SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379 / https://doi.org/10.14445/23488379/IJEEE-V10I2P115

Volume 10 Issue 2, 157-171, February 2023

Original Article

Deep Learning-Based Semantic Segmentation Models for Prostate Gland Segmentation

M. N. Rajesh[1], B. S. Chandrasekar[2]

[1, 2]*Department of Electronics and Communication Engineering, Faculty of Engineering & Technology, Jain (Deemed-to-be University), Bangalore, Karnataka, India.*

[1]*Corresponding Author: rajeshmn.mn@gmail.com*

Received: 10 January 2023 Revised: 10 February 2023 Accepted: 20 February 2023 Published: 28 February 2023

Abstract *- Prostate cancer (PCa) is one of the prevalent forms of cancer disease found in males due to the unusual development of cells. Early diagnosis of this PCa can be useful in terms of treatment and medication. Segmentation and classification of the PCa through manual observation are one of the diagnosis methods, but it is highly challenging due to complex boundaries and features. Machine learning-based semantic segmentation architecture models consume more energy and processing time and will lead to reduced scalability and reliability. In order to tackle these limitations, deep learning-based semantic segmentation architecture can be used as it has more advantages in discriminating the features of the lesions efficiently and accurately. The main aim of this work is to segment the PCa lesions accurately and efficiently. Hence, deep learning-based semantic segmentation model-based architectures such as U-Net, Linknet, and PSPNet are proposed in this research. These models are equipped with a backbone as Inception-ResNet-v2 CNN architecture for prostate cancer gland segmentation. Nearest neighbour interpolation and normalization methods are employed as the preprocessing technique for enhancing the PCa MRI images. The normalized image was taken for processing various settings of U-Net, LinkNet and PSP-Net architectures for performing segmentation and for optimizing these models, Adam, Adamax and Nadam optimizers are used. The experiment was performed using NCI-ISBI 2013 dataset. Performance analysis of the proposed models is evaluated using Intersection of Union (IoU) scores, where the LinkNet optimized with Adamax obtained a best IoU score of 0.763337802.*

Keywords *- PCa semantic segmentation, U-Net, LinkNet, PSP-Net, Inception-ResNet-v2.*

1. Introduction

PCa is one of the top cancer cases among the male population around the world, and it mostly affects the prostate gland. Segmentation and classification are the two main processes that are required to carry out performance analysis. The process of segmenting the prostate is challenging, and the specific challenges that are presented by each method are distinct from one another. The most significant difficulties that must be overcome to segment the prostate in medical images accurately include low contrast, micro-calcifications, speckle, and imaging aberrations such as shadow postures. The segmentation method involves reflecting the prostate borders in the MRI, which is particularly important for concentrating the subsequent processing on the organ of interest. This procedure involves extracting ROIs from a data set consisting of either 2D prostate or 3D prostate volumes and then segmenting the PCa [1].

When applying machine learning (ML) strategies, the process of precisely delineating the PCa from MRI might be a time-taking one. Therefore, deep learning (DL) was highlighted as the potential new technique for precision radiotherapy delivery in PCa, where accurate prostate segmentation assists in cancer identification and therapy. This is because DL has the ability to learn from data in a very large amount of detail [1].

The segmentation of images is a crucial yet challenging aspect of the image processing process. It has emerged as a central focus in the study of image comprehension in recent years. Image segmentation is the process of dividing an entire image into multiple sections that all have certain common characteristics. Generally, it is the distinguishing target process from an image's background. At the moment, techniques for segmentation are progressing in the direction of becoming faster and more precise. Image segmentation techniques that are founded on DL have shown promising results in recent years due to the rapid advancement of AI in general and DL in particular. These techniques have been particularly successful in the domain of segmenting images. DL has certain benefits in terms of accuracy and speed in segmentation in comparison to more standard ML and computer vision methods. Therefore, the application of DL to

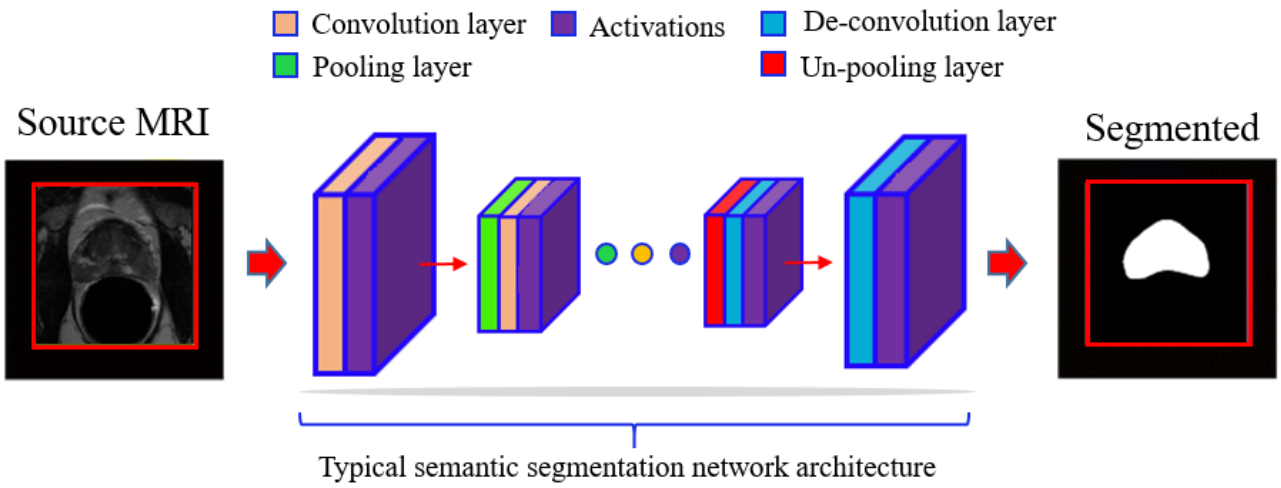


Fig. 1 Deep learning-based segmentation [5]

the segmentation of images can successfully assist physicians in confirming the infected tumor size, quantitatively evaluating the effect of both pre and post-therapy, and significantly lowering the amount of work that physicians need to undertake [2].

Magnetic resonance imaging (MRI) is a promising imaging tool for diagnosing PCa. The biopsies performed using MRI are superior to those performed using TRUS in terms of accuracy and risk to the patient. Because it offers superior soft-tissue contrast, better resolution, and it is a radiation-free technique, MRI is a particularly popular choice for identifying and staging PCa. PCa recurrence rate changes through three prostate zones known as the transitional zone (TZ), peripheral zone (PZ), and central zone (CZ), and this feature is a main consideration that went into the development of the segmentation algorithm [3]. The manual delineation of prostate glands and PCa on MR images is a time-consuming and operator-dependent task that requires skilled medical professionals—the manual delineation of prostate glands and PCa on MRI. When doing an examination of the prostate using MRI, segmentation is absolutely necessary, particularly on T2w MRI or both T2w and the related T1w images [4].

The learned CNN model compresses the underlying image using sequences of multiple convolution layers, activation, and pooling. Inverse operations have the effect of increasing the size of the compressed latent representation. The network is maintained such that it can be trained from end to end. The segmentation labels are obtained from a forward pass when the test is performed. A common procedure for segmentation that makes use of a DL model is illustrated in figure 1. The models that are used for segmentation begin by reducing the dimensions of the input image by employing a series of layers that are both convolutional and pooling [6].

This compression of the input image is done so that a better discriminative representation can be learned of the regions that need to be segmented when the latent space is made more compact. The second half of the networks is responsible for expanding the feature maps to the original image's dimensions and providing the segmentation labels for the region of interest (RoI) [7]. This paper proposes three different DL-based semantic segmentation approaches based on convolution network architectures like U-Net, LinkNet and PSP-Net. The Inception-ResNet-v2 model was implemented as a backbone network for these segmentation models for PCa lesion segmentation. Initially, image interpolation and intensity normalization were employed as preprocessing techniques for image resizing and image normalization for reducing variations in the intensity distributions of the image. Further, those processed images are applied for processing in the various sets of U-Net, LinkNet and PSP-Net architecture which is a full convolution encoder and decoder network with the skip connections among the encoder and decoder blocks using Inception-ResNet-v2 [28]. The proposed model uses hyperparameter optimization with various optimization algorithms such as Adam, Adamax, and Nadam for the segmentation techniques to enhance prostate boundary detection.

The remaining article has been sectioned as; section 2 discusses the related works done on PCa segmentations. The next section provides the implemented semantic segmentation architectures using CNN models for segmenting the lesions. Section 4 presents the proposed methodology's experimental analysis and performance analysis based on IoU Score in various settings, and section 5 concludes the work with future suggestions.

2. Literature Survey

Image segmentation is considered a classification of pixel level constraint where all the pixels were classified with one of several probable label classes, which was the final step in the development of DL models. Initially, the progression of DL approaches was directed toward issues involving image classification, followed by object detection and then segmentation of images. For instance, in cancer segmentation, all the voxels could be labelled as belonging to either the class labels of the objects of interest (target) or the backgrounds, depending on which label is assigned. Different DL-based models, such as encoder-decoder, fully convolutional, multi-scale and pyramid, recurrent neural, attention, generative, and adversarial training-based networks, have been presented for the delineation task because it is a very common task that is present in a wide variety of problem domains. These models have been developed because the delineation task is so widespread.

In the process of diagnosing PCa, it is extremely important to segment the prostate using MRI accurately. DL-based models were ineffective in earlier studies because of the disease-specific changes in the shapes and borders of the glands, as well as the complexity involved in differentiating the surrounding tissues. Utilizing combined loss and increased batch normalization function, a DL model was effectively deployed in the successful segmentation of the prostate on an MRI [8]. In order to conduct this research, the U-Net segmentation model was utilized. To get a clearer image of the segment overlap, the loss function has been altered so that it now makes use of the dice loss and the binary cross-entropy loss, both of which are derived from the next-best solution for measuring dissimilarity and similarity. The network over-fitting and vanishing gradient errors have both been minimized because of the better batch normalization with ReLU activation in the extraction of the feature process. The accuracy of this model and the amount of time it takes to process has improved as a result.

The prostate and its zones can be seen in great anatomical detail using MRI technology. It plays an essential part in a wide variety of diagnostic applications. Numerous diagnostic and therapeutic applications can be made easier thanks to automatic segmentation techniques. One example of this is the segmentation of the prostate and prostate zones using MRI. However, this is a very difficult process to accomplish due to the absence of a distinct prostate boundary, the variability of prostate tissue, and the wide range of prostate morphologies that can be found in individuals. In order to solve this issue, a Dense U-Net model was proposed in [9] to separate the prostate and its zones autonomously. The DenseNet and U-Net models were combined to create this Dense U-Net. To analyze the theory that a network could learn even when the labels were not precise, the networks were trained and tested independently on weakly annotated masks and precisely annotated masks. The prostate region is successfully detected, and the network is able to segment the gland and its various zones.

An MRI segmentation method based on Pyramid Scene Parsing Network (PSP-Net) was proposed in [10]. PSP-Net, a generally utilized architecture for producing the segmentation effects required in medical image segmentation applications, has improved its effectiveness using contrast-limited adaptive histogram equalization. This PSP-Net segmentation model worked efficiently in the PCa segmentation process utilizing MRI, and this model had a better discrimination impact than the FCN and U-Net segmentation models. In [27], three different deep learning algorithms, including U-Net, efficient residual factorized convNet (ERFNet), and efficient neural network (ENet), were offered as ways to separate the prostate using MRI. These DL networks were evaluated for their ability to segment the entire prostate gland; the results showed that ENet had the highest level of performance.

An end-to-end transfer learning-based trainable approach for segmenting cardiac MRI images was proposed in [12]. This method can be read as follows: The encoder in this model was changed with CNN network designs such as DenseNet, ResNet, and VGG. The model also made use of the Feature Pyramid Network (FPN), which is an architecture, and the U-Net architecture. The CLAHE algorithm, ROI cropping, normalization, and random training data augmentation are all components of the data training process. Because cross-entropy can only be applied on a per-pixel basis, a weighted mixture of cross-entropy was utilized as the loss function in this study. This allowed for better segmentation results than was previously possible. Exhaustive techniques of selection were used in order to determine hyperparameters like the size of the filter, learning rate, and optimizer selection for the purpose of optimizing the segmentation outcomes. After putting in place optimizers such as Adam, SGD, and Adamax for a total of one hundred epochs, the winner was determined to be Adam based on the dice score.

The majority of the currently used CNN-based medical image segmentation algorithms were initially created for the purpose of segmenting natural images. These approaches constituted the basis for the majority of the methods that are currently in use. As a result, they neglect, to a considerable extent, the distinctions that exist between the two fields, like the lower level of variability that exists in the form and appearances of the target volumes and the fewer amounts of training data that are required for medical applications. In order to solve this problem, in [13], a CNN-based technique was proposed for PCa segmentation in MRI that made use of statistical shape approaches. This approach makes predictions regarding the location of the PCa as well as the parameters of the shape model, which defines the position of the prostate surface's important points. It is also possible to use statistical shape models in order to generate additional training data, which can make the process of training big CNNs more manageable.

For the purpose of residual semantic segmentation of the prostate derived from MRI images, a VGG19-RSeg CNN model was presented in [14]. This model was created by adding residual connections to the VGG-19 fully convolutional network after it was initially established. Using semantic segmentation, or a pixel-by-pixel classification of the information contained within the input image, the VGG19-RSeg model locates the region of interest (ROI) in the images. Although a number of research have employed CNN fully in the problem of segmenting medical images, this analysis presents two additional types of residual connection—remote and neighboring—that improve the accuracy of segmentation over the basic design.

In [29], an automatic segmentation model for PCa from MRI images was developed. This model was built on U-Net and used Xception, Resnet18, and Resnet-34 as the encoder portion. Additionally, local residual connections were used. The model possesses two significant benefits. First, the model's performance has been improved thanks to the use of standard global residual connections in addition to the utilization of the connections of local residual in the decoder portion of U-Net. Second, since the encoder component uses pre-trained classification models, the convergence process to the ideal value is both accurate and rapid. This is because the process uses pre-trained weights to assist it.

3. Methodology

Medical image segmentation is an application of the image processing domain to analyze and process two-dimensional or three-dimensional images to obtain segmentation, extraction, 3D reconstruction, and human organs in 3D view, diseased bodies, and soft tissues. It does this by segmenting the image into numerous parts according to how similar or unlike each region is to the others. By utilizing this technology, medical professionals can conduct even quantitative or qualitative studies of lesions and different ROIs, significantly improving the accuracy and dependability of medical diagnosis. Presently, organs and tissues of image cells are utilized as objects in research.

In general, the segmentation of clinical image could be defined by the model of set theory: An image I is given and a similarity constraint set $C_i(i = 1,2,..)$, the segmentation of I is to acquire a part of it.

$$\cup_{x=1}^{N} R_x = I,\ R_x \cap R_y = \emptyset, \forall x \neq y, x, y \in [1, N] \quad (1)$$

In this case, R_x fulfils both sets of every pixel in the communication similarity constraint $C_i(i = 1,2,..)$, also known as image regions, R_y was subject to the same constraints. In order to differentiate between the various areas, the labels x and y are utilized. The number of regions that remain after division is denoted by the positive integer N, which must be more than 2. The step-by-step method of segmenting medical images can be broken down into the following stages:

- Acquire a data set for medical imaging, which should typically consist of the training set, validation set, and test set. When applying ML to the process of image processing, the image is frequently segmented into three distinct subsections. Particularly, the training set was utilized in the process of training the network model, the validation set was utilized in the process of adjusting the hyperparameters of the model, and the test set was utilized in the process of verifying the model's final effect.
- To increase the size of the data set, preprocessing and expansion of the image can be made, which often includes the input image's standardization. Additionally, the scaling and rotation randomly on the input image should be made.
- To segment the medical image, a proper medical image segmentation method should be used, and the segmented images should then be output.
- In order to establish and validate the efficacy of clinical image segmentation, reliable performance metrics must be established and validated. This is a fundamental component of the overall process [2].

In this section, deep semantic segmentation architectures called U-Net, LinkNet, and PSP-Net, along with encoder backbone Inception-ResNet-v2, are used to the prostate gland MR image for segmenting the prostate glands. The prostate gland was modelled to be divided into component parts based on their appearance in this design.

3.1. Image Processing

The image preprocessing technique should be used to achieve common image sizes. Images may have a variety of sizes; nevertheless, standard image sizes can be achieved by using this technique.

3.2. Nearest Neighbor Interpolation Technique

The most common interpolation method is known as the "nearest neighbour" interpolation approach. This method, as opposed to calculating an average value based on some weighting criteria or generating an intermediate value based on sophisticated procedures, simply detects the "nearest" neighbouring pixel and assumes the intensity value of that pixel's value. This method was utilized to reduce the size of an image by computing an average value for the data points contained inside the image vector based on a set of weighting criteria. At first, the NCI-ISBI 2013 data set [20] was displayed in three different sizes: 256x256, 320x320, and 384x384. Changing the image size by conducting column-wise and row-wise interpolation on the image matrix and then using the ceil function to normalize the row-wise and column-wise pixel positions is possible.

3.3. Semantic Segmentation

The process of semantic segmentation, also known as image segmentation, involves grouping together in an image those aspects of it that correspond to the same type of object. Because each pixel in an image is categorized in accordance with a category, this method can be thought of as a sort of prediction at the pixel level. To be more exact, semantic image segmentation aims to identify each pixel of the images with a corresponding class of what is being represented. This will allow the images to be broken down into their component parts. Because this task involves predicting each pixel in the image, it is frequently referred to as dense prediction. The process of semantic segmentation consists of three steps: first, classifying an object in the images, then localizing the object by locating it and drawing a bounding box around it, and finally, segmenting the image by creating a segmentation mask and grouping the pixels in a localized image together.

Semantic segmentation may basically be summed up as the classification of a specific class of image and the subsequent separation of that class from the remaining image classes by applying segmentation masks over the top of the image. The process of semantic segmentation frequently necessitates the extraction of features and representations that are capable of deriving meaningful correlations from the input images. This effectively eliminates the noise in the images. The objective here is to take an image and make output in such a way that it contains a segmentation map. The pixel value of the input image (which can range from 0 to 255) will be converted into a class label value (0, 1, 2, ... n), and that will be the output. Based on AI, the convolutional network that was employed in the process of extracting features is an encoder. The image is also down-sampled when the encoder is utilized, while the decoder is the name for the convolutional network used for up-sampling [6].

3.4. U-Net Architecture

U-Net architecture is a variation of the fully convolutional network (FCN). It was initially developed for use in medicine, and its primary goal was to locate cancers in the lungs and the brain. Encoders and decoders in the U-Net have a very similar layout. The first method, known as down-sampling, is used to extract features, while the second method, known as up-sampling, is used to up-sample the features that have been extracted using the deconvolutional layers. The main distinction that can be made between the FCN and the U-Net is that the former employs the utilization of the ultimately extracted features for the purpose of up-sampling.

In contrast, the latter makes use of something that is referred to as a shortcut link. The problem of information loss is the motivation behind creating the shortcut connection within the U-Net. When contrasted with FCN, U-Net reveals a number of distinguishing properties. Initially, both sides of the architectural structure are totally symmetrical. Secondly, the procedure of the FCN decoder was quite straightforward, requiring just one deconvolution function; thirdly, there were no comparable convolution structures. Finally, while both U-Net and FCN make use of the jump join operation, the addition function was used by FCN while U-Net uses the stack operation. Fig. 2 represents the network architecture of the U-Net system [29].

U-Net was developed in such a way that it contains both encoder and decoder blocks in its architecture. A U-Net design is formed when these individual blocks of the encoder communicate their extracted features to the individual blocks of the decoder that correspond to them. Thus, it has been discovered that as the convolutional networks process the image, its dimensions get increasingly refined. This is due to the fact that it max-pools layers at the same time, which results in the loss of information during the process. By concatenating high-level features with low-level ones, this architecture enables the network to both acquire finer information and store more information than it would be. U-Net can produce results that are both more precise and more granular as a result of the process of concatenating the information from the various blocks [16].

3.5. LinkNet Architecture

LinkNet is a light, deep neural network architecture that was developed for the purpose of performing semantic segmentation. It has applications in various domains, including augmented reality, self-driving automobiles, and more. Figure 3 illustrates the LinkNet network's underlying architecture. In this context, "conv" refers to "convolution," and "full-conv" refers to "complete" convolution. In addition, a down-sampling by a factor of 2 can be done by conducting stridden convolution. An up-sampling by a factor of 2 can be denoted using the notation *2 or the notation /2. It was decided to employ batch normalization between each convolutional layer, and then the ReLU non-linearity was applied afterwards. The encoder is located in the left half of the network depicted in figure 3, and the decoder is located in the right half. The encoder begins its operation with an initial block responsible for performing convolution on the input images using a kernel that is 7x7 in size and a stride of 2. In addition, this block is responsible for carrying out spatial max-pooling in an area measuring 3x3, with a stride of 2. The latter part of the encoder is made up of residual blocks, denoted by the code name encoder-block (*i*). Figure 4 (a) provides a detailed illustration of the layers contained within these encoder blocks. Figure 4 (b) presents further information regarding the layers that make up the decoder blocks. Encoders for modern segmentation algorithms like VGG and ResNet, which are massive in terms of the number of parameters and GFLOPs they require, use neural networks like ResNet and VGG. LinkNet employs ResNet as its encoder, which, although a relatively lighter network, nevertheless manages to outperform.

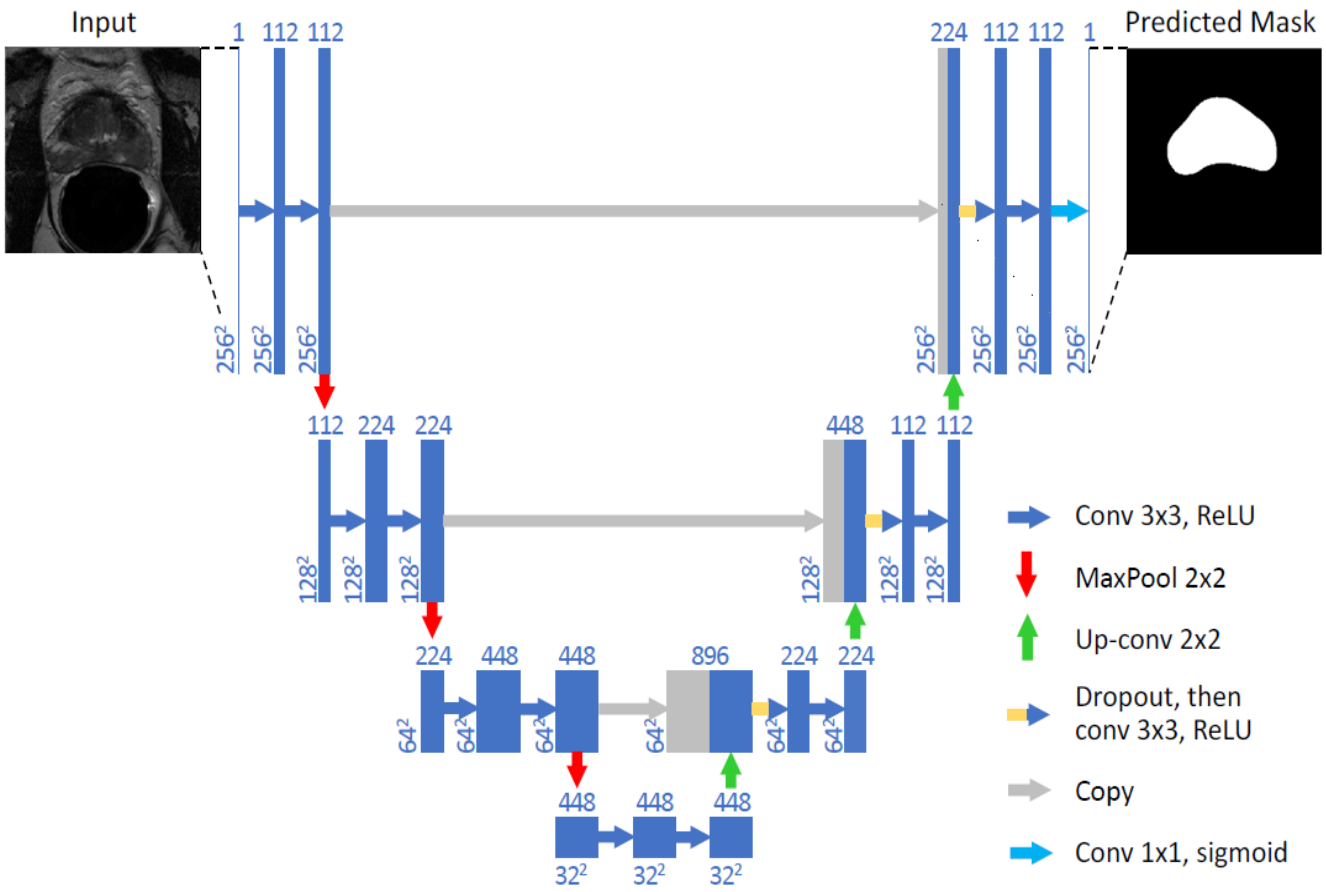


Fig. 2 Architecture of U-Net [16]

Another way to execute segmentation is by feeding the output of the encoder directly into the decoder. This is done directly using the encoder's output. In this particular implementation of LinkNet, the input of each encoder layer is instead sent directly to the output of the decoder layer that corresponds to it. By carrying out these steps, the decoder and the up-sampling operations it performs will be able to make use of the lost spatial information in its quest to retrieve it. In addition, the decoder can make use of fewer parameters due to the fact that it shares the knowledge that the encoder has gained at each and every layer [17].

3.6. PSP-Net Architecture

Pyramid Scene Parsing Network, also known as PSP-Net, is a semantic segmentation model that makes use of a pyramid parsing module (PPM). This module makes use of global context information using different-region-based context aggregation. PSP-Net performs far better than other semantic segmentation nets such as FCN, U-Net, and Deeplab when it comes to semantic segmentation. During the course of this investigation, the PSP-Net for prostate MRI segmentation was proposed. This network is mostly utilized for the semantic segmentation of prostate MRI images. PSP-network Net's architecture is depicted in Fig. 5.

PSP-Net makes use of the ReLU activation function, and the expression of the function is shown in the equation.

$$ReLU\ (x) = \max\ (x, 0) \tag{2}$$

It is clear, both from the expression of the function and the curves, that the function of ReLU does not suffer from the issue of gradient saturations, which could lead to an extension in network sparsity.

This was because the neurons whose individual variables were smaller than 0 were all fixed to 0. Through the use of the ReLU, the generalizability of both the structure and the network characteristics derived from it is increased. However, in the case of the neurons whose values are set to zero, discarding these neurons prior to training will take place when the learning rate is high. As a result, the learning rate was improved by considering both previous experience and subsequent experimentation. Figure 5 depicts the PSP-Net network architecture. PSP-Net's method for image semantic segmentation can be broken down into the following steps: After the image has been loaded into PSP-Net, the feature map of the image will be extracted using a pre-trained version of the Inception-ResNet model.

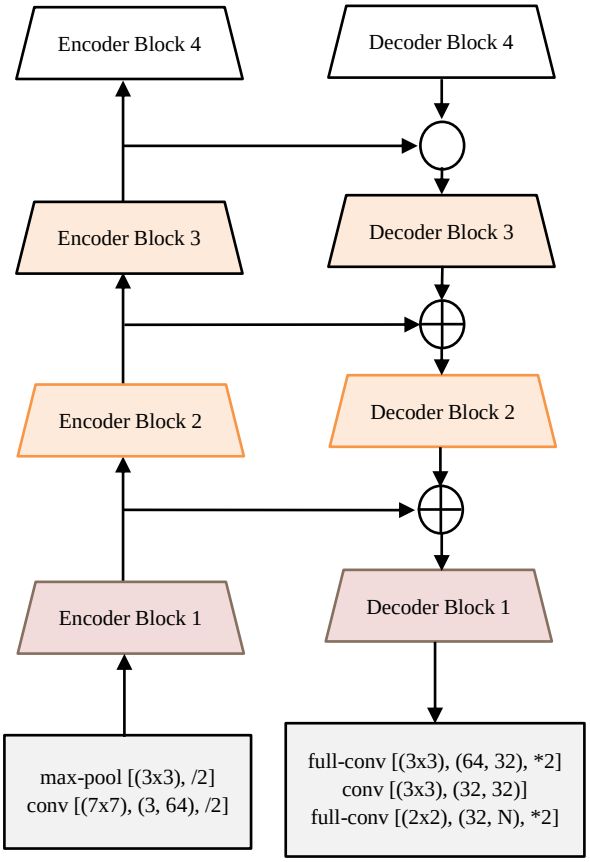


Fig. 3 Architecture of LinkNet [17]

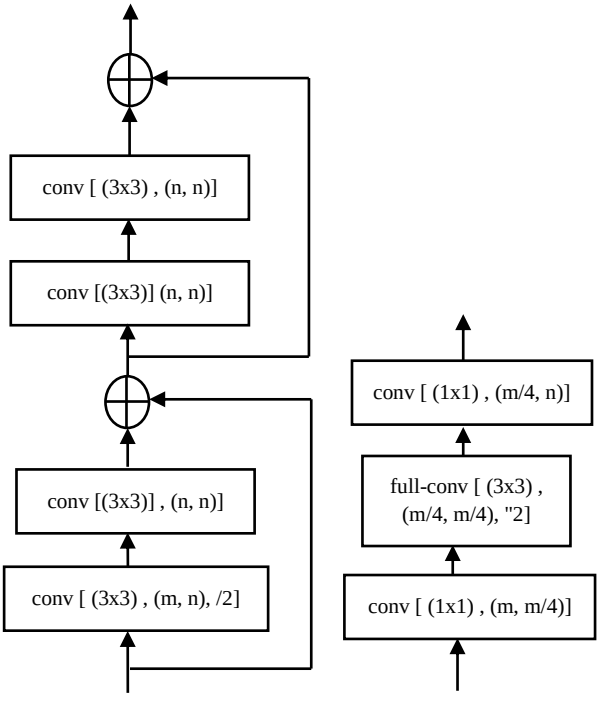


Fig. 4 Convolutional modules in (a) encoder-block (i) (b) decoder-block (i) [17]

The feature map that was acquired by the final convolution will then be sent to the PPM. This will divide the incoming feature layer into 6x6 areas, 3x3 regions, 2x2 regions, and 1x1 regions. Following this, average pooling will be performed inside each region to collect the feature information of various sub-regions, and then up-sampling will be performed. Concat is used to combine the feature information collected by Inception-ResNet-v2 and the feature information extracted from various sub-regions to produce a feature layer containing local and global context information. The final step involves classifying the feature layer using the convolution kernel softmax; thus, the prediction result for each pixel in the image is obtained [10].

3.7. Optimization of U-Net, LinkNet and PSP-Net

Optimization of the proposed segmentation models is carried out using Adam, Adamax and Nadam to reduce the error rate for the training process and parameter tuning. Adam Optimization techniques estimate the learning rate for all parameters involved in training gradients to reach an exact result with a high Dice Similarity coefficient. AdaMax is an extended version of Adam of gradient descent that standardizes the method to the infinite norms (max) and might lead to more efficient optimizations on some issues. Nadam-Nesterov-accelerated Adaptive Moment Estimation was an extended version of the Adam algorithm that includes the momentum of Nesterov and could lead to effective optimization performance of algorithms [28].

3.8. Backbone

The proposed models used Inception-ResNet-v2 pretrained model as the backbone for PCa gland segmentation. Backbone refers to a feature-extracting network usually used within an architecture. This feature extractor encodes the network input in the encoder part, from which the decoder part will be programmatically built up as part of the transfer learning process and into a certain feature representation. Hence based on the semantic segmentation model proposed herewith, for example, if U-Net is used, then it first sets the backbone with pre-trained ImageNet weights from the Keras application. Then the function builds up the decoder side by concatenating the prior outputs on the decoder side with the outputs from the related layers named in the skip connection lists and adding more convolutions. Inception-ResNet-v2 is an integration of ResNet and Inception v4. The gradient vanishing problem is avoided by using a residual connection. In this model, batch normalization is used only on top of the traditional layer. It enables the increase of the overall number of Inception blocks. The Inception blocks have convolutions with varying sizes of the same layer. These are concatenated at the end of the block—the usage of residual connection results in a drastic decrease in the time needed in training. Inception-Resnet-v2 architecture is used as a contracting path of the segmentation model [19].

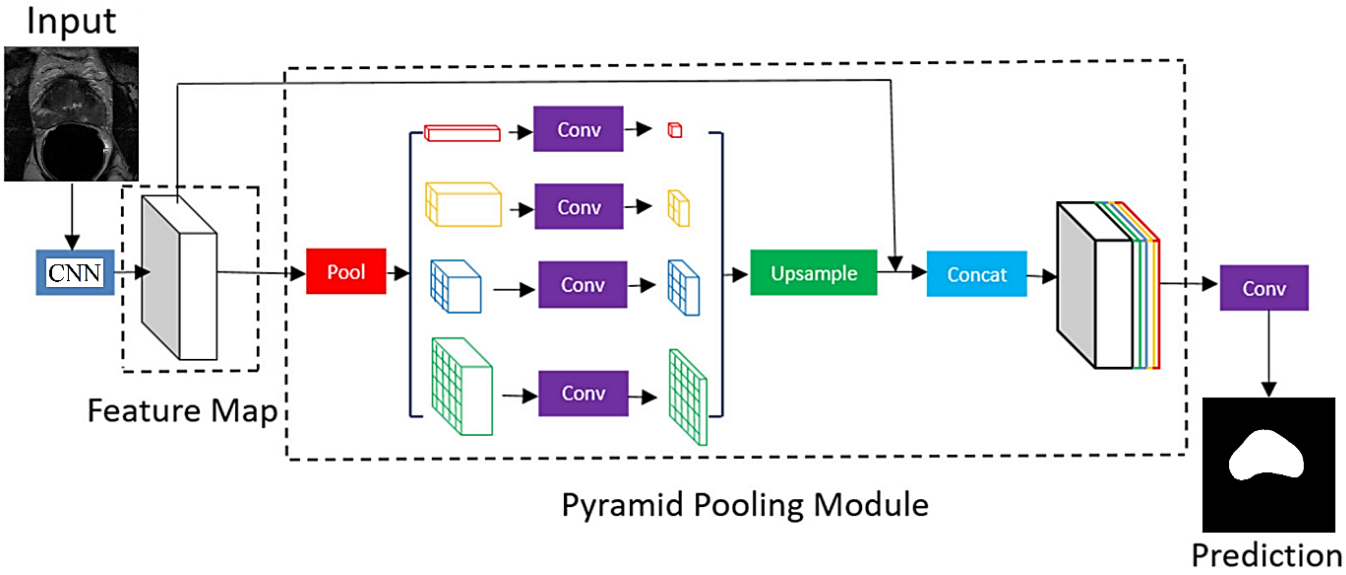


Fig. 5 Architecture of PSP-Net [10]

3.9. Loss Function

The loss function, weighted cross-entropy, was selected for the architecture of the CNN as it can differentiate between the prostate pixels and penalizes the predicted outcome of the results on gradient descent from the exact value. It is formulated as,

$$LF = -\frac{1}{n}\sum_{i=1}^{n} W_{c,i}[T_i \log p_i + (1 - T_i)\log(1 - p_i)] \quad (3)$$

Where, p_i was predicted, segmentation class, $W_{c,i}$ is the encoder weight and T_i is the target segmentation label.

Intersection-Over-Union (IOU) is a common evaluation metric for semantic image segmentation.

$$IoU = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|} \quad (4)$$

Here, *X* and *Y* were two different parts, i.e., pixel sets in the images which were 'prostate regions' and 'backgrounds' in this analysis. The sign '| |' represents the cardinal of the suitable sets.

4. Results and Discussion

Experimental results of the proposed models have been evaluated with Python program using Google Colab Notebook and with NCI-ISBI 2013 dataset containing 2276 images of various sizes as (384×384), (320×320) and (256×256) [20]. In processing the image, the dataset is portioned into a training set containing 1744, a testing set containing 261 and a validation set containing 271. The performance considering IoU measure of semantic segmentation with U-Net, LinkNet, and PSP-Net architectures in combination with a backbone network called Inception-ResNet-v2 has been evaluated while fixing hyper-parameter such as Batch Size, Validation Steps and variation in epoch values and the model was tuned with encoder weights set to 'ImageNet' and activation parameter set to sigmoid with classes set to binary and varying optimizers such as Adam, Adamax and Nadam [21-26].

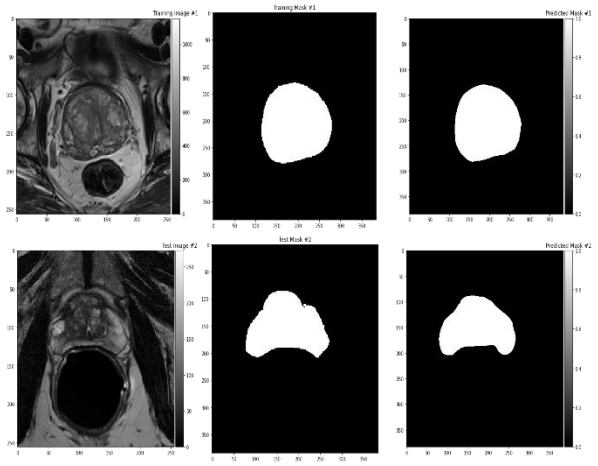
4.1. Results Obtained with U-Net

This section describes various approaches for Prostate Semantic Segmentation using U-Net architecture with backbone, along with the performance results obtained. Figure 6 displays the Original Prostate Gland Image along with the mask of ground truth that was provided against the predicted mask using U-Net with backbone and Optimizer being Adam as in figure 6 (a) U-Net with backbone and Optimizer being Adamax as in figure 6 (b) and U-Net with backbone and Optimizer being Nadam as in figure 6 (c).

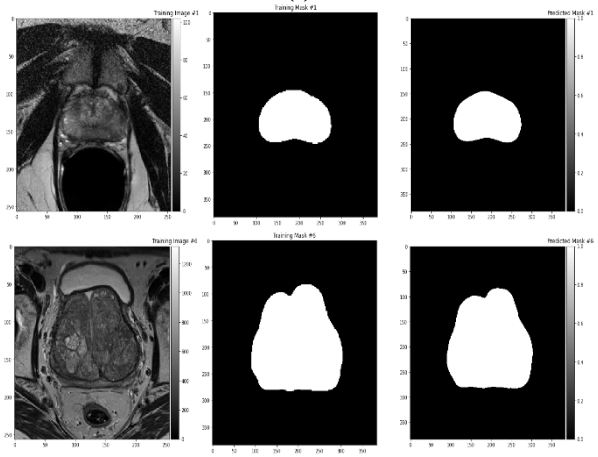
Table 1 shows IoU Score obtained when testing the U-Net model with the backbone Inception-ResNet for untrained MRI images. IoU Score obtained with Adam Optimizer is 0.733903289, with Adamax Optimizer 0.753370685 and with Nadam Optimizer giving 0.744911587. The bar chart in figure 7 shows the performance measured in terms of the IoU Score of various combinations of Optimizers with U-Net model with backbone Inception-ResNet-v2.

Table 1. IoU Score obtained with various optimizers used with U-Net model with backbone Inception-ResNet-v2

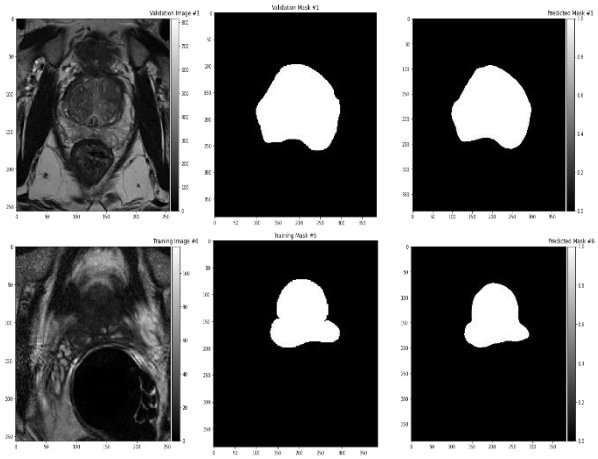
Architecture and Backbone	Adam	Adamax	Nadam
U-Net with Inception-ResNet-v2	0.7339	0.7533	0.7449



(a)



(b)



(c)

Fig. 6 Prostate MRI image along with the provided mask and the predicted mask for two trials are listed for (a) U-Net with backbone and optimizer being Adam, (b) U-Net with backbone and optimizer being adamax and (c) U-Net with backbone and optimizer being nadam

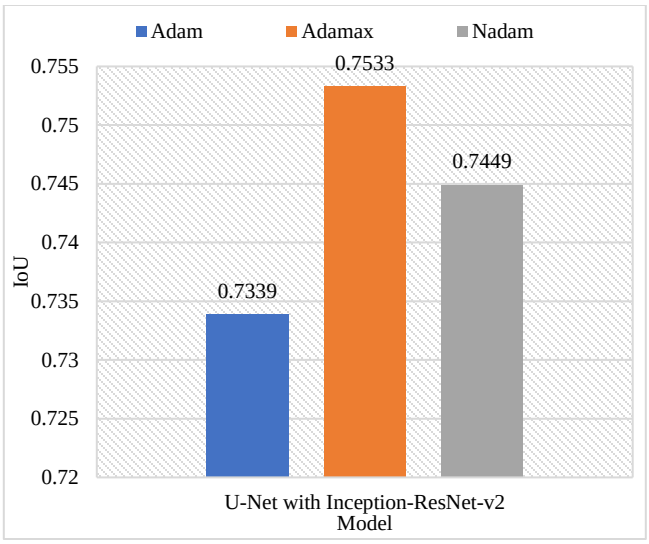
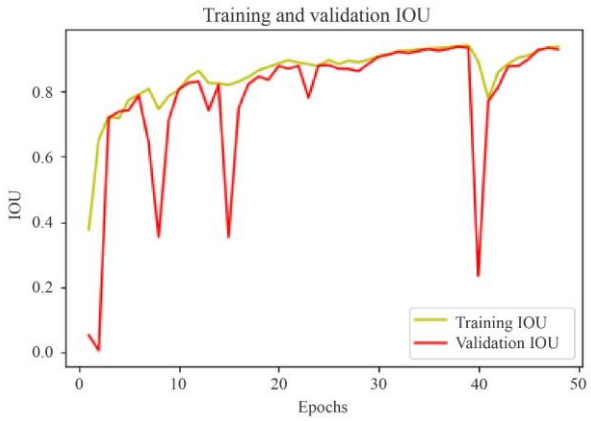
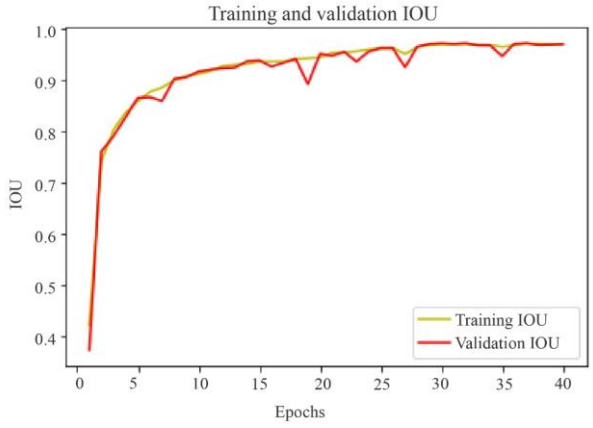


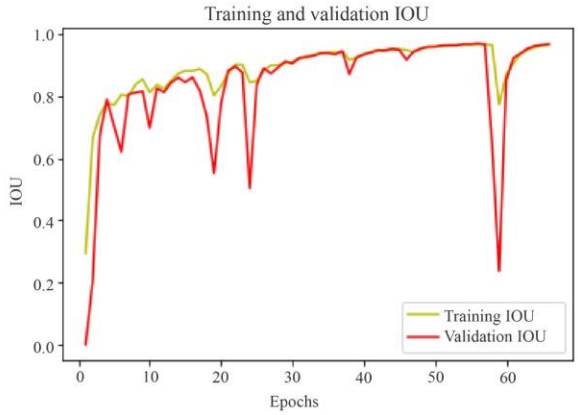
Fig. 7 Bar chart of IoU score obtained plotted with various optimizers using U-Net model with backbone Inception-ResNet



(a)



(b)



(c)

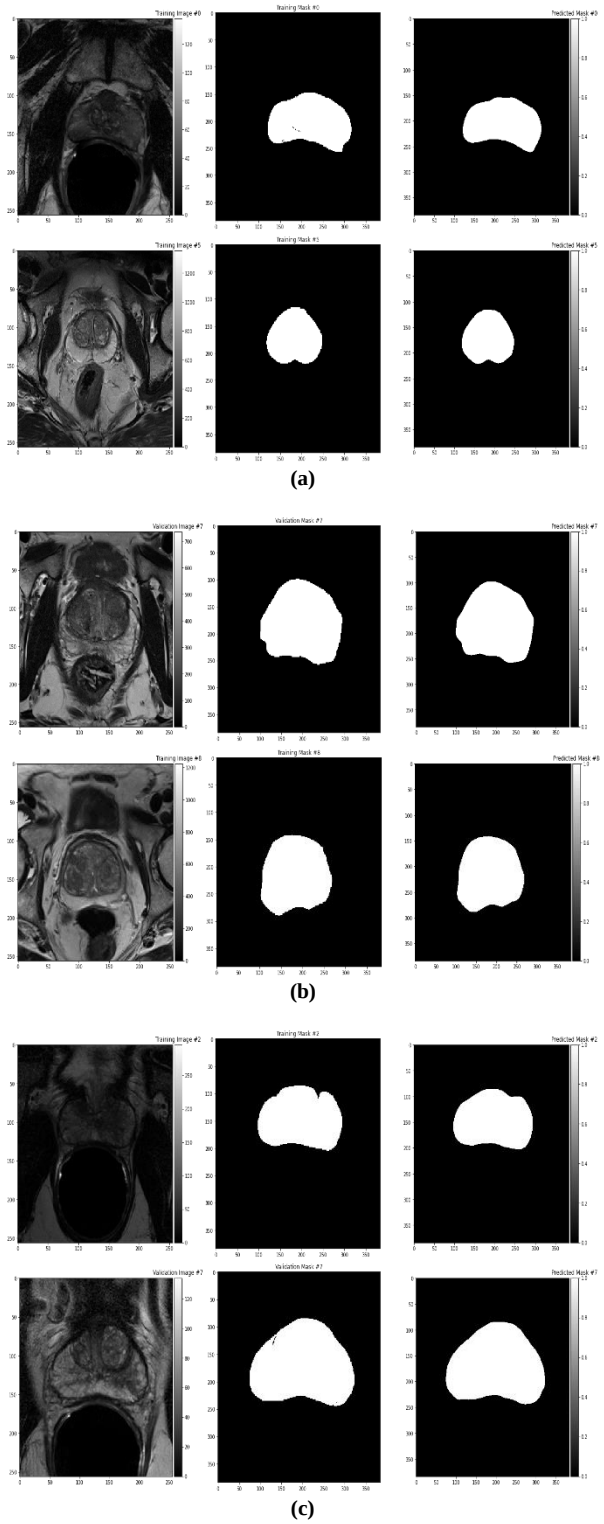
Fig. 8 Training and validation performance - IOU Vs epochs for U-Net with Backbone Inception-ResNet using a) Adam optimizer b) Adamax optimizer c) Nadam optimizer

Each model was set to 200 epochs for training, with 16 as batch size, the scheduler of the cyclic learning rate, 1e-3 initial learning rate, along with various optimizers and backbones. Performance measured as IoU Score of the U-Net Model along with backbone Inception-ResNet-v2 for training and validation as against the epochs are shown in figure 8.

It was observed that for U-Net with backbone using Adam optimizer, the training and validation IoU score went up to 0.93597, but the IoU score obtained while testing is 0.733903289. For U-Net with Adamax optimizer, the training and validation IoU score went up to 0.97182, but the IoU score obtained while testing is 0.753370685. For U-Net with Nadam optimizer, the training and validation IoU score went up to 0.96989, but the IoU score obtained while testing is 0.744911587.

4.2. Results obtained with LinkNet

This section describes various approaches for Prostate Semantic Segmentation using LinkNet architecture with backbone, along with the performance results obtained. Figure 9 displays the Original Prostate Gland Image along with the mask of ground truth that was provided against the predicted mask using LinkNet with backbone and Optimizer being Adam as in figure 9 (a) LinkNet with backbone and Optimizer being Adamax as in figure 9 (b) and LinkNet with backbone and Optimizer being Nadam as in figure 9 (c). Table 2 shows IoU Score obtained when testing the LinkNet model with the backbone for untrained MRI images. IoU Score obtained with Adam Optimizer is 0.73339371, with Adamax Optimizer 0.763337802 and with Nadam Optimizer giving 0.728948148. The bar chart in figure 10 shows the performance measured in terms of the IoU Score of various combinations of Optimizers with the LinkNet model with the backbone.



(a)

(b)

(c)

Fig. 9 Prostate MRI image along with the provided mask and the predicted mask for two trials are listed for a) LinkNet with backbone and optimizer being adam, b) LinkNet with backbone and optimizer being adamax and c) LinkNet with backbone and nadam optimizer.

Table 2. IoU Score obtained with various optimizers used with LinkNet model with backbone

Architecture and Backbone	Adam	Adamax	Nadam
LinkNet with Inception-ResNet-v2	0.7333	0.7633	0.7289

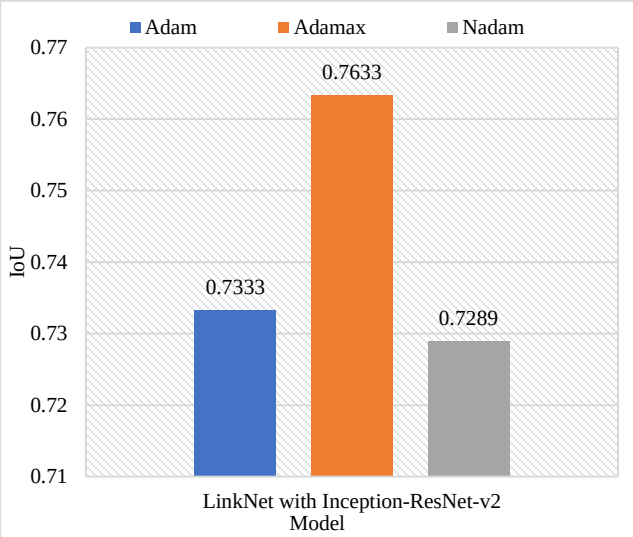
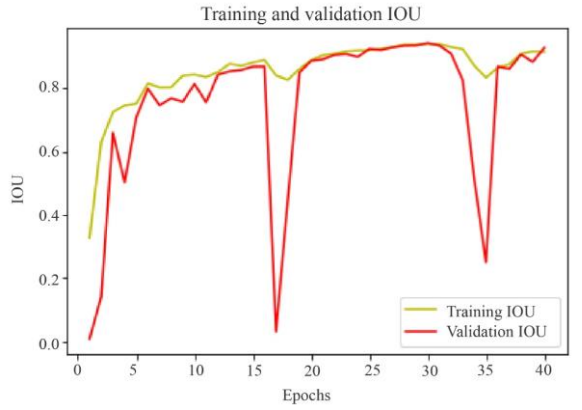
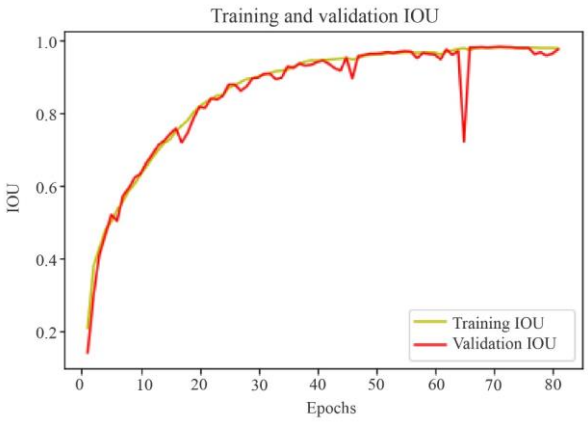


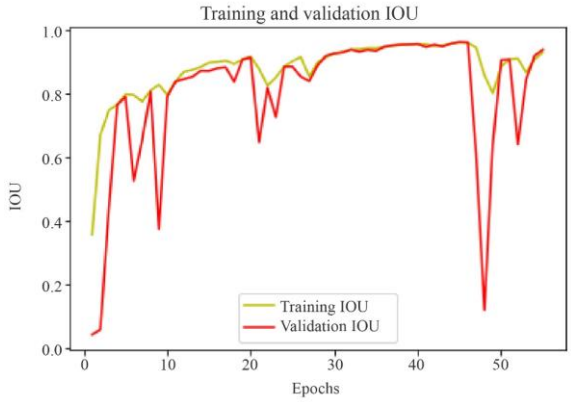
Fig. 10 Bar chart of IoU score obtained plotted with various optimizers using LinkNet model with backbone Inception-ResNet-v2



(a)



(b)



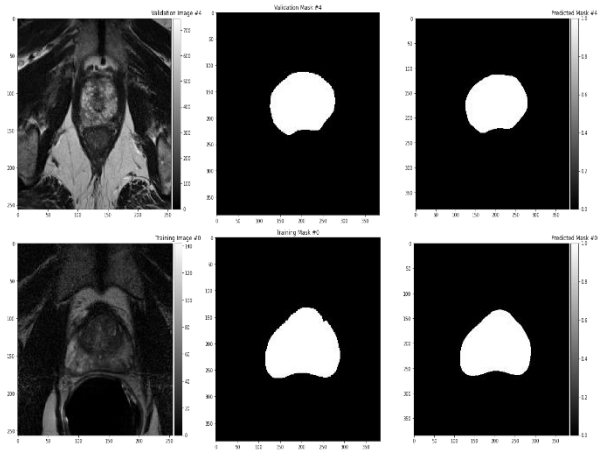
(c)

Fig. 11 Training and validation performance - IOU Vs epochs for LinkNet with backbone Inception-ResNet-v2 using a) Adam optimizer b) Adamax optimizer c) Nadam optimizer

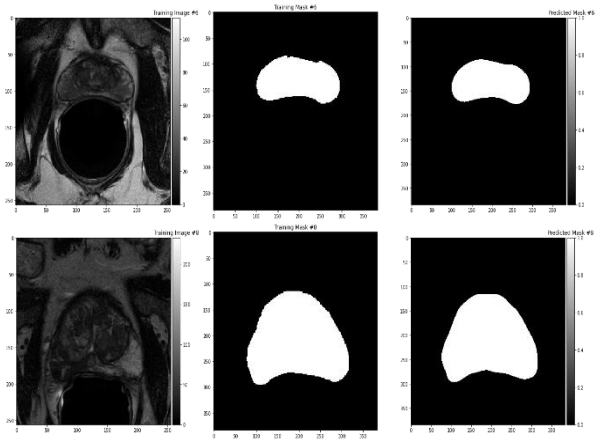
Each model was set to 200 epochs for training, with 16 as batch size, the scheduler of the cyclic learning rate, 1e-3 initial learning rate, along with various optimizers and backbones. Performance of the LinkNet Model - IoU Score obtained for Training and Validation against the epochs for the best 3 results are shown in figure 11. It was observed that for LinkNet with backbone and with Adam optimizer, the training and validation IoU score went up to 0.94219, but the IoU score obtained while testing was 0.73339371. With Adamax optimizer, the training and validation IoU score went up to 0.98033. However, the IoU score obtained while testing was 0.763337802, and with the Nadam optimizer, the training and validation IoU score went up to 0.95980, but the IoU score obtained while testing was 0.728948148.

4.3. Results obtained with PSP-Net

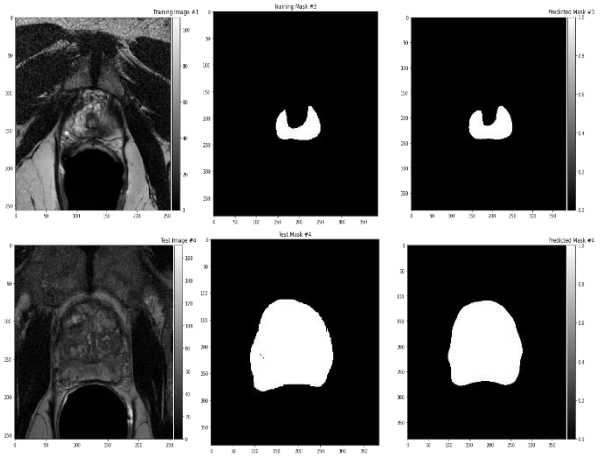
This section describes various approaches for Prostate Semantic Segmentation using PSP-Net architecture with backbone and the performance results obtained.



(a)



(b)



(c)

Fig. 12 Prostate MRI image along with the provided mask and the predicted mask for two trials are listed for (a) PSP-Net with backbone and optimizer being adam, (b) PSP-Net with backbone and optimizer being adamax and (c) PSP-Net with backbone and optimizer being nadam

Table 3. IoU Score obtained with PSP-Net model with backbone Inception-ResNet-v2

Architecture and Backbone	Adam	Adamax	Nadam
PSP-Net with Inception-ResNet-v2	0.7406	0.7152	0.7299

Figure 12 displays the original prostate gland image along with the mask of ground truth that was provided as against the predicted mask using PSP-Net with backbone and Optimizer being Adam as in Figure 12 (a) PSP-Net with backbone and Optimizer being Adamax as in Figure 12 (b) and PSP-Net with backbone and Optimizer being Nadam as in figure 12 (c).

Table 3 shows IoU Score obtained when testing the PSP-Net model with the backbone for untrained MRI images. IoU Score obtained with Adam Optimizer is 0.740674456, with Adamax Optimizer 0.715291694 and with Nadam Optimizer giving 0.729932794. The bar chart in figure 13 shows the performance measured in terms of the IoU Score of various combinations of Optimizers with the PSP-Net model with the backbone.

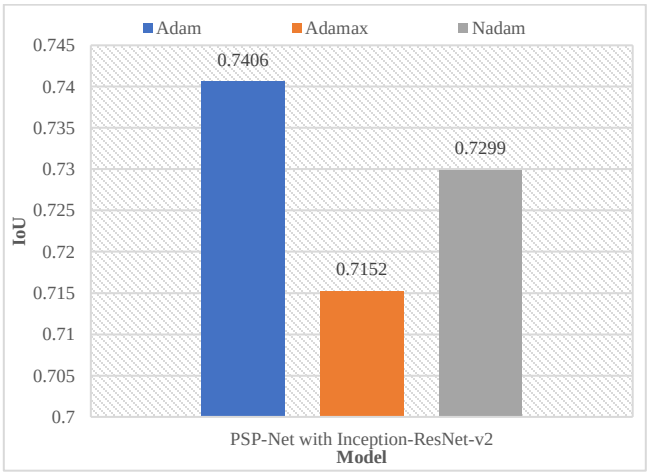
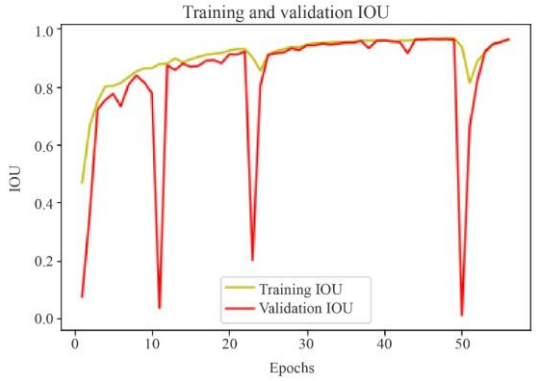
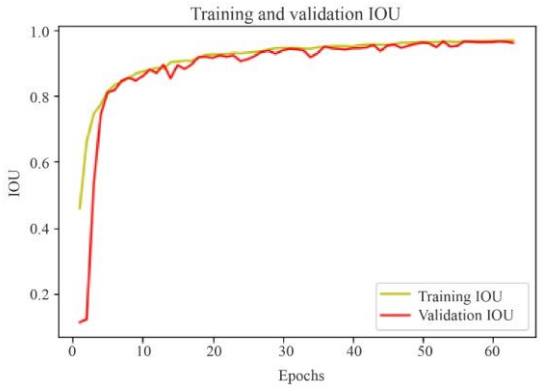


Fig. 13 Bar chart of IoU score obtained plotted with various optimizers using LinkNet model with backbone Inception-ResNet-v2

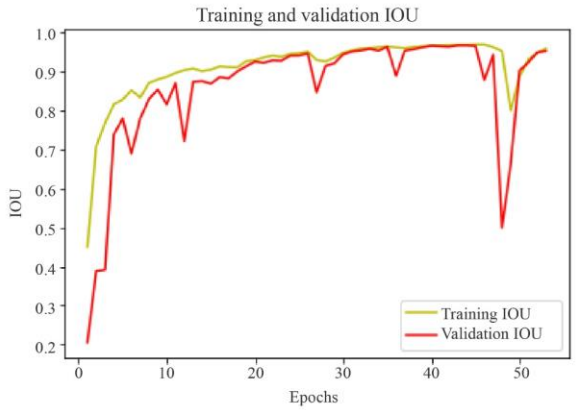
Each model was set to 200 epochs for training, with 16 as batch size, the scheduler of the cyclic learning rate, 1e-3 initial learning rate, along with various optimizers and backbones. Performance of the PSP-Net Model - IoU Score obtained for Training and Validation against the epochs for the best 3 results are shown in figure 14.



(a)



(b)



(c)

Fig. 14 Training and validation performance - IOU Vs epochs for PSP-Net with backbone Inception-ResNet-v2 using a) Adam optimizer b) Adamax optimizer c) Nadam optimizer

It was observed that for PSP-Net with backbone and with Adam optimizer, the training and validation IoU score went up to 0.96367, but the IoU score obtained while testing was 0.740674456. With Adamax optimizer, the training and validation IoU score went up to 0.96424. However, the IoU score obtained while testing is 0.715291694, and with the Nadam optimizer, the training and validation IoU score went up to 0.96587, but the IoU score obtained while testing is 0.729932794. Each of these models was pre-set to be trained for 200 epochs with16 batch sizes, the scheduler of the cyclic learning rate, 1e-3 initial learning rate to make the network architectures for extracting denser feature maps by managing the field view for proper localizations, the early saturation was defined by monitoring the loss and for every 10 epochs while restoring the best weights.

This has led to quick and effective encoder–decoder networks that build the deep representation among multi-feature images and cover good spatial data at different scales as correlated to conventional encoder–decoder networks producing the effective segmentation boundary. Based on table 4, it can be inferred that in predicting segmentation of the prostate gland with IoU score as a performance index, the LinkNet architecture with backbone and Adamax as optimizer yielded the best result of 0.763337802, followed by U-Net with Adamax as optimizer yielded a result of 0.753370685, followed by U-Net with Nadam as optimizer yielded the result of 0.744911587, followed by PSP-Net with Adam as optimizer yielding a result of 0.740674456.

Table 4. Performance evaluation of the various sets of segmentation techniques

Model	**Adam**	**Adamax**	**Nadam**
U-Net	0.7339	0.7533	0.7449
LinkNet	0.7333	0.7633	0.7289
PSP-Net	0.7406	0.7152	0.7299

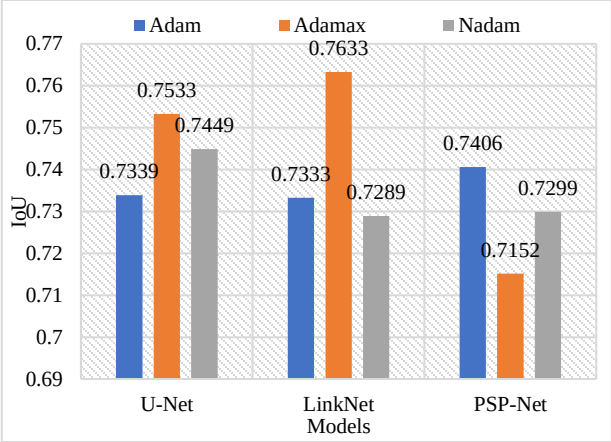


Fig. 15 Performance analysis comparison of proposed models

4.4. Comparative Analysis and Limitations

The performance of the proposed models was compared with the existing segmentation techniques called Dynamic Multi-Atlas (DMA) and DMA with Watershed, which is developed based on deep learning. These DMA and DMA + Watershed techniques were proposed for segmenting the prostate lesions in MRI images using the NCI-ISBI-2013 dataset, as discussed in [19]. Based on the obtained performances from these techniques, the performance comparison of the proposed models with the best results is validated, as shown in Table 5.

Table 5. Performance comparison of segmentation techniques

Models	**IoU**
U-Net	0.753370685
LinkNet	0.763337802
PSP-Net	0.740674456
DMA [21]	0.67±0.07
DMA + Watershed [21]	0.69±0.07

Table 5 compares the proposed models' performance with existing techniques such as DMA and DMA + Watershed for validation. The best obtained IoU values of the proposed models, such as U-Net with backbone optimized with Adamax, LinkNet with backbone optimized with Adamax, and PSP-Net with backbone optimized with Adam models, are compared. Compared to the DMA and DMA + Watershed technique, the proposed LinkNet with backbone optimized with Adamax optimizer achieved 0.763337802 as the best result in this analysis compared to the other proposed models. Following LinkNet, the U-Net and PSP-Net obtained better results than the DMA technique with 0.753370685 and 0.740674456. However, the DMA + Watershed technique has outperformed the proposed models U-Net and PSP-Net.

However, the proposed research still has some limitations. The proposed research does not use the noise elimination method that is helpful for segmenting prostate

images. Additionally, limited datasets and insufficient training are potential factors to focus on.

5. Conclusion and Future Work

Prostate gland segmentation using the deep learning semantic segmentation models such as U-Net, LinkNet and PSP-Net with backbone Inception-ResNet-v2 designed, implemented and discussed along with various optimizers and other hyperparameters setting. In this work, neural network architecture optimizers like Adam, Adamax and Nadam have been employed along with various epoch values and learning rates to improve learning and reduce the loss. Performance of the proposed model has been carried out using NCI-ISBI 2013 dataset to the IOC metric on various model settings. It was observed that LinkNet architecture with the backbone Inception-ResNet-v2 and Adamax as an optimizer yielded the best result, followed by U-Net with Adamax as an optimizer and then followed by U-Net with Nadam as an optimizer. The results obtained in this paper are based on some of the hyperparameters that are tuned to get better results. The results of the proposed models are compared with DMA and DMA + Watershed techniques, in which the proposed LinkNet model outperformed. However, future work can consider a better understanding and development of layers and tuning actuators of the semantic segmentation models adopted.

Acknowledgment

We thank the department of electronics and communication engineering, the faculty of engineering & technology, and Jain (Deemed-to-be University) for the motivation and encouragement throughout this work.

References

[1] David Gillespie et al., “Deep Learning in Magnetic Resonance Prostate Segmentation: A Review and a New Perspective,” *Electrical Engineering and Systems Science, Image and Video Processing, arXiv preprint*, 2020. [CrossRef]

[2] Xiangbin Liu et al., “A Review of Deep-Learning-Based Medical Image Segmentation Methods,” *Sustainability*, vol. 13, no. 3, pp. 1224, 2021. [CrossRef]

[3] Zia Khan et al., “Recent Automatic Segmentation Algorithms of MRI Prostate Regions: A Review,” *IEEE Access*, vol. 9, pp. 97878-97905, 2021. [CrossRef]

[4] Leonardo Rundo et al., “CNN-Based Prostate Zonal Segmentation on T2-weighted MR Images: A Cross-dataset Study,” *Neural Approaches to Dynamics of Signal Exchanges, Smart Innovation, Systems and Technologies*, vol. 151, pp. 269–280. 2020. [CrossRef]

[5] Fouzia Altaf et al., “Going Deep in Medical Image Analysis: Concepts, Methods, Challenges, and Future Directions,” *IEEE Access*, vol. 7, pp. 99540-99572, 2019. [CrossRef]

[6] Saeid Asgari Taghanaki et al., “Deep Semantic Segmentation of Natural and Medical Images: A Review,” *Artificial Intelligence Review*, vol. 54, pp. 137–178, 2021. [CrossRef]

[7] Paulo Lapa et al., “Enhancing Classification Performance of Convolutional Neural Networks for Prostate Cancer Detection on Magnetic Resonance Images: A Study with the Semantic Learning Machine,” *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pp. 381–382, 2019. [CrossRef]

[8] Sushma Shrestha et al., “A Novel Solution of Using Deep Learning for Prostate Cancer Segmentation: Enhanced Batch Normalization,” *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 21293–21313, 2021. [CrossRef]

[9] Nader Aldoj et al., “Automatic Prostate and Prostate Zones Segmentation of Magnetic Resonance Images Using DenseNet-like U-net,” *Scientific Reports*, vol. 10, no. 14315, pp. 1-17. 2020. [CrossRef]

[10] Lingfei Yan et al., “PSP Net-Based Automatic Segmentation Network Model for Prostate Magnetic Resonance Imaging,” *Computer Methods and Programs in Biomedicine*, vol. 207, no. 106211, pp. 1-8, 2021. [CrossRef]

[11] Vladimir Zaichick, and Sofia Zaichick, "Using Prostatic Fluid Levels of Zinc to Iron Concentration Ratio in Non-Invasive and Highly Accurate Screening for Prostate Cancer," *SSRG International Journal of Medical Science*, vol. 6, no. 11, pp. 24-31, 2019. [CrossRef]

[12] Taresh Sarvesh Sharan et al., “Encoder Modified U-Net and Feature Pyramid Network for Multi-class Segmentation of Cardiac Magnetic Resonance Images,” *IETE Technical Review*, vol. 39, no. 5, pp. 1-13, 2021. [CrossRef]

[13] Davood Karimi et al., “Prostate Segmentation in MRI using a Convolutional Neural Network Architecture and Training Strategy Based on Statistical Shape Models,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 8, pp. 1211–1219,018. [CrossRef]

[14] Md Sazzad Hossain, Andrew P. Paplinski, and John M. Betts, “Residual Semantic Segmentation of the Prostate from Magnetic Resonance Images,” *International Conference on Neural Information Processing*, pp. 510–521, 2018. [CrossRef]

[15] V .P. Amadi, N.D Nwiabu, and V. I. E. Anireh, "Case-Based Reasoning System for the Diagnosis and Treatment of Breast, Cervical and Prostate Cancer," *SSRG International Journal of Computer Science and Engineering* , vol. 8, no. 8, pp. 13-20, 2021. [CrossRef]

[16] Pulung Hendro Prastyo, and Amin Siddiq Sumi, “Optic Cup Segmentation using U-Net Architecture on Retinal Fundus Image,” *Journal of Information Technology and Computer Engineering*, vol. 4, no. 2, pp. 105-109, 2020. [CrossRef]

[17] Abhishek Chaurasia, and Eugenio Culurciello, “LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation,” *IEEE Visual Communications and Image Processing*, pp. 1-4, 2017. [CrossRef]

[18] Komarala Narendra Babu et al., "An Updated Review on Nanoparticles Targeting Prostate Cancer," *SSRG International Journal of Medical Science*, vol. 9, no. 6, pp. 18-31, 2022. [CrossRef]

[19] Long D. Nguyen et al., "Deep CNNs for Microscopic Image Classification by Exploiting Transfer Learning and Feature Concatenation," *IEEE International Symposium on Circuits and Systems*, pp. 1-5, 2018. [CrossRef]

[20] NCI-ISBI 2013. [Online]. Available: https://wiki.cancerimaging archive.net/display/Public/NCI-ISBI+2013+Challenge+Automated+Segmentation+of+Prostate+Structures.

[21] Hamid Moradi, and Amir Hossein Foruzan, "Integration of Dynamic Multi-Atlas and Deep Learning Techniques to Improve Segmentation of the Prostate in MR Images," *International Journal of Image and Graphics*, vol. 22, no. 4, pp. 1-17, 2022. [CrossRef]

[22] C. Narmatha, and M Surendra Prasad "A Review on Prostate Cancer Detection using Deep Learning Techniques," *Journal of Computational Science and Intelligent Technologies*,; vol. 1, no. 2, pp. 26–33, 2020 [CrossRef]

[23] M. N. Rajesh, and B. S. Chandrasekar, "Imaging Modalities used in Prostate Cancer Detection," *Journal of Computational Science and Intelligent Technologies*, vol. 2, no. 1, pp. 27–33, 2021. [CrossRef]

[24] M. N. Rajesh, B. S. Chandrasekar, and S. Shivakumar Swamy, "Feature Extraction and Analysis of Prostate Cancer MR Images," *2nd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, pp. 563-572. 2022. [CrossRef]

[25] M. N. Rajesh, B. S. Chandrasekar, and S. Shivakumar Swamy, "Prostate Cancer Detection using Radiomics-based Feature Analysis with ML Algorithms and MR Images," *SSRG International Journal of Engineering Trends and Technology*, vol. 70, no. 12, pp. 42-58, 2022. [CrossRef]

[26] M. N. Rajesh, and B. S. Chandrasekar, "Prostate Gland Segmentation using Semantic Segmentation Models U-Net and LinkNet," *SSRG International Journal of Engineering Trends and Technology*, vol. 70, no. 12, pp. 252-271, 2022. [CrossRef]

[27] Albert Comelli et al., "Deep Learning-Based Methods for Prostate Segmentation in Magnetic Resonance Imaging," *Applied Sciences*, vol. 11, no. 2, pp. 782, 2021. [CrossRef]

[28] Yulian Zhu et al., "MRI-Based Prostate Cancer Detection with High-Level Representation and Hierarchical Classification," *Medical Physics*. vol. 44, no. 3, pp. 1028–1039. 2017. [CrossRef]

[29] Ekam Singh Chahal et al., "Unet based Xception Model for Prostate Cancer Segmentation from MRI Images," *Multimedia Tools and Applications*, vol. 81, pp. 37333–37349, 2022. [CrossRef]