies abnormalities at the head CT exam level. All images under Accurate Predictions **(A–I)** have arrows added to indicate the abnormal lesion. All images under Erroneous Predictions **(J–L)** have arrows added to indicate the erroneous lesion predictions.

Supplementary Materials, the authors indicated that they also trained a separate localization model to perform segmentation for intraparenchymal, subdural, and epidural hematomas, using a set of 1,706 images labeled at the pixel level according to hemorrhage subtype. Performance metrics and example images for these segmentations were unavailable at time of publication. The authors have also made the CQ500 test set publicly available to facilitate benchmarking of algorithms developed in the future.

Subsequently, triage algorithms with the capability to classify hemorrhages by subtype were described (112, 114, 115). Lee et al. demonstrated promising classification results with three hemorrhage subtype groupings (epidural/subdural, subarachnoid, and intraparenchymal/intraventricular hemorrhage) (114). As specific hemorrhage subtypes often influence management decisions, and also have different prognostic significance, later classification models were also proposed that identify each of the five subtypes individually (112, 115).

Aside from studies performed on CT, several studies have proposed intracranial hemorrhage detection algorithms on MRI. Most recently, Nael et al. (124) developed a set of deep CNNs which were each purposed to identify a different pathology on brain MRI, with intracranial hemorrhage as one such pathology. The AUC for intracranial hemorrhage detection on the internal and external test data was 0.90 and 0.83, respectively. Rudie et al. (125) proposed a neural network system trained to diagnose 35 different neurologic diseases on brain MRI, including 5 and 3 acute intracranial hemorrhage exams in the training and test sets, respectively. Previously, Al Okashi et al. (127) proposed an ensemble learning system for hemorrhage detection on brain MRI, but describes head CT images throughout the paper as brain MRI images. Le et al. (126) proposed R-FCN as a classification model for CT/MRI images to differentiate between different hemorrhagic subtypes. However, their methods and figures also use CT scans, such that the relevance to MRI is unclear. Overall, algorithms developed for MRI in acute TBI are currently limited.

Although studies have been performed on triage algorithms, limitations remain. It is reasonable to assume that lower performance standards for radiological triage algorithms are acceptable, as all exams are ultimately reviewed by trained human experts. However, this comes with the risk that missed critical findings, particularly those with more subtle abnormalities, may be reordered to the bottom of the queue, with an extended delay in review of these studies if there is over-reliance on the algorithm (131). This time delay could be exacerbated if additional exams are added to the queue in real time and receive higher priority. Explainability also remains important; if the rationale behind critical or non-critical classifications is opaque, it could be difficult for physicians to verify and trust algorithmic outputs (132, 133). "Black box" algorithms such as these also have the potential to lengthen overall readout times if radiologists spend extra time reviewing normal studies that were flagged as abnormal. In addition, lack of explainability may result in difficulty disregarding artifacts and other questionable findings in an exam that has been flagged as abnormal. This could lead to overdiagnosis, a difficult problem noted in early studies of

the use of CAD in screening mammography (134). Finally, the identification of hemorrhage locations and sizes is also critical in order to inform neurosurgical decisions.

Localization Algorithms

Localization algorithms are trained to predict the locations of abnormal imaging findings. In these algorithms, more general exam-level classifications are also usually derived indicating the overall presence or absence of the finding of interest anywhere on the exam. Localization algorithms mitigate the explainability problem, as their outputs specify lesion locations, such as by direct annotation of the original CT images. They also leverage the quantitative nature of algorithms, as detailed localizations have the potential to produce volumetric outputs, which are difficult and time-consuming for humans to measure accurately. Although development of localization algorithms usually require the denser, pixel-level labels, which are highly time-consuming and costly to obtain, they have the potential for providing information that is useful for clinical management decisions as well as outcome prediction.

Chang et al. (27) reported an approach to the task of localizing acute intracranial hemorrhage, using a hybrid 3D/2D convolutional neural network for exam-level classification and hemorrhage segmentation derived from Mask R-CNN (135). They collected a large training set annotated with bounding-box labels around areas of hemorrhage that were verified by a board-certified radiologist. The model demonstrated excellent classification performance, achieving an AUC of 0.981 on the test set. To incorporate localization information, semi-automated pixel labels were produced for hematoma regions using level set segmentation (136). The Dice coefficients for manual and model segmentations for intraparenchymal, epidural/subdural, and subarachnoid hematomas were strong, with decreasing Dice indices in that order. Not surprisingly, the Dice coefficient for subarachnoid hemorrhage was the smallest, as this type of hemorrhage is the most difficult to segment accurately since the bleeds tend to be amorphous with poorly-defined boundaries. Hemorrhage volumes based on manual vs. model segmentations were also computed, with very strong Pearson correlation coefficients exceeding 0.95 for all three hemorrhage subtype categories. Taking a similar approach to the localization problem, Cho et al. (113) collected a pixel labeled dataset and proposed a cascaded deep learning for hemorrhage subtype identification and segmentation. The model also achieved strong performances, with a classification accuracy of 0.979 and segmentation precision of 0.802. These works were promising as they demonstrated excellent performance in classification and localization of hemorrhage, although segmentation examples for more subtle or complex hemorrhages were not available at the time of publication. In addition, the object detection and region processing requirements associated with Mask R-CNN are challenges, since hemorrhage is not discrete but is fluid and takes on highly variable morphologies that are not as well-suited to bounding-box labels. In addition, when pixel labels are obtained with the assistance of a level set algorithm rather than through manual labeling, the segmentations could be subject to the biases of the level-set algorithm assumptions, carrying

with it the possibility of reduced label quality particularly for hemorrhage that is subtle, diffuse, and/or amorphous.

Lee et al. (28) used a different approach to the localization task. Employing only image-level labels for training, the algorithm predicted locations of abnormal findings using image-level "heatmaps." The authors collected a training set as well as prospective and retrospective test sets. Each of the images was labeled according to the hemorrhagic subtype present, including intraparenchymal, intraventricular, subdural, epidural, and subarachnoid hematomas. Using only image level labels, the authors developed an understandable deep learning algorithm that detects the presence of acute intracranial hemorrhage and identifies it as one of five possible hemorrhage subtypes. The algorithm achieved excellent performance with an AUC of 0.99 on the retrospective set and 0.96 on the prospective set. In order to visualize the locations of predicted lesions, the system generated probability heatmaps that used color to highlight high-probability pixels for hemorrhage, along with the suspected hemorrhage subtype (**Figure 9**). The input-output pairing in this study is relatively uncommon. As seen earlier, the most common strategies produce predictions with the same degree of granularity as provided by the training labels. This system, in contrast, predicts the general location of hemorrhage on each image, despite training data consisting of only image-level labels. This localization allows clinicians to directly inspect and verify the predictions, avoiding the "black box" problem, and enabling clinicians to independently evaluate and potentially understand the rationale behind the algorithm's predictions, including possible reasons for error.

Kuo et al. (29) used a different approach, training a deep learning model on CT exams with hemorrhage labeled at the pixel level. The authors developed a patch-based fully convolutional neural network, which was optimized to perform joint classification and segmentation of intracranial hemorrhage. The algorithm had an identical input-output pairing (i.e., input training data and output prediction data with identical granularity of the labels). Training data consisted of 4,396 pixel-labeled CT exams. Rather than outputting heatmaps with diffuse boundaries, as seen with the Lee et al. study, this algorithm's segmentations provide high-resolution localization information (**Figure 10**, left panel). The classification performance was benchmarked against four radiologists on an independent evaluation set, in which the network outperformed two of the four radiologists, achieving an AUC of 0.991 for the classification task and a Dice coefficient of 0.75 for the segmentation task. An exploratory multiclass study was also conducted, in which the model identified and segmented different hemorrhagic subtypes. Segmentation visualizations from the multiclass study were also shown (**Figure 10**, right panel). This study was important because it included a broad selection of visualization examples, demonstrating detailed intracranial hemorrhage segmentations of intracranial hemorrhage and its subtypes. However, this approach requires a dataset annotated at the pixel level, which is highly time-consuming, expensive to obtain because it requires highly-trained human experts, and difficult to scale. Despite this, the study demonstrated the potential for deep learning algorithms to achieve expert-level classification performance
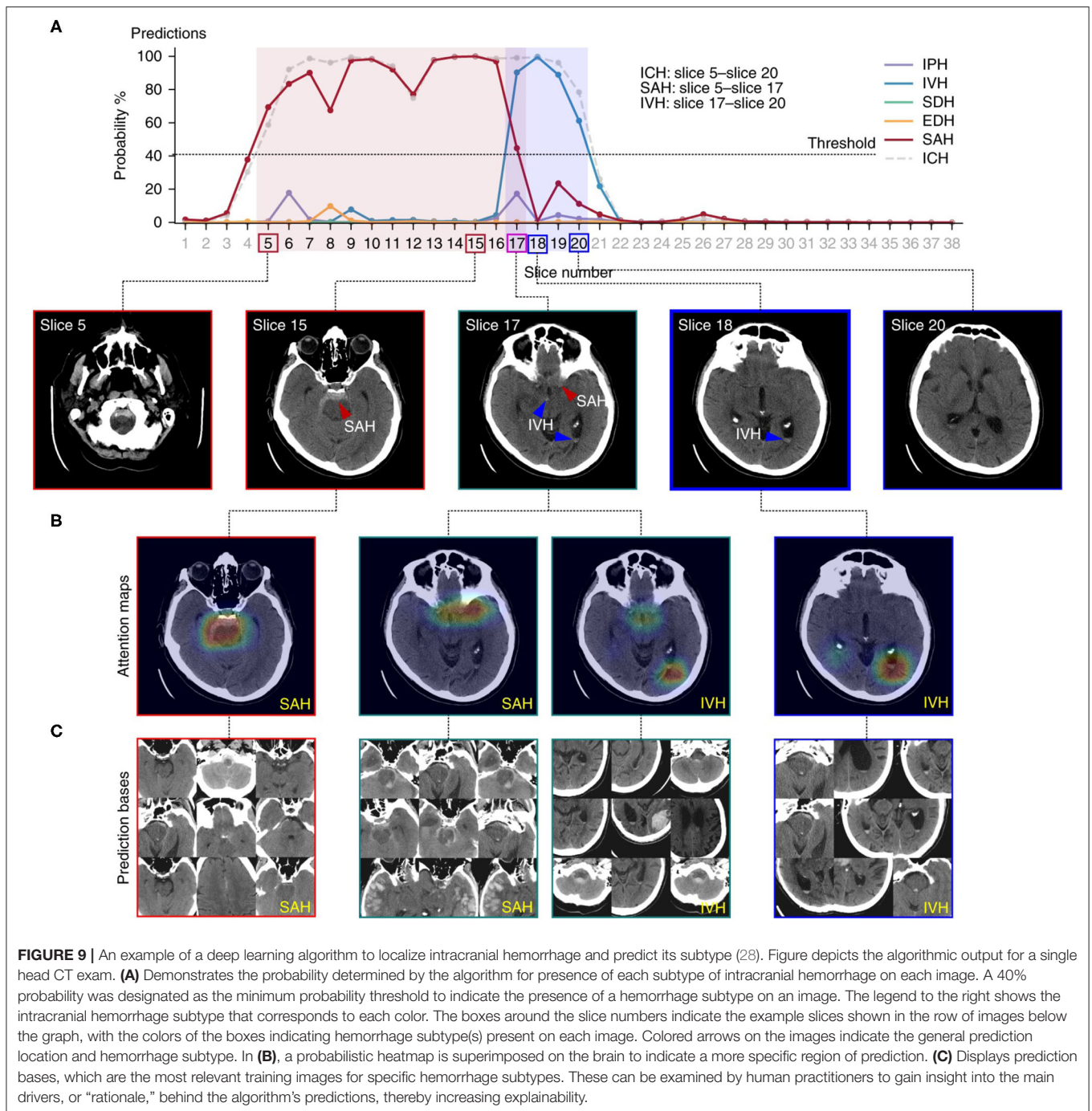
and excellent segmentation, provided densely annotated data are available.

More recently, there has been interest in the use of algorithms to quantify intracranial hemorrhage volumes. Although lesion volumes can hold important prognostic significance and have been established to correlate with mortality and functional outcome (137, 138), they are difficult to obtain accurately. Hematoma volumes are currently estimated using the ABC/2 method; however, this form of measurement assumes an elliptical hematoma and is not suitable for most types of acute intracranial hemorrhage in TBI (139). As segmentation models began to show stronger performance, interest in volumetric analysis of intracranial hemorrhage increased. Some studies focused only on computing total intracranial hemorrhage (116, 118) while others computed separate volumetric outputs for different hemorrhage subtypes (117, 119, 120). While the multiclass volumetric studies demonstrated promising results, many grouped multiple pathoanatomic lesion subtypes into the same category, although different subtypes often require very different clinical management steps. Other studies either did not report or combined results for lesion subtypes that are difficult to segment or detect, such as subarachnoid or petechial hemorrhages. Finally, many studies used small tests due to the difficulty of obtaining ground truth manual segmentations, which raises the possibility of poor generalizability. While initial results are promising, more work is needed to develop algorithms that can perform volumetric analysis in an accurate, reproducible, and comprehensive way.

SKULL FRACTURES

CT is the preferred imaging modality for the diagnosis of acute skull and facial fractures. Since non-displaced skull fractures without intracranial hemorrhage heal without intervention, automated skull fracture detection has received less attention than intracranial hemorrhage detection and is a relatively less-explored area. Automated skull fracture detection is challenging since the typical non-displaced skull fracture is a tiny feature (often < 1 mm in size) on any single CT image, and can only be differentiated from normal venous channels and sutures (140, 141) by its appearance over multiple contiguous images. Most prior attempts at skull fracture detection have used traditional morphological processing techniques. Shao and Zhao (142) used region-growing and boundary-tracing to define the skull, and optimal thresholding techniques to detect the fractures. Zaki et al. (143) used fuzzy c-means (FCM) clustering and line tracing to localize the fractures. Yamada et al. (144) used a black-hat transformation (technique to highlight dark objects of interest in a bright background) to identify the fractures. However, none of these approaches have made a transition into clinical use.

In more recent years, researchers have applied deep learning strategies to approach skull fracture detection. Chilamkurthy et al. (26) used an algorithm trained on image-level labels to perform exam-level detection of calvarial fractures, achieving an AUC of 0.96 on the CQ500 test set. Heimer et al. (145) trained a collection of deep learning models to identify skull fractures on
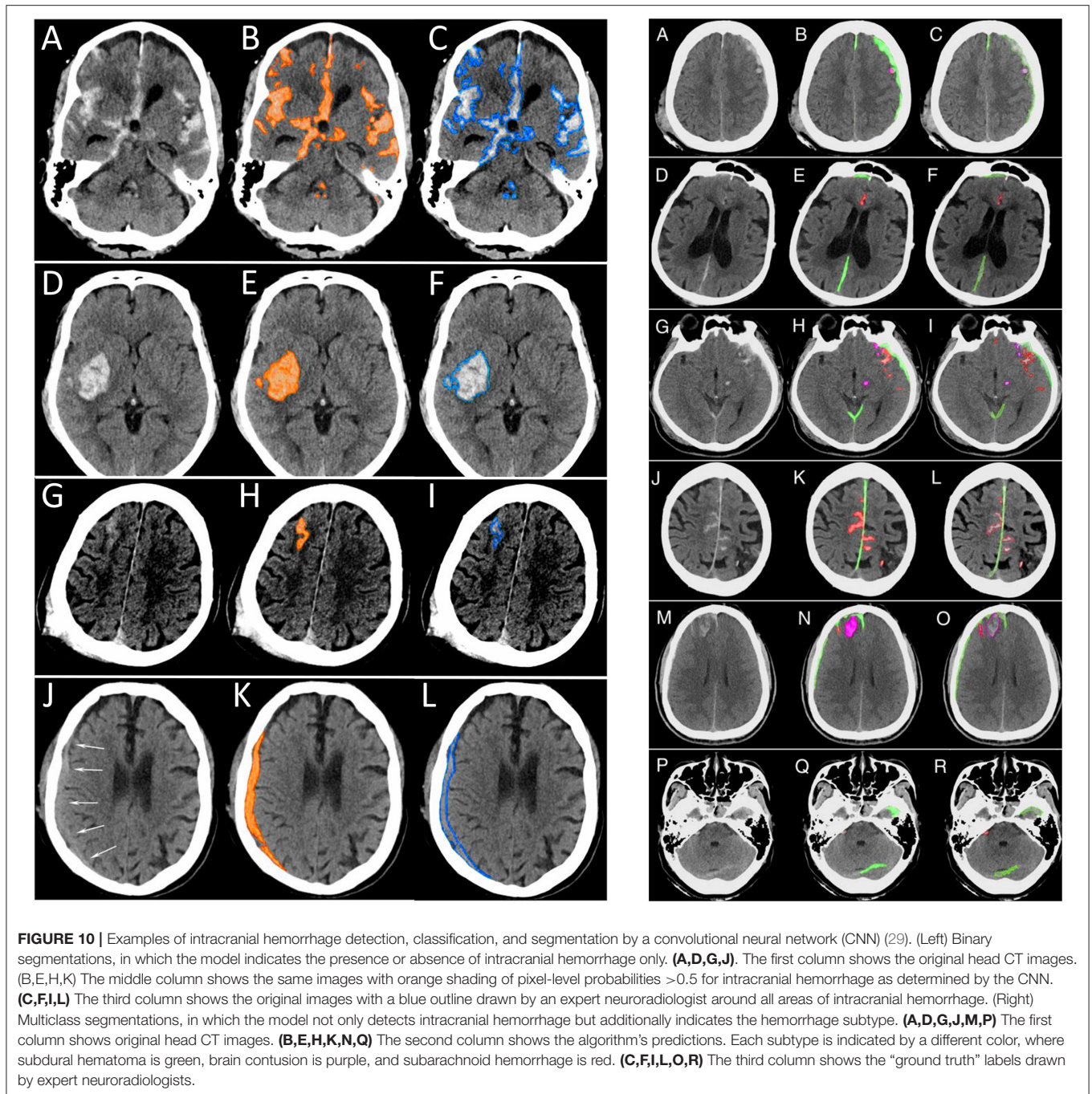
FIGURE 9 | An example of a deep learning algorithm to localize intracranial hemorrhage and predict its subtype (28). Figure depicts the algorithmic output for a single head CT exam. **(A)** Demonstrates the probability determined by the algorithm for presence of each subtype of intracranial hemorrhage on each image. A 40% probability was designated as the minimum probability threshold to indicate the presence of a hemorrhage subtype on an image. The legend to the right shows the intracranial hemorrhage subtype that corresponds to each color. The boxes around the slice numbers indicate the example slices shown in the row of images below the graph, with the colors of the boxes indicating hemorrhage subtype(s) present on each image. Colored arrows on the images indicate the general prediction location and hemorrhage subtype. In **(B)**, a probabilistic heatmap is superimposed on the brain to indicate a more specific region of prediction. **(C)** Displays prediction bases, which are the most relevant training images for specific hemorrhage subtypes. These can be examined by human practitioners to gain insight into the main drivers, or "rationale," behind the algorithm's predictions, thereby increasing explainability.

postmortem computed tomography (PMCT) exams, achieving an average AUC of 0.895 on a test set of 150 cases, half of which contained skull fracture. Ning et al. (146), developed an attention-based multi-scale architecture (AMT-ResNet) for skull fracture detection, achieving 0.903 accuracy and 0.922 recall on a test set of 1,236 images with and 1,300 without skull fracture. Kuang et al. (147) proposed another neural network architecture (Skull R-CNN) based on modification of Faster R-CNN (148) to improve detection of smaller objects, achieving a test AP of 0.60 on a small test set of 10 cases. However, studies of deep

learning models for skull fracture detection remain preliminary, with relatively small datasets and no studies to date that have demonstrated performance level or capabilities that would be consistent with practical utility in a clinical setting.

INTRACRANIAL MASS EFFECT

Intracranial mass effect occurs when space-occupying lesions (hematomas, tumors, enlarged ventricles, vasogenic, or cytotoxic edema) result in significant displacement of a portion of the

FIGURE 10 | Examples of intracranial hemorrhage detection, classification, and segmentation by a convolutional neural network (CNN) (29). (Left) Binary segmentations, in which the model indicates the presence or absence of intracranial hemorrhage only. **(A,D,J).** The first column shows the original head CT images. (B,E,H,K) The middle column shows the same images with orange shading of pixel-level probabilities >0.5 for intracranial hemorrhage as determined by the CNN. **(C,F,I,L)** The third column shows the original images with a blue outline drawn by an expert neuroradiologist around all areas of intracranial hemorrhage. (Right) Multiclass segmentations, in which the model not only detects intracranial hemorrhage but additionally indicates the hemorrhage subtype. **(A,D,G,J,M,P)** The first column shows original head CT images. **(B,E,H,K,N,Q)** The second column shows the algorithm's predictions. Each subtype is indicated by a different color, where subdural hematoma is green, brain contusion is purple, and subarachnoid hemorrhage is red. **(C,F,I,L,O,R)** The third column shows the "ground truth" labels drawn by expert neuroradiologists.

brain, or when brain swelling due to diffuse insults such as encephalitis or hypoxic-ischemic injury results in elevated intracranial pressure. Severe intracranial mass effect is an emergency and may require medication, extraventricular drain placement, or craniectomy to avoid loss of brain tissue and/or reduced perfusion of the brain. Intracranial mass effect in the setting of acute closed head injury usually takes the form of midline shift and/or downward cerebral herniation due to subdural or epidural hematoma, or large brain contusions that often develop surrounding vasogenic edema in the subacute

stage. Despite the clinical importance, however, works studying the automated detection of intracranial mass effects are limited.

The first algorithm in the previously described double-algorithm framework of Prevedello et al. (23) detected acute intracranial hemorrhage, intracranial mass effect, and hydrocephalus, achieving an AUC of 0.91. However, the three abnormalities were ordered in a hierarchical structure (acute hemorrhage > mass effect > hydrocephalus), such that each CT exam was labeled with only the highest-ranked category during training, and the reported accuracy was for all three

abnormalities collectively. Therefore, the model accuracy for intracranial mass effect alone was unknown. In addition, the representation of intracranial mass effect in the datasets was low: only 13 of the 246 training/validation cases and 5 of the 130 test cases were "positive" for intracranial mass effect. A practical problem in developing algorithms in this area is that intracranial mass effect is a relatively uncommon CT finding compared to intracranial hemorrhage, making it more difficult to collect the quantity of data needed to achieve strong performance.

As described earlier, Chilamkurthy et al. (26) developed a deep learning algorithm to simultaneously detect intracranial mass effects including midline shift, demonstrating an AUC of 0.92 on the CQ500 test set. The model was trained with 699 CT exams (320 exams positive for mass effect), for which each image was labeled by a radiologist for the presence or absence of intracranial mass effect. Finally, Monteiro et al. (119) developed a deep learning algorithm to segment vasogenic edema surrounding hemorrhagic contusions.

STROKE

Ischemic stroke, which makes up 85% of all stroke incidents, results when blood flow is hindered due to arterial blockage and can cause permanent brain tissue damage (149). TBI has been previously identified as a risk factor for ischemic stroke (150–152); head injuries may lead to cerebrovascular damage through vascular shearing mechanisms, as well as compression of the anterior cerebral arteries due to subfalcine herniation or posterior cerebral arteries due to downward cerebral herniation. Acute stroke incidence in acute TBI patients is low, affecting ~2.5% of moderate and severe TBI patients (45). As with acute intracranial hemorrhage and intracranial mass effect in acute TBI, rapid detection and treatment of acute stroke is needed to achieve favorable outcomes (153, 154). Although a comprehensive discussion of machine learning techniques for detection of acute stroke is beyond the scope of this targeted review, a brief discussion of algorithmic approaches for automated stroke detection follows.

There is a considerable body of work focusing on the algorithmic detection and segmentation of acute ischemic stroke (92, 103, 155–161), although not specifically within the context of TBI. As with computational approaches to image recognition in acute TBI, older studies used rule-based and traditional machine learning techniques (162, 163), and descriptions of deep learning approaches to this problem have appeared only in the past several years. Unlike acute TBI, acute stroke is commonly diagnosed using a number of different imaging modalities and protocols, including non-contrast head CT, CT perfusion of the brain, CT angiography of the brain and neck, and MRI. Wang et al. (164), one of the earliest studies using deep learning in this area, reported a deep symmetry CNN that achieved promising stroke segmentation results on brain MRI, though using a very small sample size of eight patients. Subsequently, Zhang et al. (103) reported a CNN that performed acute stroke segmentation on diffusion-weighted imaging (DWI) with a Dice coefficient of 0.79, using training and test sets of 90 subjects each. Recently,

Liu et al. (92) reported and publicly released a deep learning model that segmented acute ischemic stroke lesions on diffusion-weighted imaging (DWI) with a Dice coefficient of 0.76, similar to interrater agreement among human experts. The model was tested on a dataset of 2,348 DWI images, apparently the largest test set reported to date for this application. The model was also evaluated on an external dataset for generalization, and generally outperformed its peer models. Aside from applications in diagnosis of acute stroke on imaging, deep learning algorithms have also been proposed to predict stroke expansion across time (165–167) and identify ischemic stroke subtype (168, 169). As in acute TBI, these could eventually be useful for improved prognostic assessments and more personalized clinical management recommendations.

CHALLENGES AND FUTURE DIRECTIONS

Despite major recent advances in computer algorithms for analysis of images in acute TBI, there remain significant challenges and opportunities for expanded use of these algorithms in clinical settings. Firstly, while algorithms for intracranial hemorrhage detection have been developed with accuracy levels acceptable for triage, more widespread clinical use of these algorithms is unlikely until algorithms for the reliable detection of other important abnormalities such as intracranial mass effect, acute and subacute infarct (including both large-territory and small-vessel infarcts and hypoxic/ischemic injury), bony fractures, and edema have been developed. While there have been algorithms developed for the above abnormalities, they remain limited and less thoroughly explored. In addition, while a limited number of localization algorithms for acute intracranial hemorrhage have been demonstrated, the need for localization and explainability applies to other pathologies such as intracranial mass effect, acute infarct, and fractures. Algorithms need to reliably identify all intracranial hemorrhage and skull fractures, including bilateral or subtle abnormalities, as these are important for surgical management decisions in TBI, and the failure to do so can lead to devastating consequences (170, 171). High-resolution pixel-level localization also raises the possibility of objective quantitative measurements of abnormal features, which could be used to improve clinical practice guidelines in the pursuit of precision and evidence-based medicine.

Generalization of strong model performance, to institutions other than those that provided training data for the model, is also a challenge. When algorithms trained on data from one institution are applied to data from another institution, the performance degrades to varying degrees. This is attributable to differences in hardware and other technical parameters that greatly affect the appearance of the images. For an algorithm to be widely deployed, it must remain robust to variations in image appearance across institutions, as well as with ongoing technical innovations in CT scanner hardware and image post-processing techniques over time.

Aside from challenges with algorithm development, another impediment to progress in the field is the lack of dataset standardization. Research teams often collect their own datasets, which can vary by important factors such as dataset size and degree to which it accurately reflects the patient population for intended use. This lack of standardization makes it challenging to compare different models head-to-head, and can also make it difficult to measure progress in the field through rigorous benchmarking. Although CQ500 (26) was a publicly released dataset, several factors hinder more widespread use. Because CQ500 was released without the accompanying training set, research teams would still be required to collect their own training data, which would differ from CQ500 due to variations in technical hardware, scanning protocols, and image post processing techniques. Large, high-quality datasets are usually not released, due to the sensitivity of medical data and institutional restrictions.

In the clinical setting, the development of accurate algorithms for medical image recognition also opens the potential for improved prognostic models, improved metrics for monitoring disease progression, and more specific patient selection criteria for clinical research studies. Outcome studies that include imaging findings (in addition to demographic and clinical variables) as predictors often take into account the presence or absence of certain imaging abnormalities, as it is challenging to incorporate granular imaging information in a scalable and reproducible way (172, 173). The qualitative presence or absence of abnormalities can be extracted from clinical radiology reports, or from an additional dedicated radiological interpretation performed for research purposes. However, more sophisticated TBI classification schemes are a critical need in order to develop effective therapies for acute TBI (34). In TBI and other disorders, patient outcomes often vary widely based on more granular information such as number, size, subtype, and location of abnormalities, that is available but not generally accessible from imaging studies on a large scale. Models that can quantify intracranial hemorrhage volumes and predict intracranial hemorrhage subtypes have been and are continuing to be studied for this purpose (118–120). In addition, the quantitative analysis of other features, including brain parenchyma CT densities (174, 175), regional brain volumes and atrophy (176), DTI (53, 54), and deep learning-based anatomical segmentation on MRI (177, 178) will also have a future role in improved clinical management and prognostication.

RELEVANCE TO PRECISION AND EVIDENCE-BASED MEDICINE

Precision medicine aims to "transform healthcare through use of advanced computing tools to aggregate, integrate and analyze vast amounts of data… to better understand diseases and develop more precise diagnostics, therapeutics and prevention." (179). The capability to extract clinically relevant features from medical images has the potential to enable integration of imaging into precision medicine by transforming medical imaging into an increasingly quantitative and objective science.

For example, the FDA does not consider radiological interpretation, even by highly-trained human experts, to meet its definition of a "biomarker" as a "defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or biological responses to an exposure or intervention…" (128). Biomarkers must be quantitative, objective and reproducible measures (180–183). Biomarkers such as blood test results and genomic data are reproducible and quantitative in nature. In contrast, human interpretation of medical images, even when performed by trained experts, remains subjective, making it difficult to aggregate their content into precision medicine. Indeed, the FDA regards human interpretation of imaging as a "clinical outcome assessment" (COA) (128), defined as a test with results that may vary considerably due to differences in subjective interpretation by human observers, as a result of differences in their judgment and experience. Thus, the use of automated methods to extract quantitative information from medical images has the potential to accelerate the development of FDA-qualified imaging biomarker tests (184, 185), which are currently nearly non-existent in TBI. This in turn could for the first time enable aggregation of imaging biomarkers with clinical, genomic and other patient data across centers to answer important questions regarding prognosis and best treatment practices in TBI and other neurological diseases (186).

Along similar lines, despite the increasing use of clinical practice guidelines to encourage uniformity in clinical management, clinical practice guidelines that include information from radiological images continue to rely almost exclusively on human interpretation. Evidence-based medicine involves the "conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" and relies on the integration of "individual clinical expertise with the best available external clinical evidence from systematic research" (187, 188). As artificial intelligence algorithms for TBI image analysis mature, there is potential to improve practice guidelines using more reproducible data extracted from images. For example, a "midline shift" of 5 mm or more correlates with worse patient outcome (189, 190) and is one criterion used in the decision to perform decompressive craniectomy (191). Automated methods have the potential to promote the development of more advanced, granular metrics of intracranial mass effect that lead to practice guidelines that are better tailored to individual patients.

AUTHOR CONTRIBUTIONS

Both authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

1. Marshall S, Bayley M, McCullagh S, Velikonja D, Berrigan L, Ouchterlony D, et al. Updated clinical practice guidelines for concussion/mild traumatic brain injury and persistent symptoms. *Brain Injury.* (2015) 29:688–700. doi: 10.3109/02699052.2015.1004755

2. Maas AI, Stocchetti N, Bullock R. Moderate and severe traumatic brain injury in adults. *Lancet Neurol.* (2008) 7:728–41. doi: 10.1016/S1474-4422(08)70164-9

3. Bullock RCRM, Chesnut RM, Clifton G, Ghajar J, Marion DW, Narayan RK, et al. Guidelines for the management of severe head injury. *Eur J Emerg Med.* (1996) 3:109–27. doi: 10.1097/00063110-199606000-00010

4. Orrison WW, Gentry LR, Stimac GK, Tarrel RM, Espinosa MC, Cobb LC. Blinded comparison of cranial CT and MR in closed head injury evaluation. *Am J Neuroradiol.* (1994) 15:351–6.

5. Gentry LR, Godersky JC, Thompson B, Dunn VD. Prospective comparative study of intermediate-field MR and CT in the evaluation of closed head trauma. *Am J Neuroradiol.* (1988) 9:91–100.

6. Berbaum KS, Franken Jr EA, Dorfman DD, Rooholamini SA, Kathol MH, Barloon TJ, et al. Satisfaction of search in diagnostic radiology. *Invest Radiol.* (1990) 25:133–40. doi: 10.1097/00004424-199002000-00006

7. Chan T. Computer aided detection of small acute intracranial hemorrhage on computer tomography of brain. *Computer Medical Imaging Graph.* (2007) 31:285–98. doi: 10.1016/j.compmedimag.2007.02.010

8. Liao CC, Xiao F, Wong JM, Chiang IJ. Computer-aided diagnosis of intracranial hematoma with brain deformation on computed tomography. *Computer Medical Imaging Graph.* (2010) 34:563–71. doi: 10.1016/j.compmedimag.2010.03.003

9. Yuh EL, Gean AD, Manley GT, Callen AL, Wintermark M. Computer-aided assessment of head computed tomography (CT) studies in patients with suspected traumatic brain injury. *J Neurotrauma.* (2008) 25:1163–72. doi: 10.1089/neu.2008.0590

10. Bhadauria HS, Dewal ML. Intracranial hemorrhage detection using spatial fuzzy c-mean and region-based active contour on brain CT imaging. *Sign Image Video Proces.* (2014) 8:357–64. doi: 10.1007/s11760-012-0298-0

11. Scherer M, Cordes J, Younsi A, Sahin YA, Götz M, Möhlenbruch M, et al. Development and validation of an automatic segmentation algorithm for quantification of intracerebral hemorrhage. *Stroke.* (2016) 47:2776–82. doi: 10.1161/STROKEAHA.116.013779

12. Muschelli J, Sweeney EM, Ullman NL, Vespa P, Hanley DF, Crainiceanu CM. PItcHPERFeCT: primary intracranial hemorrhage probability estimation using random forests on CT. *NeuroImage.* (2017) 14:379–90. doi: 10.1016/j.nicl.2017.02.007

13. Tong HL, Fauzi MFA, Haw SC, Ng H. Comparison of linear discriminant analysis and support vector machine in classification of subdural and extradural hemorrhages. In: *International Conference on Software Engineering and Computer Systems.* Berlin; Heidelberg: Springer (2011). p. 723–34. doi: 10.1007/978-3-642-22170-5_62

14. de Toledo P, Rios PM, Ledezma A, Sanchis A, Alen JF, Lagares A. Predicting the outcome of patients with subarachnoid hemorrhage using machine learning techniques. *IEEE Trans Inform Technol Biomedicine.* (2009) 13:794–801. doi: 10.1109/TITB.2009.2020434

15. Gong T, Liu R, Tan CL, Farzad N, Lee CK, Pang BC, et al. Classification of CT brain images of head trauma. In: *IAPR International Workshop on Pattern Recognition in Bioinformatics.* Berlin; Heidelberg: Springer (2007). p. 401–8. doi: 10.1007/978-3-540-75286-8_38

16. Liao CC, Xiao F, Wong JM, Chiang IJ. A knowledge discovery approach to diagnosing intracranial hematomas on brain CT: recognition, measurement and classification. In: *International Conference on Medical Biometrics.* Berlin; Heidelberg: Springer (2008). p. 73–82. doi: 10.1007/978-3-540-77413-6_10

17. Nam GB. *From Not Working to Neural Networking.* Special Report. London: The Economist, Special Report. (2016).

18. Zhou DX. Universality of deep convolutional neural networks. *Appl Comput Harmon Anal.* (2020) 48:787–94. doi: 10.1016/j.acha.2019.06.004

19. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv [Preprint].* (2017). arXiv: 1711.05225. Available online at: https://arxiv.org/pdf/1711.05225.pdf (accessed December 25, 2017).

20. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology.* (2017) 284:574–82. doi: 10.1148/radiol.2017162326

21. Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *J Am Med Assoc.* (2017) 318:2199–210. doi: 10.1001/jama.2017.14585

22. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nat Med.* (2018) 24:1559–67. doi: 10.1038/s41591-018-0177-5

23. 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) 285:923–31. doi: 10.1148/radiol.2017162664

24. Titano JJ, Badgeley M, Schefflein J, Pain M, Su A, Cai M, et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. *Nat Med.* (2018) 24:1337–41. doi: 10.1038/s41591-018-0147-y

25. 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 Digit Med.* (2018) 1:1–7. doi: 10.1038/s41746-017-0015-z

26. 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) 392:2388–96. doi: 10.1016/S0140-6736(18)31645-3

27. 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. *Am J Neuroradiol.* (2018) 39:1609–16. doi: 10.3174/ajnr.A5742

28. 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) 3:173–82. doi: 10.1038/s41551-018-0324-9

29. Kuo W, Hne C, Mukherjee P, Malik J, Yuh EL. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. *Proc Nat Acad Sci USA.* (2019) 116:22737–45. doi: 10.1073/pnas.1908021116

30. Korley FK, Kelen GD, Jones CM, Diaz-Arrastia R. Emergency department evaluation of traumatic brain injury in the United States, 2009–2010. *J Head Trauma Rehabil.* (2016) 31:379-87. doi: 10.1097/HTR.0000000000000187

31. Head J. Definition of mild traumatic brain injury. *J Head Trauma Rehabil.* (1993) 8:86–7. doi: 10.1097/00001199-199309000-00010

32. Jagoda AS, Bazarian JJ, Bruns Jr JJ, Cantrill SV, Gean AD, Howard PK, et al. Clinical policy: neuroimaging and decisionmaking in adult mild traumatic brain injury in the acute setting. *J Emerg Nurs.* (2009) 35:e5–e40. doi: 10.1016/j.jen.2008.12.010

33. Bigler ED. Systems biology, neuroimaging, neuropsychology, neuroconnectivity and traumatic brain injury. *Front Syst Neurosci.* (2016) 10:55. doi: 10.3389/fnsys.2016.00055

34. Saatman KE, Duhaime AC, Bullock R, Maas AI, Valadka A, Manley GT. Classification of traumatic brain injury for targeted therapies. *J Neurotrauma.* (2008) 25:719–38. doi: 10.1089/neu.2008.0586

35. Ginde AA, Foianini A, Renner DM, Valley M, Camargo Jr CA. Availability and quality of computed tomography and magnetic resonance imaging equipment in US emergency departments. *Acad Emerg Med.* (2008) 15:780–3. doi: 10.1111/j.1553-2712.2008.00192.x

36. Lell MM, Kachelrieß M. Recent and upcoming technological developments in computed tomography: high speed, low dose, deep learning, multienergy. *Invest Radiol.* (2020) 55:8–19. doi: 10.1097/RLI.0000000000000601

37. Bonney PA, Briggs A, Briggs RG, Jarvis CA, Attenello F, Giannotta SL. Rate of intracranial hemorrhage after minor head injury. *Cureus.* (2020) 12:10653. doi: 10.7759/cureus.10653

38. Andriessen TM, Horn J, Franschman G, van der Naalt J, Haitsma I, Jacobs B, et al. Epidemiology, severity classification, and outcome of moderate and severe traumatic brain injury: a prospective multicenter study. *J Neurotrauma.* (2011) 28:2019–31. doi: 10.1089/neu.2011.2034

39. Yuh EL, Jain S, Sun X, Pisică D, Harris MH, Taylor SR, et al. (2021). Pathological computed tomography features associated with adverse outcomes after mild traumatic brain injury: a TRACK-TBI study with external validation in CENTER-TBI. *J Am Med Assoc Neurol.* 78:1137–48. doi: 10.1001/jamaneurol.2021.2120

40. Korley FK, Kelen GD, Jones CM, Diaz-Arrastia R. Emergency department evaluation of traumatic brain injury in the United States, 2009–2010. *J Head Trauma Rehabil.* (2016) 31:379. doi: 10.1097/HTR0000000000000187

41. Tseng WC, Shih HM, Su YC, Chen HW, Hsiao KY, Chen IC. The association between skull bone fractures and outcomes in patients with severe traumatic brain injury. *J Trauma Acute Care Surg.* (2011) 71:1611–4. doi: 10.1097/TA.0b013e31823a8a60

42. Munoz-Sanchez MA, Murillo-Cabezas F, Cayuela A, Flores-Cordero JM, Rincon-Ferrari MD, Amaya-Villar R, et al. The significance of skull fracture in mild head trauma differs between children and adults. *Child's Nervous Syst.* (2005) 21:128–32. doi: 10.1007/s00381-004-1036-x

43. Murray GD, Teasdale GM, Braakman R, Cohadon F, Dearden M, Iannotti F, et al. The European Brain Injury Consortium survey of head injuries. *Acta Neurochir.* (1999) 141:223-36. doi: 10.1007/s007010050292

44. Jacobs B, Beems T, Stulemeijer M, van Vugt AB, van der Vliet TM, Borm GF, et al. Outcome prediction in mild traumatic brain injury: age and clinical variables are stronger predictors than CT abnormalities. *J Neurotrauma.* (2010) 27:655–68. doi: 10.1089/neu.2009.1059

45. Kowalski RG, Haarbauer-Krupa JK, Bell JM, Corrigan JD, Hammond FM, Torbey MT, et al. Acute ischemic stroke after moderate to severe traumatic brain injury: incidence and impact on outcome. *Stroke.* (2017) 48:1802–9. doi: 10.1161/STROKEAHA.117.017327

46. ACR. *Appropriateness Criteria*. Available online at: http://www.acr.org/quality-safety/appropriateness-criteria/diagnostic/neurologic-imaging (accessed December 22, 2021).

47. Shih RY, Burns J, Ajam AA, Broder JS, Chakraborty S, Kendi AT, et al. ACR appropriateness criteria® head trauma: 2021 update. *J Am Coll Radiol.* (2021) 18:S13–36. doi: 10.1016/j.jacr.2021.01.006

48. American Academy of Pediatrics. The management of minor closed head injury in children. *Pediatrics.* (1999) 104:1407–15. doi: 10.1542/peds.104.6.1407

49. Haacke EM, Duhaime AC, Gean AD, Riedy G, Wintermark M, Mukherjee P, et al. Common data elements in radiologic imaging of traumatic brain injury. *J Magnet Resonan Imag.* (2010) 32:516–43. doi: 10.1002/jmri.22259

50. Mittl RL, Grossman RI, Hiehle JF, Hurst RW, Kauder DR, Gennarelli TA, et al. Prevalence of MR evidence of diffuse axonal injury in patients with mild head injury and normal head CT findings. *Am J Neuroradiol.* (1994) 15:1583–9.

51. Yuh EL, Mukherjee P, Lingsma HF, Yue JK, Ferguson AR, Gordon WA, et al. Magnetic resonance imaging improves 3-month outcome prediction in mild traumatic brain injury. *Ann Neurol.* (2013) 73:224–35. doi: 10.1002/ana.23783

52. Orman G, Kralik SF, Meoded A, Desai N, Risen S, Huisman TA. MRI findings in pediatric abusive head trauma: a review. *J Neuroimag.* (2020) 30:15–27. doi: 10.1111/jon.12670

53. Mukherjee P, Palacios E. *Advanced Structural and Functional Imaging of Traumatic Brain Injury. Youmans and Winn Neurological Surgery.* 7th ed. Philadelphia, PA: Elsevier (2017). p. 2837–42.

54. Shenton ME, Hamoda HM, Schneiderman JS, Bouix S, Pasternak O, Rathi Y, et al. A review of magnetic resonance imaging and diffusion tensor imaging findings in mild traumatic brain injury. *Brain Imag Behav.* (2012) 6:137–92. doi: 10.1007/s11682-012-9156-5

55. Yuh EL, Cooper SR, Mukherjee P, Yue JK, Lingsma HF, Gordon WA, et al. Diffusion tensor imaging for outcome prediction in mild traumatic brain injury: a TRACK-TBI study. *J Neurotr.* (2014) 31:1457–77. doi: 10.1089/neu.2013.3171

56. Palacios EM, Owen JP, Yuh EL, Wang MB, Vassar MJ, Ferguson AR, et al. The evolution of white matter microstructural changes after mild traumatic brain injury: a longitudinal DTI and NODDI study. *Sci Adv.* (2020) 6:peaaz6892. doi: 10.1126/sciadv.aaz6892

57. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage.* (2004) 23:S208–19. doi: 10.1016/j.neuroimage.2004.07.051

58. van Ginneken B. Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. *Radiol Phys Technol.* (2017) 10:23–32. doi: 10.1007/s12194-017-0394-5

59. McKinley R, Häni L, Gralla J, El-Koussy M, Bauer S, Arnold M, et al. Fully automated stroke tissue estimation using random forest classifiers (FASTER). *J Cerebr Blood Flow Metabol.* (2017) 37:2728–41. doi: 10.1177/0271678X16674221

60. Li YH, Zhang L, Hu QM, Li HW, Jia FC, Wu JH. Automatic subarachnoid space segmentation and hemorrhage detection in clinical head CT scans. *Int J Comput Assist Radiol Surg.* (2012) 7:507–16. doi: 10.1007/s11548-011-0664-3

61. Loncaric S, Dhawan AP, Broderick J, Brott T. 3-D image analysis of intra-cerebral brain hemorrhage from digitized CT films. *Comput Methods Programs Biomed.* (1995) 46:207–16. doi: 10.1016/0169-2607(95)01620-9

62. Maier O, Wilms M, von der Gablentz J, Krämer U, Handels H. Ischemic stroke lesion segmentation in multi-spectral MR images with support vector machine classifiers. *Med Imag.* (2014) 9035:903504. doi: 10.1117/12.2043494

63. Menard S. Applied logistic regression analysis. Vol. 106. Newcastle upon Tyne: Sage (2002). doi: 10.4135/9781412983433

64. Breiman L. Random forests. *Mach Learn.* (2001) 45:5–32. doi: 10.1023/A:1010933404324

65. Körding KP, Wolpert DM. Bayesian decision theory in sensorimotor control. *Trends Cogn Sci.* (2006) 10:319–26. doi: 10.1016/j.tics.2006.05.003

66. Yuille AL, Bülthoff HH. *Bayesian Decision Theory and Psychophysics.* Cambridge: Cambridge University Press (1993).

67. Likas A, Vlassis N, Verbeek JJ. The global k-means clustering algorithm. *Pattern Recognit.* (2003) 36:451–61. doi: 10.1016/S0031-3203(02)00060-2

68. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. *IEEE Intellig Syst Appl.* (1998) 13:18–28. doi: 10.1109/5254.708428

69. Bzdok D, Altman N, Kryzwinski M. Statistics versus machine learning. *Nat Methods.* (2018) 15:233-4. doi: 10.1038/nmeth.4642

70. Chong SL, Liu N, Barbier S, Ong MEH. Predictive modeling in pediatric traumatic brain injury using machine learning. *BMC Med Res Methodol.* (2015) 15:1–9. doi: 10.1186/s12874-015-0015-0

71. Hale AT, Stonko DP, Brown A, Lim J, Voce DJ, Gannon SR, et al. Machine-learning analysis outperforms conventional statistical models and CT classification systems in predicting 6-month outcomes in pediatric patients sustaining traumatic brain injury. *Neurosurg Focus.* (2018) 45:E2. doi: 10.3171/2018.8.FOCUS17773

72. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521:436–44. doi: 10.1038/nature14539

73. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition.* Piscataway, NJ: IEEE (2009). p. 248–55. doi: 10.1109/CVPR.2009.5206848

74. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst.* (2012) 25:1097–105.

75. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft coco: common objects in context. In: *European Conference on Computer Vision.* Cham: Springer (2014). p. 740–55. doi: 10.1007/978-3-319-10602-1_48

76. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis.* (2015) 115:211–52. doi: 10.1007/s11263-015-0816-y

77. Kuznetsova A, Rom H, Alldrin N, Uijlings J, Krasin I, Pont-Tuset J, et al. The open images dataset v4. *Int J Comput Vis.* (2020) 128:1956–81. doi: 10.1007/s11263-020-01316-z

78. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev.* (1958) 65:386. doi: 10.1037/h0042519

79. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. *Proc Nat Acad Sci USA.* (1982) 79:2554–8. doi: 10.1073/pnas.79.8.2554

80. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature.* (1986) 323:533–6. doi: 10.1038/323533a0

81. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys.* (1943) 5:115–33. doi: 10.1007/BF02478259

82. Hassoun MH. *Fundamentals of Artificial Neural Networks.* Cambridge, MA: MIT Press (1995). doi: 10.1109/JPROC.1996.503146

83. Jain AK, Mao J, Mohiuddin KM. Artificial neural networks: a tutorial. *Computer.* (1996) 29:31–44. doi: 10.1109/2.485891

84. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. In: *2017 International Conference on Engineering and Technology (ICET).* Piscataway, NJ: IEEE (2017). p. 1–6. doi: 10.1109/ICEngTechnol.2017.8308186

85. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Piscataway, NJ: IEEE (2015). p. 1–9. doi: 10.1109/CVPR.2015.7298594

86. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Piscataway, NJ: IEEE (2016). p. 2818–26. doi: 10.1109/CVPR.2016.308

87. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Piscataway, NJ: IEEE (2016). p. 770–8. doi: 10.1109/CVPR.2016.90

88. Mikolov T, Karafiát M, Burget L, Cernocký J, Khudanpur S. Recurrent neural network based language model. *Interspeech.* (2010) 2:1045–8. doi: 10.21437/Interspeech.2010-343

89. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* (1997) 9:1735–80. doi: 10.1162/neco.1997.9.8.1735

90. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: van Luxbourg U, Guyon I, Bengio S, Wallach H, Fergus R, editors. *Advances in Neural Information Processing Systems.* Red Hook, NY: Curran Associates (2017). p. 5998–6008.

91. 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 preprint arXiv:2010.11929.* (2020).

92. Liu CF, Hsu J, Xu X, Ramachsandran S, Wang V, Miller MI, et al. Deep learning-based detection and segmentation of diffusion abnormalities in acute ischemic stroke. *Commun Med.* (2021) 1:1–18. doi: 10.1038/s43856-021-00062-8

93. Chowdhury S, Dong X, Qian L, Li X, Guan Y, Yang J, et al. A multitask bi-directional RNN model for named entity recognition on Chinese electronic medical records. *BMC Bioinformat.* (2018) 19:75–84. doi: 10.1186/s12859-018-2467-9

94. Jagannatha AN, Yu H. Bidirectional RNN for medical event detection in electronic health records. In: *Proceedings of the conference North American Chapter Meeting. NIH Public Access, Vol. 2016.* San Diego, CA: Association for Computational Linguistics (2016). Vol. 2016. p. 473. doi: 10.18653/v1/N16-1056

95. Gao X, Qian Y, Gao A. COVID-VIT: classification of COVID-19 from CT chest images based on vision transformer models. *arXiv preprint arXiv:2107.01682.* (2021).

96. Wu M, Qian Y, Liao X, Wang Q, Heng PA. Hepatic vessel segmentation based on 3Dswin-transformer with inductive biased multi-head self-attention. *arXiv preprint arXiv:2111.03368.* (2021).

97. Barhoumi Y, Ghulam R. Scopeformer: n-CNN-ViT hybrid model for intracranial hemorrhage classification. *arXiv preprint arXiv:2107.04575.* (2021).

98. Potok TE, Schuman C, Young S, Patton R, Spedalieri F, Liu J, et al. A study of complex deep learning networks on high-performance, neuromorphic, and quantum computers. *ACM J Emerg Technol Comput Syst.* (2018) 14:1–21. doi: 10.1145/3178454

99. Jouppi NP, Young C, Patil N, Patterson D, Agrawal G, Bajwa R, et al. In datacenter performance analysis of a tensor processing unit. In: *Proceedings of the 44th Annual International Symposium on Computer Architecture.* New York, NY (2017). p. 1–12. doi: 10.1145/3079856.3080246

100. Nickolls J, Dally WJ. The GPU computing era. *IEEE micro.* (2010) 30:56–69. doi: 10.1109/MM.2010.41

101. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. *J Big Data.* (2015) 2:1–21. doi: 10.1186/s40537-014-0007-7

102. Chen XW, Lin X. Big data deep learning: challenges and perspectives. *IEEE access.* (2014) 2:514–25. doi: 10.1109/ACCESS.2014.2325029

103. Zhang R, Zhao L, Lou W, Abrigo JM, Mok VC, Chu WC, et al. Automatic segmentation of acute ischemic stroke from DWI using 3-D fully convolutional DenseNets. *IEEE Trans Med Imaging.* (2018) 37:2149–60. doi: 10.1109/TMI.2018.2821244

104. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Piscataway, NJ: IEEE (2017). p. 1492–500. doi: 10.1109/CVPR.2017.634

105. Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, et al. Residual attention network for image classification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Piscataway, NJ: IEEE (2017). p. 3156–64. doi: 10.1109/CVPR.2017.683

106. Goldstein JN, Gilson AJ. Critical care management of acute intracerebral hemorrhage. *Curr Treat Options Neurol.* (2011) 13:204–16. doi: 10.1007/s11940-010-0109-2

107. Phong TD, Duong HN, Nguyen HT, Trong NT, Nguyen VH, Van Hoa T, et al. Brain hemorrhage diagnosis by using deep learning. In: *Proceedings of the 2017 International Conference on Machine Learning and Soft Computing.* New York, NY (2017). p. 34–9. doi: 10.1145/3036290.3036326

108. Patel A, Manniesing R. A convolutional neural network for intracranial hemorrhage detection in non-contrast CT. In: *Medical Imaging 2018: Computer-Aided Diagnosis.* Bellingham, WA: International Society for Optics and Photonics (2018). Vol. 10575. p. 105751B. doi: 10.1117/12.2292975

109. Majumdar A, Brattain L, Telfer B, Farris C, Scalera J. Detecting intracranial hemorrhage with deep learning. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* Piscataway, NJ: IEEE (2018). p. 583–7. doi: 10.1109/EMBC.2018.8512336

110. Grewal M, Srivastava MM, Kumar P, Varadarajan S. Radnet: radiologist level accuracy using deep learning for hemorrhage detection in ct scans. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018).* Piscataway, NJ: IEEE (2018). p. 281–4. doi: 10.1109/ISBI.2018.8363574

111. Jnawali K, Arbabshirani MR, Rao N, Patel AA. Deep 3D convolution neural network for CT brain hemorrhage classification. In: *Medical Imaging 2018: Computer-Aided Diagnosis.* Bellingham, WA: International Society for Optics and Photonics (2018). Vol. 10575. p. 105751C. doi: 10.1117/12.2293725

112. Ye H, Gao F, Yin Y, Guo D, Zhao P, Lu Y, et al. Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. *Eur Radiol.* (2019) 29:6191–201. doi: 10.1007/s00330-019-06163-2

113. Cho J, Park KS, Karki M, Lee E, Ko S, Kim JK, et al. Improving sensitivity on identification and delineation of intracranial hemorrhage lesion using cascaded deep learning models. *J Digit Imaging.* (2019) 32:450–61. doi: 10.1007/s10278-018-00172-1

114. Lee JY, Kim JS, Kim TY, Kim YS. Detection and classification of intracranial haemorrhage on CT images using a novel deep-learning algorithm. *Sci Rep.* (2020) 10:1–7. doi: 10.1038/s41598-020-77441-z

115. Burduja M, Ionescu RT, Verga N. Accurate and efficient intracranial hemorrhage detection and subtype classification in 3D CT scans with convolutional and long short-term memory neural networks. *Sensors.* (2020) 20:5611. doi: 10.3390/s20195611

116. Arab A, Chinda B, Medvedev G, Siu W, Guo H, Gu T, 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) 10:1–12. doi: 10.1038/s41598-020-76459-7

117. Sharrock MF, Mould WA, Ali H, Hildreth M, Awad IA, Hanley DF, et al. 3D deep neural network segmentation of intracerebral hemorrhage:

Development and validation for clinical trials. *Neuroinformatics.* (2020) 2020:1–13. doi: 10.1101/2020.03.05.20031823

118. Dhar R, Falcone GJ, Chen Y, Hamzehloo A, Kirsch EP, Noche RB, et al. Deep learning for automated measurement of hemorrhage and perihematomal edema in supratentorial intracerebral hemorrhage. *Stroke.* (2020) 51:648–51. doi: 10.1161/STROKEAHA.119.027657

119. Monteiro M, Newcombe VF, Mathieu F, Adatia K, Kamnitsas K, Ferrante E, et al. Multiclass semantic segmentation and quantification of traumatic brain injury lesions on head CT using deep learning: an algorithm development and multicentre validation study. *Lancet Digit Health.* (2020) 2:e314–22. doi: 10.1016/S2589-7500(20)30085-6

120. Zhao X, Chen K, Wu G, Zhang G, Zhou X, Lv C, et al. Deep learning shows good reliability for automatic segmentation and volume measurement of brain hemorrhage, intraventricular extension, and peripheral edema. *Eur Radiol.* (2021) 31:5012–20. doi: 10.1007/s00330-020-07558-2

121. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik.* (2019) 29:102–27. doi: 10.1016/j.zemedi.2018.11.002

122. Işın A, Direkoglu C, Sah M. Review of MRI-based brain tumor image segmentation using deep learning methods. *Procedia Comput Sci.* (2016) 102:317–24. doi: 10.1016/j.procs.2016.09.407

123. Rauschecker AM, Rudie JD, Xie L, Wang J, Duong MT, Botzolakis EJ, et al. Artificial intelligence system approaching neuroradiologist-level differential diagnosis accuracy at brain MRI. *Radiology.* (2020) 295:626–37. doi: 10.1148/radiol.2020190283

124. Nael K, Gibson E, Yang C, Ceccaldi P, Yoo Y, Das J, et al. Automated detection of critical findings in multi-parametric brain MRI using a system of 3D neural networks. *Sci Rep.* (2021) 11:1–10. doi: 10.1038/s41598-021-86022-7

125. Rudie JD, Rauschecker AM, Xie L, Wang J, Duong MT, Botzolakis EJ, et al. Subspecialty-level deep gray matter differential diagnoses with deep learning and bayesian networks on clinical brain MRI: a pilot study. *Radiology.* (2020) 2:e190146. doi: 10.1148/ryai.2020190146

126. Le THY, Phan AC, Cao HP, Phan TC. Automatic identification of intracranial hemorrhage on CT/MRI image using meta-architectures improved from region-based CNN. In: *World Congress on Global Optimization.* Cham: Springer (2019). p. 740–50. doi: 10.1007/978-3-030-21803-4_74

127. Al Okashi OM, Mohammed FM, Aljaaf AJ. An ensemble learning approach for automatic brain hemorrhage detection from MRIs. In: *2019 12th International Conference on Developments in eSystems Engineering (DeSE).* Piscataway, NJ: IEEE (2019). p. 929–32. doi: 10.1109/DeSE.2019.00172

128. Food and Drug Administration and National Institutes of Health. *BEST (Biomarkers, Endpoints, and Other Tools) Resource.* Silver Spring, MD: FDA-NIH Biomarker Working Group (2016).

129. LeCun Y. *LeNet-5, Convolutional Neural Networks.* (2015). p. 14. Available online at: http://yann.lecun.com/exdb/lenet. (accessed February 14, 2022).

130. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Thirty-First AAAI Conference on Artificial Intelligence.* San Francisco, CA: AAAI (2017).

131. Ginat DT. Analysis of head CT scans flagged by deep learning software for acute intracranial hemorrhage. *Neuroradiology.* (2020) 62:335–40. doi: 10.1007/s00234-019-02330-w

132. Deeks A. The judicial demand for explainable artificial intelligence. *Columbia Law Rev.* (2019) 119:1829–50.

133. Kundu S. AI in medicine must be explainable. *Nat Med.* (2021) 27:1328. doi: 10.1038/s41591-021-01461-z

134. Fenton JJ, Lee CI, Xing G, Baldwin LM, Elmore JG. Computer-aided detection in mammography: downstream effect on diagnostic testing, ductal carcinoma in situ treatment, and costs. *J Am Med Assoc Intern Med.* (2014) 174:2032–4. doi: 10.1001/jamainternmed.2014.5410

135. He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In: *Proceedings of the IEEE International Conference on Computer Vision.* Piscataway, NJ:IEEE (2017). p. 2961−9. doi: 10.1109/ICCV.2017.322

136. Li C, Xu C, Gui C, Fox MD. Distance regularized level set evolution and its application to image segmentation. *IEEE Trans Image Proces.* (2010) 19:3243–54. doi: 10.1109/TIP.2010.2069690

137. Broderick JP, Brott TG, Duldner JE, Tomsick T, Huster G. Volume of intracerebral hemorrhage. A powerful and easy-to-use predictor of 30-day mortality. *Stroke.* (1993) 24:987–93. doi: 10.1161/01.STR.24.7.987

138. Tuhrim S, Horowitz DR, Sacher M, Godbold JH. Volume of ventricular blood is an important determinant of outcome in supratentorial intracerebral hemorrhage. *Crit Care Med.* (1999) 27:617–21. doi: 10.1097/00003246-199903000-00045

139. Webb AJ, Ullman NL, Morgan TC, Muschelli J, Kornbluth J, Awad IA, et al. Accuracy of the ABC/2 score for intracerebral hemorrhage: systematic review and analysis of MISTIE, CLEAR-IVH, and CLEAR III. *Stroke.* (2015) 46:2470–6. doi: 10.1161/STROKEAHA.114.007343

140. Connor SEJ, Tan G, Fernando R, Chaudhury N. Computed tomography pseudofractures of the mid face and skull base. *Clin Radiol.* (2005) 60:1268–79. doi: 10.1016/j.crad.2005.05.016

141. George CL, Harper NS, Guillaume D, Cayci Z, Nascene D. Vascular channel mimicking a skull fracture. *J Pediatr.* (2017) 181:326–326. doi: 10.1016/j.jpeds.2016.10.070

142. Shao H, Zhao H. Automatic analysis of a skull fracture based on image content. In: *Third International Symposium on Multispectral Image Processing and Pattern Recognition.* Bellingham, WA: International Society for Optics and Photonics (2003). Vol. 5286. p. 741–6. doi: 10.1117/12.538780

143. Zaki WMDW, Fauzi MFA Besar R. A new approach of skull fracture detection in CT brain images. In: *International Visual Informatics Conference.* Berlin; Heidelberg: Springer (2009). p. 156–67. doi: 10.1007/978-3-642-05036-7_16

144. Yamada A, Teramoto A, Otsuka T, Kudo K, Anno H, Fujita H. Preliminary study on the automated skull fracture detection in CT images using black-hat transform. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* Piscataway, NJ: IEEE (2016). p. 6437–40. doi: 10.1109/EMBC.2016.7592202

145. Heimer J, Thali MJ, Ebert L. Classification based on the presence of skull fractures on curved maximum intensity skull projections by means of deep learning. *J Forensic Radiol Imag.* (2018) 14:16–20. doi: 10.1016/j.jofri.2018.08.001

146. Ning D, Liu G, Jiang R, Wang C. Attention-based multi-scale transfer ResNet for skull fracture image classification. In: *Fourth International Workshop on Pattern Recognition.* Bellingham, WA: International Society for Optics and Photonics (2019). Vol. 11198. p. 111980D. doi: 10.1117/12.2540498

147. Kuang Z, Deng X, Yu L, Zhang H, Lin X, Ma H. Skull R-CNN: a CNN-based network for skull fracture detection. In: *Medical Imaging With Deep Learning.* Vienna: PMLR (2020). p. 382–92).

148. Ren S, He K, Girshick R, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks. *Adv Neural Inf Process Syst.* (2015) 28:91–9.

149. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, et al. Heart disease and stroke statistics−2021 update: a report from the American Heart Association. *Circulation.* (2021) 143:e254–743. doi: 10.1161/CIR.0000000000000950

150. Chen YH, Kang JH, Lin HC. Patients with traumatic brain injury: population-based study suggests increased risk of stroke. *Stroke.* (2011) 42:2733–9. doi: 10.1161/STROKEAHA.111.620112

151. Burke JF, Stulc JL, Skolarus LE, Sears ED, Zahuranec DB, Morgenstern LB. Traumatic brain injury may be an independent risk factor for stroke. *Neurology.* (2013) 81:33–9. doi: 10.1212/WNL.0b013e318297eecf

152. Wilson L, Stewart W, Dams-O'Connor K, Diaz-Arrastia R, Horton L, Menon DK, et al. The chronic and evolving neurological consequences of traumatic brain injury. *Lancet Neurol.* (2017) 16:813–25. doi: 10.1016/S1474-4422(17)30279-X

153. Saver JL. Time is brain—quantified. *Stroke.* (2006) 37:263–6. doi: 10.1161/01.STR.0000196957.55928.ab

154. Hacke W, Kaste M, Bluhmki E, Brozman M, Dávalos A, Guidetti D, et al. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. *N Engl J Med.* (2008) 359:1317–29. doi: 10.1056/NEJMoa0804656

155. Li L, Fu H, Tai CL. Fast sketch segmentation and labeling with deep learning. *IEEE Comput Graph Appl.* (2018) 39:38–51. doi: 10.1109/MCG.2018.2884192

156. Qi K, Yang H, Li C, Liu Z, Wang M, Liu Q, et al. X-net: Brain stroke lesion segmentation based on depthwise separable convolution and long-range dependencies. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Cham: Springer (2019). p. 247–55. doi: 10.1007/978-3-030-32248-9_28

157. Nishio M, Koyasu S, Noguchi S, Kiguchi T, Nakatsu K, Akasaka T, et al. Automatic detection of acute ischemic stroke using non-contrast computed tomography and two-stage deep learning model. *Comput Methods Progr Biomed.* (2020) 196:105711. doi: 10.1016/j.cmpb.2020.105711

158. Li L, Wei M, Liu B, Atchaneeyasakul K, Zhou F, Pan Z, et al. Deep learning for hemorrhagic lesion detection and segmentation on brain ct images. *IEEE J Biomed Health Informat.* (2020) 25:1646–59. doi: 10.1109/JBHI.2020.3028243

159. Liu L, Chen S, Zhang F, Wu FX, Pan Y, Wang J. Deep convolutional neural network for automatically segmenting acute ischemic stroke lesion in multi-modality MRI. *Neural Comput Appl.* (2020) 32:6545–58. doi: 10.1007/s00521-019-04096-x

160. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: hierarchical vision transformer using shifted windows. *arXiv preprint arXiv:2103.14030.* (2021).

161. Zhang Q, Yang LT, Chen Z, Li P. A survey on deep learning for big data. *Information Fusion.* (2018) 42:146–57. doi: 10.1016/j.inffus.2017.10.006

162. Takahashi N, Lee Y, Tsai DY, Matsuyama E, Kinoshita T, Ishii K. An automated detection method for the MCA dot sign of acute stroke in unenhanced CT. *Radiol Phys Technol.* (2014) 7:79–88. doi: 10.1007/s12194-013-0234-1

163. Kumar DV, Krishniah VJR. An automated framework for stroke and hemorrhage detection using decision tree classifier. In: *2016 International Conference on Communication and Electronics Systems (ICCES).* Piscataway, NJ: IEEE (2016). p. 1–6. doi: 10.1109/CESYS.2016.7889861

164. Wang Y, Katsaggelos AK, Wang X, Parrish TB. A deep symmetry convnet for stroke lesion segmentation. In: *2016 IEEE International Conference on Image Processing (ICIP).* Piscataway, NJ: IEEE (2016). p. 111–5. doi: 10.1109/ICIP.2016.7532329

165. 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) 49:1394–401. doi: 10.1161/STROKEAHA.117.019740

166. Yu Y, Xie Y, Thamm T, Gong E, Ouyang J, Huang C, et al. Use of deep learning to predict final ischemic stroke lesions from initial magnetic resonance imaging. *J Am Med Assoc Netw Open.* (2020) 3:e200772. doi: 10.1001/jamanetworkopen.2020.0772

167. Robben D, Boers AM, Marquering HA, Langezaal LL, Roos YB, van Oostenbrugge RJ, et al. (2020). Prediction of final infarct volume from native CT perfusion and treatment parameters using deep learning. *Medical Image Anal.* 59:.101589. doi: 10.1016/j.media.2019.101589

168. Fang G, Xu P, Liu W. Automated ischemic stroke subtyping based on machine learning approach. *IEEE Access.* (2020) 8:118426–32. doi: 10.1109/ACCESS.2020.3004977

169. Wu O, Winzeck S, Giese AK, Hancock BL, Etherton MR, Bouts MJ, et al. Big data approaches to phenotyping acute ischemic stroke using automated lesion segmentation of multi-center magnetic resonance imaging data. *Stroke.* (2019) 50:1734–41. doi: 10.1161/STROKEAHA.119.025373

170. Talbott JF, Gean A, Yuh EL, Stiver SI. Calvarial fracture patterns on CT imaging predict risk of a delayed epidural hematoma following decompressive craniectomy for traumatic brain injury. *Am J Neuroradiol.* (2014) 35:1930–5. doi: 10.3174/ajnr.A4001

171. Domenicucci M, Signorini P, Strzelecki J, Delfini R. Delayed post-traumatic epidural hematoma. A review. *Neurosurg Rev.* (1995) 18:109–22. doi: 10.1007/BF00417668

172. Hukkelhoven CW, Steyerberg EW, Rampen AJ, Farace E, Habbema JDF, Marshall LF, et al. Patient age and outcome following severe traumatic brain injury: an analysis of 5600 patients. *J Neurosurg.* (2003) 99:666–73. doi: 10.3171/jns.2003.99.4.0666

173. Mushkudiani NA, Engel DC, Steyerberg EW, Butcher I, Lu J, Marmarou A, et al. Prognostic value of demographic characteristics in traumatic brain injury: results from the IMPACT study. *J Neurotrauma.* (2007) 24:259–69. doi: 10.1089/neu.2006.0028

174. Cauley KA, Hu Y, Fielden SW. Head CT: toward making full use of the information the X-rays give. *Am J Neuroradiol.* (2021) 2021:A7153. doi: 10.3174/ajnr.A7153

175. Finck T, Schinz D, Grundl L, Eisawy R, Yigitsoy M, Moosbauer J, et al. Automated pathology detection and patient triage in routinely acquired head computed tomography scans. *Invest Radiol.* (2021) 56:571–8. doi: 10.1097/RLI.0000000000000775

176. Ledig C, Kamnitsas K, Koikkalainen J, Posti JP, Takala RS, Katila A, et al. Regional brain morphometry in patients with traumatic brain injury based on acute-and chronic-phase magnetic resonance imaging. *PLoS ONE.* (2017) 12:e0188152. doi: 10.1371/journal.pone.0188152

177. Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep learning for brain MRI segmentation: state of the art and future directions. *J Digit Imaging.* (2017) 30:449–59. doi: 10.1007/s10278-017-9983-4

178. Milletari F, Ahmadi SA, Kroll C, Plate A, Rozanski V, Maiostre J, et al. Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. *Comput Vis Image Understand.* (2017) 164:92–102. doi: 10.1016/j.cviu.2017.04.002

179. National Research Council. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease.* Washington, DC: The National Academies Press (2011).

180. Atkinson Jr AJ, Colburn WA, DeGruttola VG, DeMets DL, Downing GJ, Hoth DF, et al. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Therapeut.* (2001) 69:89–95. doi: 10.1067/mcp.2001.113989

181. Mayeux R. Biomarkers: potential uses and limitations. *NeuroRx.* (2004) 1:182–8. doi: 10.1602/neurorx.1.2.182

182. Strimbu K, Tavel JA. What are biomarkers? *Curr Opin HIV AIDS.* (2010) 5:463. doi: 10.1097/COH.0b013e32833ed177

183. Jain KK, Jain KK. *The Handbook of Biomarkers.* New York, NY: Springer (2010). p. 200. doi: 10.1007/978-1-60761-685-6

184. Kessler LG, Barnhart HX, Buckler AJ, Choudhury KR, Kondratovich MV, Toledano A, et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. *Stat Methods Med Res.* (2015) 24:9–26. doi: 10.1177/0962280214537333

185. RSNA.org *Quantitative Imaging Biomarkers Alliance.* (2021). Available online at: https://www.rsna.org/research/quantitative-imaging-biomarkers-alliance (accessed September 29, 2021).

186. CDC. *Precision Health: Improving Health for Each of Us and All of Us.* Centers for Disease Control and Prevention. (2021). Available online at: https://www.cdc.gov/genomics/about/precision_med.htm (accessed September 24, 2021).

187. Sackett DL, Rosenberg WM, Gray JM, Haynes RB, Richardson WS. *Evidence Based Medicine: What It Is and What It Isn't.* London: BMJ (1996). doi: 10.1136/bmj.312.7023.71

188. Sackett DL. Evidence-based medicine. In: *Seminars in Perinatology.* Philadelphia, PA: WB Saunders (1997). Vol. 21. p. 3–5. doi: 10.1016/S0146-0005(97)80013-4

189. Marshall LF, Marshall SB, Klauber MR, van Berkum Clark M, Eisenberg HM, Jane JA, et al. A new classification of head injury based on computerized tomography. *J Neurosurg.* (1991) 75:S14–20. doi: 10.3171/sup.1991.75.1s.0s14

190. 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) 57:1173–82. doi: 10.1227/01.N.E.U.0000186013.63046.6B.

191. QTIP A. *Best Practices in the Management of Traumatic Brain Injury.* Chicago, IL: ACS Committee on Trauma (2015). p. 3–23.