

Review Article

Deep Neural Networks for Medical Image Segmentation

Priyanka Malhotra ¹, **Sheifali Gupta** ², **Deepika Koundal** ³, **Atef Zaguia** ⁴,
and Wegayehu Enbeyle ⁵

¹Chitkara University Institute of Engineering and Technology, Chitkara University, Chandigarh, Punjab, India

²Chitkara University Institute of Engineering and Technology, Chitkara University, Chandigarh, Punjab, India

³Department of Systemics, School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India

⁴Department of Computer Science, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

⁵Department of Statistics, Mizan-Tepi University, Tepi, Ethiopia

Correspondence should be addressed to Wegayehu Enbeyle; wegu0202@gmail.com

Received 21 October 2021; Revised 6 January 2022; Accepted 10 January 2022; Published 10 March 2022

Academic Editor: Chinmay Chakraborty

Copyright © 2022 Priyanka Malhotra et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Image segmentation is a branch of digital image processing which has numerous applications in the field of analysis of images, augmented reality, machine vision, and many more. The field of medical image analysis is growing and the segmentation of the organs, diseases, or abnormalities in medical images has become demanding. The segmentation of medical images helps in checking the growth of disease like tumour, controlling the dosage of medicine, and dosage of exposure to radiations. Medical image segmentation is really a challenging task due to the various artefacts present in the images. Recently, deep neural models have shown application in various image segmentation tasks. This significant growth is due to the achievements and high performance of the deep learning strategies. This work presents a review of the literature in the field of medical image segmentation employing deep convolutional neural networks. The paper examines the various widely used medical image datasets, the different metrics used for evaluating the segmentation tasks, and performances of different CNN based networks. In comparison to the existing review and survey papers, the present work also discusses the various challenges in the field of segmentation of medical images and different state-of-the-art solutions available in the literature.

1. Introduction

Image segmentation involves partitioning an input image into different segments with strong correlation with the region of interest (RoI) in the given image [1, 2]. The aim of medical image segmentation [3] is to represent a given input image in a meaningful form to study the anatomy, identify the region of interest (RoI), measure the volume of tissue to measure the size of tumor, and help in the deciding the dose of medicine, planning of treatment prior to applying radiation therapy, or calculating the radiation dose. Image segmentation helps in analysis of medical images by highlighting the region of interest. Segmentation techniques can be utilized for brain tumor boundary extraction in MRI

images, cancer detection in biopsy images, mass segmentation in mammography, detection of borders in coronary angiograms, segmentation of pneumonia affected area in chest X-rays, etc. A number of medical image segmentation algorithms have been developed and are in demand as there is a shortage of expert manpower [4].

The earlier image segmentation models were based on traditional image processing approaches [3, 5] which include thresholding and edge-based and region-based techniques. In thresholding technique, pixels were allocated to different categories in accordance with the range of values where a particular pixel lies. In edge-based technique, a filter was applied to an image; it classifies the pixels as edged or nonedged in accordance with the filter output. In region-

based segmentation methods, neighbouring pixels having similar values and the groups of pixels having dissimilar values were split.

Medical image segmentation is difficult task due to various restrictions inflicted by the medical image procurement procedure, the type of pathology, and different biological variations [6]. The analysis of medical images can be done by experts and there is a shortage of medical imaging experts [7]. In the last few years, deep learning networks had contributed to the development of newer image segmentation models with improvement in performance. The deep neural networks had achieved high accuracy rates on different popular datasets. The image segmentation techniques can be broadly classified as semantic segmentation and instance segmentation. Semantic segmentation can be considered as a problem of classifying pixels. In this segmentation technique, each pixel in the image is labelled to a certain class. Instance segmentation detects and delineates each object of interest present in the input image.

The present work covers the recent literature in medical image segmentation. The work provides a review on different deep learning-based image segmentation models and explains their architecture. Many authors have worked on the review of medical image segmentation task. Table 1 gives the description of few review papers utilizing deep CNN in the field of medical image segmentation.

All the aforementioned survey literatures discuss the various deep neural networks. This survey paper does not only focus on summarizing the different deep learning approaches but also provides an insight into the different medical image datasets used for training deep neural networks and also explains the metrics used for evaluating the performance of a model. The present work also discusses the various challenges faced by DL based image segmentation models and their state-of-the-art solutions. The paper has several contributions which are as follows:

Firstly, the present study provides an overview of the current state of the deep neural network structures utilized for medical image segmentation with their strengths and weaknesses

Secondly, the paper describes the publicly available medical image segmentation datasets

Thirdly, it presents the various performance metrics employed for evaluating the deep learning segmentation models

Finally, the paper also gives an insight into the major challenges faced in the field of image segmentation and their state-of-the-art solutions

The organization of the rest of the paper is given in Table 2 [14].

2. Deep Neural Network Structures

Deep learning is the most essential approach to artificial intelligence. Deep learning algorithm uses various layers to construct an artificial neural network. An artificial neural network (ANN) consists of [52] input layer, hidden layer(s),

and output layer. The input layer of the network receives the signal, an output layer makes decision regarding the input, and between the input and output layers there are hidden layers which perform computations (shown in Figure 1). A deep neural network consists of many hidden layers between input and output layers.

This section provides a review of different deep learning neural networks employed for image segmentation task. The different deep neural network structures generally employed for image segmentation can be grouped as shown in Figure 2.

2.1. Convolutional Neural Network. A convolutional neural network or CNN (see Figure 3) consists of a stack of three main neural layers: convolutional layer, pooling layer, and fully connected layer [52, 53]. Each layer has its own role. The convolution layer detects distinct features like edges or other visual elements in an image. Convolution layer performs mathematical operation of multiplication of local neighbours of an image pixel with kernels. CNN uses different kernels for convolving the given image for generating its feature maps. Pooling layer reduces the spatial (width, height) dimensions of the input data for the next layers of neural network. It does not change the depth of the data. This operation is called as subsampling. This size reduction decreases the computational requirements for upcoming layers. The fully connected layers perform high-level reasoning in NN. These layers integrate the various feature responses from the given input image so as to provide the final results.

Different CNN models have been reported in the literature, including AlexNet [54], GoogleNet [55], VGG [56], Inception[57], SqueezeNet [58], and DenseNet [59]. Here, each network uses different number of convolutions and pooling layers with important process blocks inbetween them. The CNN models have been employed mostly for classification task. In [60], SqueezeNet and GoogleNet have been employed to classify brain MRI images into three different categories. The CNN segmentation models performance is limited by the following:

The fully connected layers in CNN cannot manage different input sizes

A convolutional neural network with a fully connected layer cannot be employed for object segmentation task, as the presence of number of objects of interest in the image segmentation task is not fixed, so the length of the output layer cannot be constant

2.1.1. Fully Convolutional Network. In fully convolutional network (FCN), only convolutional layers exist. The different existing in CNN architectures can be modified into FCN by converting the last fully connected layer of CNN into a fully convolutional layer. The model designed by [61] can output spatial segmentation map and can have dense pixel-wise prediction from the input image of full size instead of performing patch-wise predictions. The model uses skip connections which perform upsampling on feature maps from final layer and fuses it with the feature map of

TABLE 1: Description of few review papers in medical image segmentation.

Ref.	Year	Models discussed	Performance metrics	Dataset	Challenges	Remarks
[8]	2017	CNN	No coverage	No coverage	Challenges with CNN covered	Image classification, object detection, segmentation, and registration mechanisms discussed
[9]	2017	Stacked autoencoder, deep belief network, and deep Boltzmann machine	No coverage	No coverage	No coverage	—
[10]	2018	CNN, R-CNN	Image classification metrics discussed but segmentation metrics not covered	Medical image modalities covered	No coverage	All areas of medical image analysis discussed
[11]	2019	CNN, FCN, U-Net, VNet, CRN, and RNN	No coverage	Covered	Challenges and possible solutions discussed	—
[12]	2020	Supervised, weakly supervised models (RNN, U-Net)	No coverage	Covered	Challenges and possible solutions discussed	----
[13]	2021	CNN, FCN, DeepLab, SegNet, U-Net, and VNet	Covered	Covered	Challenges discussed but the solutions not discussed	—
Ours		CNN,FCN,R-CNN, fast R-CNN, faster R-CNN, mask R-CNN, U-Net, VNet, and DeepLab	Covered	Covered	Challenges and possible state-of-the-art solutions discussed	Paper provides extended coverage to the different deep neural networks for image segmentation

TABLE 2: Structure of the paper.

S. no.	Main section	Subsection
1	Introduction	Introduction and motivation literature review major contributions
2	Deep neural network structures	Artificial neural network convolutional neural network encoder-decoder models regional convolutional network deepLab model comparison, limitations, and advantages/Table 3
3	Application of deep neural network to medical image segmentation	Deep learning-based system literature review on DNN based image segmentation models for different organs summary on deep learning-based medical image segmentation methods (Table 4)
4	Medical image segmentation datasets	Types and format of dataset different types of modalities summary of medical image segmentation datasets (Table 5)
5	Evaluation metrics	Importance of metrics popular image segmentation algorithm performance metrics
6	Major challenges and state-of-the-art solutions	Dataset challenges with DL model possible solution to the problems related to dataset and DL model
7	Future direction	Motivation for further study and future research
8	Conclusion	Concluding remarks

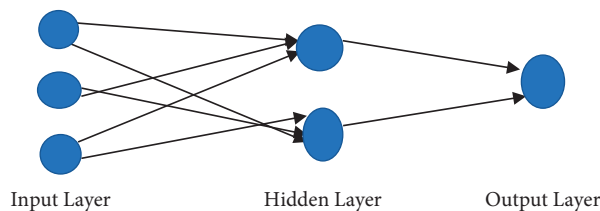


FIGURE 1: Artificial neural network (ANN) model.

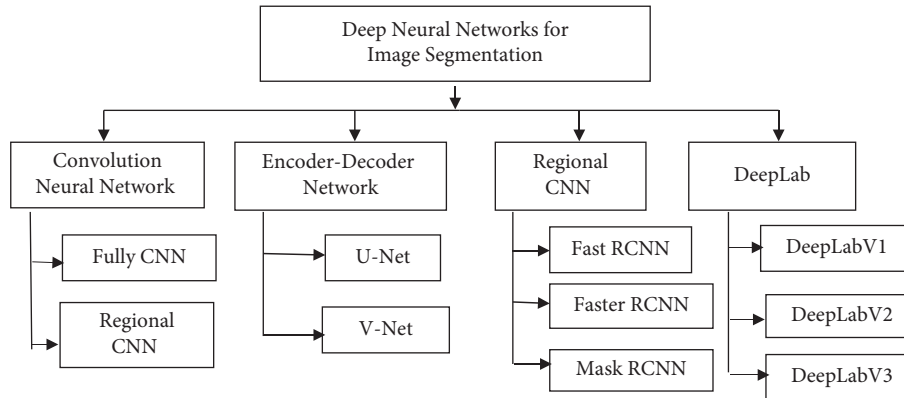


FIGURE 2: Different types of deep neural network architectures for image segmentation.

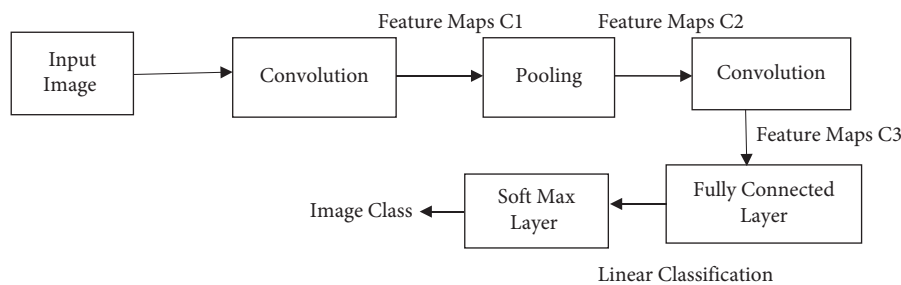


FIGURE 3: Convolutional neural network architecture.

previous layers. The model thus produces a detailed segmentation in just one go. The conventional FCN model however has the following limitations [62]:

It is not fast for real time inference and it does not consider the global context information efficiently.

In FCN, the resolution of the feature maps generated at the output is downsampled due to propagation through alternate convolution and pooling layers. This results in low resolution predictions in FCN with fuzziness in object boundaries.

An advanced FCN called ParseNet [63] has been also reported; it utilises global average pooling to attain global context. The approaches incorporating models such as conditional random fields and Markov random field into DL architecture have been also reported.

2.2. Encoder-Decoder Models. Encoder-decoder based models employ two-stage model to map data points from the input domain to the output domain. The encoder stage compresses the given input, x to latent space representation, while the decoder predicts the output from this representation. The different types of encoder-decoders based models generally employed for medical image segmentation are discussed as follows:

2.2.1. U-Net. U-Net model [64] has a downsampling and upsampling part. The downsampling section with FCN like architecture extracts features using 3×3 convolutions to

capture context. The upsampling part performs deconvolution to decrease the number of computed feature maps. The feature maps generated by downsampling or contracting part are fed as input to upsampling part so as to avoid any loss of information. The symmetric upsampling part provides precise localization. The model generates a segmentation map which categorizes each pixel present in the image.

The U-Net model offers the following advantages:

U-Net model can perform efficient segmentation of images using limited number of labelled training images

U-Net architecture combines the location information obtained from the downsampling path and the contextual information obtained from upsampling path to predict a fair segmentation map

U-Net models also have few limitations, stated as follows:

Input image size is limited to 572×572

In the middle layers of deeper UNET models, the learning generally slows down which causes the network to ignore the layers with abstract features

The skip connections of the model impose a restrictive fusion scheme which causes accumulation of the same scale feature maps of the encoder and decoder networks

To overcome these limitations, the different variants of U-Net architecture have been proposed in the literature: U-Net++ [65], Attention U-Net [66], and SD-UNet [67].

2.2.2. VNet. It is also an FCN-based model employed for medical image segmentation [68]. VNet architecture has two parts, compression and decompression network. The compression network comprises convolution layers at each stage with residual function. These convolution layers utilized volumetric kernels. The decompression network extracts feature and expands the spatial representation of low resolution feature maps. It gives two-channel probabilistic segmentation for both foreground and background regions.

2.3. Regional Convolutional Network. Regional convolutional network has been utilized for object detection and segmentation task. The R-CNN architecture presented in [69] generates region proposal network for bounding boxes using selective search process. These region proposals are then warped to standard squares and are forwarded to a CNN so as to generate feature vector map as output. The output dense layer consists of features extracted from the image and these features are then fed to classification algorithm so as to classify the objects lying within the region proposal network. The algorithm also predicts the offset values for increasing the precision level of the region proposal or bounding box. The processes performed in R-CNN architecture are shown in Figure 4. The use of basic RCN model is restricted due to the following:

It cannot be implemented in real time as it takes around 47 seconds to train the network for classification task of 2000 region proposals in a test image.

The selective search algorithm is a predetermined algorithm. Therefore, learning does not take place at that stage. This could lead to the generation of unfavourable candidate region proposals.

To overcome these drawbacks, different variants of R-CNN, fast R-CNN, faster R-CNN, and mask R-CNN have been proposed in the literature.

2.3.1. Fast R-CNN. In R-CNN, the proposed regions of image overlap and same CNN computations are carried again and again. The fast R-CNN reported by [70] is fed with an input image and a set of object proposals. The CNN then generates convolutional feature maps. After that, the ROI pooling layer reshapes each object proposal into a feature vector of fixed size. The feature vectors are sent to the last fully connected layers of the model. At the end, the computed ROI feature vector is fed to Softmax layer for predicting the class and offset values of the proposed region [71]. The fast R-CNN is slower due to the use of selective search algorithm.

2.3.2. Faster R-CNN. In R-CNN and fast R-CNN, the proposed regions were created using a process of selective search and were a slow process. So, in faster R-CNN architecture given by [72], a single convolutional network was deployed to carry out both region proposals and classification task. The model employs a region proposal network (RPN), passing the sliding window on the top of the entire

CNN feature map. For each window, it outputs K different potential boundary boxes with their respective scores representing position of object. These bounding boxes fed to fast R-CNN generate the precise classification boxes.

2.3.3. Mask R-CNN. He et al. in [73] extended faster R-CNN to present Mask R-CNN for instance segmentation. The model can detect objects in a given image and generates a high-quality segmentation mask for each object in an image. It uses RoI-Align layer to conserve the exact spatial locations of the given image. The region proposal network (RPN) generated multiple RoIs using a CNN. The RoI-Align network generates multiple bounding boxes which are warped into fixed dimensions. The warped features computed in the previous step are fed to fully connected layer so as to create classification using softmax layer. The model has three output branches with one branch computing bounding box coordinates, second branch determining associated classes, and the last branch evaluating the binary mask for each RoI. The model trains all the branches jointly. The bounded boxes are improved by employing regression model. The mask classifier outputs a binary mask for each RoI.

2.4. DeepLab Model. DeepLab model employs pretrained CNN model ResNet-101/VGG-16 with atrous convolution to extract the features from an image [74]. The use of atrous convolutions gives the following benefits:

It controls the resolution of feature responses in CNNs

It converts image classification network into a dense feature extractor without the requirement of learning of any more parameters

employs conditional random field (CRF) to produce fine segmented output

The various variants of DeepLab have been proposed in the literature including DeepLabv1, DeepLabv2, DeepLabv3, and DeepLabv3+.

In DeepLabv1 [75], the input image is passed through deep CNN layer with one or two atrous convolution layers (see Figure 5). This generates a coarse feature map. The feature map is then upsampled to the size of original image by using bilinear interpolation process. The interpolated data is applied to fully connect conditional random field to obtain the final segmented image.

In DeepLabv2 model, multiple atrous convolutions are applied to input feature map at different dilation rates. The outputs are fused together. Atrous spatial pyramid pooling (ASPP) segments the objects at different scales. The ResNet model used the atrous convolution with different rates of dilation. By using atrous convolution, information from large effective field can be captured with reduced number of parameters and computational complexity.

DeepLabv3 [20] is an extension of DeepLabv2 with added image level features to the atrous spatial pyramid pooling (ASPP) module. It also utilises batch normalization so as to easily train the network. DeepLabv3+ model combines the ASPP module of DeepLabv3 with encoder and

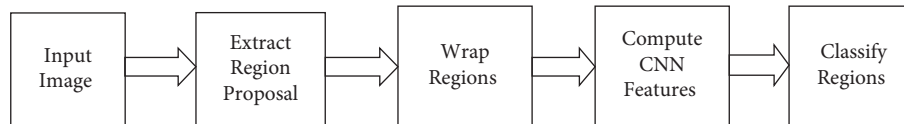


FIGURE 4: R-CNN architecture.

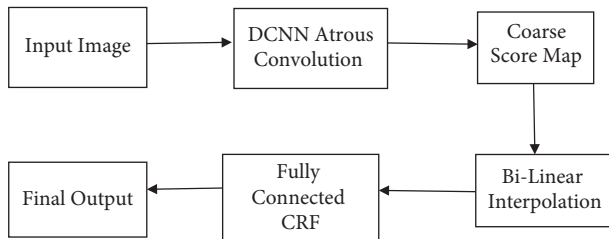


FIGURE 5: DeepLab architecture.

decoder structure. The model uses Xception model for feature extraction. The model also employed atrous and depth-wise separable convolution to compute faster. The decoder section merges the low- and the high-level features which correspond to the structural details and semantic information.

DeepLabv3+ [76] consists of an encoding and a decoding module. The encoding path extracts the required information from the input image using atrous convolution and backbone network like MobileNetv2, PNASNet, ResNet, and Xception. The decoding path rebuilds the output with relevant dimensions using the information from the encoder path.

2.5. Comparison of Different Deep Learning-Based Segmentation Methods. The different deep neural networks discussed in the above sections are employed for different applications. Each model has its own advantages and limitations. Table 3 gives a brief comparison between different deep learning-based image segmentation algorithms.

3. Applications of Deep Neural Networks in Medical Image Segmentation

Deep learning networks had contributed to various applications like image recognition and classification, object detection, image segmentation, and computer vision. A block diagram representing deep learning-based system is given in Figure 5. The first step in deep learning system consists of collecting data [77]. The collected data is then analyzed and preprocessed to be available in the format acceptable to the next block. The preprocessed data is further divided into training, validation, and testing dataset. A deep neural network-based model is selected and trained. The trained model is tested and evaluated. At the end, the analysis of the complete designed system is carried out.

This basic layout of deep learning models (shown in Figure 6) is employed in various medical applications [78] including image segmentation. In image segmentation, the objects in image are subdivided. The aim of medical image

segmentation is to identify region of interest (RoI) like tumor and lesion. The automatic segmentation of the medical images is really a difficult task because medical images are usually complex in nature due to presence of different artifacts, inhomogeneity in intensity, etc. Different deep learning models have been proposed in the literature. The choice of a particular deep learning model depends on various factors like body part to be segmented, imaging modality employed, and type of disease as different body parts and ailments have different requirements.

A 2D and 3D CNN based fully automated framework have been presented by [15] to segment cardiac MR images into left and right ventricular cavities and myocardium. The authors in [18] designed a deep CNN with layers performing convolution, pooling, normalization, and others to segment brain tissues in MR images.

Christ et al. in [30] presented a design in which two cascaded FCN were employed to segment liver and further the lesions within ROI were segmented. The final segmentation was produced by dense 3D conditional random field. Hamidian et al. in [25] converted 3D CNN with fixed field of view into a 3D FCN and generated the score map for the complete volume of CT images in one go. The authors employed the designed network for segmentation of pulmonary nodules in chest CT images. The authors concluded that by employing FCN speed of the network increases and there is fast generation of output scores. In [32], authors employed FCN for liver segmentation in CT images. In [27], authors proposed a fully convolution spatial and channel squeeze and excitation module for segmentation of pneumothorax in chest X-ray images.

Gordienko et al. [26] reported a U-Net based CNN for segmentation of lungs and bone shadow exclusion techniques on 2D CXRs images. Zhang et al. in [19] designed SDRes U-Net model, which embedded the dilated and separable convolution into residual U-Net architecture. The network was employed for segmenting brain tumor present in MR images. In [33], the authors proposed the use of Multi-ResUNet architecture for segmentation. The authors concluded that the use of Multi-ResUNet model generates better results in lesser number of training epochs as compared to the standard U-Net model. In [29], the authors segmented pneumothorax on CT images. The authors compared the performance of U-Net model with PSPNet. Ferreira [17] employed U-Net model to automatically segment heart in the short-axis DT-CMR images. The authors in [68] further designed a FCN network for segmenting 3D MRI volumes and employed a VNet based network to segment prostate in MRI images.

TABLE 3: Comparison between different image segmentation algorithms.

Deep learning algorithm	Algorithm description	Advantages	Limitations
CNN	It consists of three main neural layers, which are convolutional layers, pooling layers, and fully connected layers	(a) It is simple (b) It involves feeding segments of an image as input to the network, which labels the pixels	(a) It cannot manage different input sizes (b) Fixed size of output layer causes difficulty in segmentation task
FCN	All fully connected layers of CNN are replaced with the fully convolutional layers	The model outputs a spatial segmentation map instead of classification scores	It is hard to train a FCN model to get good performance
U-Net	It combines the location information obtained from the downsampling path and the contextual information obtained from upsampling path to predict segmentation map	It can perform efficient segmentation of images using limited number of labelled training images	(a) Input image size is limited to 572×572 . (b) The skip connections of the model impose a restrictive fusion scheme causing accumulation of the same scale feature maps of the network
VNet	It performs convolutions on each stage using volumetric kernels of size $5 \times 5 \times 5$	It can be applied to 3D data for segmentation	
R-CNN	It uses selective search algorithm to extract 2000 regions from the image called region proposals	(a) It predicts the presence of an object within the region proposals (b) It also predicts four offset values to increase the precision of the bounding box	(a) Huge amount of time is needed to train network to classify 2000 region proposals per image (b) It cannot be implemented in real time (c) Selective search algorithm is a fixed algorithm
Fast R-CNN	It uses selective search algorithm which takes the whole image and region proposals as input in its CNN architecture in one forward propagation	It improves mean average precision (mAP) as compared to R-CNN	There is high computation time due to selective search region proposal generation algorithm
Faster R-CNN	It uses region proposal network	It generates the bounding boxes of different shapes and sizes a) It is simple and flexible	There is lower computation time
Mask R-CNN	It gives three outputs for each object in the image: its class, bounding box coordinates, and object mask	b) It is current state-of-the-art technique for image segmentation task	There is high training time
DeepLabv1	a) It uses atrous convolution to extract the features from an image b) It also uses conditional random field (CRF) to capture fine details	a) There is high speed due to atrous convolution b) Localization of object boundaries improved by combining DCNNs and probabilistic graphical models	Use of CRFs makes algorithm slow
DeepLabv2	It uses <i>atrous spatial pyramid pooling (ASPP)</i> and applies multiple atrous convolutions with different sampling rates to the input feature map and fuses them together	Atrous spatial pyramid pooling (ASPP) robustly segments objects at multiple scales	There are challenges in capturing fine object boundaries
DeepLabv3	It uses atrous separable convolution to capture sharper object boundaries	It can segment sharper targets	It still needs more refinement for object boundaries
DeepLabv3+	It extends DeepLabv3 by adding a decoder module to refine the segmentation results along the object boundaries	There is better segmentation performance as compared to deepLabv3	It is a large model with number of parameters to train. So, while training on higher resolution images and batch sizes, it needs large GPU memory.

Poudel et al. in [16] developed a recurrent fully convolutional network (RFCN) to detect and segment body organ. The given design ensures fully automatic segmentation of heart in cardiac MR images. The authors concluded that the RFCN architecture reduces the computational time, simplifies segmentation pipeline, and also enables real time application. Mulay et al. in [31] presented a nested edge detection and Mask R-CNN network for segmentation of liver in CT and MR images. The input images were firstly preprocessed by

applying image enhancement so as to produce the sketch of the abdomen area. The network enhances input images for edge map. At last, the authors employed Mask R-CNN for segmenting liver from the edge maps. In [28], authors designed a CheXLocNet based on Mask R-CNN to segment area of pneumothorax from chest radiographs.

In [22], authors suggested a recurrent neural network utilizing multidimensional LSTM. The authors arranged the computations in pyramidal fashion. The authors had shown

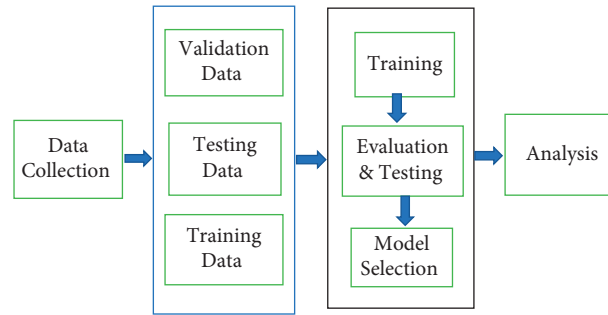


FIGURE 6: Basic layout of typical deep learning-based system.

that the PyraMiD-LSTM design can parallelize for 3D data and utilized the design for pixel-wise segmentation of MR images of brain. Table 4 summarizes the different DL based models employed for segmentation in medical images.

4. Medical Image Segmentation Datasets

Data is important in deep learning models. Deep learning models require large amount of data. The data plays an important role. It is difficult to collect the medical image data as there are data privacy rules governing collection and labelling of data and also it requires time-consuming explanation to be performed by experts [79]. The medical image datasets can be categorized into three different categories: 2D images, 2.5D images, and 3D images [2]. In 2D medical images, each information element in image is called pixels. In 3D medical images, each element is called voxel. 2.5D refers to RGB images. The 3D images are also sometimes represented as a sequential series of 2D slices. CT, MR, PET, and ultrasound pixels represent 3D voxels. The images may exist in JPEG, PNG, or DICOM format.

The medical imaging is performed in different types of modalities [2], such as CT scan, ultrasound, MRI, mammograms, positron emission tomography (PET), and X-ray of different body parts. MR imaging allows achieving variable contrast image by employing different pulse sequences. MR imaging gives the internal structure of chest, liver, brain, pelvis, abdomen, etc. CT imaging uses X-rays to obtain the information about the structure and function of the body parts. CT imaging is used for diagnosis of disease in brain, abdomen, liver, pelvis, chest, spine, and CT based angiography. Figure 7 shows MRI and CT image of brain. Mammography is a technique that uses X-rays to capture the images of the internal structure of the breast. Chest X-rays (CXR) imaging is a photographic image depicting internal composition of chest which is produced by passing X-rays through the chest and these rays are being absorbed by different amounts of different components in the chest [31]. The important publicly available medical image datasets are summarized in Table 5.

5.5. Evaluation Metrics

A metric helps in evaluating the performance of any designed model. The metrics provide the accuracy of the designed model. The popular metrics employed for assessing

effectiveness of any designed segmentation algorithm are represented in terms of the following [80]:

True positive (TP) represents that both the actual data class and the class of predicted data are true.

True negative (TN) represents that both the actual data class and the class of predicted data are false.

False positive (FP) represents that the actual data class is false while the class of predicted data is true.

False negative (FN) represents that the actual data class is true while the class of predicted data is false.

5.1. *Precision*. Precision is an evaluation metric that tells us about the proportion of input data cases that are reported to be true and represented in [81].

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

5.2. *Recall*. Recall represented in (2) gives the percentage of the total relevant results which had been correctly classified by the model [81].

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

5.3. *F1 Score*. F1 score tells about models accuracy as represented in the following equation. It is defined as the harmonic average of the precision and recall values [81]:

$$\text{F1 score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

5.4. *Pixel Accuracy*. It gives the percentage of pixels in a given input image which are correctly classified by the model [82]:

$$\text{Pixel accuracy} = \frac{\text{no. of pixels properly classified}}{\text{total number of pixels}} \quad (4)$$

5.5. *Intersection over Union*. Intersection over union (IoU) or Jaccard index [82] is a metric commonly used for checking the performance of image segmentation algorithm. It is the amount of intersecting area between the predicted

TABLE 4: Summary on deep learning-based medical image segmentation methods.

Organ	Segmented area	Model utilized	Dataset	Modality	Remarks
Cardiac	Cardiac, left, and right ventricular cavities and myocardium [15]	2D/3d CNN	ADC2017	Cardiac MR images	—
	Heart [16]	RFCN	MICCAI ² 2009 challenge dataset	Cardiac MR images	RFCN reduces computational time, simplifies segmentation, and enables real time applications
	Heart [17]	U-Net	—	DT-CMR images	U-Net automated the DT-CMR postprocessing, supporting real time results
Brain	Brain tissues [18]	2D CNN	—	Multimodal MR images	Model performance increases by employing multiple modalities
	Brain tumor [19]	SDResU-Net	—	MR images	U-Net has generalization capability
	Brain [20]	Voxel-wise residual network	MRBrainS	MRI	—
	Brain [21]	DNN	ISBI 2012 EM	TEM	—
	Pixel-wise brain segmentation [22]	MD-LSTM	MRBrainS13	Brain MR images	It can parallelize for 3D data
	Brain tumor core [23]	FCN, U-Net	—	MR images	Bounding box technique used
	Brain tumor [24]	DeepLab	—	CT images	DeepLab with conditional random fields produces high accuracy
Lungs	Pulmonary nodules [25]	3D FCN	LIDC dataset	Chest CT images	Increased speed of screening
	Lung segmentation [26]	JU-Net based CNN	JSRT	CXR	—
	Pneumothorax segmentation [27]	FC-DenseNet with SCSE module	PACS	Chest X-ray images	Spatial weighted cross-entropy loss function improves precision at boundaries
	Pneumothorax segmentation [28]	Mask R-CNN	SIIM-ACR	Chest X-ray images	Bounding box regression helps in improving classification
	Pneumothorax segmentation [29]	U-Net and PSPNet	Routine chest CT dataset	Chest CT images	—
Liver	Liver and tumor segmentation [30]	Cascaded FCN	DIRCAD dataset	CT and MRI images	Separate set of filters applied at each stage improves segmentation
	Liver segmentation [31]	HED-mask R-CNN	CHAOS challenge	CT and MR images	High segmentation accuracy obtained
	Liver segmentation [32]	FCN	MICCAI SLiver07 dataset	CT images	—
Reproductive system	Prostate [33]	VNet	—	3D MRI	—
Digestive system	Pancreas [34]	Recurrent NN (LSTM)	NIH-CT-82, ufl-mri-79	Abdominal CT and MRI images	RNN performs better than HNN and UNET
Breast	Breast masses [35]	DBN + CRF/SSVM	DDSM-BCRP, INbreast databases	Mammograms	CRF model is faster than SSVM
Eyes	Retinal blood vessels [36]	U-Net with modifications	DRIVE/STARE	Retinal images	Modification allows precise and faster segmentation of blood vessels
	Retinal blood vessels [37]	U-Net, LadderNet	DRIVE/STARE/CHASE	Retinal images	—

ADC: Alzheimer Disease Center. MICCAI: Medical Image Computing and Computer Assisted Intervention. MRBrainS: MR brain segmentation. ISBI: IEEE International Symposium on Biomedical Imaging. LIDC: Lung Image Database Consortium. JSRT: Japanese society of radiological technology. PACS: Picture Archiving and Communication System. SIIM-ACR: Society for Imaging Informatics in Medicine-American College of Radiology. DIRCAD: 3D image reconstruction for comparison of algorithm database. CHAOS: combined (CT-MR) healthy abdominal organ segmentation challenge. DDSM: digital database for screening mammography. DRIVE: digital retinal images for vessel extraction. STARE: Structural Analysis of Retinal Dataset. CHASE: Combined Healthy Abdominal Organ Segmentation Challenge.

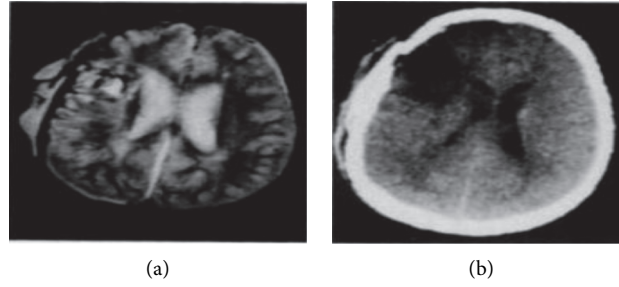


FIGURE 7: (a) MR image of brain. (b) CT scan of brain [30].

TABLE 5: Summary of medical image segmentation datasets.

Organ examined	Imaging modality	Dataset name	Dataset size	Dimensions	Image format	Segmented area	Used in reference
Brain	MRI	BraTS ¹	285	3D (240 × 240 × 155)	NIFTI	Gliomas tumor	[38]
Knee	MRI	SK110	60	3D (0.39 × 0.39 × 1.0)	NIFTI	Bones and cartilage	[39]
		OA1ZIB	507	3D (0.36 × 0.36 × 0.7)	NIFTI	Bones and cartilage	
Eyes	Retinal images	DRIVE	40	2D (768 × 584)	JPEG	Retinal vessels	[40]
	Retinal images retinal images	PALM ² STARE	1200 20	-- 700 × 605	JPEG JPEG	Lesions in pathological myopia blood vessels	[41, 42]
Abdominal area	CT	CHAOS ³	40	512 × 512	DICOM	Liver and vessels	[43]
	MRI	CHAOS	120	2D (256 × 256)	DICOM		
Chest	Chest X-ray	SIIM-ACR ⁴	—	2D (1024 × 1024)	DICOM	Pneumothorax	[44]
	Chest X-ray CT	SCR ⁵ SegTHOR	247 60	2D (2048 × 2048) -----	JPEG-----	Lungs, heart, and clavicles segmentation of heart, aorta, trachea, and esophagus	[45, 46]
Kidney	CT	KiTS ⁶	19 300	-----	NIFTI	Kidney tumor	[47]
Liver	WSI CT	PAIP	---- 50 201	3D 3D	—	Liver cancer tumor	[48]
Cardiac	MRI		30	3D		Left atrium	[49]
Lung	CT CT	Luna ⁷ DSB ⁸	16 888 1397	2D 2D	MetaImage	Nodules nucleus segmentation	[50, 51]

ACR: Society for Imaging Informatics in Medicine-American College of Radiology. BraTS: Brain Tumor Segmentation. CHAOS: Combined Healthy Abdominal Organ Segmentation Challenge. DSB: Data Science Bowl. KiTS: kidney tumor segmentation challenge. Luna: Lung Nodule Analysis. PALM: Pathologic Myopia Challenge. SCR: Segmentation in Chest Radiographs.

image segment and the ground truth mask, divided by the total area of union between the predicted segment mask and the ground truth mask:

$$\text{Dice} = \frac{|A \cap B|}{|A \cup B|}, \quad (5)$$

where A represents ground truth. B represents predicted segmentation. Mean IoU is employed for evaluating modern segmentation algorithm. Mean IoU is the average of IoU for each class.

5.6. Dice Coefficient. It is defined in the following equation and termed as twice the amount of intersection area between

the segment predicted and the ground truth divided by the total number of pixels in both the predicted segment and ground truth image [83]:

$$\text{dice} = \frac{2|A \cap B|}{|A| + |B|}. \quad (6)$$

6. Major Challenges and State-of-the-Art Solutions

The medical image segmentation field has gained advantage from deep learning, but still it is a challenging task to employ deep neural networks due to the following.

6.1. Challenges with Dataset. The different challenges related to the dataset include the following:

Limited Annotated Dataset. Deep learning network models require large amount of data. The data required for training is well annotated. The dataset plays an important role in various DL based medical procedures [84]. In medical image processing, the collection of large amounts of annotated medical images is tough [85]. Also, performing annotation on fresh medical images is tedious and expensive and requires expertise. Several large-scale datasets are publicly available. A list of few such datasets is provided in Table 2. There is still a need of more challenging datasets which can enable better training of DL models and are capable of handling dense objects. Typically, the existing 3D datasets [86] are not so large and few of them are synthetic, so more challenging datasets are required.

The size of the existing medical image datasets can be increased by (a) application of image augmentation transformations like rotating image by different angles, flipping image vertically or horizontally, cropping, and shearing image. These augmentation techniques can boost the system performance. (b) The application of transfer learning from efficient models can provide solution to the problem of limited data [87]. (c) Finally comes synthesizing data collected from various sources [87].

Class Imbalance in Datasets. Class imbalance is intrinsic in various publicly available medical image datasets. A highly imbalanced data poses great difficulty in training DL model and makes model accuracy misleading, for example, in a patient data, where the disease is relatively rare and occurs only in 10% of patients screened. The overall designed model accuracy would be high as most of the patients do not have the disease and will reach local minima [88, 89].

The problem of class imbalance can be solved by (a) oversampling the data; the amount of oversampling depends on the extent of imbalance in the dataset. (b) Second, by changing the evaluation or performance metric, the problem of dataset imbalance can be handled. (c) Data augmentation techniques can be applied to create new data samples. (d) By combining minority classes, dataset class imbalance problem can also be handled.

Sparse Annotations. Providing full annotation for 3D images is a time-consuming task and is not always possible. So, partial labelling of information slices in 3D images is done. It is really challenging to train DL model based on these sparsely annotated 3D images [85]. In case of sparsely annotated dataset, weighted loss function can be applied to the dataset. The weights for the unlabeled data in the available dataset are all set to zero, so as to learn only from the pixels which are labelled.

Intensity Inhomogeneities. In pathology images, colour and intensity inhomogeneities [90] are common. Intensity inhomogeneities cause shading over the image. It is more specific in the segmentation of MR images. Also, the TEM images have brightness variations due to presence of non-uniform support films. The segmentation process becomes tedious due to these variations.

For correcting intensity inhomogeneities [90], different algorithms are employed and many nonparametric techniques are proposed in the literature. Prefiltering operation can be employed before segmentation to remove inhomogeneities. Also, intensity inhomogeneities are taken care of by improvement in scanning devices.

Complexities in Image Texture. In medical images, there may be different artifacts present during manipulation of images. The different sensors and electronic components used for capturing images create noise in the image [11, 91]. In the captured image, gray levels can be very close to each other and there may be weak image boundaries. There may be overlap in tissues and presence of irregularities like skin lines and hair in dermoscopic images. All these complexities cause difficulty in identification of region of interest in medical images.

To remove different artefacts and noises from the image, different image enhancement techniques are used before segmentation. The image enhancement technique suppresses the noise in the image and preserves the integrity of the edges of the image.

6.2. Challenges with DL Models. The important challenging issues related to the training of DNN for robust segmentation of the medical images are as follows:

Overfitting the Model. Overfitting of the model refers to the instance when the model learn the details and regularities in training dataset with high accuracy compared with the unprocessed data instance. It mainly occurs while training the model with a small size training data [9].

Overfitting can be handled [88] by (a) increasing the size of dataset by applying augmentation techniques. (b) Dropout techniques [92] also help in handling overfitting by discarding the output of some of the random set of network neurons during each iteration.

Memory Efficient Models. Medical image segmentation models require large amount of memory [93]. In order to make these models compatible with certain devices like mobile phones, the models are required to be simplified.

Simpler models and model compression techniques can reduce memory requirements for a DL model.

Training Time. The training of deep neural network architecture needs time. In image segmentation, fast convergence of training time for deep NN is required.

The solution to this problem is (a) application of batch normalization [93]. It refers to locating the pixel values around 0 by subtracting the pixel values from the mean value of the image. It is effective in providing fast convergence. (b) Also, adding pooling layers to reduce dimension of parameters can also provide faster convergence.

Vanishing Gradient. Deep neural network faces the problem of vanishing gradient [94]. It occurs as the final gradient loss is not able to be backpropagated to earlier layers. The vanishing gradient problem is more pronounced in 3D models.

There are several solutions to the problem of gradient vanishing. (a) By upscaling the intermediate hidden layer output using deconvolution and softmax [91], the auxiliary losses and the original loss of hidden layers are combined to strengthen the gradient value. (b) Also, by carefully initializing weights [95], for the network, we can combat the problem of vanishing gradient.

Computational Complexity. Deep learning algorithm performing feature analysis needs to operate at a high level of computational efficiency. These algorithms need high performance computing devices and GPU [96]. Some of the top algorithms may require supercomputers for training the model, which may not be available. To combat these issues, the researcher has to consider the specific number of parameters to attain a limited level of accuracy.

7. Future Direction

The image segmentation techniques have come far away from manual image segmentation to automated segmentation using machine learning and deep learning approaches. The ML/DL based approaches can generate segmentation on large set of images. It helps in identification of meaningful objects and diagnosis of diseases in the images. The image segmentation techniques discussed in the paper can be explored by future researchers for application to various datasets.

The future work may include a comparative study of the different existing deep learning models discussed in the paper on the publicly available datasets. Also, different combination of layers and classifiers can be explored to improve the accuracy of image segmentation model. There is still a requirement of an efficient solution to improve performance of image segmentation model. So, the various new deep learning model designs can be explored by future researchers.

8. Conclusion

Deep learning-based automated diagnosis of diseases from medical images had become the latest area of research. In the present work, we had summarized the most popular DL based models employed for segmentation of medical images with their underlined advantages and disadvantages. An overview of the different medical image dataset employed for segmentation of diseases and the various performance metrics utilized for evaluating the performance of image segmentation algorithm is also provided. The paper also investigates the different challenges faced in segmentation of medical images using the deep networks and discusses the different state-of-the-art solutions to overcome these challenges.

With advances in technology, deep learning plays a very important role in segmentation of images. The different studies reviewed in Section 3 confirm that applications of deep neural networks in medical image segmentation task outperform the traditional image segmentation techniques. The present work will help the researchers in designing

neural network architectures in the medical field for diagnosis of disease. Also, the researchers will become aware with the possible challenges in the field of deep learning-based medical image segmentation and the state-of-the-art solutions. This review paper provides the reference material and the valuable research in the area of medical image segmentation [97].

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Taif University Researchers Supporting Project (number TURSP-2020/114), Taif University, Taif, Saudi Arabia.

References

- [1] A. Sengur, U. Budak, Y. Akbulut, M. Karabatak, and E. Tanyildizi, "A survey on neutrosophic medical image segmentation," in *Neutrosophic Set in Medical Image Analysis*, pp. 145–165, Academic Press, MA, USA, 2019.
- [2] M. Thoma, "A survey of semantic segmentation," 2016, <https://arxiv.org/abs/1602.06541>.
- [3] N. Sharma and L. M. Aggarwal, "Automated medical image segmentation techniques," *Journal of medical physics/Association of Medical Physicists of India*, vol. 35, no. 1, 2010.
- [4] L. Ng, J. Yazer, M. Abdolell, and P. Brown, "National survey to identify subspecialties at risk for physician shortages in Canadian academic radiology departments," *Canadian Association of Radiologists Journal*, vol. 61, no. 5, pp. 252–257, 2010.
- [5] S. Yuheng and Y. Hao, "Image segmentation algorithms overview," 2017, <https://arxiv.org/abs/1707.02051>.
- [6] S. D. Olabbarriaga and A. W. M. Smeulders, "Interaction in the segmentation of medical images: a survey," *Medical Image Analysis*, vol. 5, no. 2, pp. 127–142, 2001.
- [7] P. Rajpurkar, J. Irvin, R. L. Ball et al., "Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists," *PLoS Medicine*, vol. 15, no. 11, Article ID e1002686, 2018.
- [8] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [9] D. Shen, G. Wu, and H. I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017.
- [10] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical image analysis using convolutional neural networks: a review," *Journal of Medical Systems*, vol. 42, no. 11, pp. 226–313, 2018.
- [11] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: achievements and challenges," *Journal of Digital Imaging*, vol. 32, no. 4, pp. 582–596, 2019.

- [12] T. Lei, R. Wang, Y. Wan, B. Zhang, H. Meng, and A. K. Nandi, "Medical image segmentation using deep learning: a survey," 2020, <https://arxiv.org/abs/2009.13120>.
- [13] X. Liu, L. Song, S. Liu, and Y. Zhang, "A review of deep-learning-based medical image segmentation methods," *Sustainability*, vol. 13, no. 3, p. 1224, 2021.
- [14] K. Muhammad, M. S. Obaidat, T. Hussain et al., "Fuzzy logic in surveillance big video data analysis: comprehensive review, challenges, and research directions," *ACM computing surveys (CSUR)*, vol. 54, no. 3, pp. 1–33, 2021.
- [15] C. F. Baumgartner, L. M. Koch, M. Pollefeys, and E. Konukoglu, "An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation," in *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges. STACOM 2017. Lecture Notes in Computer Science*, M. Pop, Ed., vol. 10663, Cham, Springer, 2018.
- [16] R. P. Poudel, P. Lamata, and G. Montana, "Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation," in *Reconstruction, segmentation, and analysis of medical images*, pp. 83–94, Springer, Cham, 2016.
- [17] P. F. Ferreira, R. R. Martin, A. D. Scott et al., "Automating in vivo cardiac diffusion tensor postprocessing with deep learning-based segmentation," *Magnesium. Resonance in Medicine*, vol. 84, no. 5, 2020.
- [18] W. Zhang, R. Li, H. Deng et al., "Deep convolutional neural networks for multi-modality iso-intense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.
- [19] J. Zhang, L. Xiaogang, Q. Sun, Q. Zhang, X. Wei, and B. Liu, "SDResU-Net: separable and dilated residual U-net for MRI brain tumor segmentation," *Current Medical Imaging*, vol. 1, p. 15, 2019.
- [20] L. C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Re-thinking atrous convolution for semantic image segmentation," 2017, <https://arxiv.org/abs/1706.05587>.
- [21] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, pp. 2843–2851, NV, USA, December 2012.
- [22] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, "Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation," in *Proceedings of the 28th International Conference on Neural Information Processing Systems*, pp. 2998–3006, Montreal, Canada, December 2015.
- [23] G. Wang, W. Li, M. A. Zuluaga et al., "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1562–1573, 2018.
- [24] X. Gao and Y. Qian, "Segmentation of brain lesions from CT images based on deep learning techniques," in *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 10578, Article ID 105782L, 2018.
- [25] S. Hamidian, B. Sahiner, N. Petrick, and A. Pezeshk, "3D convolutional neural network for automatic detection of lung nodules in chest CT," *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134, Article ID 1013409, 2017.
- [26] Y. Gordienko, P. Gang, J. Hui et al., "Deep learning with lung segmentation and bone shadow exclusion techniques for chest X-ray analysis of lung cancer," in *International Conference on Computer Science, Engineering and Education Applications*, pp. 638–647, Springer, Cham, 2018.
- [27] Q. Wang, Q. Liu, G. Luo et al., "Automated segmentation and diagnosis of pneumothorax on chest X-rays with fully convolutional multi-scale ScSE-DenseNet: a retrospective study," *BMC Medical Informatics and Decision Making*, vol. 20, no. 14, pp. 1–12, 2020.
- [28] H. Wang, H. Gu, P. Qin, and J. Wang, "CheXLocNet: automatic localization of pneumothorax in chest radiographs using deep convolutional neural networks," *PLoS One*, vol. 15, no. 11, 2020.
- [29] W. Wu, G. Liu, K. Liang, and H. Zhou, "Pneumothorax segmentation in routine computed tomography based on deep neural networks," in *Proceedings of the 2021 4th International Conference on Intelligent Autonomous Systems (ICoIAS)*, pp. 78–83, IEEE, Wuhan, China, May 2021.
- [30] P. F. Christ, F. Ettliger, F. Grün et al., "Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks," 2017, <https://arxiv.org/abs/1702.05970>.
- [31] S. Mulay, G. Deepika, S. Jeevakala, K. Ram, and M. Sivaprakasam, "Liver segmentation from multimodal images using HED-mask R-CNN," in *Proceedings of the International Workshop on Multiscale Multimodal Medical Imaging*, pp. 68–75, Springer, Shenzhen, China, October 2019.
- [32] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P. A. Heng, "3D deeply supervised network for automatic liver segmentation from ct volumes," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 149–157, Springer, Athens, Greece, October 2016.
- [33] N. Ibtehaz and M. S. Rahman, "MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74–87, 2020.
- [34] J. Cai, L. Lu, F. Xing, and L. Yang, "Pancreas segmentation in CT and MRI images via domain specific network designing and recurrent neural contextual learning," 2018, <https://arxiv.org/abs/1803.11303>.
- [35] N. Dhungel, G. Carneiro, and A. P. Bradley, "Deep learning and structured prediction for the segmentation of mass in mammograms," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 605–612, Springer, Munich, Germany, October 2015.
- [36] T. A. Soomro, A. J. Afifi, A. Shah et al., "Impact of image enhancement technique on CNN model for retinal blood vessels segmentation," *IEEE Access*, vol. 7, Article ID 158197, 2019.
- [37] Y. Jiang, H. Zhang, N. Tan, and L. Chen, "Automatic retinal blood vessel segmentation based on fully convolutional neural networks," *Symmetry*, vol. 11, no. 9, p. 1112, 2019.
- [38] M. Islam and H. Ren, "Fully convolutional network with hypercolumn features for brain tumor segmentation," in *Proceedings of the MICCAI workshop on Multimodal Brain Tumor Segmentation Challenge (BRATS)*, pp. 108–115, Quebec, Canada, September 2017.
- [39] T. Heimann, B. J. Morrison, M. A. Styner, M. Niethammer, and S. Warfield, "Segmentation of knee images: a grand challenge," in *Proceedings of the MICCAI workshop on medical image analysis for the clinic*, pp. 207–214, Beijing, China, September 2010.
- [40] <https://drive.grand-challenge.org/>.
- [41] <https://ieee-dataport.org/documents/palm-pathologic-myopia-challenge>.

- [42] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *Journal of Medical Imaging*, vol. 6, no. 1, 2019.
- [43] <https://chaos.grand-challenge.org/Data/>.
- [44] <https://www.kaggle.com/c/siim-acr-pneumothorax-segmentation>.
- [45] <https://www.isi.uu.nl/Research/Databases/SCR/>.
- [46] Z. Lambert, C. Petitjean, B. Dubray, and S. Ruan, "SegTHOR: segmentation of thoracic organs at risk in CT images," 2019, <https://arxiv.org/abs/1912.05950>.
- [47] N. Heller, N. Sathianathen, A. Kalapara et al., "The KiTS19 challenge data: 300 kidney tumor cases with clinical context," *CT semantic segmentations, and surgical outcomes*, 2019, <https://arxiv.org/abs/1904.00445>.
- [48] <https://paip2019.grand-challenge.org/Dataset/>.
- [49] A. Andreopoulos and J. K. Tsotsos, "Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI," *Medical Image Analysis*, vol. 12, no. 3, pp. 335–357, 2008.
- [50] <https://luna16.grand-challenge.org/>.
- [51] <https://www.kaggle.com/c/data-science-bowl-2017>.
- [52] P. Malhotra, S. Gupta, and D. Koundal, "Computer aided diagnosis of pneumonia from chest radiographs," *Journal of Computational and Theoretical Nanoscience*, vol. 16, no. 10, pp. 4202–4213, 2019.
- [53] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A survey of deep learning and its applications: a new paradigm to machine learning," *Archives of Computational Methods in Engineering*, vol. 27, pp. 1–22, 2019.
- [54] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the Advances in neural information processing systems*, pp. 1097–1105, Long Beach, CA, USA, December 2012.
- [55] F. Chollet, "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, Honolulu, HI, USA, July 2017.
- [56] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <https://arxiv.org/abs/1409.1556>.
- [57] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, Las Vegas, NV, USA, June 2016.
- [58] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size," 2016, <https://arxiv.org/abs/1602.07360>.
- [59] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, Honolulu, HI, USA, July 2017.
- [60] T. Hussain, A. Ullah, U. Haroon, K. Muhammad, and S. W. Baik, "A comparative analysis of efficient CNN-based brain tumor classification models," *Generalization with deep learning: for improvement on sensing capability*, pp. 259–278, 2021.
- [61] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, MA, USA, June 2015.
- [62] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, "A review of semantic segmentation using deep neural networks," *International journal of multimedia information retrieval*, vol. 7, no. 2, pp. 87–93, 2018.
- [63] W. Liu, A. Rabinovich, and A. C. Berg, "Parsenet: looking wider to see better," 2015, <https://arxiv.org/abs/1506.04579>.
- [64] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, Munich, Germany, October 2015.
- [65] H. Cui, X. Liu, and N. Huang, "Pulmonary vessel segmentation based on orthogonal fused U-Net++ of chest CT images," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 293–300, Springer, Shenzhen, China, October 2019.
- [66] Q. Jin, Z. Meng, C. Sun, H. Cui, and R. Su, "RA-UNet: a hybrid deep attention-aware network to extract liver and tumor in CT scans," *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 1471, 2020.
- [67] C. Guo, M. Szemenyei, Y. Pei, Y. Yi, and W. Zhou, "SD-Unet: a structured Dropout U-net for retinal vessel segmentation," in *Proceedings of the 2019 IEEE 19th international conference on bioinformatics and bioengineering (bibe)*, pp. 439–444, IEEE, Athens, Greece, October 2019.
- [68] F. Milletari, N. Navab, and S. A. Ahmadi, "V-net: fully convolutional neural networks for volumetric medical image segmentation," in *Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, IEEE, Stanford, CA, USA, October 2016.
- [69] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, Columbus, OH, USA, June 2014.
- [70] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, Santiago, Chile, December 2015.
- [71] P. Malhotra and E. Garg, "Object detection techniques: a comparison," in *Proceedings of the 2020 7th International Conference on Smart Structures and Systems (ICSSS)*, pp. 1–4, IEEE, Chennai, India, July 2020.
- [72] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," in *Proceedings of the Advances in neural information processing systems*, pp. 91–99, Montreal, Canada, December 2015.
- [73] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, Venice, Italy, October 2017.
- [74] 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.
- [75] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 801–818, Munich, Germany, September 2018.
- [76] H. Harkat, J. Nascimento, and A. Bernardino, "Fire segmentation using a DeepLabv3+ architecture," *Image and Signal Processing for Remote Sensing XXVI*, vol. 11533, Article ID 115330M, 2020.
- [77] A. Khan, A. Sohail, U. Zahoor, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural

- networks,” *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, 2020.
- [78] D. S. Kermany, M. Goldbaum, W. Cai et al., “Identifying medical diagnoses and treatable diseases by image-based deep learning,” *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.
- [79] S. S. Yadav and S. M. X. Jadhav, “Deep convolutional neural network based medical image classification for disease diagnosis,” *Journal of Big Data*, vol. 6, no. 1, 2019.
- [80] L. J. Muhammad, E. A. Algehyne, S. S. Usman, A. Ahmad, C. Chakraborty, and I. A. Mohammed, “Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset,” *SN computer science*, vol. 2, no. 1, pp. 1–13, 2021.
- [81] M. Vakili, M. Ghamsari, and M. Rezaei, “Performance analysis and comparison of machine and deep learning algorithms for IoT data classification,” 2020, <https://arxiv.org/abs/2001.09636>.
- [82] E. Shelhamer, J. Long, and T. Darrell, “Fully convolutional networks for semantic segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017.
- [83] J. Bertels, T. Eelbode, M. Berman et al., “Optimizing the Dice score and Jaccard index for medical image segmentation: theory and practice,” in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 92–100, Springer, Shenzhen, China, October 2019.
- [84] P. N. Srinivasu, A. K. Bhoi, R. H. Jhaveri, G. T. Reddy, and M. Bilal, “Probabilistic deep Q network for real-time path planning in censorious robotic procedures using force sensors,” *Journal of Real-Time Image Processing*, vol. 18, no. 5, pp. 1773–1785, 2021.
- [85] O. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: learning dense volumetric segmentation from sparse annotation,” in *Proceedings of the International conference on medical image computing and computer-assisted intervention*, pp. 424–432, Springer, Athens, Greece, October 2016.
- [86] D. N. Levin, C. A. Pelizzari, G. T. Chen, C. T. Chen, and M. D. Cooper, “Retrospective geometric correlation of MR, CT, and PET images,” *Radiology*, vol. 169, no. 3, pp. 817–823, 1988.
- [87] S. Bhattacharya, P. K. R. Maddikunta, Q. V. Pham et al., “Deep learning and medical image processing for coronavirus (COVID-19) pandemic: a survey,” *Sustainable Cities and Society*, vol. 65, Article ID p.102589, 2021.
- [88] J. Merkow, A. Marsden, D. Kriegman, and Z. Tu, “Dense volume-to-volume vascular boundary detection,” in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 371–379, Springer, Athens, Greece, October 2016.
- [89] S. Bhattacharya, P. K. R. Maddikunta, S. Hakak et al., “Antlion re-sampling based deep neural network model for classification of imbalanced multimodal stroke dataset,” *Multimedia Tools and Applications*, 2020.
- [90] S. Roy, A. Carass, P. L. Bazin, and J. L. Prince, “Intensity inhomogeneity correction of magnetic resonance images using patches,” in *Medical Imaging 2011: Image Processing*, vol. 7962, Article ID 79621F, 2011.
- [91] Y. Guo and A. S. Ashour, “Neutrosophic sets in dermoscopic medical image segmentation,” in *Neutrosophic Set in Medical Image Analysis*, pp. 229–243, Academic Press, MA, USA, 2019.
- [92] H. Cui, H. Zhang, G. R. Ganger, P. B. Gibbons, and E. P. Xing, “Geeps: scalable deep learning on distributed gpus with a gpu-specialized parameter server,” in *Proceedings of the Eleventh European Conference on Computer Systems*, pp. 1–16, Lisbon, Portugal, July 2016.
- [93] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [94] S. Ioffe and C. Szegedy, “Batch normalization: accelerating deep network training by reducing internal covariate shift,” 2015, <https://arxiv.org/abs/1502.03167>.
- [95] K. Kamnitsas, C. Ledig, V. F. Newcombe et al., “Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation,” *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.
- [96] E. Goceri, “Challenges and recent solutions, for image segmentation in the era of deep learning,” in *Proceedings of the 2019 ninth international conference on image processing theory, tools and applications (IPTA)*, pp. 1–6, IEEE, Istanbul, Turkey, November 2019.
- [97] B. V. Ginneken, B. T. H. Romeny, and M. A. Viergever, “Computer-aided diagnosis in chest radiography: a survey,” *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1228–1241, 2001.