Volumetric Segmentation of Brain Regions From MRI Scans Using 3D Convolutional Neural Networks

FARHEEN RAMZAN[1], MUHAMMAD USMAN GHANI KHAN[1], SAJID IQBAL[2], (Member, IEEE), TANZILA SABA[3], (Senior Member, IEEE), AND AMJAD REHMAN[3], (Senior Member, IEEE)

[1]Intelligent Criminology Research Laboratory, National Center of Artificial Intelligence, Al-Khawarizmi Institute of Computer Science (KICS), University of Engineering and Technology (UET) Lahore, Lahore 54000, Pakistan
[2]Department of Computer Science, Bahauddin Zakariya University, Multan 60700, Pakistan
[3]Artificial Intelligence and Data Analytics (AIDA) Laboratory, College of Computer and Information Sciences (CCIS), Prince Sultan University, Riyadh 11586, Saudi Arabia

Corresponding author: Sajid Iqbal (sajid.iqbal@bzu.edu.pk)

ABSTRACT Automated brain segmentation is an active research domain due to the association of various neurological disorders with different regions of the brain, to help medical professionals in prognostics and diagnostics. Traditional techniques like atlas-based and pattern recognition-based methods led to the development of various tools for automated brain segmentation. Recently, deep learning techniques are outperforming classical state-of-the-art methods and gradually becoming more mature. Consequently, deep learning has been extensively employed as a tool for precise segmentation of brain regions because of its capability to learn the intricate features of the high-dimensional data. In this work, a network for the segmentation of multiple brain regions has been proposed that is based on 3D convolutional neural networks and utilizes residual learning and dilated convolution operations to efficiently learn the end-to-end mapping from MRI volumes to the voxel-level brain segments. This research is focused on the segmentation of up to nine brain regions including cerebrospinal fluid, white matter and gray matter as well as their sub-regions. Mean dice scores of 0.879 and 0.914 have been achieved for three and nine brain regions, respectively by using the data from three different sources. Comparative analysis shows that our network gives better dice scores for most of the brain regions than state-of-the-artwork. Moreover, the mean dice score of 0.903, obtained for eight brain regions segmentation with MRBrains18 dataset, is better than 0.876 which was achieved in the previous work.

INDEX TERMS Brain segmentation, convolutional neural networks, magnetic resonance imaging, volumetric segmentation.

I. INTRODUCTION

The domain of medical imaging analysis encompasses a variety of tasks ranging from tumor detection, tumor segmentation to organ and multi-organ segmentation. Segmentation amongst all has gained a reputation as one of the leading problems in this area as it helps in the detection, analysis and treatment of the organ or tissue-related problems [1]. Brain segmentation is important for the analysis of different brain regions as volume, surface area and

The associate editor coordinating the review of this manuscript and approving it for publication was Jinjia Zhou.

morphology of them have found to be linked with various neurological disorders such as Parkinson's and Alzheimer's diseases [2]–[4]. The precise segmentation of different brain regions and tissues is usually a prerequisite for the detection and diagnosis of various neurological disorders. The importance of brain segmentation can be realized through the fact that various conferences such as MICCAI [5] hold challenges for it, and they have continuously been held for so many years now. Over the past years, many techniques have been experimented to segment brain for finding the most accurate -results and even today it stands to be a hard task.

Manual segmentations are very time consuming and susceptible to variability, so researchers have moved towards the development of automated techniques. Initially, atlas-based segmentation or pattern-recognition based methods were extensively used for segmentation problems, but mostly atlas-based ones [6]–[15]. For atlas-based methods, the image at hand is mapped onto pre-defined atlas for that specific task. Although, atlas-based methods obtain good segmentation results and are generally robust to certain anomalies, however, their dependability on population-specific atlases might limit their applicability to the dataset that is not well represented by the atlas. Due to this, it becomes difficult to segment brain tissue/region types accurately. Moreover, atlas-based methods are suboptimal if the patient population in the dataset is significantly different from the atlas. In this case, these approaches are unable to perform well due to variability in brain morphology among patients. To overcome these limitations, pattern recognition approaches were proposed that use spatial, intensity or other information in atlas space as features for the segmentation of different regions [13], [16]–[18]. These afore-mentioned methods require explicit information or features for tasks at hand.

In view of the limitations of manual, atlas-based or pattern recognition-based segmentation methods, researchers have moved towards using deep learning architectures. Deep learning based segmentation methods are capable of self-learning and can generalized well over large amounts of data [19], [20]. Deep learning architectures are gradually becoming more mature and are outperforming classical state-of-the-art methods. The performance of deep learning methods depends on the availability of training datasets to achieve generalization.

Convolutional Neural Networks (CNN) have known to be useful in various computer-vision tasks including but not limited to image recognition, object detection, classification and segmentation. CNNs learn the features in a hierarchal manner through multiple convolutions across multiple layers and do not need any predefined features or spatial information. The convolutional layers learn the spatial spacing and their generalizability through training [21]–[23]. So, CNNs have been used by various researchers for the segmentation of brain regions by varying the architecture as per task. By deep learning methods, the brain has been successfully segmented into three major regions: Cerebrospinal Fluid (CSF), Gray Matter (GM) and White Matter (WM). Many sub-regions 8, 25, 134, etc. have also been segmented whilst using variations of CNNs [24]–[26].

Deep learning has most prevalently been used for the two-dimensional images but as medical images are acquired in three-dimensional volumes, 3D CNNs have become the latest technique for the segmentation problems in the medical imaging domain [25], [27]. 3D images tend to incorporate spatial information which gets lost if 2D patches of those volumes are used. Hence this study presents a deep learning technique based on 3D CNN for segmentation of three and nine regions of the brain including WM, GM, CSF and their sub-regions.

II. RELATED WORK

Initially, research was focused on the segmentation of three major regions of the brain i.e. GM, WM and CSF. Tools like FreeSurfer [28] and FSL [29] have been used for the segmentation of the afore-mentioned regions using atlas-based methods. However, due to their time consumption, researchers have moved towards deep learning. Even within deep learning, two-dimensional architectures have been used extensively. So, for brain segmentation, the research started with the 2D CNNs leading to the usage of 3D architectures.

In this regard, a hybrid architecture of CNNs was proposed by de Brebisson *et al.* [26], which used three 2D patches and one 3D patch to segment 134 anatomical regions of the brain. 3D spatial information around each voxel, as well as 2D information in each plane (axial, coronal and sagittal), was utilized. Convolutional layers along with max pooling layers were utilized to extract and downscale features that were eventually combined through fully-connected and softmax layers. Also, to ensure refined segmentation, centroid distances between regions and voxels were used. The initial segmentation of the image was learned through the first network, then centroid distances were computed and the image along with centroids was given as input to full network for results and again centroid was re-calculated using the refined segmentation. The improved performance was achieved with the help of this additional input of centroids. In this work, 0.725 dice score was achieved for MICCAI multi-atlas labeling challenge dataset.

2D CNNs were used by Zhang *et al.* [30] to segment CSF, GM and WM of MRI scans of infants since the distinction between these regions of infants is not distinct as adults. In research presented by Moeskops *et al.* [31], volumes from three different regions of the body were used for segmentation whilst using a single network. An average dice score of around 0.80 was achieved by segmenting 6 tissues from MICCAI 2012 multi-atlas labeling challenge. In another study, 2D CNNs were utilized by Moeskops *et al.* [24] to segment 8 regions of the brain. Five different datasets for volumes of infants, young adults and aging adults were used. Their system was able to classify eight sub-regions and three major regions of the brain better than any previous work.

The usage of 3D CNNs in brain segmentation tasks has not been very prevalent except in recent years. The concept of residual learning into a 3D network was extended by Chen *et al.* [27], who introduced 3D residual neural networks. Residual learning strengthens the feature representation by using skip connections and adding the output of the preceding layer to its succeeding layer to enhance performance. This network produced a better dice score than that of the top ten teams on multi-modality volumes of MRBrains 2013 challenge. 2D, 2.5D and 3D patches of volumes were utilized by Milletari *et al.* [32] for segmentation. The feature vector was produced by the fully-connected patch of each dimension. The patch and the distance of voxel from where the patch was collected to the corresponding centroid in volume were stored in the database. To segment a new

instance, the fully-connected features were used to identify K-nearest neighbors in the database based on feature vectors. After identifying the neighbors, the distances stored in the database were used to perform localized segmentations. Better results were achieved than VNet for brain segmentation task.

A 3D CNN was employed by Wachinger *et al.* [25] to perform segmentation of structured MRI scans by using multi-task learning and Condition Random Fields (CRF) to enhance performance. Better results were achieved than FreeSurfer and FSL for segmentations of 25 regions with a mean dice score of 0.897. Another 3D CNN based network architecture was presented by Wong *et al.* [33] that used exponential logarithmic loss function to segment 19 brain regions. The skip connections and deep supervision were employed in the network to improve the efficiency of 3D segmentations. The exponential logarithmic loss function helped to learn the brain regions that differ in their sizes, morphology and complexity.

AssemblyNet was developed by Coupé *et al.* [34] for the segmentation of 132 brain regions. This network was made of two assemblies of U-Nets that shared knowledge among the neighboring U-Nets. The features learned by the first assembly were refined by the second assembly. The majority voting scheme was used to obtain the final decision. The model was evaluated on the MRI images from three different datasets and 0.733 dice score was obtained. Similarly, patch-based 3D CNN was employed by Ganaye *et al.* [35] to segment eight regions of the brain. The network consisted of encoding and decoding layers for feature extraction and label reconstruction, respectively. To learn more robust features, transition layers and batch-normalization were used in every convolution layer. T1, T2-FLAIR and T1-IR scans were used to train and evaluate the model. Their model ranked 1st in the MRBrainS18 challenge at the MICCAI 2018. The literature review on the segmentation of brain regions using deep learning techniques has been summarized in Table 1.

Besides brain tissue and region segmentation, there are other studies on deep learning-based semantic segmentation for medical images. In this regard, Yang *et al.* [49] presented a method for the segmentation of left atrium (LA) and pulmonary veins using the Late Gadolinium-Enhanced Cardiac MRI (LGE-CMRI). The method was based on deep learning utilizing convolutional long-short term memory (convLSTM) based sequential learning and dilated residual learning to segment both heart sub-structures from LGE-MRI images obtaining a dice score of 0.897±0.053. Similarly, a recent study [50] was presented for the automatic segmentation of high-intensity scar tissue and left atrium anatomy in the 3D LGE-CMR images of atrial fibrillation patients. For the segmentation of both scar and left atrium, a method based on a multi-view-two-task (MVTT) recursive attention model was proposed, which consists of three subnetworks incorporating multi-view learning, convLSTM and attention mechanism. The mean dice scores of 93% and 87% were obtained for LA anatomy and scar, respectively.

Computer-aided diagnosis of brain tumors is also an active field of study in medical imaging. In this regard, over the years research has been carried out for not only the classification and detection [51]–[53] of brain tumor types but also their segmentation [54], [55] using deep learning methods. An automatic method for brain tumor segmentation from 3D MRI images was proposed by Dong *et al.* [56]. The developed method was based on U-Net based deep CNN and was evaluated on Multimodal Brain Tumor Image Segmentation (BRATS 2015) datasets achieving promising results. Prostate segmentation is another research area in medical imaging based semantic segmentation. Liu *et al.* [57] presented a deep learning based algorithm for automatic prostate zonal segmentation. The developed algorithm was based on CNN and was able to segment the peripheral zone (PZ) and prostatic transition zone (TZ) on T2-weighted MRI images. The model was developed using MRI scans obtained from 250 patients and tested on 63 patients. The dice similarity coefficient were 0.74±0.08 and 0.86±0.07 for PZ and TZ, respectively. The results were comparable to other methods for prorate zones segmentation.

Medical images are acquired in three-dimensional volumes, therefore, 3D CNN has been prevalently employed for the automatic segmentation tasks. 2D CNN fail to extract volume and context information from adjacent slices which may be useful for precise segmentation. Some studies [50], [58] utilized the attention mechanism to learn intra and inter-slice features and context information from 2D slices to perform medical imaging segmentation tasks. On the other hand, the use of 3D CNN for three-dimensional images tend to incorporate spatial information which gets lost if 2D patches of those volumes are used. 3D CNN work by using three-dimensional convolutional kernels to make predictions from the volumetric patch of a scan. Their ability to extract inter-slice information lead to improved performance in various studies [59]–[61]. Similarly, the mechanism of dilated convolutions [62] and residual learning [63] have been utilized to perform segmentation tasks with improved performance. Dilated convolutions aggregate multi-scale contextual information without losing resolution. Considering the advantages of these techniques, our aim in this study is to combine dilation and residual learning mechanisms in a 3D CNN to perform volumetric segmentation.

The benefit that deep learning has provided besides time-consumption is that it made it feasible to segment more than a hundred anatomical regions of the brain as they follow a hierarchal feature learning. Nevertheless, even if multi-region segmentation of the brain has become possible, the precision is still a problem even with three basic regions as the dice score is usually near 90 but not over that. So, this paper focuses on the segmentation tasks of three and nine regions of the brain. Considering all the previous research and the amount of post or parallel processing that has to be employed for refined segmentation, we intend to demonstrate that with 3D CNNs, dilated convolutions and residual learning, better results can be achieved.

TABLE 1. A literature review on brain regions segmentation using deep learning techniques.

Study	Dataset	Sample Size	Brain Regions	Methods	Results and Evaluation
(de Brebisson and Montana, 2015)[26]	MICCAI 2012	35 MRI	134	2D CNN	MDC = 0.725
(Zhang et al., 2015)[30]	Private	10 MRI	3	2D CNN	MDC = 85.0
(Chen et al., 2016)[27]	MICCAI MRBrainS	15 MRI	3	3D CNN, VoxResNet	M_Score = 39
(Moeskops et al., 2016a)[24]	Private	22 MRI	8	2D CNN	MDC = 0.87, 0.82, 0.84, 0.86, 0.91
(Milletari et al., 2017)[32]	Private	55 MRI, US 34	14	2D, 2.5D, 3D CNN	MDC = 0.85
(Wachinger et al., 2018)[25]	MICCAI Multi-atlas labeling	30 MRI	25	3D CNN, Deep NAT	MDC = 0.897
(Ganaye et al., 2018)[35]	MICCAI 2012	35 MRI	135	Multi-Resolution (2D, 3D) CNN	MDC = 0.748
(Kushibar et al., 2018)[36]	MICCAI 2012, IBSR 18	35+18 MRI	14, 32	2.5D CNN	MDC = 0.869, 0.843
(Wong et al., 2018)[33]	Private	43 MRI	19	3D CNN, ELL	MSD = 0.82
(Kushibar et al., 2019)[37]	MICCAI 2012, IBSR 18	35+18 MRI	14	2.5 D CNN	MDC = 0.792, 0.742
(Roy et al., 2019)[38]	MALC-15 [40][35], ADNI-29, CANDI-13 [41][26], IBSR-18	15+29+13+18 MRI	33	Bayesian QuickNAT	MDC = 0.88, 0.78, 0.91, 0.84
(Jog et al., 2019)[41]	Buckner, Siemens13, GE14	20+13+14+(13+24) MRI	13	PSACNN	MDC = 0.94
(Coupé et al., 2019)[34]	OASIS, colin27 cohort, CANDI	45+19 MRI	132	AssemblyNet, 3D U-Net, NNTL	MDC = 0.733
(Chen et al., 2019)[42]	ISLES-2018, MRBrainS-2013	94+62+5 (DWI, T1, T1-IR, T1-FLAIR) MRI, CT	3	OctopusNet, Hyper fusion	MDC = 0.829
(Xie and Wen, 2019)[43]	BrainWeb, MRBrainS	(T1, T2, PD, FLAIR, IR) MRI	3	LSTM-MA	MDC = 0.9866, 0.8702
(Li et al., 2019)[44]	MRBrainS	5 (T1, T1-IR, T2-FLAIR) MRI	3	MMAN	DC = GM 0.864, WM 0.897, CSF 0.849
(Dolz et al., 2019)[45]	iSEG 2017, MRBrainS 2013	T1, T1-IR, T2-FLAIR	3	HyperDense-Net, 3D CNN	DC = GM 0.86, WM 0.89, CSF 0.83
(Luna and Park, 2019)[46]	MRBrainS18	30 (T1, T1-IR, T2-FLAIR) MRI	8	3D CNN, 3D U-Net	MDC = GM 0.86, BG 0.83, WM 0.882, WMH 0.65, CSF 0.837, VT 0.93, CB 0.94, BS 0.91
(Ahn et al., 2019)[47]	MRBrainS18	7 (Tl, Tl-IR, T2-FLAIR) MRI	8	2D CNN, ResNet, DenseNet, U-Net	MDC = 0.876
(Anand and Anand, 2020)[48]	ADNI, OASIS, MRBrainS13, MRBrainS18	1096+918+5+7 (T1, T1, -IR, T2-FLAIR) MRI	13	Deeplab-V3+ Segmentation Model, DCNN	OIU = 0.7744, 0.7720

MDC = Mean Dice Coefficient, M_Score = Rank DC + Rank HD + Rank AVD, ELL = Exponential Logarithmic Loss, PSACNN = Pulse Sequence Adaptive Convolutional Neural Network, NNTL = Nearest Neighbor Transfer Learning, CT = Computed Tomography, LSRM-MA = LSTM with Multi-Modality and Adjacency Constraint, MMAN = Multi-modality Aggregation Network, IOU = Intersection-Over-Union.

III. METHODOLOGY

Two segmentation tasks were performed in this work. The first task was the segmentation of the brain into three regions including GM, WM and CSF. And the second task was to segment brain into nine regions namely WM, CSF, Cortical Gray Matter (cGM), Brain Stem (BS), Cerebellum (CB), Basal Ganglia (BG), White Matter Lesions (WML), Ventricles (VT) and Infarction (INF). This section discusses the data sources and the deep learning architecture used for this study.

A. DATASET

MRI brain scans were acquired from various databases and were preprocessed to be used for segmentation tasks. In the dataset, there were a total of 90 MRI scans; 34 scans from Alzheimer's Disease Neuroimaging Initiative (ADNI) [64], [65], 21 scans from MRBrain18 challenge [66] and 35 scans from MICCAI 2012 challenge [67].

For three brain regions, a segmentation dataset of 76 MRI scans was prepared that contained images from ADNI,
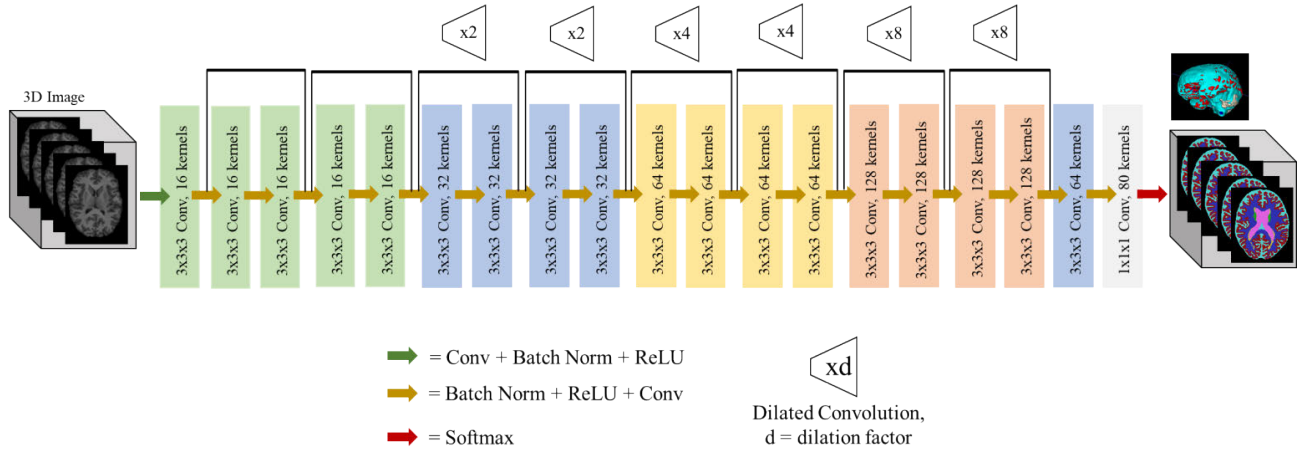
FIGURE 1. The proposed network architecture for 3D image segmentation.

MICCAI 2012 challenge and T1 weighted images from MRBrain18 challenge. Since MRI scans provided by MRBrain18 challenge contained segmented images, however, the remaining dataset was manually segmented to obtain labels for three brain segments. For this purpose FSL tool-box [29] was used and segmented data was generated in two steps. Firstly, automated brain extraction was performed by using FSL-BET [68], [69] for extracting brain voxels and removing non-brain regions/tissues such as skull and neck tissues from images. Secondly, brain scans were segmented into 3 regions (CSF, WM and GM) using FSL-FAST tool [70]. An algorithm based on expectation-maximization and hidden Markov random field was used by FSL-FAST to produce three segmentations of the whole brain by correcting non-homogeneities in intensity values. After this, the segmentations generated by FSL-FAST were manually corrected using ITK-Snap tool [71]. The details of the dataset for 3 brain segments have been presented in Table 2.

TABLE 2. Dataset for 3 brain regions segmentation.

Dataset	Image Count	Sequence Properties	Modality	Preprocessing
ADNI	34	170x256x256 @1x1x1mm TR=2300, TE=min full echo	MP-RAGE T1-Weighted	Brain extraction and brain segmentation
MICCAI 2012 challenge	35	256x334x256 @1x1x1mm TR=10, TE=6	MP-RAGE T1-Weighted	Brain extraction and brain segmentation
MRBrains 18 challenge	7	256x256x192 @1x1x1mm, TR=7.9, TE=4.5	T1-Weighted	No
Total		76		

For nine brain regions, the dataset of MRBrains18 challenge was used, as it provided MRI scans from 7 subjects in 5 different modalities (FLAIR, IR, registered IR, T1 weighted and registered T1 weighted). However, data from 3 modalities (T1 weighted, registered T1 weighted and FLAIR) was used in this study. There were 21 images in total along with their

segmentation (cortical gray matter, cerebrospinal fluid in the extra-cerebral space, white matter, basal ganglia, ventricles, brain stem, white matter lesions, cerebellum and infarction). For the segmentation task of nine brain regions, registered T1 weighted and FLAIR images were used. Data augmentation was also employed that included rotation by $-10°$ to $10°$, translation by 0.9 to 1.1 and scaling by -10% to 10%. Data augmentation was performed by using SimpleITK [72] python library. The details of the dataset for 9 regions of the brain have been given in Table 3.

TABLE 3. Dataset for 9 brain regions segmentation.

Dataset	Image Count	Sequence Properties	Modality	Preprocessing
	7	240x240x48 @0.958x0.958 x3mm, TR = 7.9, TE = 4.5	Registered T1 Weighted	No
	7	240x240x48 @0.958x0.958 x3mm, TR = 11000, TE = 125	FLAIR	No
MRBrains18	21	---	Registered T1 Weighted, FLAIR	Rotation 10 °, Translation 10%, Scaling 1.1
	21	---	Registered T1 Weighted, FLAIR	Rotation -10 °, Translation -10%, Scaling 0.9
Total		56		

B. PROPOSED DEEP LEARNING ARCHITECTURE: HIGH3DSEGNET

In this work, a method is presented for the volumetric image segmentation of cortical and subcortical regions of the brain. Our method is based on 3D convolutional neural networks and utilizes the concepts of skip connections, residual learning and dilated convolutions for efficiently learning the end-to-end mappings from MRI volumes to voxel-level brain segments. The architecture of the proposed network has been shown in Figure 1.

The network contains 19 convolutional layers. In each convolutional layer, $3 \times 3 \times 3$-voxel convolutions are applied, except for the last layer. Due to the small 3D convolutional kernels, the network has relatively fewer parameters. In the first five convolutional layers, 16 convolutional kernels are employed to learn low-level features from images. To learn features at multiple scales, dilated convolutions [73] are applied by gradually increasing the dilation factor as the layers go deeper.

Unlike the previous volumetric segmentation networks, for example, 3D U-net [59], that uses down- and up-sampling to learn hierarchical image features increasing computational cost; in this study, dilated convolutions has been used to compute features with high spatial resolution. To mathematically describe the dilated convolution for input feature map I with N channels and to up-sample the convolutional kernels with the dilation factor d, the output feature map O is generated as:

$$O_{x,y,z} = \sum_{n=0}^{N-1}\sum_{i=0}^{2}\sum_{j=0}^{2}\sum_{k=0}^{2} W_{i,j,k,n} I_{(x+id),(y+jd),(z+kd),n} \quad (1.1)$$

where x, y, z denote spatial locations of volumes and W denotes kernels. The dilated convolution conserves spatial resolution of the images and creates a receptive field of $(2d+1)^3$ voxels. Therefore, in the later layers dilated convolutions are applied and kernels have been dilated by a factor of 2 in the layers six to nine, dilated by a factor of 4 in the layers ten to thirteen and dilated by a factor of 8 in the layers fourteen to seventeen, respectively. The convolutional layers, that employ dilated kernels, learn middle and high-level features from images.

Residual connections [63], [74] have been applied between two consecutive convolutional layers throughout the network to allow the fusion of features from different scales and to improve the information propagation. To mathematically describe this, if the input from the L-layer to the residual block is x_L, then the output x_{L+1} from the residual block is:

$$x_{L+1} = x_L + f(x_L, w_L) \quad (1.2)$$

where $f(x_L, w_L)$ is the non-linear function. By stacking residual blocks, the output x_N form the last layer can be denoted by:

$$x_N = x_L + \sum_{i=L}^{N-1} f(x_i, w_i) \quad (1.3)$$

Within each residual block, each convolutional layer is associated with a batch-normalization layer [75] and a Rectified Linear Unit (ReLU) layer, arranged in a pre-activation order [63]. The second-last layer is a convolutional layer without any dilation and employs 64 kernels. The last layer in the network applies $1 \times 1 \times 1$ convolutions. The learned features from the last layer are passed to the softmax layer to output predicted probabilities over all labels. The mean dice coefficient [61] was used as a loss function for the volumetric image segmentation task. Let the image volume

is denoted by $\{x_v\}_{v=1}^{V}$ and the L-label segmentation map is denoted by $\{y_v\}_{v=1}^{V}$, where V represents number the of voxels and $y_v \in \{1, 2, 3, \ldots, L\}$ then the mean dice coefficient can be expressed as:

$$D(\{x_v\}, \{y_v\}) = \frac{1}{L}\sum_{l=1}^{L} \frac{2\sum_{v=1}^{V} \delta(y_v = l) Soft\max_l(x_v)}{\sum_{v=1}^{V} [\delta(y_v = l)]^2 + \sum_{v=1}^{V} [Soft\max_l(x_v)]^2} \quad (1.4)$$

where δ denotes Dirac delta function, $Soft\max_l(x_v)$ represents the softmax classification score of x_v over the lth-class. During training, the mean dice coefficient is maximized.

IV. RESULTS AND DISCUSSION

Our network was trained and evaluated on a combination of three datasets i.e. MICCAI 2012 challenge, ADNI and MICCAI MRBrainS18 challenge. Furthermore, two segmentation tasks were performed i.e. segmentation of 3 brain regions and segmentation of 9 brain regions. The network architecture for volumetric segmentation of MRI scans was designed in the same way as shown in Figure 1. The implementation was done using Python and especially TensorFlow and NiftyNet [76] library. All the experiments were conducted on an Intel(R) Core i5 system with 16GB RAM and a ZOTAC 11GB GPU.
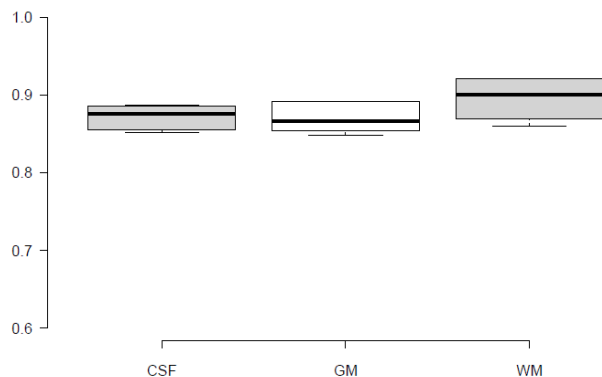
A. SEGMENTATION RESULTS

For the segmentation task of three brain regions, there were total samples of 76 images from three different sources, as shown in Table 2. For the segmentation task of nine brain regions, there were total samples of 56 images that included augmented images and images of 2 different modalities from the MICCAI MRBrainS18 challenge as illustrated in Table 3. In the training process of the network, all the weights were randomly initialized in a normal distribution with mean 0 and standard deviation 1. The learning rate was initialized with a value as 1e-03 and Adam optimizer was used. The dataset was split for training, validation and evaluation with 70%, 20% and 10% ratios, respectively. The network showed better convergence and speedy learning capabilities. However, since the network did not down-sample the inputs and number of kernel filters were increased by a certain factor, space complexity was higher. In the experiments, patch size of $88 \times 88 \times 88$ was used. The quantitative results in terms of Dice Score (DS), Jaccard Score (JC), Symmetric Volume Difference (SVD) and Volumetric Overlap Error (VOE) of each brain structure obtained for both segmentation tasks have been shown in Table 4.

The proposed method was evaluated on MRI images from different sources. Since the training data was limited and there were chances of over-fitting. Firstly, suitable hyper-parameters of the network were searched in various experiments. After, parameter tuning, the network was trained using the dataset and optimal hyper-parameters. The training of
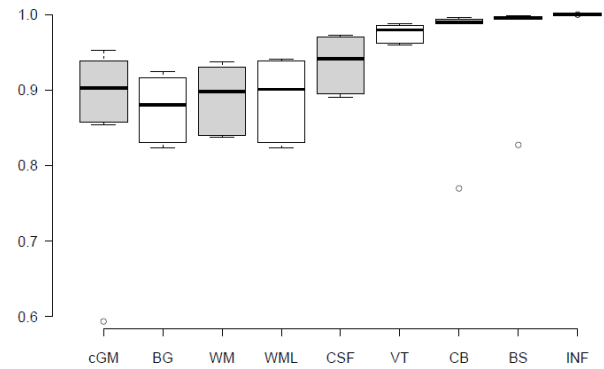
TABLE 4. Results for segmentation tasks of 3 and 9 brain regions.

Regions	DS	JS	SVD	VOE
Dice scores for 3 brain regions segmentation task				
Cerebrospinal Fluid (CSF)	0.896	0.812	0.104	0.188
Gray Matter (GM)	0.872	0.773	0.128	0.227
White Matter (WM)	0.872	0.773	0.128	0.227
Average Dice Score	**0.879**	**0.784**	**0.121**	**0.216**
Dice scores for 9 brain regions segmentation task				
White Matter Lesions (WML)	0.844	0.730	0.156	0.270
Cerebrospinal Fluid (CSF)	0.906	0.828	0.094	0.172
Cortical Gray Matter (cGM)	0.879	0.784	0.121	0.216
White Matter (WM)	0.864	0.761	0.136	0.239
Basal Ganglia (BG)	0.839	0.723	0.161	0.277
Ventricles (VT)	0.937	0.881	0.063	0.119
Cerebellum (CB)	0.973	0.947	0.027	0.053
Brain Stem (BS)	0.982	0.965	0.018	0.035
Infraction (INF)	0.999	0.998	0.001	0.002
Average Dice Score	**0.914**	**0.842**	**0.086**	**0.158**

FIGURE 3. Box plot for segmentation dice scores of 9 brain regions (WML = White Matter Lesions, CSF = Cerebrospinal Fluid, cGM = Cortical Gray Matter, WM = White Matter, BG = Basal Ganglia, VT = Ventricles, CB = Cerebellum, BS = Brain Stem, INF = Infraction).

the network was performed until the dice coefficient stopped improving and validation loss stopped decreasing. After this, the testing data was used for evaluation and prediction. The accuracy of segmentation was evaluated by using mean dice score measure. The dice score for each sample in the testing dataset was calculated and an average is computed across all the samples in the testing dataset. The box plots in Figure 2 and Figure 3 represent the distribution of dice scores of each brain structure over the validation dataset for three and nine regions segmentation tasks, respectively.


FIGURE 2. Box plot for segmentation dice scores of 3 brain regions (CSF = Cerebrospinal Fluid, GM = Gray Matter, WM = White Matter).

For the quantitative evaluation of segmentation results, the box plots were observed to identify the outliers and distribution of prediction accuracies. For the segmentation task of three brain regions, the dice scores were relatively better without any outliers. This shows that their accuracies were pretty much close to each other and our model performed considerably well on predicting all the testing samples. For nine brain regions, the results were quite different. Cerebellum, brain stem and ventricles segmented well across all the testing images. These three regions achieved high dice scores with a few outliers. The infraction was the region obtaining the highest dice score without any outlier. Although white

matter lesions, cerebrospinal fluid, cortical gray matter, white matter and basal ganglia were the brain structures with most outliers, their dice score was relatively better.

The WHM structure was most challenging to learn among all the regions. Since these are lesions or abnormalities in WM and they don't have predefined size, shape or location. Most number of outliers are in WML class, however, the average dice score of this structure is better. Our method worked best on larger regions such as cerebellum and ventricles and achieved relatively low scores with irregular and smaller regions such as WML. However, our segmentation accuracy even for smaller regions was encouraging, thus overall dice score across all the testing samples and overall the brain structures was 0.914. If traditional categorical cross-entropy measure was used, it would have given more importance to large brain structures that would have resulted in under- or over-segmentation. The use of the dice coefficient helped to avoid such a problem. That's why our proposed method was able to perform better segmentation of WML even though it had the lowest dice score among all the brain regions.
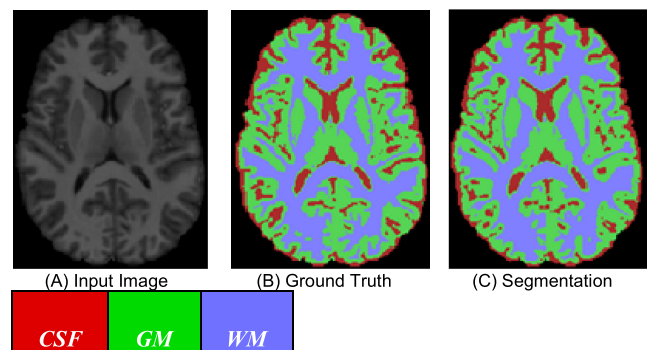
FIGURE 4. Segmentation Results for 3 brain regions; (A) Input image (B) ground truth (C) Segmentation; (CSF = Cerebrospinal Fluid, GM = Gray Matter, WM = White Matter).

The brain images and their corresponding regions segmented by our method are illustrated in Figure 4 and Figure 5.
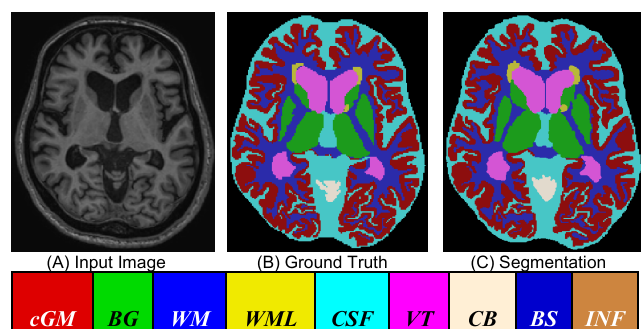
FIGURE 5. Segmentation results for 9 brain regions; (A) Input image (B) ground truth (C) Segmentation; (WML = White Matter Lesions, CSF = Cerebrospinal Fluid, cGM = Cortical Gray Matter, WM = White Matter, BG = Basal Ganglia, VT = Ventricles, CB = Cerebellum, BS = Brain Stem, INF = Infraction).

An input image ground truth and predicted segmentation have been shown. From the input data and ground truths, it was observed that the brain regions have very complex morphology and great variations across slices. These morphological variations are more evident in the brain images of different subjects. Brain regions not only have diverse appearances but also vary in size. A small tissue or region that appears in fewer slices is difficult to locate and segment automatically. Moreover, the intensity values of voxels also overlap significantly across brain tissues. This diversity of brain regions across slices and subjects makes the segmentation task very challenging.

Leveraging the advantages of 3D CNNs combined with the techniques of dilated convolutions and residual learning, our method uses the spatial information of voxels and effectively segments the brain regions despite their size. In the segmentation of three brain regions, it was noticed that our method accurately detected the borders of the brain tissues and accurately classified the voxels into GM, CSF and WM. Similarly, in the segmentation of nine brain regions, the brain regions were effectively segmented. However, there were a few false positives and false negatives in the results. For example, a few voxels belonging to some regions, particularly WM, BG and VT are different from the ground truth. While, CSF, cGM and WML tissues are quite learned accurately.

Notice that, the dice scores of some regions such as WML were relatively low compared to larger regions. However, they were predicted quite well by our model. The reason for this is their irregular morphology that affects the learning. Due to this, the dice scores vary significantly across brain scans, resulting in low overall dice scores. However, due to the small size of these brain structures, there were a few incorrect voxels predicted and hence fewer false positives and false negatives. Contrary to this, the predicted results of the large regions such as WM and VT contain more misclassified voxels and appear slightly different from their ground truths. Although their dice scores are better with fewer outliers, their large sizes require accurate prediction of a greater number of voxels than the smaller brain structures. cGM is the largest brain structure that has been predicted accurately.

These discrepancies in the results arise due to the overlapping and similar nature of voxel intensities. The voxels in the adjacent brain tissues have nearly similar coordinates. Some of the brain structures don't have predefined shapes or locations or both. Moreover, data is from different sources and subjects. These images have been acquired with different scanners and from subjects with different brain disorders. For example, the dataset acquired from ADNI contained scans of Alzheimer's patients. The scans from MRBrains18 were acquired from multiple patients with different health conditions such as dementia, Alzheimer's and diabetes as well as healthy subjects. These images contain varying degrees of white matter lesions and atrophy. The age of subjects was also different as patients with neurodegenerative disorders such as dementia and Alzheimer's are mostly of age >70. Moreover, neuro-degenerative disorders greatly affect the brain morphology and shapes of brain structures. The model had to learn not only the complex brain structures but also the variations associated with the brains of healthy or diseased patients. Despite, a few incorrect results, our method performed quite well for most of the brain regions and small brain regions have been learned accurately.

B. COMPARATIVE ANALYSIS

The achieved results were compared with other methods proposed in the previous studies. Although there are many tools and methods for the segmentation of GM, WM and CSF and most of them use atlas-based methods. For the comparison of results, those studies were identified that have used deep learning techniques and segmented the same brain regions that are under consideration in this study.

For the segmentation task of three brain regions, our results were compared with three recent studies. In this respect, Zhang *et al.* [30] presented a method based on 2D CNNs to segment infant brain scans into three tissue types using multi-modality MR images. The proposed network architecture employed convolutions, pooling and other operations in several layers to learn low- and high-level features. Another method was presented by Nie *et al.* [77] to segment infant brain scans into three tissue types. The method was based on employing fully convolutional networks (FCN) for the segmentation of three modalities of MR images including T1, T2 and fractional anisotropy. Initially, the network was trained with images of each modality separately. And then, the features learned from each modality were fused together to get the final segmentation maps.

While, Nguyen *et al.* [78] used 3D CNNs and Gaussian mixture models (GMM) to segment GM, WM and CSF. Firstly, GMM identified voxels that were easy to classify based on their intensity values. While CNN identified the voxels that had overlapping intensities. The method was evaluated on IBSR 18 dataset and better results were obtained for GM and WM but significantly lower dice score for CSF. Although, the aforementioned studies obtained better results, however, small datasets from healthy subjects were used for

evaluation that might limit the applicability of these methods in clinical practices for disease diagnosis and prognosis.

For this study, a relatively large dataset from three different sources was obtained from healthy as well as patients with dementia, Alzheimer's and diabetes. Dice scores of 0.872, 0.872 and 0.896 were obtained for WM, GM and CSF, respectively by using our method. As compared to previous studies, our method didn't achieve the best dice scores for GM and WM. Since, Alzheimer's and dementia patients are characterized by WM microstructural and ischemic changes as well as cortical atrophy [79], [80], [81]. As a result of this, our dataset contained MRI scans with varying levels of WM lesion load and GM atrophy making precise segmentation of these regions a challenging task. Our method achieved reasonable dice scores for these two regions and can help to segment brain tissues of the patients that have been affected with neurodegenerative diseases. The diversity in the dataset makes the model more generalized and suitable for clinical applications related to neurological disorders. Our method succeeded to achieve the best dice score for CSF and overall dice score of 0.879 as compared to other methods. The detailed results and comparative analysis for the segmentation results of three brain structures in terms of dice scores are presented in Table 5.

TABLE 5. Comparison of results for 3 brain regions segmentation.

Study	Methods	Dice Score				Sample Size	Data Source
		WM	GM	CSF	Avg.		
(Zhang et al., 2015)[30]	2D CNN	0.864	0.852	0.835	0.850	10 MRI	Private data
(Nie et al., 2016)[77]	FCN	0.887	0.873	0.855	0.871	10 MRI	Private data
(Nguyen et al., 2017)[78]	3D CNN, GMM	**0.90**	**0.91**	0.79	0.866	18 MRI	IBSR 18
Proposed work	3D CNN, High3DSeg Net	0.872	0.872	**0.896**	**0.879**	76 MRI	ADNI, MICCA I 2012, MRBrains18

In order to perform comparative analysis for the segmentation task of nine brain regions, the previous studies were identified. It was observed that most of the work was on the segmentation of three brain tissues. A few studies have subdivided the three brain regions into sub-regions. We considered those studies for comparison that have worked on brain regions similar to ours. One such study was conducted by Moeskops *et al.* that presented a method for the automatic segmentation of brain tissues. The method was based on multi-scale CNNs and used T1, T2-FLAIR and T1-IR images as input. The dataset from the MRBrains13 challenge was used for performing brain segmentation. Although the data provided by the challenge contains three brain segments, however, for their study, these segments were subdivided into 8 regions including WM, cGM, CB, BS, WMH, Lateral Ventricular Cerebrospinal Fluid (lvCSF), Basal Ganglia and Thalami (BGT) and Peripheral Cerebrospinal Fluid (pCSF). Their method achieved an average dice score of 0.67.

Moreover, the dice score of 0.85 for cGM, 0.87 for WM, 0.93 for CB, 0.92 for BS, 0.82 for BGT, 0.93 for lvCSF and 0.76 for pCSF were achieved. For the comparison purpose, we considered the dice scores of cGM, WM, BS and CB only.

Another method was proposed by Luna *et al.* for the automatic segmentation of eight brain regions from MRI images using patch-based 3D CNNs. The network consisted of encoding and decoding layers for feature extraction and label reconstruction, respectively. To learn more robust features, transition layers and batch-normalization were used in every convolution layer. T1, T2-FLAIR and T1-IR scans were used to train and evaluate the model using the dataset from MICCAI 2018 challenge. Similarly, Ahn *et al.* [47] also worked on the segmentation task of eight brain regions. Their method used the attention module and CNN based approach. The network consisted of compression and intensity modules to improve feature representation and spatial attention of pixels. The proposed method was evaluated using MRBrainS18 dataset.

Luna *et al.* achieved an average dice score of 0.855 and Ahn *et al.* achieved 87.60. Since both studies considered eight brain tissue classes, to have a fair comparison, the results of eight brain regions were considered. The average dice score of eight brain regions with our method was 0.903 which was higher than both studies. Moreover, the dice scores of 0.879 for cGM, 0.839 for BG, 0.864 for WM, 0.844 for WML, 0.906 for CSF, 0.937 for VT, 0.973 for CB, 0.982 for BS and 0.999 for INF were obtained by using our method. By observing the results, it was found that our method obtained best dice scores for cGM, BG, CSF, VT, CB and BS. The results of other brain structures such as WM and WML were also encouraging making average dice scores over all the brain structures higher than previously achieved with any method. The analysis of results for the segmentation task of nine brain regions is presented in Table 6. Our results were compared with the common brain regions in both studies. The comparative analysis is illustrated in Figure 6.

In this study, a method has been proposed for the automatic segmentation of various brain structures using 3D CNNs. The model has been evaluated for the segmentation task of three and nine brain regions. Contrary to the previous studies that have segmented eight brain structures, an additional brain structure i.e. infraction was also considered for segmentation. It was shown that our brain structures segmentation approach can be employed to include INF as an additional segmentation class. Unlike, other methods, the inclusion of INF in our evaluation didn't decrease the performance, this structure was learned quite effectively and obtained the highest dice score of all the other brain structures. Moreover, another brain region WML that is challenging to segment, has also been included in the evaluation. Thus, performing segmentation of brain tissues, structures and abnormalities at the same time.

In this study, images from three different modalities (T1 weighted, registered T1 weighted and FLAIR) were used. It was shown that our method is not limited to the input modality of brain scans and can effectively work with other

TABLE 6. Comparison of results for 9 brain regions segmentation.

Study	Methods	Dice Score									Data Source
		cGM	BG	WM	WML	CSF	VT	CB	BS	INF	
(Moeskops et al., 2018)[82]	2D CNN	0.840	---	0.870	---	---	---	0.910	0.890	---	MRBrains13
(Luna and Park, 2019)[46]	3D CNN, U-Net	0.860	0.834	**0.882**	0.652	0.837	0.931	0.939	0.905	---	MRBrainS18
(Ahn et al., 2019)[47]	2D CNN, U-Net	0.853	**0.839**	0.859	**0.879**	0.830	0.918	0.930	0.897	---	MRBrainS18
Proposed Work	3D CNN, HighRes3D Net	**0.879**	**0.839**	0.864	0.844	**0.906**	**0.937**	**0.973**	**0.982**	**0.999**	MRBrains18
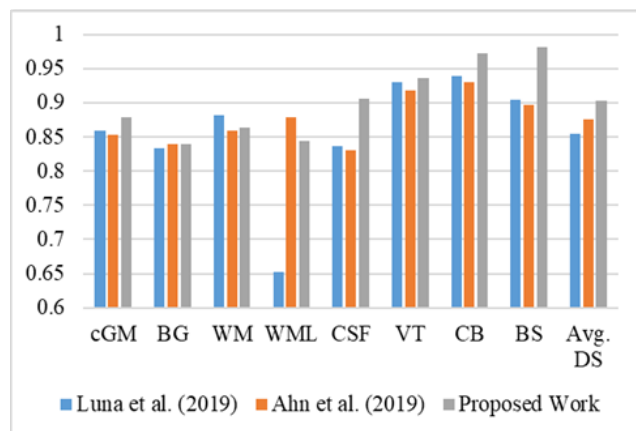
FIGURE 6. Comparative analysis of brain regions segmentation with the previous studies (WML = White Matter Lesions, CSF = Cerebrospinal Fluid, cGM = Cortical Gray Matter, WM = White Matter, BG = Basal Ganglia, VT = Ventricles, CB = Cerebellum, BS = Brain Stem, Avg. DS = Average Dice Score).

MR modalities. In terms of abnormality detection, it was observed that our model automatically detected lesions and infraction despite the image volume belongs to a healthy or a diseased patient. In clinical applications, a method that can automatically detect and segment brain structures and lesions would be beneficial to increase the sensitivity and precision of diagnosis. Additionally, data generated by such an automatic method can be used to improve the automatic segmentation method.

The data used in this study for the segmentation task of nine brain regions (Table 3) was anisotropic, scans having a voxel size of 0.958mm × 0.958mm × 3.0mm. The use of 3D CNNs is advantageous to learn features from isotropic images. Due to the low quality of anisotropic images, it is difficult for the network to learn features representing the dataset. Despite the anisotropic nature of scans used in this study, our method performed well for most of the brain regions. Similarly, motion and other artefacts were present in the dataset, especially in ADNI scans. The brain scans with artefacts were removed from the study in order to limit the influence of these artefacts on the performance of our model.

The evaluation of the model was performed on MRI images from different healthy and diseased subjects including patients with memory impairments. The segmentation results

revealed that our approach can accurately predict brain structures with varying degrees of abnormalities, morphologies and scanners. The proposed model obtained the highest dice scores for 6 out of 8 brain regions. The average dice score for the eight brain regions obtained with our model was higher than the previous work on MRBrains18 dataset. Moreover, the development of an automatic segmentation method for brain tissues as well as abnormalities can facilitate the diagnosis of brain disorders especially in the aging population and can help to identify the disease related biomarkers for prevention and treatment.

V. CONCLUSION

This research targeted medical imaging segmentation problems, specifically to segment brain into various regions including cerebrospinal fluid, white matter and gray matter. Unlike, traditional segmentation methods in medical imaging that are based on graph theory or atlas-based techniques, this work was directed towards the use of a 3D deep learning segmentation algorithm. In this regard, a 3D CNN based network was employed for two segmentation tasks i.e. 3 and 9 brain regions. For 3 brain regions segmentation task, dice scores of 0.872, 0.872 and 0.896 were obtained for WM, GM and CSF, respectively. While, for the segmentation task of 9 brain regions, dice scores of 0.879, 0.839, 0.864, 0.844, 0.906, 0.937, 0.973, 0.982 and 0.999 were achieved for cGM, BG, WM, WML, CSF, VT, CB, BS and INF, respectively. Our network obtained mean dice scores of 0.879 and 0.914 for three and nine brain regions respectively. Moreover, a mean dice score of 0.903 obtained for eight brain regions was better than the previous research. This indicates the promise of developing medical image segmentation algorithms using 3D deep learning techniques. In the future, we intend to improve our network to achieve better results for all the brain regions under consideration.

ETHICAL APPROVAL

We declare that all human and animal studies have been approved by the Medical University of South Carolina Institutional Review Board and have therefore been performed under the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

REFERENCES

[1] B. van Ginneken, C. M. Schaefer-Prokop, and M. Prokop, "Computer-aided diagnosis: How to move from the laboratory to the clinic," *Radiology*, vol. 261, no. 3, pp. 719–732, 2011.

[2] J. R. Petrella, R. E. Coleman, and P. M. Doraiswamy, "Neuroimaging and early diagnosis of Alzheimer disease: A look to the future," *Radiology*, vol. 226, no. 2, pp. 315–336, Feb. 2003.

[3] M. Hutchinson and U. Raff, "Structural changes of the substantia Nigra in Parkinson's disease as revealed by MR imaging," *Amer. J. Neuroradiol.*, vol. 21, no. 4, pp. 697–701, Apr. 2000.

[4] F. Ramzan *et al.*, "A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks," *J. Med. Syst.*, vol. 44, no. 2, pp. 1–16, Feb. 2020.

[5] N. Heller, J. Rickman, C. Weight, and N. Papanikolopoulos, "The role of publicly available data in MICCAI Papers from 2014 to 2018," Tech. Rep., Aug. 2019.

[6] M. Prastawa, J. H. Gilmore, W. Lin, and G. Gerig, "Automatic segmentation of MR images of the developing newborn brain," *Med. Image Anal.*, vol. 9, no. 5, pp. 457–466, Oct. 2005.

[7] N. I. Weisenfeld and S. K. Warfield, "Automatic segmentation of newborn brain MRI," *NeuroImage*, vol. 47, no. 2, pp. 564–572, Aug. 2009.

[8] F. Shi, Y. Fan, S. Tang, J. H. Gilmore, W. Lin, and D. Shen, "Neonatal brain image segmentation in longitudinal MRI studies," *NeuroImage*, vol. 49, no. 1, pp. 391–400, Jan. 2010.

[9] P. A. Habas, K. Kim, F. Rousseau, O. A. Glenn, A. J. Barkovich, and C. Studholme, "Atlas-based segmentation of developing tissues in the human brain with quantitative validation in young fetuses," *Hum. Brain Mapp.*, vol. 31, no. 9, pp. 1348–1358, Sep. 2010.

[10] M. Kuklisova-Murgasova, P. Aljabar, L. Srinivasan, S. J. Counsell, V. Doria, A. Serag, I. S. Gousias, J. P. Boardman, M. A. Rutherford, A. D. Edwards, J. V. Hajnal, and D. Rueckert, "A dynamic 4D probabilistic atlas of the developing brain," *NeuroImage*, vol. 54, no. 4, pp. 2750–2763, Feb. 2011.

[11] V. Srhoj-Egekher, M. J. N. L. Benders, M. A. Viergever, and I. Išgum, "Automatic neonatal brain tissue segmentation with MRI," *Proc. SPIE*, vol. 8669, 2013, Art. no. 86691K.

[12] M. J. Cardoso, A. Melbourne, G. S. Kendall, M. Modat, N. J. Robertson, N. Marlow, and S. Ourselin, "AdaPT: An adaptive preterm segmentation algorithm for neonatal brain MRI," *NeuroImage*, vol. 65, pp. 97–108, Jan. 2013.

[13] P. Anbeek, I. Išgum, B. J. M. van Kooij, C. P. Mol, K. J. Kersbergen, F. Groenendaal, M. A. Viergever, L. S. de Vries, and M. J. N. L. Benders, "Automatic segmentation of eight tissue classes in neonatal brain MRI," *PLoS ONE*, vol. 8, no. 12, Dec. 2013, Art. no. e81895.

[14] A. Makropoulos, I. S. Gousias, C. Ledig, P. Aljabar, A. Serag, J. V. Hajnal, A. D. Edwards, S. J. Counsell, and D. Rueckert, "Automatic whole brain MRI segmentation of the developing neonatal brain," *IEEE Trans. Med. Imag.*, vol. 33, no. 9, pp. 1818–1831, Sep. 2014.

[15] L. Wang, F. Shi, G. Li, Y. Gao, W. Lin, J. H. Gilmore, and D. Shen, "Segmentation of neonatal brain MR images using patch-driven level sets," *NeuroImage*, vol. 84, pp. 141–158, Jan. 2014.

[16] H. A. Vrooman, C. A. Cocosco, F. van der Lijn, R. Stokking, M. A. Ikram, M. W. Vernooij, M. M. B. Breteler, and W. J. Niessen, "Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification," *NeuroImage*, vol. 37, no. 1, pp. 71–81, Aug. 2007.

[17] L. Wang, Y. Gao, F. Shi, G. Li, J. H. Gilmore, W. Lin, and D. Shen, "LINKS: Learning-based multi-source IntegratioN frameworK for segmentation of infant brain images," *NeuroImage*, vol. 108, pp. 160–172, Mar. 2015.

[18] P. Moeskops, M. J. N. L. Benders, S. M. Chiţă, K. J. Kersbergen, F. Groenendaal, L. S. de Vries, M. A. Viergever, and I. Išgum, "Automatic segmentation of MR brain images of preterm infants using supervised classification," *NeuroImage*, vol. 118, pp. 628–641, Sep. 2015.

[19] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "Deep learning for brain MRI segmentation: State of the art and future directions," *J. Digit. Imag.*, vol. 30, no. 4, pp. 449–459, Aug. 2017.

[20] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift Medizinische Phys.*, vol. 29, no. 2, pp. 102–127, May 2019.

[21] 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.

[22] M. I. Razzak, S. Naz, and A. Zaib, *Deep Learning for Medical Image Processing: Overview, Challenges and the Future*. Cham, Switzerland: Springer, 2018, pp. 323–350.

[23] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[24] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders, and I. Isgum, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1252–1261, May 2016.

[25] C. Wachinger, M. Reuter, and T. Klein, "DeepNAT: Deep convolutional neural network for segmenting neuroanatomy," *NeuroImage*, vol. 170, pp. 434–445, Apr. 2018.

[26] A. De Brébisson and G. Montana, "Deep neural networks for anatomical brain segmentation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Oct. 2015, pp. 20–28.

[27] H. Chen, Q. Dou, L. Yu, and P.-A. Heng, "VoxResNet: Deep voxel-wise residual networks for volumetric brain segmentation," Aug. 2016, *arXiv:1608.05895*. [Online]. Available: https://arxiv.org/abs/1608.05895

[28] B. Fischl, "FreeSurfer," *NeuroImage*, vol. 62, no. 2, pp. 774–781, Aug. 2012.

[29] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *NeuroImage*, vol. 62, no. 2, pp. 782–790, Aug. 2012.

[30] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, Mar. 2015.

[31] P. Moeskops, J. M. Wolterink, B. H. M. van der Velden, K. G. A. Gilhuijs, T. Leiner, M. A. Viergever, and I. Išgum, *Deep Learning for Multi-Task Medical Image Segmentation in Multiple Modalities*. Cham, Switzerland: Springer, 2016, pp. 478–486.

[32] F. Milletari, S.-A. Ahmadi, C. Kroll, A. Plate, V. Rozanski, J. Maiostre, J. Levin, O. Dietrich, B. Ertl-Wagner, K. Bötzel, and N. Navab, "Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound," *Comput. Vis. Image Understand.*, vol. 164, pp. 92–102, Nov. 2017.

[33] K. C. L. Wong, M. Moradi, H. Tang, and T. Syeda-Mahmood, *3D Segmentation With Exponential Logarithmic Loss for Highly Unbalanced Object Sizes*. Cham, Switzerland: Springer, 2018, pp. 612–619.

[34] P. Coupé, B. Mansencal, M. Clément, R. Giraud, B. D. de Senneville, V.-T. Ta, V. Lepetit, and J. V. Manjon, "AssemblyNet: A novel deep decision-making process for whole brain MRI segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Jun. 2019, pp. 466–474.

[35] P.-A. Ganaye, M. Sdika, and H. Benoit-Cattin, "Towards integrating spatial localization in convolutional neural networks for brain image segmentation," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 621–625.

[36] K. Kushibar, S. Valverde, S. González-Villà, J. Bernal, M. Cabezas, A. Oliver, and X. Lladó, "Automated sub-cortical brain structure segmentation combining spatial and deep convolutional features," *Med. Image Anal.*, vol. 48, pp. 177–186, Aug. 2018.

[37] K. Kushibar, S. Valverde, S. González-Villà, J. Bernal, M. Cabezas, A. Oliver, and X. Lladó, "Supervised domain adaptation for automatic sub-cortical brain structure segmentation with minimal user interaction," *Sci. Rep.*, vol. 9, no. 1, p. 6742, Dec. 2019.

[38] A. G. Roy, S. Conjeti, N. Navab, and C. Wachinger, "Bayesian QuickNAT: Model uncertainty in deep whole-brain segmentation for structure-wise quality control," *NeuroImage*, vol. 195, pp. 11–22, Jul. 2019.

[39] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *J. Cogn. Neurosci.*, vol. 19, no. 9, pp. 1498–1507, Sep. 2007.

[40] D. N. Kennedy, C. Haselgrove, S. M. Hodge, P. S. Rane, N. Makris, and J. A. Frazier, "CANDIShare: A resource for pediatric neuroimaging data," *Neuroinformatics*, vol. 10, no. 3, pp. 319–322, Jul. 2012.

[41] A. Jog, A. Hoopes, D. N. Greve, K. Van Leemput, and B. Fischl, "PSACNN: Pulse sequence adaptive fast whole brain segmentation," *NeuroImage*, vol. 199, pp. 553–569, Oct. 2019.

[42] Y. Chen, J. Chen, D. Wei, Y. Li, and Y. Zheng, "OctopusNet: A deep learning segmentation network for multi-modal medical images," in *Proc. Int. Workshop Multiscale Multimodal Med. Imag.*, Jun. 2019, pp. 17–25.

[43] K. Xie and Y. Wen, "LSTM-MA: A LSTM method with multi-modality and adjacency constraint for brain image segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 240–244.

[44] J. Li, Z. L. Yu, Z. Gu, H. Liu, and Y. Li, "MMAN: Multi-modality aggregation network for brain segmentation from MR images," *Neurocomputing*, vol. 358, pp. 10–19, Sep. 2019.

[45] J. Dolz, K. Gopinath, J. Yuan, H. Lombaert, C. Desrosiers, and I. Ben Ayed, "HyperDense-Net: A hyper-densely connected CNN for multi-modal image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 5, pp. 1116–1126, May 2019.

[46] M. Luna and S. H. Park, *3D Patchwise U-Net With Transition Layers for MR Brain Segmentation*. Cham, Switzerland: Springer, 2019, pp. 394–403.

[47] S. Ahn, T. D. Bui, and J. Shin, "Compression and intensity modules for brain MRI segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, 2019, pp. 719–722.

[48] A. Anand and N. Anand, *Fast Brain , Volumetric Segmentation from T1 MRI Scans*. Cham, Switzerland: Springer, 2020, pp. 402–415.

[49] G. Yang, J. Chen, Z. Gao, H. Zhang, H. Ni, E. Angelini, R. Mohiaddin, T. Wong, J. Keegan, and D. Firmin, "Multiview sequential learning and dilated residual learning for a fully automatic delineation of the left atrium and pulmonary veins from late gadolinium-enhanced cardiac MRI images," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 1123–1127.

[50] G. Yang, J. Chen, Z. Gao, S. Li, H. Ni, E. Angelini, T. Wong, R. Mohiaddin, E. Nyktari, R. Wage, L. Xu, Y. Zhang, X. Du, H. Zhang, D. Firmin, and J. Keegan, "Simultaneous left atrium anatomy and scar segmentations via deep learning in multiview information with attention," *Future Gener. Comput. Syst.*, vol. 107, pp. 215–228, Jun. 2020.

[51] S. Iqbal, M. U. G. Khan, T. Saba, and A. Rehman, "Computer-assisted brain tumor type discrimination using magnetic resonance imaging features," *Biomed. Eng. Lett.*, vol. 8, no. 1, pp. 5–28, Feb. 2018.

[52] M. A. Khan *et al.*, "Brain tumor detection and classification: A framework of marker-based watershed algorithm and multilevel priority features selection," *Microsc. Res. Techn.*, vol. 82, no. 6, pp. 909–922, Jun. 2019.

[53] T. Saba, A. Sameh Mohamed, M. El-Affendi, J. Amin, and M. Sharif, "Brain tumor detection using fusion of hand crafted and deep learning features," *Cogn. Syst. Res.*, vol. 59, pp. 221–230, Jan. 2020.

[54] S. Iqbal, M. U. Ghani, T. Saba, and A. Rehman, "Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN)," *Microsc. Res. Techn.*, vol. 81, no. 4, pp. 419–427, Apr. 2018.

[55] S. Iqbal *et al.*, "Deep learning model integrating features and novel classifiers fusion for brain tumor segmentation," *Microsc. Res. Techn.*, vol. 82, no. 8, pp. 1302–1315, Aug. 2019.

[56] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks," in *Medical Image Understanding and Analysis* (Communications in Computer and Information Science), vol. 723. Cham, Switzerland: Springer, 2017, pp. 506–517.

[57] Y. Liu, K. Sung, G. Yang, S. A. Mirak, M. Hosseiny, A. Azadikhah, X. Zhong, R. E. Reiter, Y. Lee, and S. S. Raman, "Automatic prostate zonal segmentation using fully convolutional network with feature pyramid attention," *IEEE Access*, vol. 7, pp. 163626–163632, 2019.

[58] J. Tian, L. Liu, Z. Shi, and F. Xu, "Automatic Couinaud segmentation from CT volumes on liver using GLC-UNet," in *Machine Learning in Medical Imaging* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11861. Cham, Switzerland: Springer, 2019, pp. 274–282.

[59] 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—-MICCAI* (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9901, no. 8. Cham, Switzerland: Springer, 2016, pp. 424–432.

[60] E. Gibson, F. Giganti, Y. Hu, E. Bonmati, S. Bandula, K. Gurusamy, B. Davidson, S. P. Pereira, M. J. Clarkson, and D. C. Barratt, "Automatic multi-organ segmentation on abdominal CT with dense V-Networks," *IEEE Trans. Med. Imag.*, vol. 37, no. 8, pp. 1822–1834, Aug. 2018.

[61] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[62] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. 4th Int. Conf. Learn. Represent. (ICLR)*, Nov. 2016, pp. 1–13.

[63] K. He, X. Zhang, S. Ren, and J. Sun, *Identity Mappings in Deep Residual Networks*. Cham, Switzerland: Springer, 2016, pp. 630–645.

[64] C. R. Jack *et al.*, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *J. Magn. Reson. Imag.*, vol. 27, no. 4, pp. 685–691, 2008.

[65] C. R. Jack, R. C. Petersen, Y. C. Xu, P. C. O'Brien, G. E. Smith, R. J. Ivnik, B. F. Boeve, S. C. Waring, E. G. Tangalos, and E. Kokmen, "Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment," *Neurology*, vol. 52, no. 7, pp. 1397–1397, Apr. 1999.

[66] A. M. Mendrik *et al.*, "MRBrainS challenge: Online evaluation framework for brain image segmentation in 3T MRI scans," *Comput. Intell. Neurosci.*, vol. 2015, pp. 1–16, Dec. 2015.

[67] B. Landman and S. Warfield, "MICCAI 2012 workshop on multi-atlas labeling," in *Proc. Med. Image Comput. Comput. Assist. Intervent. Conf.*, Apr. 2012, pp. 1–164.

[68] S. M. Smith, "Fast robust automated brain extraction," *Hum. Brain Mapping*, vol. 17, no. 3, pp. 143–155, 2002.

[69] M. Jenkinson, M. Pechaud, and S. Smith, "BET2: MR-based estimation of brain, skull and scalp surfaces," in *Proc. 11th Annu. Meeting Org. Hum. Brain Mapping*, 2005, p. 167.

[70] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.

[71] P. A. Yushkevich and G. Gerig, "ITK-SNAP: An intractive medical image segmentation tool to meet the need for expert-guided segmentation of complex medical images," *IEEE Pulse*, vol. 8, no. 4, pp. 54–57, Jul. 2017.

[72] *SimpleITK*. Accessed: Sep. 27, 2018. [Online]. Available: http://www.simpleitk.org/

[73] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[74] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[75] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Feb. 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167

[76] E. Gibson, W. Li, C. Sudre, L. Fidon, D. I. Shakir, G. Wang, Z. Eaton-Rosen, R. Gray, T. Doel, Y. Hu, T. Whyntie, P. Nachev, M. Modat, D. C. Barratt, S. Ourselin, M. J. Cardoso, and T. Vercauteren, "NiftyNet: A deep-learning platform for medical imaging," *Comput. Methods Programs Biomed.*, vol. 158, pp. 113–122, May 2018.

[77] D. Nie, L. Wang, Y. Gao, and D. Shen, "Fully convolutional networks for multi-modality isointense infant brain image segmentation," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 1342–1345.

[78] D. M. H. Nguyen, H. T. Vu, H. Q. Ung, and B. T. Nguyen, "3D-brain segmentation using deep neural network and Gaussian mixture model," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 815–824.

[79] M. Weiler, F. Agosta, E. Canu, M. Copetti, G. Magnani, A. Marcone, E. Pagani, M. L. F. Balthazar, G. Comi, A. Falini, and M. Filippi, "Following the spreading of brain structural changes in Alzheimer's disease: A longitudinal, multimodal MRI study," *J. Alzheimer's Disease*, vol. 47, no. 4, pp. 995–1007, Aug. 2015.

[80] G. S. Alves, "CSF β-amyloid and white matter damage: Unravelling the neuropathology of Alzheimer's disease," *J. Neurol., Neurosurg. Psychiatry*, vol. 89, no. 4, p. 329, Apr. 2018.

[81] H. Jang, H. Kwon, J.-J. Yang, J. Hong, Y. Kim, K. W. Kim, J. S. Lee, Y. K. Jang, S. T. Kim, K. H. Lee, J. H. Lee, D. L. Na, S. W. Seo, H. J. Kim, and J.-M. Lee, "Correlations between gray matter and white matter degeneration in pure Alzheimer's disease, pure subcortical vascular dementia, and mixed dementia," *Sci. Rep.*, vol. 7, no. 1, pp. 1–9, Dec. 2017.

[82] P. Moeskops, J. de Bresser, H. J. Kuijf, A. M. Mendrik, G. J. Biessels, J. P. W. Pluim, and I. Išgum, "Evaluation of a deep learning approach for the segmentation of brain tissues and white matter hyperintensities of presumed vascular origin in MRI," *NeuroImage, Clin.*, vol. 17, pp. 251–262, Jan. 2018.

FARHEEN RAMZAN received the bachelor's and master's degrees in computer science from the University of Engineering and Technology (UET) Lahore, Pakistan, in 2016 and 2019, respectively. She is currently working as a Research Officer with the Bioinformatics Research Laboratory (BRL), Al-Khawarizmi Institute of Computer Science (KICS), UET Lahore. She has authored few articles in international refereed journals. Her research interests include medical image processing, bioinformatics, genomics, digital signal and image processing, and deep learning.

MUHAMMAD USMAN GHANI KHAN received the Ph.D. degree from Sheffield University, Sheffield, U.K., in 2012. During his Ph.D. degree, he worked on statistical modeling for machine vision signals, specifically generating language descriptions of video streams. He is currently working as an Associate Professor with the Department of Computer Science, University of Engineering and Technology, Lahore, Pakistan. His research interest includes natural language processing using statistical approaches, with applications, such as information extraction from speech and speech summarization. His recent work is concerned with multimedia, incorporating text, and audio and visual processing into one framework.

SAJID IQBAL (Member, IEEE) received the Ph.D. degree from the University of Engineering and Technology, Lahore, Pakistan, in 2019. He is currently working as an Assistant Professor with the Department of Computer Science, Bahauddin Zakariya University, Multan, Pakistan. At present, he has authored more than 15 publications and one book. His research interests include medical image processing, natural language processing, and statistical modeling for AI-based solutions.

TANZILA SABA (Senior Member, IEEE) received the Ph.D. degree in document information security and management from the Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia, in 2012. She won the best student award in the Faculty of Computing UTM for 2012. She is currently serving as an Associate Professor and Associate Chair of Information Systems Department, College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia. Her primary research focus in recent years is medical imaging, pattern recognition, data mining, MRI analysis, and soft-computing. She has above one hundred publications that have around 1800 citations with h-index 28. Her mostly publications are in biomedical research published in ISI/SCIE indexed. Due to her excellent research achievement, she is included in Marquis Who's Who (S & T) 2012. She is currently an editor and reviewer of reputed journals and on the panel of TPC of international conferences. She has full command of a variety of subjects and taught several courses at the graduate and postgraduate levels. On the accreditation side, she is a skilled lady with ABET and NCAAA quality assurance. She is the leader of the Artificial Intelligence and Data Analytics Research Lab at PSU and active professional members of ACM, AIS and IAENG organizations. She is the PSU WiDS (Women in Data Science) ambassador at Stanford University. She has been nominated as a Research Professor with PSU, since September 2019.

AMJAD REHMAN (Senior Member, IEEE) received the Ph.D. degree in image processing and pattern recognition from Universiti Teknologi Malaysia, Malaysia, in 2010. During his Ph.D., he proposed novel techniques for pattern recognition based on novel features mining strategies. He is currently conducting research under his supervision for three Ph.D. students. He is the author of dozens of papers published in international journals and conferences of high repute. His keen research includes information security, data mining and documents analysis, and recognition.

• • •