3D Deep Learning on Medical Images: A Review

Satya P. Singh[1,4], Lipo Wang[2], Sukrit Gupta[3], Haveesh Goli[3], Parasuraman Padmanabhan[1,4] and Balázs Gulyás[1,4,5]

[1] Lee Kong Chian School of Medicine, Nanyang Technological University, 608232 Singapore,

[2] School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore

[3] School of Computer Science and Engineering, Nanyang Technological University, 639798, Singapore

[4] Cognitive Neuroimaging Centre, Nanyang Technological University, 636921 Singapore,

[5] Department of Clinical Neuroscience, Karolinska Institute, 17176 Stockholm, Sweden.

ABSTRACT The rapid advancements in machine learning, graphics processing technologies and availability of medical imaging data has led to a rapid increase in use of deep learning models in the medical domain. This was exacerbated by the rapid advancements in convolutional neural network (CNN) based architectures, which were adopted by the medical imaging community to assist clinicians in disease diagnosis. Since the grand success of AlexNet in 2012, CNNs have been increasingly used in medical image analysis to improve the efficiency of human clinicians. In recent years, three-dimensional (3D) CNNs have been employed for analysis of medical images. In this paper, we trace the history of how the 3D CNN was developed from its machine learning roots, give a brief mathematical description of 3D CNN and the preprocessing steps required for medical images before feeding them to 3D CNNs. We review the significant research in the field of 3D medical imaging analysis using 3D CNNs (and its variants) in different medical areas such as classification, segmentation, detection, and localization. We conclude by discussing the challenges associated with the use of 3D CNNs in the medical imaging domain (and the use of deep learning models, in general) and possible future trends in the field.

INDEX TERMS CNN, Machine learning, 3D Deep Learning, 3D Medical Imaging, 3D Convolutional Neural Networks.

I. INTRODUCTION

Medical images have varied characteristics depending on the target organ and the suspected diagnosis. Common modalities used for medical imaging include X-ray, computed tomography (CT), diffusion tensor imaging (DTI), positron emission tomography (PET), magnetic resonance imaging (MRI), and functional MRI (fMRI) [1]–[4]. In the past thirty years, these radiological image acquisition technologies have improved enormously in terms of acquisition time, image quality, resolution [5]–[9] and have become more affordable. Despite improvements in hardware, all radiological images require subsequent image analysis and diagnosis by trained human radiologists. Besides the significant time and economic costs involved in training radiologists, radiologists also suffer from limitations due to lack of experience, time and fatigue. This becomes especially significant because of an increasing number of radiological images due to aging population and more prevalent scanning technologies that put additional stress on radiologists. This puts focus on automated machine learning algorithms that can play a crucial role in assisting clinicians in alleviating their onerous workloads.

Historical methods for automated classification of images involved extensive rule-based algorithms or manual feature handcrafting [11]–[16], that were time-consuming, had poor generalization capacity and required domain knowledge. All this changed with the advent and demonstrated success of Convolutional Neural networks (CNNs) that were devoid of any manual feature handcrafting, required little preprocessing and are translation-invariant [17]. In CNNs, low-level image features are extracted by the initial layers of filters, and progressively higher features are learnt by successive layers before classification.

The commonly seen X-ray is an example of a two-dimensional (2D) medical image. The machine learning of these medical images is no different from CNNs applied to classify natural images in recent years, for example the ImageNet Large Scale Visual Recognition Competition [10]. With decreasing computational costs and more powerful graphics processing (GPU) units available, it has become

possible to analyze three-dimensional (3D) medical images, such as CT and MRI scans [10]. These scans give a detailed three-dimensional image of human organs and can be used to detect infection, cancers, traumatic injuries, and abnormalities in blood vessels and organs.

II. DEEP LEARNING BACKGROUND

Deep learning refers to learning patterns in data samples using neural networks containing multiple interconnected layers of artificial neurons [18]. An artificial neuron by analogy to a biological neuron is something that takes multiple inputs, performs a simple computation, and produces an output. This simple computation has the form of a linear function of the inputs followed by an activation function (usually non-linear) denoted by $f(.)$. Examples of some commonly used non-linear activation functions are the hyperbolic tangent (tanh), sigmoid transformation and the rectified linear unit (ReLU) and their variants [19]. The development of deep learning can be traced back to Walter Pitts and Warren McCulloch (1943). This was followed by significant advancements due to the the development of the backpropagation model (1960), convolutional neural networks (1979), LSTM (long short-term memory) (1997), ImageNet (2009), AlexNet (2011) [20], [21]. In 2014, Google presented GoogleNet (Winner of ILSVRC 2014 challenge) [22] which introduced the concept of inception modules (Fig. 2) that drastically reduced the computational complexity of the CNN. Deep learning is essentially a reincarnation of the artificial neural network where we stack layer upon layer of artificial neurons. Using the outputs of the terminal layers built on the outputs of previous layers, one can start to describe arbitrarily complex patterns. In the CNN [23], network features are generated by convolving kernels in a layer with the outputs of the previous layers, such that the first hidden layer kernels perform convolution on the input images. While the features captured by the initial hidden layers are generally in the form of shapes, curves, or edges the deeper hidden layers capture more abstract and complex features.

A complete CNN comprises four basic components: 1) Local receptive field, 2) Sharing weights, 3) Pooling, and 4) Fully connected (FC) layers. Deep CNN architecture is constructed by stacking several convolutional layers and pooling layers and one or so fully connected layers at the end of the network. Suppose we have a layer of $M \times M$ neurons followed by the convolutional layer ω with filter size $a \times b$. In order to compute the per-nonlinearity input to $(i,j)^{th}$ unit in layer $'\ell'$, $x_{i,j}^{\ell}$, we add up the weight contribution from the previous layer as follows:
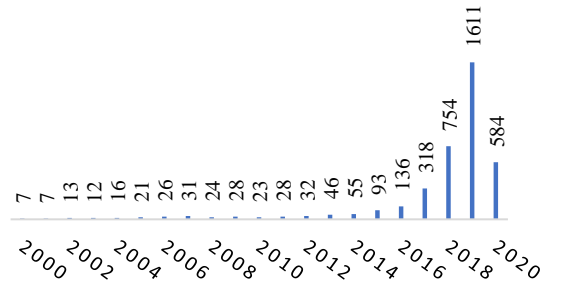
$$x_{i,j}^{\ell} = \sum_{a=0}^{n_i-1} \sum_{b=0}^{n_j-1} \omega_{a,b} \times x_{(i+a)(j+b)}^{\ell} + b_{i,j} \qquad (1)$$

Here, $b_{i,j}$ is the shared value of biases contribution. The output of the $(i,j)^{th}$ unit in the $'\ell^{th}'$ convolutional layer is given as follows:
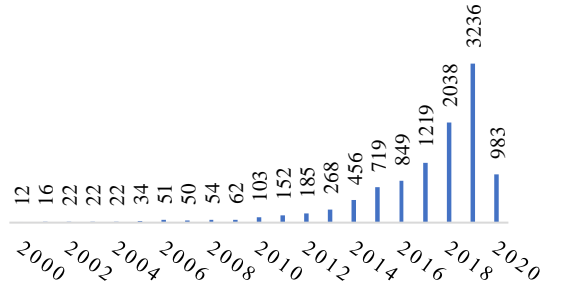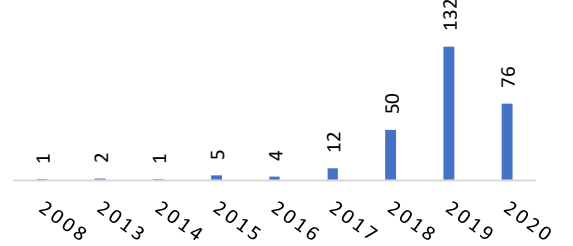

Fig. 1. Graphical illustration of results for different queries in title and abstracts on PubMed (as on 20th March 2020).

$$y_{i,j}^{\ell} = f(x_{i,j}^{\ell}) \qquad (2)$$

The size of the output of the convolutional layer becomes $(M - n + 1) \times (M - n + 1)$.

In a very short span of time, deep learning has become an alternative to several machine learning algorithms that were traditionally used in medical imaging. We did a search of the different terms used in medical imaging literature to understand the trend of usage of deep learning in medical imaging applications. We searched for 'machine learning + medical' in the title and abstract in PubMed publication database (*on 20th March 2020*). We came across a predictable trend of using more and more similar data to different approaches (Fig. 1). We observed a similar trend for the query 'deep learning + medical', albeit with few publications before 2015. However, while searching for the query '3D deep learning + medical' in the title and abstract, we see a different
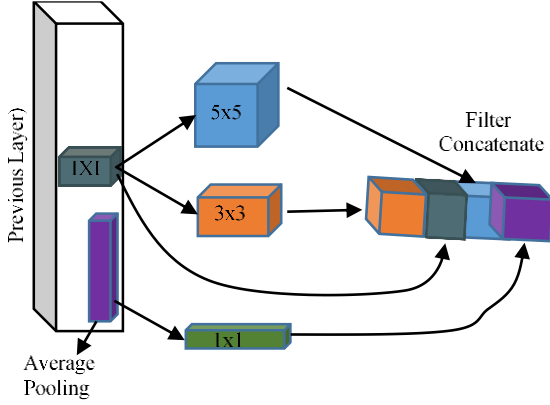
Fig. 2 Working example of the Inception Module in GoogleNet

scenario. An exponential increase can be seen for 'deep learning' and '3D deep learning' after the year 2015 and 2017 onwards, respectively. This signifies that while there was not much work in the domain a few years ago, there is fast rise in the number of publications related to deep learning for both 2D and 3D images.

III. 3D CONVOLUTIONAL NEURAL NETWORK

While 1D CNN can extract spectral features from the data, 2D CNN can extract spatial features from the input data. On the other hand, 3D CNN can take advantages of both 1D and 2D CNN by extracting both spectral and spatial features simultaneously from the input volume. These 3D CNN features are very useful in analyzing the volumetric data in medical imaging. The mathematical formulation of 3D CNN is very similar to 2D CNN with one extra dimension added. A basic architecture of 3D CNN is shown in Fig. 3. We briefly discuss the mathematical background of 3D CNN.

Convolutional Layer: The basic definition, principle, and working equation of 3D CNN same as 1D or 2D CNN. We only add an extra dimension (depth) to the working equation of 2D CNN. Suppose we have a layer of $M \times M \times D$ neurons followed by the convolutional layer ω with filter size $a \times b \times c$. In order to compute the nonlinear input $x_{i,j,k}^{\ell}$ to $(i, j, k)^{th}$ unit in layer ℓ, we add up the weight contribution from the previous layer as follows:

$$x_{i,j,k}^{\ell} = \sum_{a=0}^{n_i-1} \sum_{b=0}^{n_j-1} \sum_{c=0}^{n_k-1} \omega_{a,b,c} x_{(i+a)(j+b)(k+c)}^{\ell} + b_{i,j,k} \quad (3)$$

Pooling Layer: Each convolutional layer in 3D CNN may contain a pooling layer. Pooling layer simply takes multiple voxels (rectangular pixels in case of 2D CNN) and produces a single output to the input of the next layer by taking the average or maximum of the group of input voxels (pixels for 2D).

In backward pass, the CNN adjusts its weights and parameters according to the output by calculating the error by means of some loss functions, e (other names are cost function and error

function) and backpropagating the error with some rules towards the input. The loss is calculated by taking the partial derivative of e w.r.t. the output of each neuron in that layer such as $\partial e / y_{i,j,k}^{\ell}$ for the output, $y_{i,j,k}^{\ell}$ of $(i, j, k)^{th}$ unit in layer ℓ. Chain rule allow us to write the add up the contribution of each variable as follows:

$$\frac{\partial e}{\partial x_{i,j,k}^{\ell}} = \frac{\partial e}{\partial y_{i,j,k}^{\ell}} \frac{\partial f(y_{i,j,k}^{\ell})}{\partial x_{i,j,k}^{\ell}} = \frac{\partial e}{\partial y_{i,j,k}^{\ell}} f'(x_{i,j,k}^{\ell}) \quad (4)$$

Weights in the previous convolutional layer can be updated by backpropagating the error to the previous layer according to the following equation:

$$\frac{\partial e}{\partial y_{i,j,k}^{\ell-1}} = \sum_{a=0}^{n-1} \sum_{b=0}^{n-1} \sum_{c=0}^{n-1} \frac{\partial e}{\partial x_{(i-a),(j-b),(k-c)}^{\ell}} \frac{\partial x_{(i-a),(j-b),(j-b)}^{\ell}}{\partial y_{i,j,k}^{\ell-1}} \quad (5)$$

$$= \sum_{a=0}^{n-1} \sum_{a=0}^{n-1} \sum_{b=0}^{n-1} \frac{\partial e}{\partial x_{(i-a),(j-b),(k-c)}^{\ell}} \omega_{a,b,c} \quad (6)$$

Eq. (6) allows us to calculate the error for the previous layer. Also, above eq. makes sense for the only those points which are n times away from each side of the input data. This situation can be avoided by simply adding the zero padding to the end of each side of the input volume.

IV. 3D MEDICAL IMAGING PRE-PROCESSING

The preprocessing of the image dataset before feeding the CNN or other classifiers is important for all types of imaging modalities. However, it is more relevant for 3D medical imaging as the whole volume must be fed to the 3D CNN. Several preprocessing steps are recommended for the medical images before they are fed as input to the deep neural network model, such as 1) artifact removal, 2) normalization, 3) slice timing correction (STC), 4) image registration, and 5) bias field correction. Although all the steps through 1) to 5) help in getting reliable results, STC and image registration are very important in the case of 3D medical images (especially MR and CT images). Artifact removal and normalization are the most commonly performed preprocessing steps across modalities. We briefly discuss the above pre-processing steps.

A. REMOVING IMAGE ARTIFACTS

The first part of any preprocessing pipeline is the removal of artifacts. Removal of extra-cerebral tissues is highly recommended before analyzing the T1 or T2 weighted MRI, and DTI modalities for brain images. fMRI data often contains transient spikes artifacts or is slowed over drift time. Thus, the principal component analysis technique can be used to look at
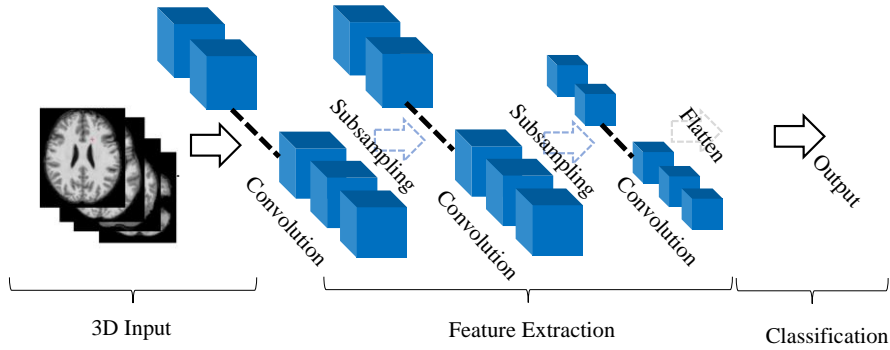
Fig. 3 Simple systematic architecture of a typical 3D CNN

these spike related artifacts [3], [24], [25]. Before feeding the data for preprocessing to an automated pipeline, a manual check is also advisable. For example, if the input T1 anatomical data is large in size, FSL BET command will not performs proper brain region extraction (Fig. 4) and if we use images with artifacts for the popular fMRI preprocessing tool fMRIprep [26], it fails as well. Therefore, to remove these extra neck tissue, we should perform other necessary steps for proper preprocessing.
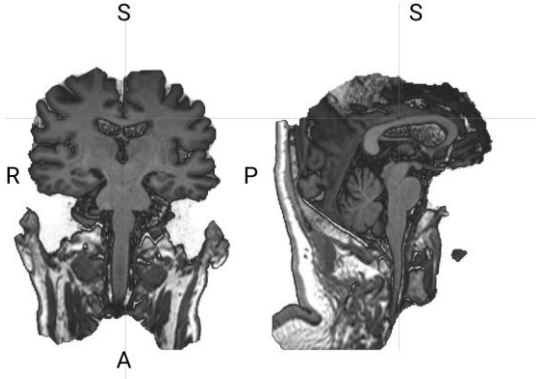

Fig. 4 The residual artifacts in T1 weighted MRI during brain segmentation using fsl BET tool command. One should need to be careful about these residuals before feeding the data to fully automatic pipeline for preprocessing.

B. NORMALIZATION

The brain and other body parts for imaging of every person can vary in shape and size. Hence it is advisable to normalize brain scans before further processing. [4], [27]–[30]. Due to the characteristics of MRI, essentially, the same scanning device can have different intensities even in the same patient's medical images. Since scanning of the patient may be performed in different light conditioning, intensity normalization also plays an important role in the performance of 3D CNN. Additionally, typically with CNN, each input channel (i.e. sequence) is normalized to have zero mean and unit variance within the training set. Parameter normalization within the CNN also affects the CNN performance.

C. SLICE TIME CORRECTION

In creating the volumetric representation of the brain, we often sample several slices in the brain during each individual repetition time (TR). However, each slice is typically sampled at slightly different time points as we acquire them sequentially [31], [32]. Hence, even though the 3D brain volume should be scanned instantaneously, in practical terms there is always some delay in sampling the first and the last slice. This is a key problem that needs to be considered and accounted for before performing any further analysis like classification, or segmentation.

In this regard, STC is frequently employed for adjusting the temporal misalignment and is widely utilized by a range of software such as SPM and FSL [33]. Several types of techniques have been proposed based on data interpolation methods for STC, including cubic spline, linear, and CNC interpolation [34]. In general, the STC methods based on interpolation techniques can be grouped as scene-based and object-based. In the scene-based approach, the interpolated pixel intensity is revealed by the pixel intensity of a slice. Although the interpolation techniques are sub-standard, they are relatively simple, intuitive, and easy to implement. On the other hand, the object-based methods have much better accuracy and are reliable but are computationally expensive. Subsequently, cubic spline and other polynomials were also found in medical image interpolation. Essentially, all these strategies perform strength averaging of the neighboring pixels without forming any feature deformation. Therefore, the resultant in-between pieces have blurring negative effects within the object boundary. Cubic interpolation is the standard technique selected in BrainVoyager [35] software.

D. IMAGE REGISTRATION

Medical imaging is becoming increasingly multimodal, whereby images of the same patient from different modalities are acquired to give information about different organ features. Additionally, situations also arise where multiple images of the same patient and location are acquired with

different orientations. It becomes necessary to match the images by visual comparison, in this case [36]. This alignment or registration of the images to a standard template can also be automated making it easier to locate repetitive locations where abnormalities due to a condition occur. The image alignment not only makes it easier to manually analyze images and locate lesions or other abnormalities, but also makes it easier to train a 3D CNN on these images [37]–[39].

E. BIAS FIELD CORRECTION

MRI images are corrupted by a low-frequency and smooth bias field signal that is produced by MRI scanners, affecting pixel intensities to fluctuate [40], [41]. Bias field usually appears due to improper image acquisition as well as the scanner, and influences machine learning algorithms that perform classification and segmentation using pixel intensities. It is, therefore, important to either remove the bias field artifacts from sample images or incorporate this artifact into the model before training a model on these images.

V. APPLICATIONS IN 3D MEDICAL IMAGING

A. SEGMENTATION

For several years, machine learning and artificial algorithms have been facilitating radiologists in the segmentation of medical images such as breast cancer mammograms, brain tumor, brain lesion, finding lung nodules, etc. Segmentation not only helps expert focus on specific regions in the medical image, but also helps expert radiologists in quantitative assessment, and planning further treatment. Several researchers have contributed to the use of 3D CNN in medical image segmentation. Here, we focus on the most important related work of medical image segmentation using 3D CNN.

Brain tumor/lesion/substructures: Lesion segmentation is probably the most challenging task in medical imaging because lesions are rather small in most of the cases. Also, there are considerable variations in their sizes in different scans which can create an imbalance the training samples. In this regard, recognizable work is *Deep Medic* [42], which was also the winner of the *ISLES 2015* competition. In DeepMedic, a 3D CNN architecture has been introduced for automatic brain lesion segmentation, which gives a state-of-the-art performance on 3D volumetric brain scans. The multiresolution approach has been utilized to include the local as well as the spatial contextual information. The network gives a 3D map of where the network believes the lesions are located. DeepMedic was implemented on datasets where patients suffered from traumatic brain injuries due to accidents and was also shown to work well for classification and detection problems in head images to detect brain tumor. This work was extended by Kamnitsas et al. [43] for the *BRATS 2016 challenge,* where the authors exploit the advantages of residual connections in 3D CNN. The results were impressive and were in the top 20 teams with median Dice scores of 0.898 (whole tumor, WT), 0.75 (tumor core, TC) and 0.72

(enhancing core, EC). In accordance with DeepMedic, Casamitjana et al. [46] proposed a 3D CNN to process the entire 3D volume in a single pass in making predictions.

Besides constraints with acquiring enough training samples, class imbalance also pervades in the medical imaging domain, whereby samples of the diseased patients are hard to come by. This issue is further exacerbated in problems related to tumor or lesion segmentation, because the size of tumors or lesions are usually small compared to the whole scan volume. In this context, Zhou et al. [44] proposed 3D CNN (3D variant of FusionNet) for brain tumor segmentation on *BRATS 2018 challenge*. The authors split the multiclass tumor segmentation problem into three separate segmentation tasks for the deep 3D CNN model i.e. i) Coarse segmentation for whole tumor, ii) Refined segmentation for Wavelet transform (WT) and intra class tumor, and iii) precise segmentation for brain tumor. Their model was ranked first for *BRATS 2015* dataset and third (among 64 teams) on the *BRATS 2017* validation dataset. Ronneberger et al. proposed the *U-Net* architecture for segmentation of 2D biomedical images [45]. They make use of up-sampling layers which in turn enable the architecture to be used for segmentation besides classification. However, the original U-Net network was not too deep as there was a single pooling after the convolution layer. Additionally, this only analyzed 2D images and did not fully exploit the spatial and texture information that can be obtained from the 3D volumetric. To solve these issues, Chen et al. [47] proposed a separable 3D U-Net for brain tumor segmentation. On *BRATS 2018 challenge* dataset, they achieved Dice scores of 0.749 (EC), 0.893 (WT) and 0.830 (TC). Kayalıbay et al. [48] presented a modified 3D U-Net architecture for brain tumor segmentation where they introduce some nonlinearity in the traditional U-Net architecture by inserting residual blocks during up-sampling, thus facilitating the gradients to flow easily. The proposed architecture also intrinsically handles the class imbalance problem due to use of Jaccard loss function. However, the proposed architecture was computationally expensive owing to the large receptive field size used. Isensee et al. [49] proposed a 3D U-Net architecture which consists of a perspective collection pathway for brain tumor segmentation. The strategy encodes progressively abstract interpretations of the input as we move deeper and adds a localization pathway that recombines these interpretations with features for lower layers. By hypothesizing that semantic features are easy to learn and process, Peng et al. [50] presented a multi scale 3D U-Net for brain tumor segmentation. Their model consists of several 3D U-Net blocks for capturing long distance spatial resolutions. The up sampling was done at different resolutions in order to capture meaningful features. On the *BRATS 2015 challenge* dataset they achieved 0.893 (WT), 0.830 (TC), 0.742 (EC).
While brain tumor or lesion segmentation are used in to detect glioblastoma, brain stroke or traumatic brain injuries, multiple deep learning solutions are being proposed for segmentation of brain lobes or deep brain structures. Milletari et al. [51]

Table I Important developments in 3D CNN for Brain tumor/Lesion segmentation on BRAST Challenges.

Ref.	Methods	Data	Task	Performance evaluation
Zhou et al. [44]	3D variant of FusionNet (One-pass Multi-task Network (OM-Net))	BRATS 2018	brain tumor segmentation	91.59 (WT), 82.74 (TC), 80.73(EC)
Chen et al. [47]	Separable 3D U-Net	BRATS 2018	--do--	0893(WT), 0.830(TC), 0.742(EC)
Peng et al. [50]	Multi-Scale 3D U-Nets	BRATS 2015	--do--	085(WT), 0.72(TC), 0.61(EC)
Kayalıbay et al. [48]	3D U-Nets	BRATS 2015	--do--	085 (WT), 0.872(TC), 0.61(EC)
Kamnitsas et al. [43]	11 layers deep 3D CNN	BRATS 2015 and ISLES 2015	--do--	0.898 (WT), 0.75 (TC), 0.72(EC)
Kamnitsas et al. 2016 46 [42]	3D CNN in which features extracted by 2D CNNs	BRATS 2017	--do--	0.918 (WT), 0.883(TC), 0.854 (EC)
Casamitjana et al. [46]	3D U-Net followed by fully connected 3D CRF	BRATS 2015	--do--	91.74(WT), 83.61(TC), 76.82(EC)
Isensee et al.53 [49]	3D U-Nets	BRATS 2017	--do--	085(WT), 0.74(TC), 0.64(EC)

combined a Hough voting approach with 2D, 2.5D, and 3D CNN to segment volumetric data of MRI scans. However, these networks still suffer from the class imbalance problem. In [52], a 3D CNN was implemented for subcortical brain structure segmentation in MRI and this study was based on the effect of the size of kernels in a network. In [53], the authors applied 3D U-Net for dense volume segmentation. However, this network was not entirely in 3D because it used 2D annotated slices for the training of the network. Sato et al. [54] proposed 3D deep network for segmentation of head CT volume. Some important developments in 3D CNN for Brain tumor/lesion segmentation applications on BRATS Challenges are summarized in Table I.

Other organs: Liver cancer is one of the major causes of cancer deaths worldwide. Therefore, a reliable and computerized liver tumor segmentation technique are strongly needed to assist the expert radiologist and doctors for hepatocellular carcinoma identification and management. Duo et al. [58] presented a fully connected 3D CNN for liver segmentation from 3D CT scans. The same network was also tested on whole heart and vessel segmentation. Further, 3D U-Net was applied in liver segmentation problems [59]. In [60], *3D ResNet* was used for liver segmentation using the coarse to fine approach. Some other similar approaches for segmentation of the liver can be found in [32], [61]–[63]. In this sequence, another work based on 2D *DenseUnet* and hierarchical diagnosis approach (*H-DensNet*) for segmentation of liver lesions were presented in [64]. This network secured the first position in the *LiTS 2017 Leaderboard*. The network was tested on 3DIRCADs database and achieved state-of-the-art outcomes compared to other very well-established liver segmentation approaches. They achieved 98.2% and 93.7% accuracy on Dice for liver and tumor segmentation respectively.

3D CNNs are also being used in segmentation of knee structures. In [55], Ambellan et al. proposed a technique with 3D Statistical Shape Models along with 2D in addition to 3D CNN's to accomplish an effective and precise segmentation of knee structures. In [56], the the authors suggest a 3D CNN to segment cervical tumors on 3D PET images. Their architecture uses prior information constraint spatial information for segmentation purpose. Authors claim highly precise results for segmenting cervical tumors on the 3D PET. In [57], the authors propose 3D convolution kernels for learning filter coefficients and spatial filter offsets simultaneously for 3D CT multi-organ segmentation work. The outcomes were compared with U-Net architectures. Authors claim that their architecture needs less trainable parameters and storage while obtaining high quality.

B. CLASSIFICATION

Classification of diseases using deep learning technologies on medical images has gained a lot of traction in the last few years. For neuroimaging, the major focus of 3D deep learning has been on detecting diseases from anatomical images. Several studies have focused on detecting dementia and its different variants from different imaging modalities including functional MRI. Alzheimer's Disease (AD) is the most common form of dementia, usually linked with the pathological amyloid depositions, structural-atrophy and metabolic variations in the chemistry of the brain. The timely diagnosis of AD plays an important role in slowing, avoiding, and preventing the progression of the disease.

Yang et al. [28] visualized the 3D CNN trained for classifying AD in terms of AD features which can be a very good step in understanding the behavior of each layer of 3D CNN. They proposed three types of visual inspection approaches: 1) based on sensitivity analysis, 2) 3D class activation mapping, and 3)

Table II Important developments in3D CNN for classification task in medical imaging.

Ref.	Task	Model	Data	Performance measures
Yang et al. [28]	AD classification	3D VggNet, 3D Resnet	MRI scans from ADNI dataset (47 AD, 56 NC)	0.863 AUC using 3D VggNet and 0.854 AUC using 3D Resnett
Kruthika et al. [67]	--do--	3D capsule network, 3D CNN	MRI scans from ADNI dataset (345 AD, NC, 605, and 991MCI)	Acc. for AD/MCI/NC 89.1%
Feng et al. [68]	--do--	3D CNN + LSTM	PET + MRI scans from ADNI dataset (93 AD, 100 NC)	Acc. 65.5% (sMCI/NC), 86.4% (pMCI/NC), and 94.8 % (AD/NC)
Wegmayr et al. [69]	--do--	3D CNN	ADNI and AIBL data sets, 20000 T1 scans	Acc. 72% (MCI/AD), 86 % (AD/NC), and 67 % (MCI/NC)
Oh et al. [66]	--do--	3D CNN +transfer learning	MRI scans from ADNI dataset (AD 198, NC 230, pMCI 166, and sMCI 101) at baseline.	74% (pMCI/sMCI), 86% (AD/NC), 77% (pMCI/NC)
Parmar et al. [65]	--do--	3D CNN	fMRI scans from ADNI dataset (30 AD, 30 NC)	Classification acc. 94.85 % (AD/NC)
Nie et al. [70]	Brain tumor	3D CNN with learning supervised features	Private adat 69 patient (T1 MRI, fMRI and DTI)	Classification acc. 89.85 %
Amidi et al. [103]	Protein shape	2-layer 3D CNN	63,558 enzymes from PDB data sets	Classification acc. 78%
Zhou et al. [71]	Breast cancer	Weakly supervised 3D CNN	Private, 1537 female	Classification acc. 78% 83.7%

3D weighted gradient weighted mapping. Authors explains how visual inspection can improve the accuracy and the possible improvements in deciding the 3D CNN architecture. In this work, some well-known baseline 2D deep architectures such as VGGNet and ResNet were converted to their 3D counterparts and classification of AD using MRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) was performed. In [67], the authors trained an auto-encoder to derive an embedding from input features from 3D patches extracted from the preprocessed MRI scans downloaded from the ADNI dataset and demonstrated an improvement in results in comparison to 2D approaches available in the literature. In [68], authors stacked recurrent neural network (long short-term memory) layers on 3D CNN layers for AD classifications using PET and MRI data. The 3D fully connected CNN layers obtained deep feature representations and the LSTM was applied on these features for performance improvement. In [69], a deep 3D CNN has been researched on a sizeable dataset for classification of AD. Gao et al. [75] show 87.7% accuracy in classification of AD, lesion, and normal aging by implementing 7 layers deep 3D CNN on 285 volumetric CT head scans from Navy General hospital, China. In this study, the authors also compared their results from 3D CNN with hand crafted features of 3D scale invariant Fourier transform (SIFT) and show that the proposed 3D CNN approach gives around four percent higher classification accuracy.

Besides detecting AD using head MRI (or other modalities), multiple studies have been performed for detecting diseases from varied organs in the body. Nie et al. [70] take advantage of the 3D aspect of MRI through training a 3D CNN to evaluate the survival in patients going through high-grade gliomas. Zhou et al. [71] proposed a weakly supervised 3D CNN for breast cancer detection. However, there were several limitations with the study: 1) the data was selective in nature,

2) The proposed architecture was only able to detect the tumor with high probability, and 3) only structural features were used for the experiments. Jnawali et al. [30] demonstrated the performance of 3D CNN in the classification of CT brain hemorrhage scans. The authors constructed three versions of 3D architectures based on CNNs. Two of these architectures are 3D versions of the VggNet and GoogleNet. This unique research was done on a large private dataset and about 87.8% accuracy was demonstrated. In [76] Ker et al. developed a 3 layer shallow 3D CNN for brain hemorrhage classification. The proposed network was giving state-of-the-art results with small training time compared to 3D VGGNet and 3D GoogleNet. Ha et al. [72] modify 2D U-Net into 3D CNN to quantify the breast MRI fibro-glandular tissue (FGT) and background parenchymal enhancement (BPE). In [58], Nie et al. proposed a multi-channel structure of 3D CNN for survival time prediction of Glioblastoma patients using multi-modal head images (T1 weighted MRI and diffusion tensor imaging, DTI). Recently, in [73], the author presented a hybrid model for classification and prediction of LNM in head and neck cancer. They combined the outputs of MaO-radiomics and 3D CNN architecture by using an ER fusion strategy. In [74], the authors presented a 3D CNN for predicting the maximum standardized uptake value of lymph nodes in patients suffering from cancer using CT images from a PET/CT examination. We summarized some important developments in 3D deep learning models for classification task in medical imaging in Table II.

C. DETECTION
Cerebral Microbleeds (CMBs) are small foci of chronic blood hemorrhages that can occur in the normal brain due to structural abnormalities of small blood vessels in the brain. Due to the differential properties of blood, MRI can detect CMBs. However, detecting cerebral micro-hemorrhages in brain tissue is a difficult and time-consuming task for

radiologists, and recent studies employed 3D deep architectures to detect CMBs. and. Dou et al. [77] proposed a two-stage fully connected 3D CNN architecture to detect CMBs from the dataset of MRI susceptibility weighted images (SWI). The network reduced many false positive candidates. For training purposes, multiple 3D cubes were extracted from the preprocessed dataset. This study also examines the effect of the size of 3D patches on network performance. This study also focuses on the higher performance of 3D architectures in the detection of CMBs in comparison to their 2D architectures such as Random Forest and 2D-CNN-SVM. Dou et. al. further employed a fully 3D CNN to detect microscopic areas of a brain hemorrhage on MRI brain scans [78]. This method had a sensitivity of 93% and outperformed prior methods of detection. Standvoss et al. [79] detected CMBs in traumatic brain injury. In their study, the authors prepared three types of 3D architectures with varying depth i.e. 3, 5 and 8 layers. These models were quite simple and straight forward, with the overall best accuracy of 87%. The drawback of these studies was that they utilized a small dataset for training the network. In [80], the author presented a 3D CNN to forecast route and radius of an artery at any given point in a cardiac CT angiography image which depends on the local image patch. This approach has the capacity to precisely and effectively figure out the path and radius of coronary arteries according to details extracted through the image files.

D. LOCALIZATION

Localization of biological architectures is a basic requirement for various initiatives in medical image investigation. Localization might be a hassle-free process for the radiologist, but it is usually a hard task for NNs that are vulnerable to variation in medical images induced by dissimilarities in the image acquisition process, structures, and pathological differences among patients. Generally, a 3D volume is required for the localization in medical images. Several techniques treat the 3D space as an arrangement of 2D orthogonal planes. Wolterink et al. [81] detected coronary artery calcium scoring in coronary CT angiography using a CNN based architecture. De Vos et al. [82] introduced localization technique using a solitary CNN, and 2D CT image slices (chest CT, cardiac CT, and abdomen CT) as input. Although, this work was related to a 3D localization approach, but they didn't use 3D CNN in a real sense. In addition, the approach depended heavily on the accurate recognition of biological structures. Huo et al. [83] utilized the properties of a 3D fully connected CNN and presented a spatially localized atlas network tiles (SLANT) model for whole brain segmentation on high-resolution multi-site images.

Intervertebral discs (IVDs) are modest joint parts that are located in between surrounding vertebrae and the localization of IVDs, are usually important for spine disease analysis and measurement. In [84], the authors presented a 3D detection of multiple brain structures in fetal neuro-sonography using fully connected CNN and named it VP-Nets. They explained that

the proposed strategy requires a comparatively less amount of data for training and can learn from coarsely annotated 3D data. Recently, a 3D CNN based on regression has been introduced in [31] to assess the degree of enlarged perivascular spaces (EPVS) through 2000 basal ganglia scans from 3D head MRI. In [85], the authors reported the human level efficiency of 3D CNN in landmark detection in clinical 3D CT data. In [86], Saleh et al. proposed a 3D CNN based regression models for 3D pose estimation of anatomy using T2 weighted imaging. They showed that the proposed network offers fine initialization for optimization-based techniques to increase the capture range of slice-to-volume registration. Xiaomeng et al. [87] presented fully connected, accurate and automatic 3D deep architecture for localization and segmentation of IVDs using multimodal MR images. The work shows state-of-the-art performance in *MICCAI-2016 challenge* for IVDs localization and segmentation section with dice score 91.2% for IVD segmentation.

VI. CHALLENGES AND CONCLUSIONS

It takes a large number of training samples to train deep learning models [42], [88], [89]. This is further strengthened by the recent successes of deep learning models trained on large datasets like the ImageNet. However, it is still ambiguous whether deep learning models can successfully work with smaller datasets, as in the case of medical images. The ambiguity is caused by the nature and characteristics of medical images. For example, the images from the ImageNet dataset possess large variations in their appearance (e.g. light, intensity, edges, color, etc.) [23], [25], [90]–[92] since the images were taken at different angles and distances and have several different features that are completely different from medical images. Therefore, networks needed to learn meaningful representations of these images require huge training parameters and thus training samples. However, in case of medical images, there is much less variation in comparison to traditional image datasets [93]. In this regard, the process of fine-tuning of 3D CNN models which are already trained on natural image dataset can be applied to medical image [23], [25], [90]–[92], [94], [95]. This process, known as *transfer learning*, has been successfully applied to many areas of medical imaging.

Regardless of their high computational complexity, 3D deep networks have shown incredible performance in diverse domains. 3D deep networks require large number of training parameters which becomes more severe in the case of 3D medical images where the depth of the image volume varies roughly from 20 to 400 slices per scan [9], [25], [70], [96] with each scan containing very fine and important information about the patient. Usually, high-resolution scan volumes are of the size of 512x512 and need to be downsampled before being fed to the 3D network in order to reduce the computational cost. Researchers generally use interpolation techniques to reduce the overall size of these medical image volumes but on the cost of significant information loss. There are also

restrictions on the resizing of the medical image volume without loss of significant information. This is still an unexplored area and there is further research scope.

Most of the 3D deep network architectures involve basic convolution or modifications of convolution layers. Although the number of trainable parameters of convolutional layers are independent of the input size, but the number of trainable parameters in the subsequent fully connected layers depend on the output of the convolution layers. This often leads to intractable models due to large number of trainable weights in the case when input images are fed into 3D CNN models without any downsampling. However, this issue is not the case with 2D images, that have smaller latent representations learnt by convolution filters. This makes it harder (and more GPU intensive) to train 3D deep networks based on CNNs. The inception module by GoogleNet can be further explored in the concern of computational complexity in 3D medical image analysis.

As mentioned earlier, the depth of medical image volumes approximately varies between 20 and 400. For 3D CNN, we put the whole volume as the input to the 3D CNN. In most of the cases, only a few slices show abnormalities and therefore a lot of unnecessary volumes are fed to the model for most of the cases. However, for most cases we have labels for the entire scan and not for each image slice. Therefore, methods that choose what data to feed into a model can be investigated.

Indeed, in the deep learning context, learning the right features might sound unconventional because we cannot be sure if the models learn features that are indeed discriminating for the condition or just overfit on some specific features for the given dataset. CNNs can handle raw image data and they do not need handcrafted and designing the features [18], [90]. It is the responsibility of CNN to discover the right features from the data. While CNNs have made *encoding* the raw features in a latent space very convenient, it is very important to understand whether the CNN learnt features that are generalizable across datasets. Machine learning models often overfit on train samples, whereby they only perform well on the test samples from the training dataset. This issue is acute in case of medical imaging applications where there are issues with scanner variability, scan acquisition settings, subject demography and heterogeneity in disease characteristics across subjects. Therefore, it is important to *decode* the trained network using model interpretability approaches and validate the important features learnt by the network [97]. It also becomes important to report testing results with an external dataset whose samples were not used for training. However, this may not always be possible because of paucity of datasets for training and testing.

Finally, the ultimate challenge is to go beyond a human-level performance. Researchers are working on reaching human-level performance for many tasks (known as Artificial General Intelligence) [24], [42], [98], [99]. However, the lack of labelled images, the high costs involved in labeling the datasets, the lack of consensus among experts in the assigned labels [27], [100], [101] are some present challenges that face the field. These issues force us to consider using reliable data augmentation methods and generate samples with known ground-truths. In this regard, generative adversarial networks (GAN) [102], especially CycleGANs for cross-modal image synthesis, offer a viable approach for synthesizing data and have been used to produce pseudo images that are highly similar to the original dataset.

REFERENCES
[1] K. Doi, "Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential," *Comput. Med. Imaging Graph.*, vol. 31, no. 4–5, pp. 198–211, 2007, doi: 10.1016/j.compmedimag.2007.02.002.

[2] A. S. Miller, B. H. Blott, and T. K. hames, "Review of neural network applications in medical imaging and signal processing," *Med. Biol. Eng. Comput.*, vol. 30, no. 5, pp. 449–464, Sep. 1992, doi: 10.1007/BF02457822.

[3] M. P. Siedband, "Medical imaging systems," *Med. Instrumentation-Application Des.*, pp. 518–576, 1998.

[4] J. Prince and J. Links, "Medical imaging signals and systems," *Med. Imaging*, pp. 315–379, 2006, doi: 0132145189.

[5] R. S. Shapiro, J. Wagreich, R. B. Parsons, A. Stancato-Pasik, H. C. Yeh, and R. Lao, "Tissue harmonic imaging sonography: Evaluation of image quality compared with conventional sonography," *Am. J. Roentgenol.*, vol. 171, no. 5, pp. 1203–1206, Nov. 1998, doi: 10.2214/ajr.171.5.9798848.

[6] K. Matsumoto, M. Jinzaki, Y. Tanami, A. Ueno, M. Yamada, and S. Kuribayashi, "Virtual Monochromatic Spectral Imaging with Fast Kilovoltage Switching: Improved Image Quality as Compared with That Obtained with Conventional 120-kVp CT," *Radiology*, vol. 259, no. 1, pp. 257–262, Apr. 2011, doi: 10.1148/radiol.11100978.

[7] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–4544, Oct. 2007, doi: 10.1118/1.2789499.

[8] D. Marin *et al.*, "Low-Tube-Voltage, High-Tube-Current Multidetector Abdominal CT: Improved Image Quality and Decreased Radiation Dose with Adaptive Statistical Iterative Reconstruction Algorithm—Initial Clinical Experience," *Radiology*, vol. 254, no. 1, pp. 145–153, Jan. 2010, doi: 10.1148/radiol.09090094.

[9] J. J.-B. Thibault, K. K. D. Sauer, C. A. C. C. A. Bouman, J. H.-M. Physics, U. 2007, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Wiley Online Libr.*, vol. 34, no. 11, pp. 4526–4544, Oct. 2007, doi: 10.1118/1.2789499.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *ImageNet Classification with Deep Convolutional Neural Networks*, 2012, pp. 1097–1105.

[11] S. C. H. Hoi, R. Jin, J. Zhu, and M. R. Lyu, "Batch mode active learning and its application to medical image classification," in *23rd international conference on machine learning*, 2006, pp. 417–424, doi: 10.1145/1143844.1143897.

[12] M. M. Rahman, P. Bhattacharya, and B. C. Desai, "A Framework for Medical Image Retrieval Using Machine Learning and Statistical Similarity Matching Techniques With Relevance

Feedback," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 1, pp. 58–69, 2007, doi: 10.1109/TITB.2006.884364.

[13] M. Wernick, Y. Yang, J. Brankov, G. Yourganov, and S. Strother, "Machine Learning in Medical Imaging," *IEEE Signal Process. Mag.*, vol. 27, no. 4, pp. 25–38, Jul. 2010, doi: 10.1109/MSP.2010.936730.

[14] E. Criminisi, A., Shotton, J., & Konukoglu, "Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning," *Found. Trends® Comput. Graph. Vision,* vol. 7, no. 2–3, pp. 81–227, 2012.

[15] S. P. Singh and S. Urooj, "An Improved CAD System for Breast Cancer Diagnosis Based on Generalized Pseudo-Zernike Moment and Ada-DEWNN Classifier," *J. Med. Syst.*, vol. 40, no. 4, p. 105, Apr. 2016, doi: 10.1007/s10916-016-0454-0.

[16] S. Urooj, S. S.-C. for S. Global, and undefined 2015, "Rotation invariant detection of benign and malignant masses using PHT," *ieeexplore.ieee.org*.

[17] S. Ji, W. Xu, M. Yang, and K. Yu, "3D Convolutional Neural Networks for Human Action Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013, doi: 10.1109/TPAMI.2012.59.

[18] D. Shen, G. Wu, and H.-I. Suk, "Deep Learning in Medical Image Analysis.," *Annu. Rev. Biomed. Eng.*, vol. 19, no. 1, pp. 221–248, Jun. 2017, doi: 10.1146/annurev-bioeng-071516-044442.

[19] S. H. Wang, P. Phillips, Y. Sui, B. Liu, M. Yang, and H. Cheng, "Classification of Alzheimer's Disease Based on Eight-Layer Convolutional Neural Network with Leaky Rectified Linear Unit and Max Pooling," *J. Med. Syst.*, vol. 42, no. 5, p. 85, May 2018, doi: 10.1007/s10916-018-0932-7.

[20] V. K. A. M. Z. T. T. M. Y. C. W. S. H. M. V. E. B. C. A. A.A.S. and Asari, "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," *arXiv Prepr. arXiv1803.01164*, 2018.

[21] J. Boroumand, M., Fridrich, "Deep learning for detecting processing history of images," *Electron. Imaging*, vol. 7, pp. 1–9, 2018.

[22] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, doi: 10.1109/CVPR.2015.7298594.

[23] A. Krizhevsky, Ii. Sulskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.*, vol. 60, no. 6, pp. 84–90, 2012, doi: 10.1145/3065386.

[24] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep Learning Applications in Medical Image Analysis," *IEEE Access*, pp. 1–1, 2018, doi: 10.1109/ACCESS.2017.2788044.

[25] J. Burt, "Volumetric quantification of cardiovascular structures from medical imaging," 9,968,257, 2018.

[26] O. Esteban *et al.*, "fmriprep: A Robust Preprocessing Pipeline for fMRI Data — fmriprep version documentation," *Nat. Methods*, pp. 111–116, 2019.

[27] A. Alansary *et al.*, "Fast Fully Automatic Segmentation of the Human Placenta from Motion Corrupted MRI," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 589–597, doi: 10.1007/978-3-319-46723-8_68.

[28] C. Yang, A. Rangarajan, and S. Ranka, "Visual Explanations From Deep 3D Convolutional Neural Networks for Alzheimer's Disease Classification," *arXiv Prepr. arXiv1803.02544.*, Mar. 2018.

[29] D. K. Jones *et al.*, "Spatial Normalization and Averaging of Diffusion Tensor MRI Data Sets," *Neuroimage*, vol. 17, no. 2, pp. 592–617, Oct. 2002, doi: 10.1006/nimg.2002.1148.

[30] K. Jnawali, M. Arbabshirani, and N. Rao, "Deep 3D convolution neural network for CT brain hemorrhage classification," in *Medical Imaging 2018: Computer-Aided Diagnosis, spiedigitallibrary.org*, 2018, p. 105751C.

[31] F. Dubost, H. Adams, G. Bortsova, and M. Ikram, "3D Regression Neural Network for the Quantification of Enlarged Perivascular Spaces in Brain MRI," *arXiv Prepr. arXiv1802.05914*, 2018.

[32] C. Lian, M. Liu, J. Zhang, X. Zong, W. Lin, and D. Shen, "Automatic Segmentation of 3D Perivascular Spaces in 7T MR Images Using Multi-Channel Fully Convolutional Network," *Elsevier*, pp. 5–7, 2018.

[33] R. Pauli, A. Bowring, R. Reynolds, G. Chen, T. E. Nichols, and C. Maumet, "Exploring fMRI Results Space: 31 Variants of an fMRI Analysis in AFNI, FSL, and SPM," *Front. Neuroinform.*, vol. 10, 2016, doi: 10.3389/fninf.2016.00024.

[34] D. Parker, X. Liu, and Q. R. Razlighi, "Optimal slice timing correction and its interaction with fMRI parameters and artifacts," *Med. Image Anal.*, vol. 35, pp. 434–445, 2017, doi: 10.1016/j.media.2016.08.006.

[35] R. Goebel, "BrainVoyager - Past, present, future," *Neuroimage*, vol. 62, no. 2, pp. 748–756, Aug. 2012, doi: 10.1016/j.neuroimage.2012.01.083.

[36] F. Maes, A. Collignon, D. Vandemeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *ieeemi*, vol. 16, no. 2, pp. 187–198, 1997.

[37] M. J.B.A. and V. M. A., "A survey of medical image registration," *Med Image Anal*, vol. 2, no. 1, pp. 1–36, 1998, doi: http://dx.doi.org/10.1016/S1361-8415(01)80026-8.

[38] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Interpolation Artefacts in Mutual Information Based Image Registration," *Comput. Vis. Image Underst.*, vol. 77, no. 2, pp. 211–232, 2000.

[39] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes, "A comparison of similarity measures for use in 2-D-3-D medical image registration," *Med. Imaging, IEEE Trans.*, vol. 17, no. 4, pp. 586–595, 1998, doi: 10.1109/42.730403.

[40] and T. M. Ahmed, Mohamed N., Sameh M. Yamany, Nevin Mohamed, Aly A. Farag, "A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Trans. Med. Imaging*, vol. 21, no. 3, pp. 193–199, 2002.

[41] C. Li, C. Xu, A. W. Anderson, and J. C. Gore, "MRI tissue classification and bias field estimation based on coherent local intensity clustering: A unified energy minimization framework," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, vol. 5636 LNCS, pp. 288–299, doi: 10.1007/978-3-642-02498-6_24.

[42] K. Kamnitsas *et al.*, "DeepMedic for Brain Tumor Segmentation," in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 2016, pp. 138–149, doi: 10.1007/978-3-319-55524-9_14.

[43] K. Kamnitsas *et al.*, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017, doi: 10.1016/j.media.2016.10.004.

[44] C. Zhou, C. Ding, X. Wang, Z. Lu, and D. Tao, "One-pass Multi-task Networks with Cross-task Guided Attention for Brain Tumor Segmentation," *IEEE Trans. Image Process.*, pp. 1–1, 2020, doi: 10.1109/TIP.2020.2973510.

[45] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9351, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[46] A. Casamitjana, S. Puch, A. Aduriz, & Sayrol, E., and V. Vilaplana, "3d convolutional networks for brain tumor segmentation," in *MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS)*, 2016, pp. 65–68.

[47] W. Chen, B. Liu, S. Peng, J. Sun, and X. Qiao, "S3D-UNET: Separable 3D U-Net for brain tumor segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11384 LNCS, pp. 358–368, doi: 10.1007/978-3-030-11726-9_32.

[48] B. Kayalibay, G. Jensen, and P. van der Smagt, "CNN-based Segmentation of Medical Imaging Data," *arXiv Prepr. arXiv1701.03056*, Jan. 2017.

[49] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H.

10

Maier-Hein, "Brain tumor segmentation and radiomics survival prediction: Contribution to the BRATS 2017 challenge," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 10670 LNCS, pp. 287–297, doi: 10.1007/978-3-319-75238-9_25.

[50]   S. Peng, W. Chen, J. Sun, and B. Liu, "Multi-Scale 3D U-Nets: An approach to automatic segmentation of brain tumor," *Int. J. Imaging Syst. Technol.*, vol. 30, no. 1, pp. 5–17, Mar. 2019, doi: 10.1002/ima.22368.

[51]   F. Milletari *et al.*, "Hough-CNN: Deep Learning for Segmentation of Deep Brain Regions in MRI and Ultrasound," *Comput. Vis. Image Underst.*, vol. 164, pp. 92–102, 2017.

[52]   J. Dolz and C. Desrosiers, "3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study," *Neuroimage*, 2017.

[53]   Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in *Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 424–432, doi: 10.1007/978-3-319-46723-8_49.

[54]   D. Sato *et al.*, "A primitive study on unsupervised anomaly detection with an autoencoder in emergency head CT volumes," in *Medical Imaging 2018: Computer-Aided Diagnosis*, 2018, p. 60, doi: 10.1117/12.2292276.

[55]   F. Ambellan, A. Tack, M. Ehlke, and S. Zachow, "Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative," *Med. Image Anal.*, vol. 52, pp. 109–118, 2019, doi: 10.1016/j.media.2018.11.009.

[56]   and J. W. Chen, Liyuan, Chenyang Shen, Zhiguo Zhou, Genevieve Maquilan, Kevin Albuquerque, Michael R. Folkert, "Automatic PET cervical tumor segmentation by deep learning with prior information," in *Physics in medicine and biology*, 2019, p. 111, doi: 10.1117/12.2293926.

[57]   M. P. Heinrich, O. Oktay, and N. Bouteldja, "OBELISK-Net: Fewer layers to solve 3D multi-organ segmentation with sparse deformable convolutions," *Med. Image Anal.*, vol. 54, pp. 1–9, May 2019, doi: 10.1016/j.media.2019.02.006.

[58]   Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, and … J. Q., "3D deeply supervised network for automated segmentation of volumetric medical images," *Med. Image Anal.*, vol. 41, pp. 40–54, 2017.

[59]   G. Zeng, X. Yang, J. Li, L. Yu, P. A. Heng, and G. Zheng, "3D U-net with multi-level deep supervision: Fully automatic segmentation of proximal femur in 3D MR images," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10541 LNCS, pp. 274–282, doi: 10.1007/978-3-319-67389-9_32.

[60]   Z. Zhu, Y. Xia, W. Shen, E. K. Fishman, and A. L. Yuille, "A 3D Coarse-to-Fine Framework for Automatic Pancreas Segmentation," *arxiv.org*, 2017.

[61]   X. Yang, C. Bian, L. Yu, D. Ni, and P. A. Heng, "Hybrid loss guided convolutional networks for whole heart parsing," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 10663 LNCS, pp. 215–223, doi: 10.1007/978-3-319-75541-0_23.

[62]   H. R. Roth, H. Oda, X. Zhou, N. Shimizu, and Y. Yang, "Computerized Medical Imaging and Graphics An application of cascaded 3D fully convolutional networks for medical image segmentation," *Comput. Med. Imaging Graph.*, vol. 66, no. October 2017, pp. 90–99, 2018, doi: 10.1016/j.compmedimag.2018.03.001.

[63]   L. Yu, X. Yang, J. Qin, and P. A. Heng, "3D FractalNet: Dense volumetric segmentation for cardiovascular MRI volumes," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10129 LNCS, pp. 103–110, doi: 10.1007/978-3-319-52280-7_10.

[64]   X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P. A. Heng, "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Liver Tumor Segmentation from CT Volumes," *IEEE Trans. Med. Imaging*, 2018, doi: 10.1109/TMI.2018.2845918.

[65]   H. S. Parmar, B. Nutter, R. Long, S. Antani, and S. Mitra, "Deep learning of volumetric 3D CNN for fMRI in Alzheimer's disease classification," in *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 2020, vol. 11317, p. 11, doi: 10.1117/12.2549038.

[66]   K. Oh, Y. C. Chung, K. W. Kim, W. S. Kim, and I. S. Oh, "Classification and Visualization of Alzheimer's Disease using Volumetric Convolutional Neural Network and Transfer Learning," *Sci. Rep.*, vol. 9, no. 1, pp. 1–16, Dec. 2019, doi: 10.1038/s41598-019-54548-6.

[67]   K. R. Kruthika, Rajeswari, and H. D. Maheshappa, "CBIR system using Capsule Networks and 3D CNN for Alzheimer's disease diagnosis," *Informatics Med. Unlocked*, vol. 14, pp. 59–68, Jan. 2019, doi: 10.1016/j.imu.2018.12.001.

[68]   C. Feng *et al.*, "Deep Learning Framework for Alzheimer's Disease Diagnosis via 3D-CNN and FSBi-LSTM," *IEEE Access*, vol. 7, pp. 63605–63618, 2019, doi: 10.1109/ACCESS.2019.2913847.

[69]   V. Wegmayr, S. Aitharaju, and J. Buhmann, "Classification of brain MRI with big data and deep 3D convolutional neural networks," *Med. Imaging 2018 Comput. Diagnosis*, p. 63, 2018, doi: 10.1117/12.2293719.

[70]   D. Nie *et al.*, "3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 212–220.

[71]   J. Zhou *et al.*, "Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images," *J. Magn. Reson. Imaging*, p. jmri.26721, Mar. 2019, doi: 10.1002/jmri.26721.

[72]   R. Ha *et al.*, "Fully Automated Convolutional Neural Network Method for Quantification of Breast MRI Fibroglandular Tissue and Background Parenchymal Enhancement," *Journal of Digital Imaging*, vol. 32, no. 1. pp. 141–147, 03-Feb-2019, doi: 10.1007/s10278-018-0114-7.

[73]   L. Chen *et al.*, "Combining many-objective radiomics and 3-dimensional convolutional neural network through evidential reasoning to predict lymph node metastasis in head and neck cancer," *Phys. Med. Biol.*, vol. 64, no. 7, p. 075011, Mar. 2019, doi: 10.1088/1361-6560/ab083a.

[74]   H. Shaish, S. Mutasa, J. Makkar, P. Chang, L. Schwartz, and F. Ahmed, "Prediction of lymph node maximum standardized uptake value in patients with cancer using a 3D convolutional neural network: A proof-of-concept study," *Am. J. Roentgenol.*, vol. 212, no. 2, pp. 238–244, Feb. 2019, doi: 10.2214/AJR.18.20094.

[75]   X. Gao, R. Hui, and Z. T. Biomedicine, "Classification of CT brain images based on deep learning networks," *omputer methods programs Biomed. Elsevier*, vol. 138, pp. 49–56, 2017.

[76]   J. Ker, S. P. Singh, Y. Bai, J. Rao, T. Lim, and L. Wang, "Image Thresholding Improves 3-Dimensional Convolutional Neural Network Diagnosis of Different Acute Brain Hemorrhages on Computed Tomography Scans," *Sensors*, vol. 19, no. 9, p. 2167, May 2019, doi: 10.3390/s19092167.

[77]   Q. Dou, H. Chen, L. Yu, L. Zhao, and … J. Q., "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1182–1195, 2016.

[78]   Q. Dou *et al.*, "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1182–1195, May 2016, doi: 10.1109/TMI.2016.2528129.

[79]   K. Standvoss *et al.*, "Cerebral microbleed detection in traumatic brain injury patients using 3D convolutional neural networks," in *Medical Imaging 2018: Computer-Aided Diagnosis*, 2018, p. 48, doi: 10.1117/12.2294016.

[80]   I. Wolterink, J. M., van Hamersvelt, R. W., Viergever, M. A., Leiner, T., & Išgum, "Coronary Artery Centerline Extraction in Cardiac CT Angiography," *Med. Image Anal.*, vol. 51, pp. 46–60, 2019, doi: 10.1016/j.media.2018.10.005.

[81] J. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, *Automatic coronary calcium scoring in cardiac CT angiography using convolutional neural networks*, vol. 9349. 2015.

[82] B. D. De Vos, J. M. Wolterink, P. A. De Jong, T. Leiner, M. A. Viergever, and I. Išgum, "ConvNet-Based Localization of Anatomical Structures in 3-D Medical Images," *IEEE Trans Med Imaging*, vol. 36, no. 7, pp. 1470–1481, 2017.

[83] Y. Huo *et al.*, "3D whole brain segmentation using spatially localized atlas network tiles," *Neuroimage*, 2019, doi: 10.1016/j.neuroimage.2019.03.041.

[84] R. Huang, W. Xie, and J. Alison Noble, "VP-Nets: Efficient automatic localization of key brain structures in 3D fetal neurosonography," *Med. Image Anal.*, vol. 47, pp. 127–139, 2018, doi: 10.1016/j.media.2018.04.004.

[85] A. Q. O'Neil *et al.*, "Attaining human-level performance for anatomical landmark detection in 3D CT data," *arxiv.org*, 2018.

[86] S. S. Mohseni Salehi, S. Khan, D. Erdogmus, and A. Gholipour, "Real-Time Deep Pose Estimation With Geodesic Loss for Image-to-Template Rigid Registration," *IEEE Trans. Med. Imaging*, vol. 38, no. 2, pp. 470–481, Feb. 2019, doi: 10.1109/TMI.2018.2866442.

[87] X. Li *et al.*, "3D multi-scale FCN with random modality voxel dropout learning for Intervertebral Disc Localization and Segmentation from Multi-modality MR Images," *Med. Image Anal.*, vol. 45, pp. 41–54, Apr. 2018, doi: 10.1016/j.media.2018.01.004.

[88] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *J. Big Data*, vol. 2, no. 1, p. 1, Dec. 2015, doi: 10.1186/s40537-014-0007-7.

[89] X. Chen, X. W., & Lin, "Big data deep learning: challenges and perspectives," *IEEE access*, no. 2, pp. 514–525, 2014.

[90] A. Vedaldi and K. Lenc, "MatConvNet - Convolutional Neural Networks for MATLAB," in *23rd ACM international conference on Multimedia*, 2015, pp. 689–692, doi: 10.1145/2733373.2807412.

[91] J. Duncan and N Ayache, "Medical image analysis: Progress over two decades and the challenges ahead," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 85–106, 2000.

[92] J. K. Iglehart, "Health Insurers and Medical-Imaging Policy — A Work in Progress," *N. Engl. J. Med.*, vol. 360, no. 10, pp. 1030–1037, Mar. 2009, doi: 10.1056/NEJMhpr0808703.

[93] L. Wang, Y. Wang, and Q. Chang, "Feature selection methods for big data bioinformatics: A survey from the search perspective," *Methods*, vol. 111. pp. 21–31, 2016, doi: 10.1016/j.ymeth.2016.08.014.

[94] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 8150 LNCS, no. PART 2, pp. 246–253, doi: 10.1007/978-3-642-40763-5_31.

[95] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006, doi: 10.1162/neco.2006.18.7.1527.

[96] C. B. Binnie, *Functional Brain Imaging*, vol. 52, no. 9. 1989.

[97] S. Gupta, Y. H. Chan, J. C. Rajapakse, and the A. D. N. Initiative, "Decoding brain functional connectivity implicated in AD and MCI," *bioRxiv*, p. 697003, Jul. 2019, doi: 10.1101/697003.

[98] J. Seward, "Artificial general intelligence system and method for medicine that determines a pre-emergent disease state of a patient based on mapping a topological module," 2018.

[99] T. Huang, "Imitating the brain with neurocomputer a 'new' way towards artificial general intelligence," *Int. J. Autom. Comput.*, vol. 14, no. 5, pp. 520–531, 2017.

[100] S. Shigeno, "Brain evolution as an information flow designer: the ground architecture for biological and artificial general intelligence," *Brain Evol. by Des.*, pp. 415–438, 2017.

[101] N. Mehta and M. V Devarakonda, "Machine Learning, Natural Language Programming, and Electronic Health Records: the next step in the Artificial Intelligence Journey?," *The Journal of allergy and clinical immunology*. 2018, doi: 10.1016/j.jaci.2018.02.025.

[102] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, vol. 3, no. January, pp. 2672–2680, doi: 10.3156/jsoft.29.5_177_2.

[103] A. Amidi, S. Amidi, D. Vlachakis, V. Megalooikonomou, N. Paragios, and E. I. Zacharaki, "EnzyNet: enzyme classification using 3D convolutional neural networks on spatial representation," *peerj.com*, 2017, doi: 10.7717/peerj.4750.

SATYA P SINGH(M'14) received the bachelor's degree in Electronics & Telecommunication Engineering from IETE, New Delhi, India, Master's degree in Electrical and Electronics Engineering from the YMCA University of Science and Technology, India, and Ph.D. degree in Electrical Engineering from Gautam Buddha University, India. Currently, he is Post-Doc Research Fellow in Lee Kong Chian School of Medicine at Nanyang Technological University, Singapore. His research interests include artificial intelligence in healthcare and development of medical imaging algorithms for computer-aided diagnosis of life-threatening diseases.

LIPO WANG received the bachelor's degree from the National University of Defense Technology, China, and the Ph.D. degree from Louisiana State University, USA. His research interest is intelligent techniques with applications to optimization, communications, image/video processing, biomedical engineering, and data mining. He is a (co-)author of 300 papers, of which 100 are in journals. He has co-authored two monographs and (co-)edited 15 books. He holds a U.S. patent in neural networks and a patent in systems. He was a keynote speaker for 30 international conferences. He is/was an associate editor/editorial board member of 30 international journals, including three the IEEE Transactions, and a guest editor for 10 journal special issues. He was a member of the Board of Governors of the International Neural Network Society (for 2 terms), the IEEE Computational Intelligence Society (CIS, for 2 terms), and the IEEE Biometrics Council. He received the Asia Paci_c Neural Network Assembly (APNNA) Excellent Service Award. He served as CIS Vice President for Technical Activities, the Chair of Emergent Technologies Technical Committee, and the Chair of Education Committee of the IEEE Engineering in Medicine and Biology Society (EMBS). He was the President of an APNNA. He was the Founding Chair of both the EMBS Singapore Chapter and CIS Singapore Chapter. He serves/served as the chair/committee member of over 200 international conferences.

Mr. Sukrit Gupta obtained a bachelor's in engineering in Computer Science from Punjab Engineering College, India. He has been a PhD Candidate in Computer Science at NTU, Singapore since 2016. He works in the areas of deep learning, complex network analysis and neuroimaging. His work has mainly been focused on the detection of brain functional modules and using brain topological features as disease biomarkers, thus obtaining more efficient deep neural networks with better diagnostic accuracy.

Haveesh Goli is a graduate research student with the School of Computer Science and Engineering (SCSE) at Nanyang Technological University (NTU), Singapore. Currently, he is pursuing Master of Engineering (M.Eng) in Computer Science from NTU. He received bachelor's degree in Computer Science and Engineering from Mahindra École Centrale, Hyderabad, India. He is also a part-time research

assistant in SCSE-NTU. Haveesh's research is focused in the areas of Brain structural analysis and Deep Neural Networks.

Dr Parasuraman Padmanabhan has vast experience in establishing different *in vivo* imaging modalities (Optical, MRI, PET, SPECT and CT) for monitoring different cellular events in mice, rats and Non-Human Primates. He mainly focuses on developing new imaging assays for studying different cellular signal transduction networks in living animals. He has more than thirty years of research and teaching experience gained from several leading universities, such as Cornell and Stanford University, and also from one of the premier A*STAR R&D Research Institute SBIC and currently he is Head of Operation in Centre for Neuroimaging Research at NTU (CeNReN).

Prof. Balázs Zoltán Gulyás is Professor of Translational Neuroscience and Director of the Cognitive Neuroimaging Centre at Lee Kong Chian School of Medicine, NTU. He has made pioneering contributions to functional human brain mapping with positron emission tomography (PET), in particular to the localisation of cortical areas in the human brain related to perceptual, cognitive and memory functions. More recently in his research he focuses on molecular neuroimaging with PET, with special regard to neurological and psychiatric diseases, neuropsychiatric drug development and molecular imaging probe development, and structural-functional neuroimaging using multimodal approaches (MEG, MRI, EEG, fNIRS, etc.) to interrogate the human brain's morphology and functions in normal and altered states.