Artificial Intelligence, Machine Learning and Deep Learning in Neuroradiology: Current Applications

Artificial intelligence is rapidly expanding in all medical fields and especially in neuroimaging/neuroradiology (more than 5000 articles indexed on PubMed in 2021) however, few reviews summarize its clinical application in diagnosis and clinical management in patients with neurological diseases. Globally, neurological and mental disorders impact 1 in 3 people over their lifetime, so this technology could have a strong clinical impact on daily medical work. This review summarizes and describes the technical background of artificial intelligence and the main tools dedicated to neuroimaging and neuroradiology explaining its utility to improve neurological disease diagnosis and clinical management.

KEYWORDS: Deep learning • Artificial intelligence • Machine learning • Neuroradiology • Neuroimaging

Introduction

The ever-increasing number of diagnostic tests requires rapid reporting without reducing diagnostic accuracy and this could lead to misdiagnosis. In this context, the recent exponential increase in publications related to artificial intelligence (AI) and the central focus on artificial intelligence at recent professional and scientific radiology meetings underscores the importance of artificial intelligence to improve neurological disease diagnosis and clinical management.

Currently, there are many well-known applications of AI in diagnostic imaging, however, few reviews summarized its applications in Neuroimaging/neuroradiology [1]. Thus, we aim at providing a technical background of AI and an overview of the current literature on the clinical applications of AI in Neuroradiology/Neuroimaging highlighting current tools and rendering a few predictions regarding the near future.

Technical Background of AI

Any computer technique that simulates human intelligence is considered AI. AI is composed of Machine Learning (ML) and Deep Learning (DL).

ML designs systems to learn and improve from experience without being preprogrammed based on statistical data using computer technology. ML uses observations and data which are taken as examples to create some models and algorithms which are then used to make future decisions. In ML, some "ground truth" exists, which is used to train the algorithms. One example is a collection of brain CT scans that a neuroradiologist has classified into different groups (ie, haemorrhage versus no haemorrhage). The goal is to design software to learn automatically and independently of any human intervention or assistance for further intended decisions. DL, representing ML processing, instead applies artificial convolutional neural networks (CNNs) to accelerate the learning process [2, 3]. CNN's are non-linear structures of statistical data organized as modelling tools. They can be used to simulate complex relationships between inputs and outputs using several steps (layers) of nonlinear transformations, which other analytic functions cannot represent [2].

CNNs can be trained to classify an image based on its characteristics through the observation of different images. More specifically, DL can identify common features in different images to use them as a classification model. For example, DL can be trained to find common features in variable images with and without a given pathology to discriminate between both entities. Consequently, it is possible to determine a specific diagnosis without human intervention and therefore there is some potential to improve both the time efficiency and the productivity of radiologists. A strength of DL is that its learning is based on growing experience. As a result, DL has enormous potential because it could update its response models by collecting data from large databases such as the Internet or the Image Storage and Communication System (PACS). However, a limit is that as a consequence, algorithm performance depends largely on both the quantity and quality of data on which

Danilo Caudo*

Department of Biomedical Sciences and Morphological and Functional Imaging, University Hospital Messina, Messina, Italy

*Author for correspondence danycaudo@gmail.com

it is trained [4]. For example, an algorithm for tumour detection trained on data set in which there is no occipital tumour is likely to have a higher error rate for tumours in that location. For a more complete description of DL, the reader is directed to the paper by Montagnon et al. [5].

■ Image acquisition and image quality improvement

Deep learning methods can be used to perform image reconstruction and improve image quality. AI can "learn" standard MR imaging reconstruction techniques, such as Cartesian and non-Cartesian acquisition schemes [6]. Additionally, deep learning methods could be applied to improve image quality. If low and high-resolution images are available, a deep mesh can be used to improve the resolution [7]. This has already been applied to CT imaging to improve resolution in low-dose CT images [8]. Another approach to improve image quality is to use MR images acquired at different magnetic field strengths and coupled from the same anatomy [9].

AI is also able to reduce image acquisition times, Fthis is especially useful in the case of DTI sequences, where the need for more angular directions extends the examination beyond what many patients can tolerate. A deep learning approach can reduce imaging duration by 12 times by predicting final parameter maps (fractional anisotropy, mean diffusivity, and so on) from relatively few angular directions [10]. There are also studies in which DL has increased the signal-to-noise ratio in arterial spin-labelling (ASL) sequences to improve image quality [11]. Finally, some applications of AI could improve resolution and image enhancement by providing a better resolution and signal-to-noise ratio reducing the dose of contrast needed to provide diagnostic images [2].

■ Clinical AI applications in neuroradiology/neuroimaging

Recently, 37 AI applications were reviewed in the domain of Neuroradiology/Neuroimaging from 27 vendors offering 111 functionalities [12]. These AI functionalities mostly support radiologists and extend their tasks. Interestingly, these AI applications are designed for just one pathology, such as ischemic stroke (35%), intracranial haemorrhage (27%), dementia (19%), multiple sclerosis (11%), or brain tumour (11%) to mention the most common [12]. In our review, we found miscellaneous clinical applications of AI in neuroradiology/ neuroimaging ranging from the detection and classification of anomalies on imaging to the prediction of outcomes with disease quantification by estimating the volume of anatomical structures, the burden of lesions, and the volume of the tumour

[13]. In particular, regarding detecting tools, primary emphasis has been placed on identifying urgent findings that enable worklist prioritization for abnormalities such as intracranial haemorrhage [14].

 acute infarction [20–23], large-vessel occlusion [24, 25], aneurysm detection [26–28], and traumatic brain injury [29–30] on non-contrast head CT.

Other AI detecting tools are in brain degenerative disease, epilepsy, oncology, degenerative spine disease (to detect the size of the spinal canal, facet joints alterations, disc herniations, size of conjugation foramina, and in scoliosis the Cobb angle), fracture detection (vertebral fracture such as compression fracture), and in multiple sclerosis to identify disease burden over time and predicting disease activity. In glioma, some DL algorithms were tested to predict glioma genomics.

Regarding segmentation tools, we found tools able to segment vertebral disc, vertebral neuroforamina, and vertebral body for degenerative spine disease, brain tumour volume in the neuro-oncological field, and white and grey matter in degenerative brain diseases [16,17].

In the following paragraphs, we report the main diseases where AI is useful with the more significant relative studies [18,19].

■ Intracranial haemorrhage

Intracranial haemorrhage detection has been widely studied as a potential clinical application of AI [15,30] able to work as an early warning system, raise diagnostic confidence, and classify haemorrhage types. In particular, in these studies, Kuo et al.[14] developed a robust haemorrhage detection model with an area under the receiver operating characteristic curve (ROC-AUC) of 0.991. They trained a CNN on 4396 CT scans for classification and segmentation concurrently on training data which were labelled pixel-wise by attending radiologists. Their test set consisted of 200 CT scans obtained at the same institution at a later time. They report a sensitivity of 96% and a specificity of 98%.

Ker developed a 3-dimensional (3D) convolutional neural network (CNN) model to detect 4 types of intracranial haemorrhage (subarachnoid haemorrhage, acute subdural

haemorrhage, intraparenchymal hematoma, and brain polytrauma haemorrhage). They used a data set consisting of 399 locally acquired CT scans and experimented with data augmentation methods as well as various threshold levels (ie, window levels) to achieve good results. They measured performance as a binary comparison between normal and one of the 4 haemorrhage types and achieved ROC-AUC values of 0.919 to 0.952. The RSNA 2019 Brain CT Hemorrhage Challenge was another milestone, in which a data set of 25 312 brain CT scans were expert-annotated and made available to the public. The scans were sourced from 3 institutions with different scanner hardware and acquisition protocols. The submitted models were evaluated using logarithmic loss, and top models achieved excellent results on this metric (0.04383 for the first-place model). However, it is difficult to directly compare this result to other studies that utilize the ROC-AUC of haemorrhage versus no haemorrhage as their performance metric. Approved commercial software for haemorrhage detection now exist on the market and has been evaluated in clinical settings. Rao evaluated Aidoc (version 1.3, Tel Aviv) as a double reading tool for the prospective review of radiology reports. They assessed 5585 non-contrast CT scans of the head at their institutions which were reported as being negative for haemorrhage and found 16 missed haemorrhages (0.2%), all of which were small haemorrhages. The software also flagged 12 false positives. The Aidoc software was also tested as a triage tool by Ginat in which the software, evaluating 2011 non-contrast CT head scans, contained both false positive and false negative findings of haemorrhage. The study reports sensitivity and specificity of 88.7% and 94.2% for haemorrhage detection, respectively. The author however described a benefit of false-positive flags for haemorrhage as these studies sometimes contained other hyperattenuating pathologies. On the flip side, the author reports a drawback in flagging in patient scans in which a haemorrhage is stable or even improving, which may unnecessarily prioritize nonurgent findings.

■ Aneurysm detection

Detecting unruptured intracranial aneurysms has significant clinical importance considering that they account approximately for 85% of non-traumatic subarachnoid haemorrhages and their prevalence is estimated at approximately 3%. MRI with Time-of-flight angiography sequences (TOF-MRA) is the modality of choice for aneurysm screening, as it does not involve ionizing radiation nor intravenous contrast agents. Deep learning has been used

to detect aneurysms on TOF-MRA. Published methods demonstrate high sensitivity but poor specificity, resulting in multiple false positives per case. Although this fact necessitates a close review of all aneurysms flagged by the software, the resulting models may nevertheless be useful as screening tools. Ueda tested a DL model on 521 scans from the same institution as well as 67 scans from an external data set. DL model achieved 91% to 93% sensitivity with a rate of 7 false positives per scan. Despite the high false positive rate, the authors found that this model helped them detect an additional 4.8% to 13% of aneurysms in the internal and external test data sets. In another study, Faron et al., trained and evaluated a CNN model on a data set of TOF-MRA scans on 85 patients. CNN achieved 90% sensitivity with a rate of 6.1 false positives per case. Similarly, Nakao trained a CNN on a data set of 450 patients with TOF-MRA scans on 3-Tesla magnets only. They were able to achieve a better result at 94% sensitivity with 2.9 false positives per case. Yang used a different approach, in which they produced 3D reconstructions of the intracranial vascular tree using TOF-MRA images. They annotated aneurysms on the 3D projections and used them to train several models. In this manner, they achieved a good discriminatory result between healthy vessels and aneurysms.

■ Stroke

In stroke imaging 3 components of stroke imaging are explored: large vessel occlusion (LVO) detection, automated measurement of core infarct volume and the Alberta Stroke Program Early CT Score (ASPECTS) , and infarct prognostication.

The timely detection of STROKE is critical in brain ischemic treatment, in this context AI has shown the potential in reducing the time to diagnosis. In particular, there are some AI applications able to detect LVO. You et al. developed an LVO detection model using clinical and imaging data (non-contrast CT scans of the head). AI detects the hyperdense Middle Cerebral Artery (MCA) sign which is a finding suggestive of the presence of an MCA thrombus. The AI, based on the U-Net architecture, was trained and tested on a local data set of 300 patients. It achieved a sensitivity and specificity of 68.4% and 61.4%, respectively. On NCCT, an SVM algorithm detected the MCA dot sign in patients with acute stroke with high sensitivity (97.5%). A neural network that incorporated various demographic, imaging, and clinical variables in predicting LVO outperformed or

equalled most other prehospital prediction scales with an accuracy of 0.820. A CNN-based commercial software, Viz-AI-Algorithm v3.04, detected proximal LVO with an accuracy of 86%, a sensitivity of 90.1%, a specificity of 82.5, AUC of 86.3% (95% CI, 0.83–0.90; P # .001), and intraclass correlation coefficient (ICC) of 84.1% (95% CI, 0.81–0.86; P # .001), and Viz-AI-Algorithm v4.1.2 was able to detect LVO with high sensitivity and specificity (82% and 94%, respectively). Unfortunately, no study has yet shown whether AI methods can accurately identify other potentially treatable lesions such as M2, intracranial ICA, and posterior circulation occlusions.

Establishing infarct volumes is important to triage patients for appropriate therapy. AI has been able to establish core infarct volumes on MRI sequences through automatic lesion segmentation. One reported limitation was the reliance on FLAIR and T1 images that do not fully account for the timing of stroke occurrence. Another limitation was a tendency to overestimate the volume of small infarcts and underestimate large infarcts compared with manual segmentation by expert radiologists and difficulty in distinguishing old versus new strokes. Discrepancies in volumes were attributed to nondetectable early ischemic findings, partial volume averaging, and stroke mimics on CT.

ASPECTS is an important early predictor of infarct core for middle cerebral artery (MCA) territory ischemic strokes. It assesses 10 regions within the MCA territory for early signs of ischemia and the resulting score ranges from 0 to 10, where 10 indicates no early signs of ischemia, while 0 indicates ischemic involvement in all 10 regions. The score is currently a key component in the evaluation of the appropriateness of offering endovascular thrombectomy. Several commercial AI applications perform automated ASPECTS evaluation and they have been assessed in clinical settings. In particular, Goebel. compared Frontier ASPECTS Prototype (Siemens Healthcare GmbH) and e-ASPECTS (Brainomix) to 2 experienced radiologists and found that e-ASPECTS showed a better correlation with expert consensus. Guberina compared 3 neuroradiologists with e-ASPECTS and found that the neuroradiologists had a better correlation with infarct core as judged on subsequent imaging than the software. Maegerlein compared RAPID ASPECTS (iSchemaView) to 2 neuroradiologists and found that the software showed a higher correlation with expert consensus than each neuroradiologist individually. Accuracy varies

widely and depends on the software and chosen ground truth. An interesting result suggested by one study, however, was that the RAPID software produced more consistent results when the image reconstruction algorithm was varied compared to human readers. The interclass correlation coefficient between multiple reconstruction algorithms was 0.92 for RAPID, 0.81 to 0.84 for consultant radiologists, and 0.73 to 0.79 for radiology residents.

Prognostication in Stroke treatment is critical to detect patients who are most likely to benefit from treatment considering the risks related. For this reason, AI has been studied as a tool for predicting post-treatment outcomes. In this context [23] developed a CNN model to predict post-treatment infarct core based on initial pre-treatment magnetic resonance imaging (MRI). The authors used a locally acquired data set of 222 patients, 187 of whom were treated with tPA. The model was evaluated using a modified version of the ROC-AUC, where the true positive rate was set to the number of voxels correctly identified as positive, the true negative rate was set to the number of voxels correctly identified as negative, and so on. The reported modified ROC-AUC is 0.88. In another study, Nishi developed a U-Net model to predict clinical post-treatment outcomes using pretreatment diffusion-weighted imaging on patients who underwent mechanical thrombectomy. The clinical outcome was defined using the modified Rankin Scale (mRS) at 90 days after the stroke. The outcomes were categorized as "good" (mRS < 2) and "poor" (mRS > 2). After training on a data set of 250 patients, the model was validated on a data set of 74 patients and found to have a ROC-AUC of 0.81.

■ Multiple sclerosis

In multiple sclerosis deep learning has been investigated as a tool to estimate disease burden and predict disease activity through MRI imaging. MRI is used to assess disease burden over time but this requires comparison with prior scans, which can be burdensome and error-prone when the number of lesions is large. Nair evaluated a DL algorithm for MS lesion identification on a private multicenter data set of 1064 patients diagnosed with the relapsing-remitting variant, containing a total of 2684 MRI scans. In this study, the DL performance was worst with small lesions. The algorithm was tested on 10% of the data set where it achieved a ROC-AUC of 0.80 on lesion detection. In another study, Wang trained a CNN on 64 magnetic resonance (MR) scans to detect MS lesions which were able to

achieve a sensitivity of 98.77% and specificity of 98.76% for lesion detection, respectively.

Regarding predicting disease activity Yoo et al. they developed a CNN that combined a data set of 140 patients, who had onset of the first demyelinating symptoms within 180 days of their MR scan, with defined clinical measurements. CNN achieved a promising result with a ROC-AUC of 0.746 for predicting progression to clinically definite MS.

Another application of AI in MS regards efforts made to reduce gadolinium use where possible considering emerging evidence that repeated administrations of gadolinium-based contrast agents lead to their deposition in the brain. In particular, Narayana et al. used DL to predict lesion enhancement based on their appearance on non-contrast sequences (precontrast T1-weighted imaging, T2-weighted imaging, and fluid-attenuated inversion recovery). They used a data set of 1008 patients with 1970 MR scans acquired on magnets from 3 vendors. DL achieved a ROC-AUC of 0.82 on lesion enhancement prediction, suggesting that this approach may help reduce contrast use.

■ Fracture detection

Regarding fracture detection tools we found that Tomita tested a DL model to detect osteoporotic vertebral fractures in a data set of 1432 CT scans. The outcome was a binary classification of whether or not a fracture was present. Using 80% of the data set for training, a ROC-AUC of 0.909 to 0.918 was achieved with an accuracy of 89.2%. This was found to be equivalent to radiologists on the same data set. In a similar study by Bar. CNN was trained with a data set of 3701 CT scans of the chest and/or abdomen to detect vertebral compression fractures. The model was able to detect vertebral compression fractures with 89.1% accuracy, 83.9% sensitivity, and 93.8% specificity.

Furthermore, Derkatch used a data set of 12 742 dual-energy X-ray absorptiometry scans to train a binary classifier for the detection of vertebral compression fractures, 70% of the data set was used for training, which yielded a ROC-AUC of 0.94. The optimal threshold achieved a sensitivity of 87.4% and a specificity of 88.4%.

Brain Tumor

For brain tumours there AI application for segmentation, that can be used as a stand-alone clinical tool, such as in contouring targets for radiotherapy, or it can also be used to extract tumours as a preliminary step for further downstream ML tasks, such as diagnosis, pre-surgical planning, follow-up and tumour grading. Unfortunately, there is a limit to the AI segmentation, usually, only a minority of voxels represents tumours and the majority represents healthy tissue, however, in a recently published study, Zhou. Trained an AI model with a publicly available MRI data set of 542 glioma patients and they were able to tackle this limit. Their results demonstrate excellent performance with a Dice score of 0.90 for the whole tumour (entire tumour and white matter involvement) and 0.79 for tumour enhancement. Another AI clinical application for brain tumours is predicting glioma genomics Isocitrate dehydrogenase (IDH) mutations that are important prognosticators. Multiple studies investigated the prediction of IDH mutation status from MRI. Zhao et al. published a meta-analysis of 9 studies totalling 996 patients (published in 2019). The largest data set used for training had 225 patients. These studies developed binary classification models and had a ROC-AUC of 0.89 (95% CI: 0.86-0.92). Pooled sensitivities and specificities were 87% (95% CI:76-93) and 90% (95% CI: 72-97), respectively. Since then, another study was published by Choi et al. using a larger MRI data set of 463 patients. It showed excellent results with ROC-AUC, sensitivity, and specificity of 0.95, 92.1%, and 91.5%, respectively. This model used a CNN to segment and as a feature extractor to predict IDH mutation risk Haubold. used 18F-fluoro-ethyl-tyrosine positron emission tomography (PET) combined with MRI to predict multiple tumour genetic markers using a data set of 42 patients before biopsy or treatment. They trained 2 different classical ML models and used biopsy results as the ground truth. They achieved a ROC-AUC of 0.851 for predicting the ATRX mutation, 0.757 for the MGMT mutation, 0.887 for the IDH1 mutation, and 0.978 for the 1p19q mutation.

Degenerative Brain Disease

Regarding dementias numerous AI networks have been trained from large longitudinal datasets such as the Alzheimer's Disease Neuroimaging Initiative (ADNI), resulting in many diagnostic DL tools for Alzheimer's Disease (AD), such as models using 18F fluorodeoxyglucose (FDG) PET76 and structural MRI of the hippocampus to predict AD onset from 1 to 6 years in advance. Furthermore, AI may assist in the diagnosis of dementia types. AI can differentiate AD from Lewy body and Parkinson's dementia. Similarly, other AI tools can differentiate

between mild cognitive impairment (MCI) and AD. In addition to diagnosis, AI can also probe neurobiology. New ML techniques such as Subtype and Stage Inference (SuStaIn) have provided novel neuroimaging and genotype data-driven classifications of diagnostic subtypes and progressive stages for AD and frontotemporal dementia (FTD). SuStaIn has localized distinct regional hotspots for atrophy in different forms of familial FTD caused by mutations in genes.

In other studies, DL has integrated MRI, neurocognitive, and APOE genotype information to predict conversion from MCI to AD. Combining several AI systems (including structural MRI and amyloid PET) may augment the diagnosis and management of the complex natural history of AD. In the future, the integration of AI tools for imaging with AI systems designed to examine serum amyloid markers, mortality prediction from clinicians' progress notes and assessments of cognition, and postmortem immunohistochemistry images, may improve many facets of care in neurodegenerative disease. In Huntington's Disease, an autosomal dominant movement disorder, diagnosis, and management may be enhanced by incorporating CAG repeat length data with CNN developed for caudate volumetry, and objective gait assessment. Such multi-approaches may improve risk stratification, progression monitoring, and clinical management in patients and families.

■ Epilepsy

The use of AI in the diagnosis of epilepsy could improve the diagnostic capabilities of this condition as the symptoms are not specific and often overlap with other conditions. In particular, the integration of anamnestic, clinical, electroencephalographic and imaging information is fundamental for an accurate diagnosis and subtype differentiation. Neuroimaging plays an important role in both diagnosis and follow-up and prognosis. In particular, structural magnetic resonance imaging (sMRI) can help identify cortical abnormalities (eg temporal mesial sclerosis, focal cortical dysplasia [FCD], neoplasms, etc.), while functional magnetic resonance imaging (fMRI), emission tomography positron imaging (PET) and magnetoencephalography (MEG) can help localize brain dysfunction.

Park. (2020) used an SVM classifier on bilateral hippocampi. The model obtained an area under the receiver operating characteristic curve (AUC) of 0.85 and an accuracy of 85% in differentiating epileptic patients from healthy controls, better than human evaluators, in many cases mesial sclerosis. it is subtle and often invisible. Such cases can lead to a misdiagnosis and consequently delay the surgical treatment. Therefore, recent machine learning models have been proposed to identify MRI-negative patients and lateralize foci. For example, Mo et al. (2019) used an SVM classifier based on clinically empirical features, achieving 88% accuracy in detecting MRI-negative patients and an AUC of 0.96 in differentiating MRI-negative patients from controls. The most important feature was the degree of blurring of the grey-white matter at the temporal pole. Similarly, Beheshti et al. (2020) used an SVM to diagnose epileptic patients for mesial sclerosis and lateralize foci in a cohort of 42 MRI-negative PET-positive patients. Focusing on FLAIR, a simple and widely available sequence, the authors extracted signal strength from regions of interest (ROIs) a priori. The model achieved 75% accuracy in differentiating right and left epileptics from controls. The best performance was obtained in identifying the right epilepsy, with an accuracy of 88% and an AUC of 0.84. The most important ROIs were the amygdala, the inferior, middle and superior temporal gyrus and the temporal pole.

However, analyzing only the temporal lobes may not reveal a more global pathology, for this reason, Sahebzamani. (2019), using unified segmentation and an SVM classifier, found that whole brain features are more diagnostic than hippocampal features alone (94% vs 82% accuracy). In particular, global contrast and white matter homogeneity were found to be the most important, along with the clustering tendency and grey matter dissimilarity. In particular, the best performances were obtained based on the mean sum of the whole brain's white matter. Another possibility of AI on sMRI is to lateralize the temporal epileptic focus. In a study with an SVM, the combination of hippocampus, amygdala, and thalamic volumes was more predictive of the laterality of the epileptic focus. The combined model achieved 100% accuracy in patients with mesial sclerosis (Mahmoudi et al., 2018). Furthermore, Gleichgerrcht et al. (2021) used SVM deep learning models to diagnose and lateralize temporal epileptic focus based on structural and diffusion-weighted MRI ROI data. The models achieved an accuracy of 68–75% in diagnosis and 56–73% in lateralization with diffusion data. Based on the sMRI data, ipsilateral hippocampal volumes were the most important for prediction performance. Based on the dwMRI data, ipsilateral tracked beams had the highest predictive weight. Machine learning

techniques have also been used to diagnose cortical dysplasia, the most common cause of medically refractory epilepsy in children and adults second most common cause (Kabat and Król, 2012). In one study, Wang et al. (2020) trained a CNN to exploit the differences in texture and symmetry on the boundary between white matter and grey matter. The model achieved a diagnostic accuracy of 88%. Similarly, Jin et al. (2018) trained surface-based morphometry and a non-linear neural network model. Based on six characteristics of a 3D cortical reconstruction, the model achieved an AUC of 0.75. The contrast of intensity between grey matter and white matter, local cortical deformation and local cortical deformation of cortical thickness were the most important factors for classification. Notably, the model worked well in three independent epilepsy centres. However, data extraction limits have been reported regarding variations in image quality between different clinical sites of the tests.

Another possibility is to combine features derived from sMRI with those derived from fMRI for a variety of clinical applications, such as diagnosing epilepsy and predicting the development of epilepsy following traumatic events. In one study, Zhou et al. (2020) found that the combination of fMRI and sMRI functions was more useful than either modality alone in identifying epileptic patients. In another study, Rocca et al. (2019) used Random Forest and SVM models to help predict the development of seizures following temporal brain injuries. The highest AUC of 0.73 was achieved with a random forest model using functional characteristics. Therefore, additional studies that directly compare the additive utility of fMRI and sMRI, perhaps using framework models, may be useful. In summary, a variety of machine learning approaches have been used for the automated analysis of sMRI data in epilepsy. Given the limited number of publicly available sMRI datasets, models tend to be trained on small single-centre cohorts, this and the lack of external validation limits the interpretation of widespread clinical utility. Future work with larger, multicenter data sets is needed. Functional MRI, a method that measures changes in blood flow to assess and map the magnitude and temporospatial characteristics of neural activity, is gaining popularity in the field of epilepsy (Sidhu et al., 2018). A growing body of evidence suggests that epilepsy is likely characterized by complex and dynamic changes in the way neurons communicate (i.e., changes in neural networks and functional connectivity), both locally and globally. AI applied to fMRI is useful in recent epilepsy studies. Mazrooyisebdani et

al. (2020) used an SVM to diagnose temporal epilepsies based on functional connectivity characteristics derived from graph theory analysis. The model achieved an accuracy of 81%. Similarly, Fallahi et al. (2020) constructed static and dynamic matrices from fMRI data to derive measurements of global graphs. Then, the most important characteristics were selected using random forest and the classification was performed with SVM. The use of dynamic features led to better accuracy than the use of static features (92% versus 88%) in the lateralization of temporal epilepsy. In another study, Hekmati et al. (2020) used fMRI data to quantify mutual information between different cortical regions and insert these quantities into a four-layered perceptual classifier. The model achieved 89% accuracy in locating seizure foci. Finally, Hwang et al. (2019) used LASSO feature selection to extract functional connectivity features, which were then used to train an SVM, linear discriminant analysis, and naive Bayes classifier to diagnose temporal epilepsy. The best accuracy of 85% was achieved with the SVM model. Recent work by Bharat et al. (2019) with machine learning provided further evidence that epilepsy arises from impaired functional neural networks. Using resting-state fMRI (rs-fMRI), the researchers were able to identify connectivity networks specific to temporal epilepsy. The model differentiated temporal epilepsy patients from healthy controls with 98% accuracy and 100% sensitivity. The networks were also found to be highly correlated with disease-specific clinical features and hippocampal atrophy. Although this evidence provides new proof of concept for the existence of specific epilepsy networks, future work is needed, as impaired functional activity can occur secondary to the effects of antiepileptic drugs or a variety of other confounding factors. In summary, fMRI-based machine learning can be used to identify complex alterations in functional neural networks in the epileptic brain and further exploit these differences for classification purposes. In many cases of epilepsy, structural and functional anomalies of the network probably coexist. Current machine learning models with fMRI are limited by small sample sizes, probably because there are few publicly available data sets. However, fMRI is increasingly being integrated into routine clinical practice, particularly for lateralization before surgery (Szaflarski et al., 2017). Recent studies have shown that fMRI may be specifically useful in pre-surgical lateralization (Bauer et al., 2014). With technological advances and further methodological refinements, fMRI could become the standard of care in epilepsy

and AI will be increasingly used to assist in diagnostic and prognostic tasks. Another field of AI application is diffusion tensor imaging (DTI), which has advantages in detecting subtle structural abnormalities of epileptogenic foci. Machine learning in DTI can use for classification improving the diagnosis and treatment of epilepsy, particularly when used for pre-surgical planning and post-surgical outcome prediction.

Degenerative Spine Disease

The rate of MRI examinations is stressfully increased due to the significant number of patients suffering from degenerative spine disease. Consequently, radiologists face a work overload who need to evaluate numerous parameters (size of the spinal canal, facet joints, disc herniations, size of conjugation foramina, etc) in all spinal levels in a short time. Accordingly, different DL and ML algorithms that can automatically classify spinal pathology may help to reduce patient waiting lists and examination costs. In this context, Jamaludin et al. evaluated an automatic disc disease classification system that yielded an accuracy of 96% compared to radiologist assessment. Notably, the main sources of limitation were either poor scan quality or the presence of transitional lumbosacral anatomy. Furthermore, Chen evaluated a DL tool to measure Cobb angles in spine radiographs for patients with scoliosis. They used a data set of 581 patients and were able to achieve a correlation coefficient of r ¼ .903 to 0.945 between the DL-predicted angle and the ground truth. Regarding spine segmentation, some models have been developed with good results. Huang et al. Achieved intersection over-union scores of 94.7% for vertebrae and 92.6% for disc segmentations on sagittal MR images using a training set of 50 subjects and a test set of 50 subjects. Whitehead trained a cascade of CNNs to segment spine MR scans using a data set of 42 patients for training and 20 patients for testing. They were able to achieve Dice scores of 0.832 for discs and 0.865 for vertebrae. In this context DL has been used to answer research questions, for example, Gaonkar used DL to look for potential correlations between the cross-sectional area of neural foramina and patient height and age, showing that the area of neural foramina is directly correlated with patient height and inversely correlated with age

■ AI Tools for Ultrasound in Neuroimaging

AI applications in neuroimaging and ultrasound (US) are mostly focused on the identification of anatomical structures such as nerves. A large number of algorithms have shown to be able to segment US images for these aims. For example, Kim et al. developed a neural network that accurately and effectively segments the median nerve. To train the algorithm and evaluate the model, 1,305 images of the median nerve of 123 normal subjects were used. However, the proposed neural network yielded more accurate results in the wrist datasets, rather than forearm images, with a precision respectively of 90.3% and 87.8%. Different studies showed that AI may help to automatically segment nerve and blood vessels to facilitate ultrasound-guided regional anaesthesia. Automated medical image analysis can be trained to recognize the wide variety of appearances of the anatomical structures and could be used to enhance the interpretation of anatomy by facilitating target identification (e.g., peripheral nerves and fascial planes). For example, a model has been well developed for peripheral nerve block in the adductor canal. In this model, the sartorius and adductor longus muscles, as well as the femur, were first identified as landmarks. The optimal block site is chosen as the region where the medial borders of these two muscles align. The femoral artery is labelled as both a landmark and a safety structure. The saphenous nerve is labelled as a target. AI-applications is assist the operator in identifying the nerve and the correct target site for the block.

Conclusions

This review explores important recent advances in ML and DL within neuroradiology. There have been many published studies exploring AI applications in neuroradiology, and the trend is accelerating. AI applications may cover multiple fields of neuroimaging/neuroradiology diagnostics, such as image quality improvement, image interpretation, classification of disease, and communication of salient findings to patients and clinicians. The DL tools show an outstanding ability to execute specific tasks at a level that is often compared to those of expert radiologists. In this context, AI may indeed have a role in enhancing radiologists' performance through a symbiotic interaction which is going to be more likely mutualistic. However, the existing AI tools in neuroradiology/neuroimaging have been trained for single tasks so far. This means that an algorithm trained to detect stroke would not be able to show similar accuracy to detect and classify brain tumours and vice-versa. This is a great limit since patients often suffer from multiple pathologies, a complete AI assessment

that integrates all different algorithms would be favourable. As far as we know, ML or DL models which are capable of simultaneously performing multiple interpretations have not yet been reported. We believe that this technology development may represent the key requirement to shift AI from an experimental tool to an indispensable application in clinical practice. AI algorithms for combined analysis of different pathologies should also warrant an optimized and efficient integration into the daily clinical routine. Furthermore, rigorous validation studies are still needed before these technological developments can take part in clinical practice, especially for imaging modalities such as MRI and CT, for which the accuracy of DL models highly depends on the type of scanner used and protocol performed. In addition, the reliability of AI techniques requires the highest validation also considering the legal liabilities that radiologists would hold for their usage and results. Nowadays, only specific DL applications have demonstrated accurate performance and may be integrated into the clinical workflow under the supervision of an expert radiologist. In particular, AI algorithms for intracranial haemorrhage, stroke, and vertebral compression fracture identification may be considered suitable for application in daily clinical routines. Other tasks, such as glioma genomics identification, stroke prognostication, epilepsy foci identification and predicting clinically definite MS, have shown significant progress in the research domain and may represent upcoming clinical applications in the not-so-distant future. However, the majority of AI algorithms show that there is still a range of inaccuracies for example in labelling anatomical structures, especially in the context of atypical or complex anatomy. Moreover, another challenge will be to ensure the presence of highly skilled practitioners, since machine learning systems are not guaranteed to outperform human performance and these systems should not be relied upon to replace the knowledge of doctors. Probably, the ongoing development of DL in neuroradiology/neuroimaging will significantly influence the work of future radiologists and other specialists, which will need a specific AI education to begin during residential training, to deeply understand the mechanisms and potential pitfalls. Furthermore, knowledge of AI could be an opportunity to improve training in radiology and other specialities. AI can assign specific cases to trainees based on their training profile, to promote consistency in the trainees' individual experiences, and, in the context of anaesthetic procedures, to facilitate an easier understanding of anatomy.

References

1. McDonald RJ, Schwartz KM, Eckel LJ *et al*. The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. *Acad Radiol.* 22, 1191-1198 (2015).

2. Hosny A, Parmar C, Quackenbush J *et al*. Artificial intelligence in radiology. *Nat Rev Cancer.* 18, 500-510 (2018).

3. D'Angelo T, Caudo D, Blandino A *et al*. Artificial intelligence, machine learning and deep learning in musculoskeletal imaging: Current applications. *J Clin Ultrasound JCU* (2022).

4. Gebru T, Morgenstern J, Vecchione B *et al*. Datasheets for Datasets (2021).

5. Montagnon E, Cerny M, Cadrin-Chênevert A *et al*. Deep learning workflow in radiology: a primer. *Insights Imaging.* 11, 22 (2020).

6. Zhu B, Liu JZ, Cauley SF *et al*. Image reconstruction by domain-transform manifold learning. *Nature.* 555, 487-492 (2018).

7. Dong C, Loy CC, He K *et al*. Learning a Deep Convolutional Network for Image Super-Resolution. (2014)

8. Chen H, Zhang Y, Kalra M *et al*. Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network. *IEEE Trans Med Imaging.* 36, 2524-2535 (2017).

9. Bahrami K, Shi F, Zong X *et al*. Reconstruction of 7T-Like Images From 3T MRI. *IEEE Trans Med Imaging* 35, 2085-2097 (2016).

10. Golkov V, Dosovitskiy A, Sperl JI *et al*. q-Space Deep Learning: Twelve-Fold Shorter and Model-Free Diffusion MRI Scans. *IEEE Trans Med Imaging.* 35, 1344-1351 (2016).

11. Cui J, Gong K, Han P *et a.l* Super Resolution of Arterial Spin Labeling MR Imaging Using Unsupervised Multi-Scale Generative Adversarial Network. (2020)

12. Olthof AW, Van Ooijen PMA, Rezazade Mehrizi MH *et al*. Promises of artificial intelligence in neuroradiology: a systematic technographic review. *Neuroradiology.* 62, 1265-1278 (2020).

13. Ginat, DT. Analysis of head CT scans flagged by deep learning software for acute intracranial hemorrhage. *Neuroradiology.* 62, 335-340 (2020).

14. Kuo W, Häne C, Mukherjee P *et al*. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. *Proc Natl Acad Sci USA.* 116, 22737-22745 (2019).

15. Chang PD, Kuoy E, Grinband J *et al*. Hybrid 3D/2D Convolutional Neural Network for Hemorrhage Evaluation on Head CT. *AJNR Am J Neuroradiol.* 39, 1609-1616 (2018).

16. Cho J, Park KS, Karki M *et al*. Improving Sensitivity on Identification and Delineation of Intracranial Hemorrhage Lesion Using Cascaded Deep Learning Models. *J Digit Imaging.* 32, 450-461 (2019).

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

18. Ye H, Gao F, Yin Y *et al*. Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. *Eur Radiol.* 29, 6191-6201 (2019).

19. Chilamkurthy S, Ghosh R, Tanamala S *et al*. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet Lond Engl.* 392, 2388-2396 (2018).

20. Yu Y, Xie Y, Thamm T *et al*. Use of Deep Learning to Predict Final Ischemic Stroke Lesions From Initial Magnetic Resonance Imaging. *JAMA Netw Open.* 3, e200772 (2020).

21. Ho KC, Scalzo F, Sarma KV *et al*. Predicting ischemic stroke tissue fate using a deep convolutional neural network on source magnetic resonance perfusion images. *J Med Imaging Bellingham Wash.* 6, 026001 (2019).

22. J H, Jg Y, H P *et al*. Machine Learning-Based Model for Prediction of Outcomes in Acute Stroke. *Stroke.* 50 (2019).

23. Nielsen A, Hansen MB, Tietze A *et al*. Prediction of Tissue Outcome and Assessment of Treatment Effect in Acute Ischemic Stroke Using Deep Learning. *Stroke.* 49,1394-1401 (2018).

24. Murray NM, Unberath M, Hager GD *et al*. Artificial intelligence to diagnose ischemic stroke and identify large vessel occlusions: a systematic review. *J Neurointerventional Surg.* 12, 156-164 (2020).

25. Alawieh A, Zaraket F, Alawieh MB *et al*. Using machine learning to optimize the selection of elderly patients for endovascular thrombectomy. *J Neurointerventional Surg.* 11, 847-851 (2019).

26. Sichtermann T, Faron A, Sijben R *et al*. Deep Learning-Based Detection of Intracranial Aneurysms in 3D TOF-MRA. *AJNR Am J Neuroradiol.* 40, 25-32 (2019).

27. Park A, Chute C, Rajpurkar P *et al*. Deep Learning-Assisted Diagnosis of Cerebral Aneurysms Using the HeadXNet Model. *JAMA Netw Open.* 2, e195600 (2019).

28. Stember JN, Chang P, Stember DM *et al*. Convolutional Neural Networks for the Detection and Measurement of Cerebral Aneurysms on Magnetic Resonance Angiography. *J Digit Imaging.* 32, 808-815 (2019).

29. Stone JR, Wilde EA, Taylor BA *et al*. Supervised learning technique for the automated identification of white matter hyperintensities in traumatic brain injury. *Brain Inj.* 30, 1458-1468 (2016).

30. Jain S, Vyvere TV, Terzopoulos V *et al*. Automatic Quantification of Computed Tomography Features in Acute Traumatic Brain Injury. *J Neurotrauma.* 36, 1794-1803 (2019).