2D Semantic Segmentation of the Prostate Gland in Magnetic Resonance Images using Convolutional Neural Networks

Silvia Patricia Vacacela Sarango

A thesis submitted in partial fulfillment for the
degree of Master of Intelligent Systems.

Faculty of Systems Engineering
Escuela Politécnica Nacional

Supervisor: Marco Enrique Benalcázar Palacios

June 2020

SUPERVISOR'S APPROVAL

As director of the 2D Semantic segmentation of the prostate gland in magnetic resonance images using convolutional neural networksdegree work developed by Silvia Patricia Vacacela Sarango, student of the Master's degree in Computing, having supervised the completion of this work and made the corresponding corrections, I approve the final drafting of the written document to continue with the corresponding procedures to support oral defense.

Marco Enrique Benalcázar Palacios
SUPERVISOR

AUTHORSHIP DECLARATION

I, Silvia Patricia Vacacela Sarango, declare under oath that the work described here is my responsibility; that has not been previously submitted for any degree or professional qualification; and, that I have consulted the bibliographic references included in this document.

Escuela Politécnica Nacional (EPN) may use the rights corresponding to this work, as established by the Intellectual Property Law, by its Regulations and by current institutional regulations.

Silvia Patricia Vacacela Sarango

DEDICATED TO

I dedicate this thesis to my parents who gave me life, education, support and advice. To my fellow students, my teachers and friends, especially my tutor who without their help I could never have done this thesis. I thank all of them from the bottom of my soul. For all of them I make this dedication.

Contents

List of Tables

Abstract

Convolutional Neural Network (CNNs) is one of the most commonly used methods for automatic prostate segmentation in Magnetic Resonance Images (MRI). However, the results of the literature review in this work show that there is one research on the segmentation of the two main zones of the prostate gland (peripheral and central zones). This work proposed two different 2D semantic segmentation models to address the segmentation of the prostate in its main zones. The first model (Model-A) used an encoder-decoder architecture composed by a global and local U-net. The former segments the whole prostate, whereas the latter segments the central zone. The second model (Model-B) used an encoder-classifier architecture, where pixel by pixel classification was used to achieve the same segmentation goal. Moreover, model B used a Vggnet that was fine tuned. The Prostate-3T collection of the NCI-ISBI 2013 Challenge database was use to evaluate the performance of each model. The experimental results show a superior segmentation performance for Model A ($DSC = 96.79\% \pm 0.15\%$ and $IoU = 93.79\% \pm 0.29\%$) compared to Model B ($DSC = 92.50\% \pm 1.19\%$ and $IoU = 86.13\% \pm 2.02\%$).

Keywords: 2D Semantic segmentation, Magnetic Resonance Images MRI, Central Gland CG, Peripheral Zone PZ, Prostate, U-net, VGGnet.

Chapter 1

Introduction

Prostate cancer is one of the major public health worldwide problem that affects men. The Global Cancer Statistics in 2018 showed that Prostate Cancer is the most frequently diagnosed cancer in 105 countries, and the second leading cause of death among men in 46 countries [1]. Diagnosis of prostate cancer is performed by visual analysis of tissue samples obtained through biopsy from the patient. The Transrectal Ultrasound Guided Biopsy (TRUS-guided biopsy) is the standard procedure [2, 3] to detect prostate cancer. In a first time biopsy, the chance of detecting prostate cancer trough TRUS-guided biopsy is approximately 30–40% [4, 5]. However, this procedure presents some disadvantages such as: risk of life-threatening Gram-negative sepsis; does not enable sampling of the entire gland, omitting parts of the prostate; and the diagnostic can yield to false-negative findings [6].

An alternative to TRUS-guided biopsy is to add Multi-parametric Magnetic Resonance Images (mp-MRIs) for guiding this procedure [7]. The use of mp-MRIs helps to check for the presence or absence of prostate cancer without the need of biopsy sampling. Consequently, clinicians can target biopsy needles to suspicious areas rather than attempting to random samples of the whole prostate. However, this approach requires the urologist to outline or segment the prostate border and target [6] which is challenging due to: a) noise present in medical images; b) voxel intensities variation; c) finite image resolution; and d) anatomy variation of prostate among different individuals. The aforementioned difficulties in prostate segmentation suggest that some degree of uncertainty is expected for both manual and auto-segmentation [8]. However, a manual segmentation of prostate is a tedious and time-consuming task that has a high degree of interobserver variability, and requires training and medical experience [9]. Therefore, automatic segmentation (or auto-segmentation) is preferable as it would address the same problems of the manual segmentation and it can improve the speed and accuracy of segmentation.

In recent years, machine learning approach has used to solve problems in different domains classification, object detection, recognition and segmentation [10, 11]. The Artificial Neural Network is one of the approaches to machine learning, which have been widely used in automatic segmentation of medical images. Therefore, we car-

ried out a Systematic Literature Review (SLR): "Prostate Segmentation in Magnetic Resonance Images using Artificial Neural Networks". In fact, the results of the SLR shown that 74% of the studies used convolutional neural networks (CNNs), whereas 23% combined CNNs with other segmentation techniques, and 3% used feed-forward neural networks. Also, an finding was that there was only one research[12] on the segmentation of the prostate gland in its two main zones: the central gland and the peripheral zone . The research [12] used a Feed-Forward Artificial Neural Network of three layers for segmentation, there is still room for improvement in the problem of prostate segmentation in its main parts. In fact, the lack of extraction of samples from the central gland, which contains 30% of adenocarcinoma, is related to failure in the diagnose of cancer [13]. Also, SLR results shown that the type of segmentation common used was 2-dimensional 2D. Moreover, the standard architecture as U-net and VGG-net were the most used. The U-net used encoder and decoder blocks. On the other hand, the VGG-net used an encoder block followed by a classifier.

From these results of the SRL, the "2D semantic segmentation of the prostate gland in its two main parts: the central and the peripheral zone in magnetic resonance images using convolutional neuronal networks" was raised. The main contributions of research work are:

1. Two models for 2D semantic segmentation of the prostate gland in Magnetic Resonance Images using convolutional neuronal networks were to be developed and compared. One of the models was based on an encoder-decoder architecture, whereas the other model was based on an encoder-classifier architecture.

2. The images used in this research are of the non-ERC MRI type, which were more challenging for 2D semantic segmentation because they do not provide a reference point. These images were from the dataset NCI-ISBI 2013, which were the only one that has segmentation labels for the main areas of the prostate.

3. This work was aimed to perform a comparison of the performance achieved by the proposed models with respect to the results found in scientific literature.

The manuscript is structured as follows. Chapter 2 presents a SLR about prostate segmentation in MRI images using artificial neural networks or a combination with other techniques. Chapter 3 outlines materials and methodology needed to develop the proposed models and solve the research task. Results and analysis of models are presented in Chapter 4. Finally, conclusion and recommendations of this research work are presented in Chapter 5.

1.1 Goals and hypothesis

The main and specific objectives are detailed in this section. In addition, the hypothesis of this research work is detailed.

- **General goal**

  - Develop two models for 2D semantic segmentation of the prostate gland in its two main parts: the central and the peripheral zone in Magnetic Resonance Images using convolutional neural networks.

- **Specific goals**

  - Design and implement a model (model-A) for 2D semantic segmentation of the prostate gland in its two main parts: the central and the peripheral zone in MRIs using convolutional neural networks with an encoder-decoder architecture.

  - Design and implement a model (model-B) for 2D semantic segmentation of the prostate gland in its two main parts: the central and the peripheral zone in MRIs using convolutional neural networks with an encoder-classifier architecture.

  - Test models A and B to estimate their performance on central and peripheral zone segmentation on the public database "NCI-ISBI Challenge.

  - Compare the performance of models A and B with respect to literature.

- **Hypothesis**

  - The use of Convolutional Neural Networks for 2D semantic segmentation of the prostate gland in its two main parts (central and peripheral zones) on MRIs using the encoder-decoder architecture (model A) improves over the encoder-classifier architecture (model B).

Chapter 2

Systematic Literature Review

In automatic prostate segmentation, there are some traditional methods in literature such as: Atlas-based models, deformable models, threshold-based models, region-based models and other methods based on machine learning [10]. Artificial Neural Networks (ANN) have become a popular machine learning approach for solving problems in different domains [11]. Moreover, in recent years, artificial neural networks have been widely applied in medical image segmentation. Some authors have submitted reviews [14, 15, 16] that compare and show the benefits and limitations of various segmentation techniques. Other surveys evaluate and analyze different segmentation techniques for prostate with respect to different imaging modalities (e.g. TRUS, MRI and computed tomography CT) [17, 18]. Therefore, the Systematic Literature Review (SLR) about prostate segmentation in MRI images using artificial neural networks or a combination with other techniques is necessary. The main contributions/goals of this Chapter are:

- To determine the model's structure and type of segmentation obtained through Artificial Neural networks in prostate segmentation on MRI images.

- To retrieve the standard architectures used in prostate segmentation on MRI images and determine which are the most used.

- To determine the most common metrics used in the segmentation prostate.

- To determine which databases have been used to train and test models for prostate segmentation in MRI images using artificial neural networks.

2.1   Methods

This SLR followed the procedure indicated by Kitchenham [19]. Through the SLR process is identified research gaps in prostate segmentation, as well as document the procedure followed for replicability.

According to the methodology proposed by Kitchenham, the first phase consists in planning, which composed of the following steps: research questions, search process, inclusion and exclusion criteria, quality assessment, data extraction, and data analysis. The next phase conducts the review, which comprised of the following steps: research identification, study selection, study quality assessment, data extraction and monitoring progress, and data synthesis.

2.1.1 Planning

The most important pre-review activities are defining the research questions(s) that the SLR addressed and the review protocol to be followed. The latter comprises different stages such as: the search process to identify the primary studies; the definition of inclusion and exclusion criteria to classify the collected studies; the assessment of metrics to know the quality of the selected studies; the definition of the data to be extracted from accepted studies; and the data analysis stage that synthesize the data and answers the questions [19].

Research questions

Specifying the research question(s) was the most important part of any SLR. The research question drove the entire systematic review methodology in the processes of search, data extraction and data analysis [19]. This SLR was driven by a general question and four specific questions as indicated next:

RQ: What models have been proposed to segment the prostate using artificial neural networks in MRI images?

- RQ1. What model structure and type of segmentation based on artificial neural networks have been used to segment the prostate in MRI images?

- RQ2. What standard architectures of artificial neural networks have been used to segment the prostate in MRI images?

- RQ3. What metrics and results have been obtained using the proposed models to segment the prostate in MRI images?

- RQ4. What databases have been used in developing and testing the models to segment the prostate in MRI images?

Search process

The search process identified primary studies that addressed the research questions using an unbiased search strategy which should be as transparent as possible and documented in sufficient detail [19]. The repositories used for the search process was shown in Table 2.1. To reduce bias, arXiv open-access repository of electronic pre-prints was also included; even though it was not full pair reviewed. The studies

identified as potentially relevant were those that address prostate segmentation using artificial neural networks on MRI . Inclusion and exclusion criteria were applied to the relevant studies to find most relevant literature.

Source	Acronym
ACM Digital Library	ACM
IEEE Xplore	IEEE
Nature	Nature
PubMed (Medline)	PubMed
Science Direct	ScienceDirect
Springer	Springer
Arxiv	Arxiv

Table 2.1: Repositories

Inclusion and exclusion criteria

Systematic reviews required explicit inclusion and exclusion criteria to assess each primary study. These inclusion and exclusion criteria [20] were based on the research question and should capture studies of interest. The exclusion criteria was used to remove irrelevant papers (i.e. papers that did not address any aspect of the research questions) found as a result of the initial electronic searches. In Table 2.2, the inclusion and exclusion criteria were presented.

Inclusion	Exclusion
Prostate segmentation studies using convolutional or feedforward Neural Networks are included. Also, mix of convolutional or feedforward Neural Networks with other segmentation techniques are included.	Duplicate reports of the same study (when several reports of a study exist in different journals the most complete version of the study was included in the review)
Prostate segmentation studies based on MRI images are included.	All studies that do not contain MRI images.
Documents written in English are included because of the scientific relevance of this language.	All studies that do not segment the prostate.
Papers of journals or conferences published between January 1st, 2015 and February 28th, 2019.	Studies without results.

Table 2.2: Inclusion and exclusion criteria

Quality assessment

Quality assessments (QA) were usually based on "quality instruments" which were checklists of factors that need to be evaluated for each study [19]. The studies selected by using the inclusion and exclusion criteria need to be evaluated respect to their quality. To establish the quality of the selected studies, the following questions were proposed:

- QA1. Is the database used in the study specified?

- QA2. Is the architecture (structure, work flow) of the model specified?

- QA3. Is the methodology described to replicate the study?

- QA4. Are the results (with performance metrics) of the studies indicated?

Each question can be answered with Yes (Y) = 1, Partially (P) = 0.5 and No (N) = 0. Evaluations are pair reviewed and results are averaged. The different answers for each question are presented next:

QA1: Y (yes), the database(s) was/were explicitly defined in the study; P (Partly), the database(s) was/were implicit; N (no), the database(s) was/were not defined and it can not be readily inferred. QA2: Y (yes), the architecture of the model was explicitly defined in the study; P (Partially), it shows some parts of architecture or the architecture was implicit; N (no), the architecture was not defined, and it can not be inferred. QA3: Y (yes), the methodology (procedure, methods) was explicitly described, and it can be replicate the study; P (Partially), it describes some parts of methodology, or the methodology was implicit; N (no), the methodology was not defined, and it can not be replicate the study. QA4: Y (yes), results are indicated, and it shows performance metric of results in the study; P (Partially), results are implicit; N (no), results are not defined.

Data extraction

The objective of this stage was to design data extraction forms to accurately record relevant information from the accepted studies that help to answer the research question. These forms were defined and piloted to reduce the bias [19]. The extracted information was organized in the form of data items. For this review, the data items extracted from each study are shown below:

- The source and full reference

- Segmentation Challenges

- Segmented area

- MRI protocol

- Image Plane

- Modalities

- Type of MRI

- Structure

- Data Augmentation

- Pre-processing

- Feature extraction

- Classifier (Feedforward, CNN, combinate with other)

- Post-Processing

- Segmentation type

- Model on which it was based

- Performance evaluated metrics

- Result of metric

- Type of database

- Database public

16

Data synthesis

Data synthesis involved to collate and summarize the extracted information from the accepted studies. Data extraction forms from previous section were considered to answer the research questions [19]. Some of these items are complementary: segmentation challenge, segmented area and parameters of the MRI (as protocol, image plane, modality and type). All these complementary items were collected. To answer the research questions, the following items were considered:

- Define the model structure with pre-processing, feature extraction, classifier (Feedforward, CNN, mixed), post-Processing (addressing RQ1).

- What technique of data augmentation was used in the segmentation process? (addressing RQ1).

- Segmentation type (addressing RQ1).

- The segmented area and the model on which it was based (addressing RQ2).

- The metrics used in the segmentation prostate (addressing RQ3).

- The result of the most used metric in prostate segmentation (addressing RQ3).

- The studies that use private and public databases (addressing RQ4).

- Type of databases (public or private databases) (addressing RQ4).

- Detail of database MRI images parameters such as: acquisition protocol, image plane, type of image (addressing RQ4).

2.1.2 Conducting the review

Once the protocol has been defined, the review can be started. In this phase, the studies were collected by running the search string on the selected repositories. Then, inclusion and exclusion criteria were used to select the most relevant studies. Also, the quality instrument was implemented and executed on the selected studies. Next, the information was extracted from each of the final accepted studies [19].
The StArt (State of the Art through Systematic Review) software was used to conduct the review. This software aims to help the researcher, giving support to each of the stages of a Systematic Review.

Identification of research

A search string was designed to identify the primary documents for the review. It was generated considering the population and intervention aspects of the research question [19]. In this research, the population was related to magnetic resonance images of the prostate, whereas the intervention was related to the means to achieve the segmentation, which were the artificial neural networks. Although, the structure

of the search string was modified depending on the repository where it was used, the same keywords were used. For instance, the Springer repository was very sensitive and required two search strings. The structure of the search string used for each repository is shown in Table 2.3.

A total of 114 studies were retrieved by using the proposed search strings in each of the repositories. The distribution of studies in each repository was shown in Figure 2.1. The ACM repository has 4 studies that were equivalent to 4% of the total. Similarly, the IEEE repository had 23 studies (20%). For the PubMed repository, 18 studies were found (16%). In Springer repository, 38 studies were obtained (33%). In Nature repository, 11 studies were found (11%). In the Science Direct repository, 12 studies (11%). Finally, in the ArXiv repository, there were 8 studies (7%).
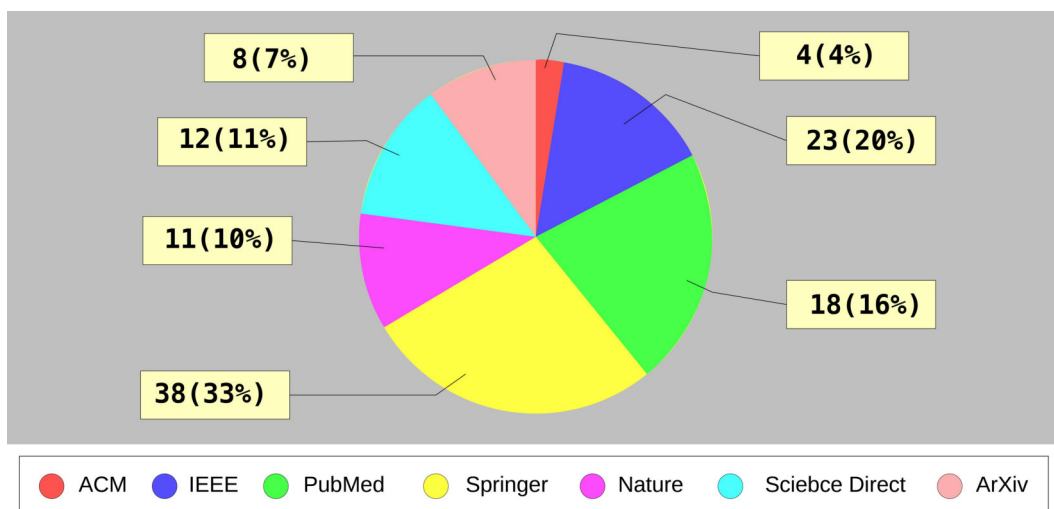
Figure 2.1: Search result in the repositories

Study Selection

Papers that address prostate segmentation using artificial neural networks (or a combination) in MRI were identified as potentially relevant. Then, duplicated papers were removed leaving 105 studies to be analyzed through the inclusion and exclusion criteria, previously defined. The inclusion and exclusion criteria were applied by reading the title, the abstract and the keywords. Consequently, the initial 114 primary studies were reduced to a total of 30 documents. This is presented in Figure 2.2 where the total number of papers obtained by using the removal of duplicates and the inclusion and exclusion criteria were presented. As a result, a total of 12 relevant studies were found in the IEEE repository; Similarly, 5 relevant studies were found for Springer and ArXiv repositories, while in Science Direct, only 2 relevant studies were found. The ACM and Nature repositories did not have any relevant studies for this literature review.

Repositories	Search string
ACM	(+Prostate +Segmentation "Magnetic Resonance Images" "Magnetic Resonance Image" "MR Images" "MR Image" "MRI" "MRIs" "MR" "convolutional neural networks" "convolutional neural network" "CNN" "CNNs" "feedforward neural networks" "feedforward" "Deep Learning" "semantic segmentation" "deep network" "autoencoder")
IEEE	Prostate AND Segmentation AND ("Magnetic Resonance Images" OR "Magnetic Resonance Image" OR "MR Images" OR "MR Image" OR "MRI" OR "MRIs" OR "MR") AND ("convolutional neural networks" OR "convolutional neural network" OR "CNN" OR "CNNs" OR "feedforward neural networks" OR "feedforward" OR "Deep Learning" OR "semantic segmentation" OR "deep network" OR "autoencoder")
Nature	that contain these terms: Prostate AND Segmentation AND ("Magnetic Resonance Images" OR "Magnetic Resonance Image" OR "MR Images" OR "MR Image" OR "MRI" OR "MRIs" OR "MR") AND ("convolutional neural networks" OR "convolutional neural network" OR "CNN" OR "CNNs" OR "feedforward neural networks" OR "feedforward" OR "Deep Learning" OR "semantic segmentation" OR "deep network" OR "autoencoder") publication date: 2015-2019
PubMed	Prostate AND Segmentation AND ("Magnetic Resonance Images" OR "Magnetic Resonance Image" OR "MR Images" OR "MR Image" OR "MRI" OR "MRIs" OR "MR") AND ("convolutional neural networks" OR "convolutional neural network" OR "CNN" OR "CNNs" OR "feedforward neural networks" OR "feedforward" OR "Deep Learning" OR "semantic segmentation" OR "deep network" OR "autoencoder")
Science Direct	Find articles with these terms: ("convolutional neural networks" OR "convolutional neural network" OR "CNN" OR "CNNs" OR "feedforward neural networks" OR "feedforward" OR "Deep Learning" OR "semantic segmentation" OR "deep network" OR "autoencoder") Year(s): 2015-2019 Title, abstract or keywords: Prostate AND Segmentation AND ("Magnetic Resonance Images" OR "Magnetic Resonance Image" OR "MR Images" OR "MR Image" OR "MRI" OR "MRIs" OR "MR")
Springer	with at least one of the words: ("Magnetic Resonance Images" OR "Magnetic Resonance Image" OR "MR Images" OR "MR Image" OR "MRI" OR "MRIs" OR "MR") where the title contains: Prostate Segmentation ("convolutional neural networks" OR "convolutional neural network" OR "CNN" OR "CNNs" OR "feedforward neural networks" OR "feedforward" OR "Deep Learning" OR "semantic segmentation" OR "deep network" OR "autoencoder") Start year and End year: 2015 - 2019
	with at least one of the words: ("Magnetic Resonance Images" OR "Magnetic Resonance Image" OR "MR Images" OR "MR Image" OR "MRI" OR "MRIs" OR "MR") where the title contains: Segmentation of the prostate Start year and End year: 2015 - 2019
ArXiv	Prostate Segmentation (Title) AND: "convolutional neural networks" OR "convolutional neural network" OR "CNN" OR "CNNs" OR "feedforward neural networks" OR "feedforward" OR "Deep Learning" OR "semantic segmentation" OR "deep network" OR "autoencoder" (All fields) From: 2015 to: 2019

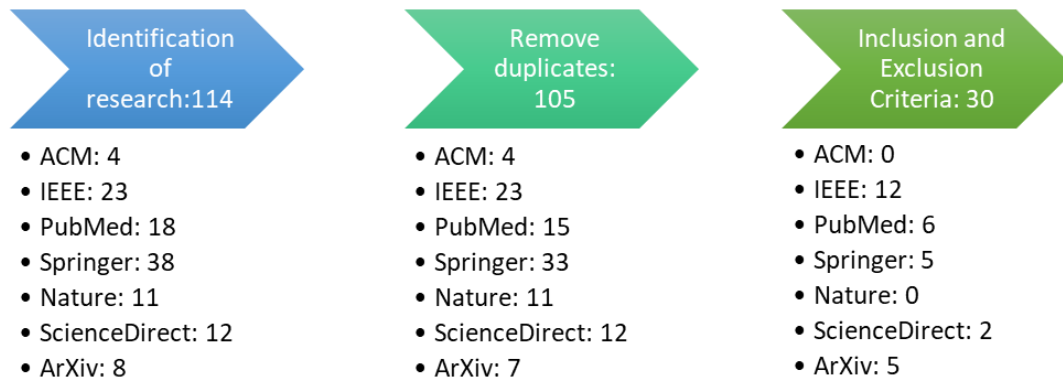Table 2.3: Search string depending of repository

Figure 2.2: Study selection process. Once the primary documents were retrieved by use of the search string, duplicates were removed and finally the inclusion and exclusion criteria were used to find most relevant documentation

Study quality assessment

The quality assessment instrument consisted in scoring each study according to a set of predefined quality questions whose answers are scored, as indicated in Section 2.1.1. The final weighted score is considered to evaluate the quality of the study. Studies that do not contain any of the proposed questions are rejected [19]. For the 30 relevant studies obtained through study selection, the quality instrument was applied. As a result, one study was rejected due to low score and three studies were duplicated. A total of 26 studies were obtained to extract information in the next stage. The implementation the quality instrument in each study is shown in Table 2.4, which contains the list of authors, the evaluated checklists and the score of each study.

Data extraction and monitoring progress

In Section 2.1.1 the data items to be extracted from the 26 accepted studies were defined. The data items were organized in lists of categories when needed. For instance, segmentation types can be 2 dimensional (2D), or 3 dimensional (3D). Also, the segmented area can be: Peripheral Zone (PZ), Central Gland (CG), Transition Zone (TZ), and whole prostate. Databases were also organized in categories according to their origin: public, private, or both(merge). The MRI images according to their type can be: T2-weighted (T2W), Diffusion-Weighted Imaging (DWI), Apparent Diffusion Coefficient (ADC),Coherence Enhanced Diffusion (CED) and pulse sequence images or their derivatives. The MRI protocol can be: with endorectal coil (ERC), without endorectal coil (non-ERC), both, undefined. The image modalities can be: monoparametric, multiparametric, undefined. Data extracted and organized in the data extraction forms was used to analyze information from the accepted documents and answer the research questions.

Author	QA1	QA2	QA3	QA4	Score
benalcazar2014 [12]	Y	P	Y	Y	3,5
yan2016 [21]	P	P	Y	Y	3
milletari2016 [22]	P	P	P	Y	2,5
mun2017 [23]	P	Y	P	Y	3
sun2017 [24]	P	P	P	Y	2,5
cheng2017 [25]	Y	P	Y	Y	3,5
Clark2017 [26]	Y	Y	Y	Y	4
zhu2017 [27]	Y	Y	Y	Y	4
meyer2018 [28]	Y	Y	Y	Y	4
gibson2018 [29]	P	P	Y	Y	3
jia2018 [30]	Y	P	Y	Y	3,5
hossain2018 [31]	P	P	Y	Y	3
ji2018 [32]	P	Y	Y	Y	3,5
Gelder2018 [33]	Y	P	Y	Y	3,5
Chen2018 [34]	Y	P	Y	Y	3,5
jit2018 [35]	Y	P	Y	Y	3,5
to2018 [36]	Y	Y	Y	Y	4
zhu2018 [37]	Y	Y	Y	Y	4
he2018 [38]	P	P	P	Y	2,5
brosch2018 [39]	P	P	Y	Y	3
jia2018_Atlas [40]	Y	P	Y	Y	3,5
yan2018 [10]	Y	P	Y	Y	3,5
Tian2018 [41]	Y	P	P	Y	3
zhan2019 [42]	Y	Y	Y	Y	4
Karimi2019 [43]	P	Y	Y	Y	3,5
bo2019 [44]	Y	P	Y	Y	3,5

Table 2.4: Quality assessment implementation

Data synthesis

The research questions that guide this review were answered from the extracted data by using the extraction forms (see appendix A). The data analysis was presented in the Result Section 2.2.

2.2 Result

In this section, the results of the research questions were shown.

2.2.1   RQ1. What model structure and type of segmentation based on artificial neural networks have been used to segment the prostate in MRI images?

Data augmentation was an important pre-processing step for training and validation on models based on artificial neural networks. Due to the limited number of prostate MRI images in public and private databases. Many authors have used different data augmentation techniques such as: flipping, random rotations, channel shifting, vertically

and horizontally shift, Gaussian noise addition, zooming, cropping, scale transformation, and isotropic expansions. Figure 2.3 shown the number of papers where an specific data augmentation technique is used. The most widely used data augmentation techniques were: random rotation and flipping. The flipped technique reflected an image. Random rotation technique consisted in rotating an image randomly. It should be noticed that only 6 studies did not define a data augmentation technique.
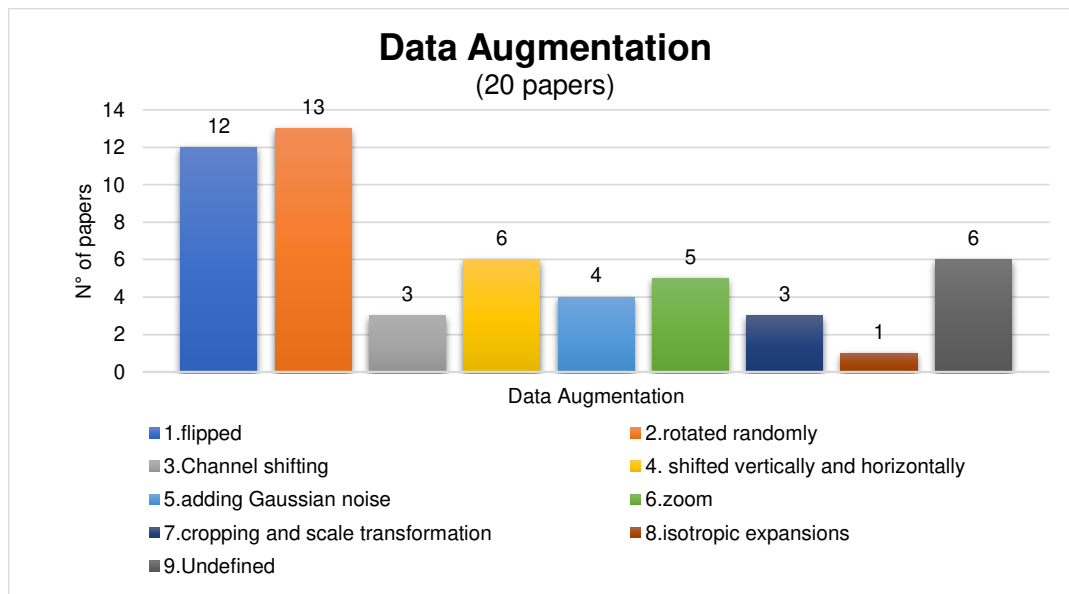

Figure 2.3: Data Augmentation

Defining a model structure was very complex. Therefore, global processes were grouped to form different model configurations. The model structure was formed: pre-processing, feature extraction, segmentation and post-processing.

Figure 2.4, literal (a) shown that 19 studies used pre-processing techniques such as: normalization, sub-volume center cropping, uniform voxel, contrast adjustment, patches extraction, resized or re-scaled, filtering, equalization, removal of slices without prostate, and/or classifier. The most common pre-processing technique among the gathered studies was the normalization.

Literal (b) in Figure 2.4 presented the feature extraction group which can be automatic or manual. An encoder was used in automatic extraction, whereas manual extraction relies on a feature vector. A total of 25 studies use automatic extraction, whereas only one study used manual extraction.

Literal (c) presented the segmentation process, which used a decoder or classifier to segment the prostate gland. A total of 16 studies used a decoder, whereas the remaining 10 used a classifier. The decoders used in the 16 articles are based on: U-net, V-net, DenseNet, VGGNet, ResNet and undefined. U-net was one of the most used standard architectures for prostate segmentation using a decoder. The 10 remaining studies that used a classifier for prostate segmentation used: VGGNet, AlexNet, and undefined. The most used standard architecture for prostate segmentation using a

classifier was not defined.

Literal (d) shown that 12 studies used post-processing techniques such as: largest connected component, thresholding, boundary refinement, small component removal, hole filling, mathematical morphology, and others. The most common techniques used in post-processing are: the largest connected component, thresholding, and boundary refinement.

(a) Pre-processing

(b) Feature Extraction

(c) Segmentation

(d) Post-processing
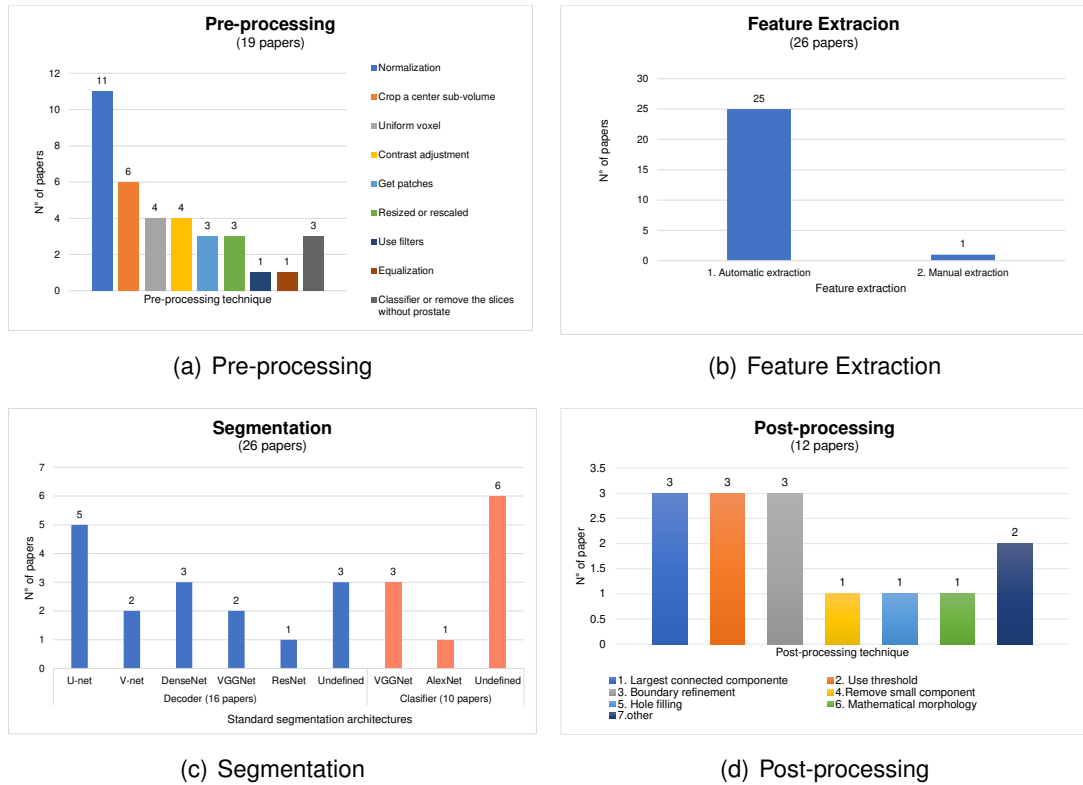
Figure 2.4: Model structure

Appendix A presents the extracted information from the 26 accepted documents. Data extracted includes: year, type and area of segmentation, parameters type, plane and acquisition of MRI images, type and name of the public database. Table 2.5, extracted from appendix A, presents the segmentation type (2D or 3D) performed by year. As indicated in the aforementioned table, 14 studies used 2D segmentation, whereas 3D segmentation was used in 12 studies. The most used segmentation type in literature corresponds to 2D segmentation.

Year	2D	3D
2014	1	0
2016	0	2
2017	4	1
2018	7	8
2019	2	1
Total	14	12

Table 2.5: Segmentation type used in prostate segmentation

2.2.2 RQ2. What standard architectures of artificial neural networks have been used to segment the prostate in MRI images?

To address the question about the standard architectures of ANN used in prostate segmentation, it was important to establish the segmented area explored in literature. Table 2.6, built from the extracted data in appendix A. In this table was compared the prostate areas segmented and the number of studies published by year. It can be concluded that most studies focus on the whole prostate gland segmentation because there are 25 studies from years 2014 to 2019, which dedicated to this type of segmentation. Whereas only 1 study focused on the central and peripheral zone in 2014. Also, The transition and peripheral zones had 1 studied, respectively. we can be concluded that most of the studies focused to segment the whole gland of prostate.

	Gland	Gland + TZ	Gland + TZ + PZ	CZ + PZ
2014	0	0	0	1
2016	2	0	0	0
2017	5	1	0	0
2018	15	0	1	0
2019	3	0	0	0
	25	1	1	1

Table 2.6: Parts of the prostate segmented in literature

Figure 2.5 presented the different standar architectures used in prostate segmentation. A total of 9 (35%) studies did not specify the standard model. The remaining 17 (65%) studies used DenseNet, Vgg-Net, ResNet and AlexNet models. Even though, the aforementioned models was generally used in classification, they can also be used in segmentation tasks. In this case, a pixel-by-pixel classification was performed to produce a prediction map (i.e. input image (patches) and the class prediction). Semantic segmentation through the U-net (19%) and V-net (8%) models were common in literature. Furthermore, U-net was the most used of its kind in the reviewed accepted papers. In this case, semantic segmentation performed a pixel-wise prediction where the output label map has the same size as the input image.

2.2.3 RQ3. What metrics and results have been obtained using the proposed models to segment the prostate in MRI images?

The extraction stage of the literature review found the following metrics used in the 26 accepted documents: Dice Similarity Coefficient (DSC), Hausdorff Distance) (HD), Average Boundary Distance (ABD), Relative Absolute Volume Difference (RAVD), 95%- Hausdorff- Distance (95HD), Intersection over Union (IoU), Average minimum surface-to-surface distance (AVGDIST), Relative Volume Difference (RVD), symmetric Boundary Distance (BD), MSD, CentroidDistance (CD) and precision-recall. Figure 2.6 shows that the metrics most used in literature are: DSC (25 studies), HD (10 studies), and ABD (7studies). The remaining metrics are less 6 studies. As depicted in Figure 2.6, DSC metric was the most widely used metrics in literature for prostate
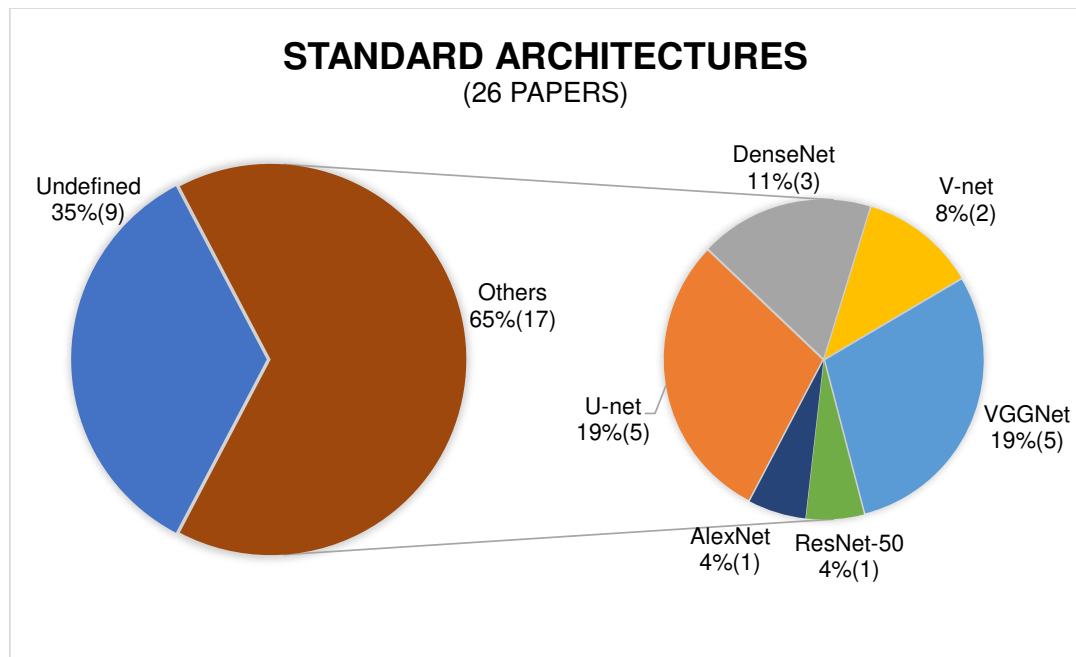
Figure 2.5: Standard architectures used in prostate segmentation

segmentation. Figure 2.7 presented the results achieved in literature by the authors that used the DSC metric to evaluate segmentation performance in public and private databases. Because, these studies used different databases and standard structures, it was difficult to establish a general comparison between them. However, the highest and lowest DSC achieved in literature corresponds to 94.57% and 84%, respectively. Only one study (displayed in green in Figure 2.7) used DSC on a merged database of public and private databases. This study achieved a DSC of 85%.

2.2.4   RQ4. Which databases have been used for developing and testing models to segment the prostate in MRI images?

A very important feature of the prostate databases was the anatomical plane of the MRI images. These planes can be axial, coronal and sagittal, as shown in Figure 2.8 where it can be observed that the most common plane used corresponds to the axial plane. However, only one study used the three type of planes.

The MRI images acquired by using endorectal coil (ERC) present high resolution [45]. The type of MRI images acquisition for each of the 26 studies is detailed in Table 2.7. The results of MRI images acquisition system show that 14 studies used MRI images with ERC; 2 studies used MRI images without ERC; 6 studies used MRI images with and without ERC; and 4 studies did not define the use of ERC. Therefore, it can be concluded that the MRI images acquisition without endorectal coil is the most widely used in literature. The use of MRI images without ERC leads to a greater challenge because there is no reference point. Also, the MRI images without ERC have a lower resolution compared to MRI images with ERC.
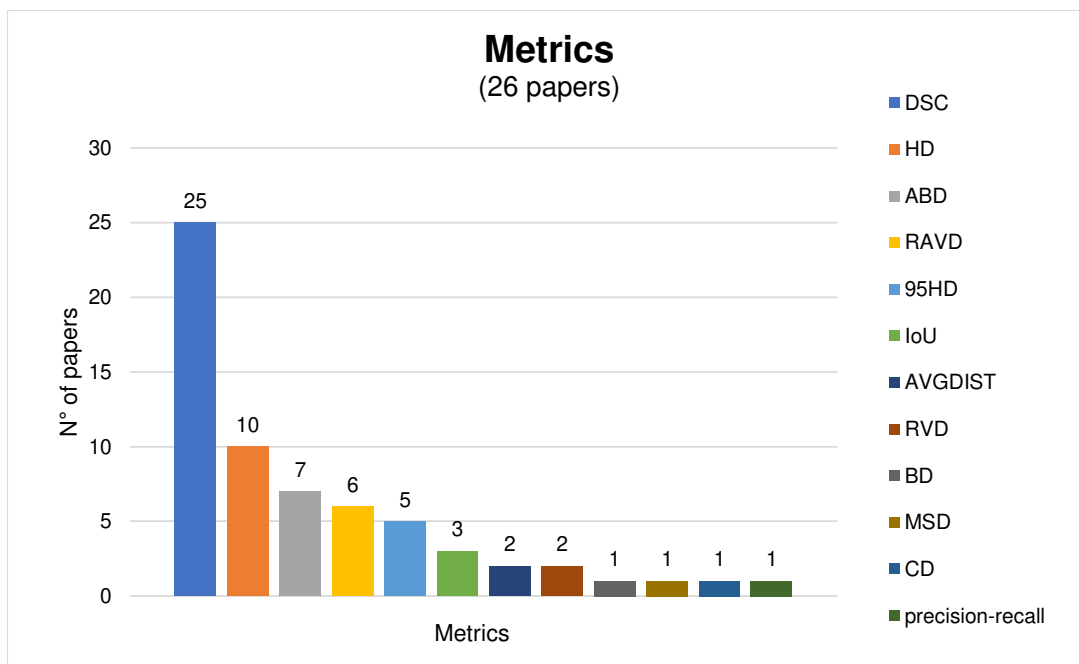
Figure 2.6: Metrics used in prostate segmentation
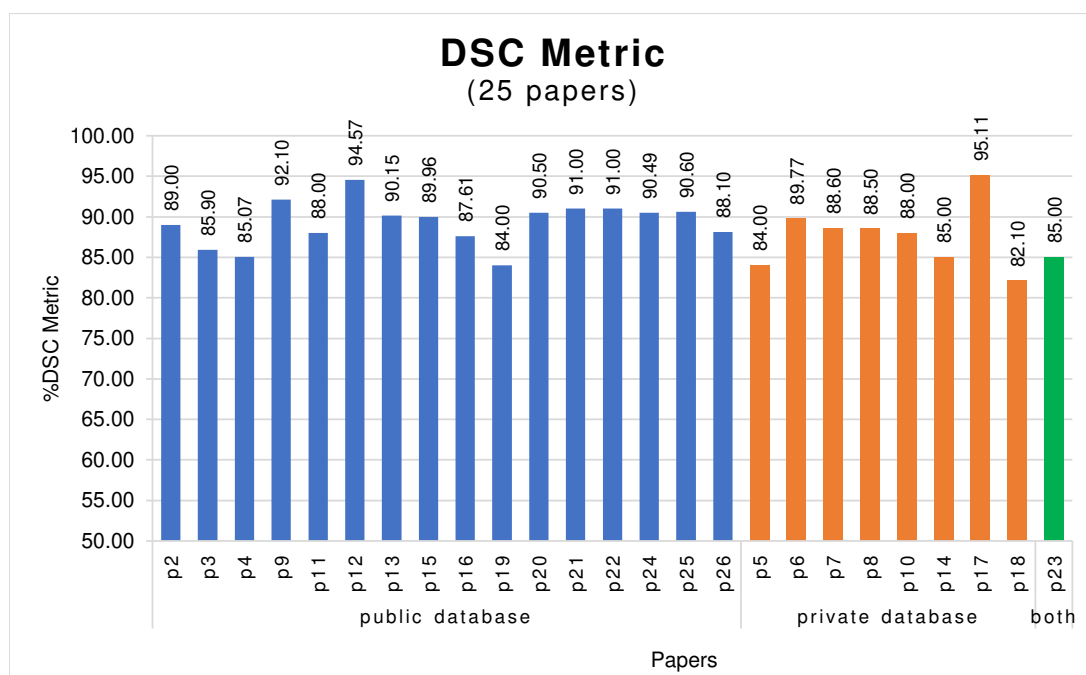
Figure 2.7: The DSC metric used in prostate segmentation

The prostate MRI images techniques and protocols vary across centers and clinical indication, but there are individual images sequences commonly used [46] like: T2W, DWI, ADC, etc. The use of one or more MRI images sequences in database is referred with the number of parameters. If the database has only one MRI images sequence is denominated as monoparameteric. The databases that have more than 2 MRI images sequences are called multiparametric. Figure 2.9 shown the parameter number and
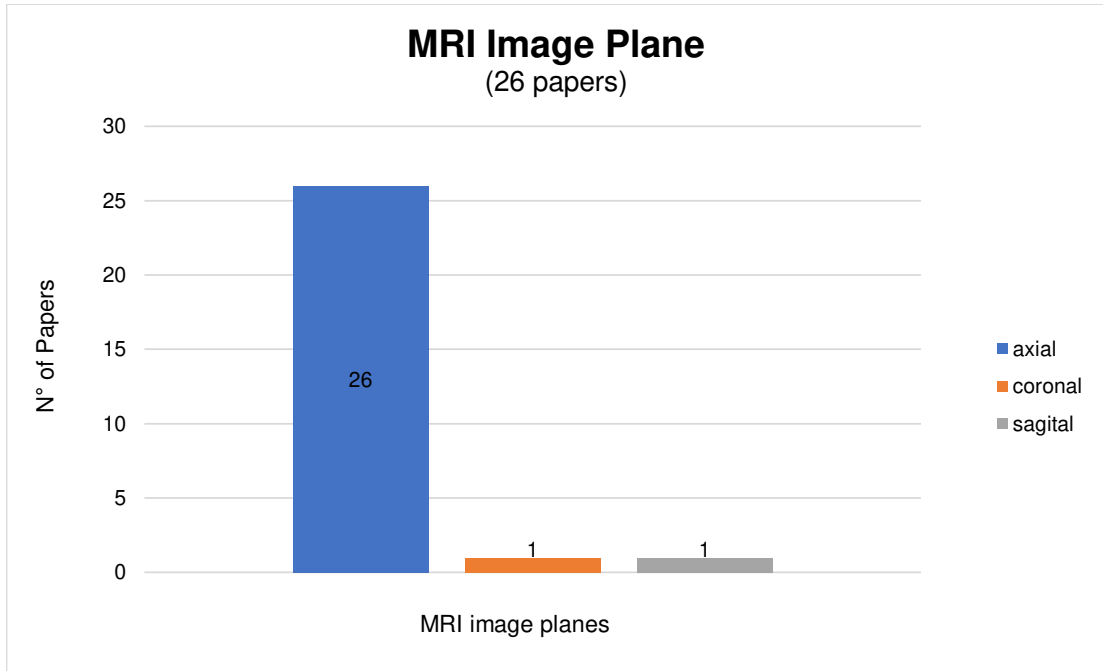
Figure 2.8: MRI images plane used in prostate segmentation

Endorectal coil	Num paper
with ERC	14
non-ERC	2
both	6
undefined	4
Total	26

Table 2.7: MRI images acquisition using endorectal coil

MRI images sequences used in prostate segmentation. As a result, 22 studies used one parameter; 3 studies used multiparametric images; and one study did not define the parameter number. Also, The studies of one parameter use MRI images sequences T2W. whereas multiparametric studies use MRI images sequences T2W, DWI, ADC, CED and DERIVATE.

Regarding the database, this SLR shown that public and private databases have been used in prostate segmentation. One benefit of the public databases is the free access to download the information by the research community and consequently, they permit a comparison of different research proposals. However, a disadvantage of this type of database is the number of samples, which is limited. On the other hand, private databases usually have a mean of 140 MRI volume, but they cannot be freely accessed by the research community which makes difficult to replicate and compare results.

Using the information extracted from the 26 accepted papers and presented in Appendix A. Table 2.8 was built to present the type of database used and the number of papers. It shows that 17 studies use public databases, wheras 8 use private data. Only 1 study used both type of databases. Consequently, we can be said that public databases were
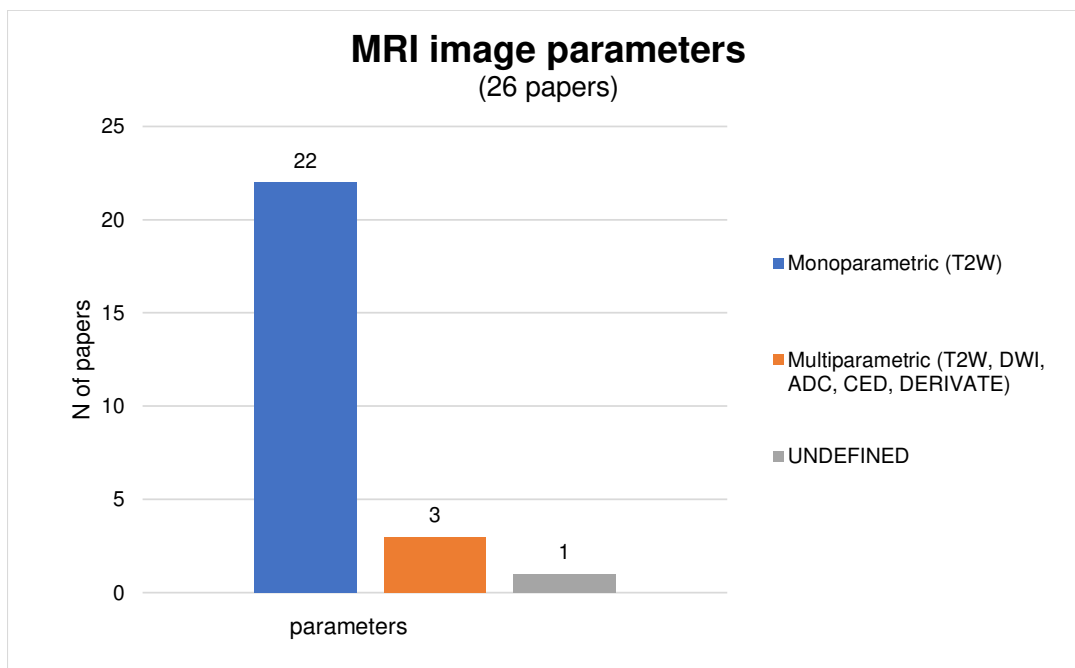
Figure 2.9: MRI images parameter and sequences used in prostate segmentation

the most used in literature for prostate segmentation studies.

Database	Number of paper
private	8
public	17
both	1
Total	26

Table 2.8: Databases used in prostate segmentation

The following are the public databases used in the 17 studies identified in literature: NCI-ISBI 2013, SPIE-AAPM-NCI PROSTATEx and PROMISE12. The NCI-ISBI 2013 [47] database was designed mainly to the development of automatic segmentation methods of the Central Gland and Peripheral Zone of the prostate gland. The SPIE-AAPM-NCI PROSTATEx [48] database was a retrospective set of prostate MRI studies. All studies include T2W, proton density-weighted (PDW), dynamic contrast enhanced (DCE), and DWI. The PROMISE12 [49] database contains multi-center and multi-vendor MRI images with different acquisition protocols. For each of the MRI images, a reference for the prostate segmentation is included. Figure 2.2.4 shows that of the 17 studies: 2 studies use the NCI-ISBI 2013 database, 1 study uses the SPIE-AAPM-NCI PROSTATEx database, and 14 studies use the PROMISE12 database. The most used database in whole prostate gland is PROMISE12.

Public Databases
(17 papers)
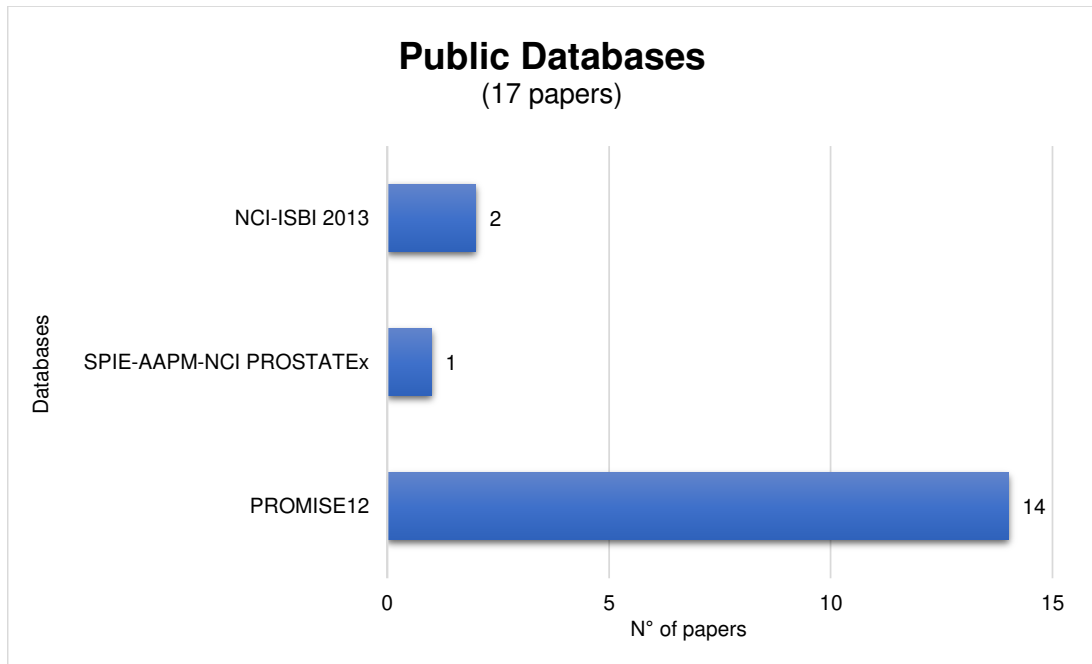
2.3 Discussion

This section presented an analysis of the literature review findings with respect to the research question. Thus, this analysis contained key points about the type of segmentation (RQ1), the standard model (RQ2), the performance metrics (RQ3) and the public databases (RQ4). These points are detailed below:

- Segmentation type: 2D, and 3D segmentation types were found in literature. The segmentation type was very important because it allowed to know how the prostate segmentation problem is being addressed by the researcher.

- RQ1 Segmented area: The segmentation area describes the zone of the prostate that a particular study has focused on for segmentation. According to this literature review, the most of the studies segment the whole prostate gland.

- RQ2 Standard architectures: These were the different models based on artificial neural networks that was used in prostate segmentation. This literature shows that the U-Net decoder model was the most used in literature. Also, convolutional neural networks used in classification are used in segmentation by using a patches.

- RQ3 Metric and result: There were different types of metrics used in prostate segmentation. One of the most used metrics was the DSC, which was implemented in 25 studies. For the purpose of comparison between the different studies, this metric has been considered.

- RQ4 Public database: PROMISE12 was the most used public database as the results of this review show. Databases were important to train and validate the

segmentation model. For ease of comparison, public databases are preferable because of their open access.

Figure 2.10 presented a comparison of the performance of different standard architectures on the PROMISE database with respect to different segmentation types by using the DSC metric. Eight studies considered 2D segmentation of whole prostate used U-Net, VGG16, VGG19 and undefined standard architecture. The best result is achieved by VGG19 with an average DSC coefficient of 94.57%. However, it must be noticed that the authors did not provide the standard deviation of the DSC result achieved in 2D segmentation by VGG19. Regarding 3D segmentation, 6 studies addressed this type of segmentation by using the models: V-net, AlexNet, and undefined standard architecture. The best result in these studies achieved a DSC coefficient of 91% and corresponds to an undefined standard architecture.

The same analysis was carried out for the public database SPIE-AAPM-NCI PROSTA-TEx, where one study was found. This study used 3D segmentation and U-net as the standard architecture. It achieved a DSC coefficient of 92.1%. For the NCI-ISBI 2013 database, a study was found to focus on 3D segmentation and it used ResNet-50 as the standard model. The performance achieved corresponds to 88% on the DSC coefficient. It should be noted that there was another study that uses this database, but the study did not use the DSC metric and because of that, it was excluded.
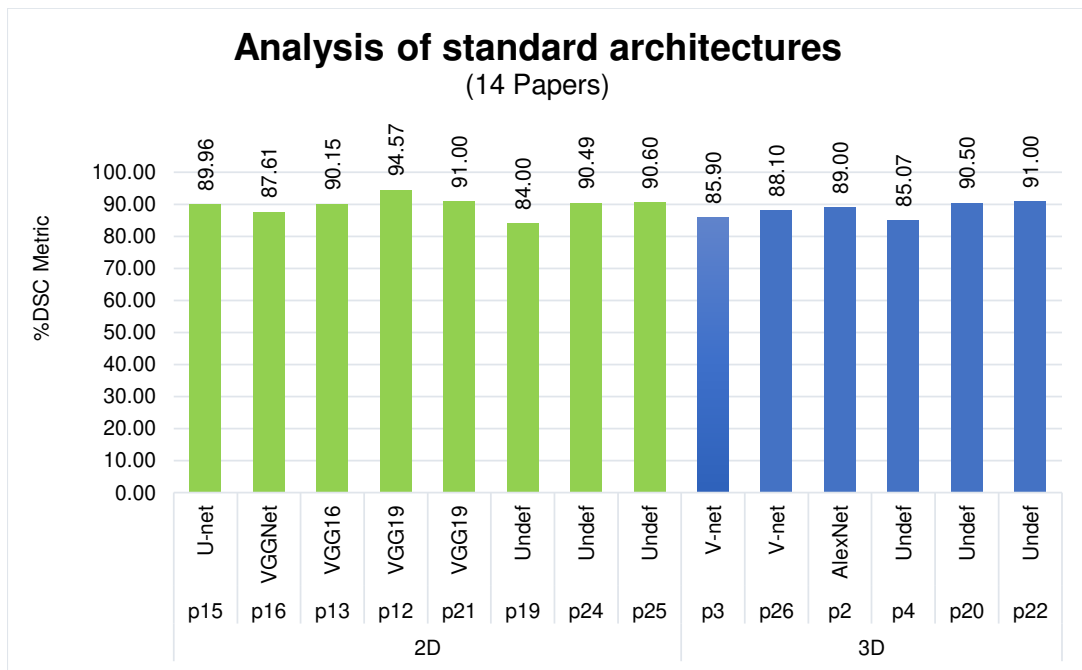

Figure 2.10: Analysis of standard architectures in PROMISE12

2.4 Conclusions

- Due to the limited number of prostate MRI images in the databases, many studies used data augmentation techniques to increase the number of images. The most common techniques found in literature were: flipping and random rotation.

- The model structure was comprised of: pre-processing, feature extraction, segmentation and post-processing. The most commonly used techniques in pre-processing was normalization and center cropping. Although other techniques exist. Most of the works in literature used automatic feature extraction through encoder. Only one study used manual feature extraction. The segmentation process can use a decoder or a classifier based on standard architectures. According to this literature review, a U-net was the most commonly used standard architecture for prostate segmentation. Finally, according to this literature review, the most common techniques for post-processing were the largest connected component, thresholding, and refinement of limits.

- 2D segmentation in MRI images is the leading segmentation technique.

- Table 2.6 shows that most of the studies focus on the task of whole prostate segmentation. This review found only one study focused on the segmentation of the prostate gland in its two main areas (CG and the PZ). Therefore, it is an open field for research to implement segmentation methods and techniques for the main areas of the prostate.

- The standard architectures used for prostate segmentation in MRI images were: U-net, DenseNet, V-net, VGG-net, ResNet, AlexNet and undefined. These standard architectures are based on convolutional neural networks.

- Twelve Different metrics were found in prostate segmentation. However, DSC is the most common metric in literature.

- The type of acquisition of the MRI image was important because it defined the quality of the image. For instance, MRI with ERC are high quality images of the surrounding area and with a reference point. Moreover, Table 2.7 shows that only 2 studies use MRIs without endorectal coil. Because of that, it can be said that it is a challenge to work with images without ERC for prostate segmentation and its main areas.

- Most of the studies used public databases for prostate segmentation. Among the databases that have been found to be used in literature are: NCI-ISBI 2013, SPIE-AAPM-NCI PROSTATEx and PROMISE12. Only NCI-ISBI 2013 had a segmentation reference of the CG and PZ.

- Models based on U-net, denseNet, V-net, VGGnet, and Alexnet have not been used to segment the prostate gland and its main areas on the NCI-ISBI 2013.

- As a general conclusion, it can be affirmed that prostate segmentation remains a current challenging problem for computer vision and machine learning because of the following problems: prostate variation, interscan variability, intra-scan variations, the prostate boundary, image noise, the size of the prostate area (see appendix B) and the limited number of images.

Chapter 3

Methodology and materials:

In this section, the materials and methodology are described which were used to develop this work.

3.1 Materials

The database and evaluation metric were used as materials. The database was used to train, validate and test the models. Regarding the metric in Chapter 2 determined that DSC metric is the most used in SLR. Therefore, DSC metric is used to evaluate the models. Also, other metrics are specific for the evaluation.

3.1.1 Database

The National Cancer Institute (NCI) in collaboration with the International Society for Biomedical Imaging (ISBI) "NCI-ISBI 2013" database was used in this work for prostate segmentation in its main parts. This database was chosen because its access is public. The database contains 60-case prostate MRI 3D series with markups that define the Central Gland and Peripheral Zone. The 60-case were acquired as T2-weighted MRI axial pulse sequences. Half of the cases were obtained at 1.5T (Philips Achieves) with an endorectal receiver coil and the other half at 3T (Siemens TIM) with a surface coil [47]. Despite 1.5T MRI images with endorectal receiver coil provide a reference point for prostate segmentation, 3T MRI images with a surface coil are more commonly used. Therefore, in this work is used 3T MRI images collection. 3T-collection consists of 30 transverse volumes accompanied by their respective manual segmentation's or gold standard volumes. The number of slices per volume of this database varies between 15 and 24. Each slice contains images of 320×320 pixels.

3.1.2 Metric

The metrics used to evaluate the models are DSC, IoU, precision, recall, and accuracy. The DSC metric is the most used in the SLR. However, DSC metric is a variant of the IoU metric. Also, the DSC metric gives twice the weight to the intersection area.

Therefore, DSC and IoU metrics are used. Finally, the metrics precision, recall, and accuracy were used to compare with the proposed model in the literature. Each of the metrics is detailed below:

The dice similarity coefficient measures the spatial overlap between the ground truth of the prostate and the final segmentation. where $|X|$ and $|Y|$ in (3.1) denote the region of the predicted segmentation and reference segmentation respectively.

$$DSC = \frac{2|X \cap Y|}{|X \cup Y|} = \frac{2TP}{2TP + FP + FN} \tag{3.1}$$

The Jaccard coefficient, Jaccard similarity coefficient or Intersection over Union is defined as the intersection over union of two sets. In equation (3.2), X and Y are two different regions, i.e. sets of pixels of the predicted segmentation and reference segmentation respectively. The modulus sign '$|$ $|$' defines the cardinal of the corresponding sets.

$$IoU = \frac{|X \cap Y|}{|X \cup Y|} = \frac{TP}{TP + FP + FN} \tag{3.2}$$

Other metrics were specific, such as precision, recall and accuracy (see equation 3.3). These metrics are obtained from the confusion matrix. These metrics are used in classification tasks.

$$
\begin{aligned}
Precision &= \frac{TP}{TP + FP} \\
Recall &= \frac{TP}{TP + FN} \\
Accuracy &= \frac{TP + TN}{TP + FP + FN + TN}
\end{aligned}
\tag{3.3}
$$

3.2 Methodology

The development of research proposed was guided mainly by the phases that are carried out in machine learning.

3.2.1 Design, training, validation and testing of models

The necessary steps are defined to create the architecture of models A and B. Each step has a procedure that is specified below:

4.2.1.1 Model-A: Encoder-Decoder Architecture

The proposed model-A was based on U-net [50]. U-net is an encoder-decoder architecture. The encoder capture semantic/contextual information, whereas the decoder recover spatial information. The model-A implements two U-net networks: global U-Net and local U-Net. The input of the global and local U-net is a set of 2D MRI image. The global U-net was used to segment the whole prostate gland. On the other hand, the local U-net was used to segments the central zone.

Global U-Net architecture

The steps defined to create the global U-Net architecture were: a) pre-processing, b)Data augmentation, c) feature extraction and segmentation, and d) post-processing.

(a) Pre-processing was used to extract region of interest (ROI) images of three different sizes (151×151, $M \times N$ and mask's size). The techniques used in pre-processing were: normalization, sub-volume central crop, histogram equalization and non prostate slice removal. The operations were carried out in a Matlab program. Figure 3.1 presents the ROI extraction process result. The different ROI process is detailed below:

- 151×151 *MRI Images:* first, MRI images volumes (OBSERVED) and masks (IDEAL) of 320×320 are read. Second, the X and Y coordinates of the ROI were found by a pattern search approach based on the fact that the prostate gland is located between the pelvis bones (see Table 3.1). Third, a resulting ROI of 151×151 was obtained by adding 75 pixels to the right, left, up and down directions from the X, Y coordinates of ROI (see Table 3.1 and appendix C). Next, slices without prostate were removed from the set. After, MRI images were normalized and saved in matrix format for later use. Finally, all this process was repeated on each volume of the database. As, final result ROI of 151×151 is presented in Sub-figure 3.1(a))

Num.	Process
1.1	The data type of MRI images was changed from uint16 to uint8 (see appendix C Sub-figure (a)).
1.2	The intensity of the values were adjusted according to the image's limits.
1.3	Histogram equalization was used to enhance contrast of the MRI image.
1.4	In appendix C Sub-figure (b), an opening morphological operation with a ball structuring element of radius 5 and maximum height offset of 5 was applied followed by the operations of erosion and close. The latter operations used the same structuring element but with a radius of 4 and a maximum height offset of 3. Finally, dilation and opening morphological operations were applied using the same structuring element with a radius of 3 and maximum offset height of 3.
1.5	The minimum pixel values of the MRI image was set to zero.
1.6	The regions and holes of MRI image were filled with an 8 connectivity (see appendix C Sub-figure (c))
1.7	The MRI image was increased in size which added 80 pixels to the left and right of it.
1.8	The MRI image was binarized with a threshold of 0.2.
1.9	Hough's Circular Transformation algorithm was used to locate the pelvic bones in the MRI image, because they have a circular shape. The radius r was $27 \leq r \leq 30$, as indicated in appendix C Sub-figure (d). Next, the locations of the left and right pelvic bones were averaged and the midpoint between them was used to find the X, Y coordinates of ROI.

Table 3.1: Getting the X and Y coordinates of the ROI

- $M \times N$ *MRI Images:* From, the ROI of 151×151 were generated the ROI of $M \times N$ size, where $M \leq 151$ and $N \leq 151$, by searching the intensity of the prostate's pixels. The steps followed to obtain $M \times N$ ROIs are described in Table 3.2 and appendix D. Finally, the MRI images were normalized and slices without the prostate gland were removed. Slices containing the gland were saved in Matrix format for later use. This procedure is repeated on all the volumes of the database. As, final result ROI of $M \times N$ is presented in Sub-figure 3.1(b))

Num.	Process
2.1	Slices without prostate were removed.
2.2	A volumetric central area of 30x30 was obtained and the pixel values were averaged.
2.3	The region's maximum pixel value was set to 0.
2.4	The volume of MRI images were normalized to a range between 0 and 1.
2.5	Intensity values were adjusted in volumetric MRI images. Notice that the previous sub-steps (2.1 to 2.5) used volume of MRI images, whereas the sub-steps (2.6 to 2.16) were processed slice to slice in volumetric MRI images, as depicted in appendix D Sub-figure (a).
2.6	Pixel values lower than 0.2 are set to 0, whereas pixel values higher than 0.8 are set to 1 on each MRI image slice.
2.7	Erosion morphological operation was used with a disk structuring element of radius 2.
2.8	Connectivity 8 was used to fill holes and image regions.
2.9	Opening morphological operation with a disk structuring element of radius 15 was applied.
2.10	Pixel values lower than 0.21 and higher than 0.31 were set to 0 and 1, respectively on the MRI image slice.
2.11	Image regions and holes were filled with an eight connectivity.
2.12	Erosion and opening morphological operations were used with an structuring element of type disk and radius 3 and 2, respectively.
2.13	The intensity values were adjusted.
2.14	opening morphological operation with a 22 radio disk structuring element was used, as indicated in appendix D Sub-figure (b).
2.15	The MRI image slice was verified to have non-zero pixel values. The centered ROI was highlighted to calculate the binary mask. Pixel values higher than 0 are set to 1 (see appendix D Sub-figure (c)).
2.16	All calculated masks were added one by one. The sum of the calculated mask was normalized in ranges of [0-1] and binarized with a threshold of 0.3; which adjusted the mask to the region of interest.
2.17	Opening morphological operation with a radio 30 disk structuring element was performed over the adjusted mask.
2.18	Coordinates X1, Y1, X2, Y2 were obtained to cut the adjusted ROI adjusted out of the mask, as indicated in appendix D Sub-figure (d).
2.19	Using the X1, Y1, X2 and Y2 coordinates the $M \times N$ ROI is extracted out of the 151×151 image, as indicated in appendix D Sub-figure (e).

Table 3.2: Getting the $M \times N$ ROI

- *Adjusted to the Mask MRI images:* This procedure obtained a ROI image ad-

justed to the dataset's binary mask. First, the 320×320 MRI images volumes (OBSERVED) and masks (IDEAL) were read. Second, mask coordinates containing the whole prostate were extracted (see Table 3.3 and appendix E). Then, normalization of the resulting MRI images and removal of slices without containing the prostate gland was performed. Finally, the MRI images containing were saved in matrix format and the procedure repeated among all the cases in the database. As, final result ROI of Adjusted to the Mask is presented in Sub-figure 3.1(c))

Num.	Process
3.1	320×320 mask volumes (IDEAL) were read and normalized to a range between 0 and 1 (Figure E.1(b).
3.2	All masks were added one by one (Figure E.1(c).
3.3	Pixels values greater than zero were fixed to one. As a result, the ROI adjusted mask was obtained.
3.4	Coordinates X1, Y1, X2, Y2 adjusted mask ROI were obtained (see Figure E.1(d).
3.5	Using the X1, Y1, X2, Y2 coordinates, a ROI sub-volume was cut out of the original MRI (see Figure E.1(e).

Table 3.3: Getting ROI Adjusted to the Mask

(a) 151 x 151 MRI Images

(b) $M \times N$ MRI Images

(c) MRI images adjusted to the mask
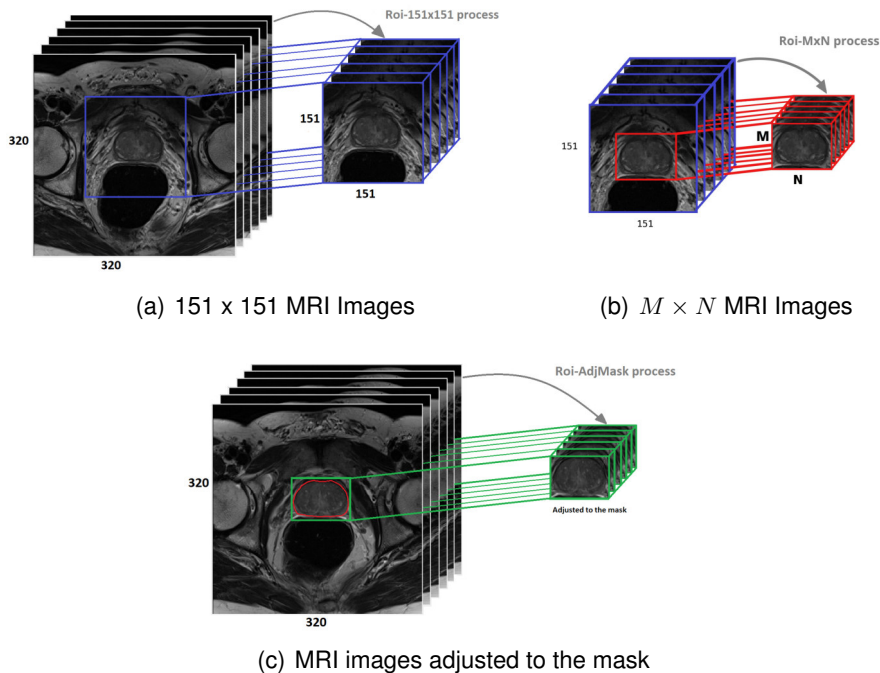
Figure 3.1: Different sizes ROI extraction: a) 151x 151 ROI patch is obtained from the 320×320 MRI images volume; b) M x N ROI patch, where $M \leq 151$, $N \leq 151$ and not necessarily $M = N$ c) ROI patch is obtained using the database mask.

(b) Data augmentation: The original dataset contains 578 slices of MRI images. However, only 421 MRI image slices contain the prostate gland. In order to increase the number of samples of MRI images slices, data augmentation techniques were used.

Among these techniques, the following were used: flipping, $\pm$ 45 degree random rotation, zoom and zoom with random rotation. The data augmentation increment for the different ROI is detailed below:

- 151×151 *MRI Images:* Through data augmentation, the 151 X 151 MRI slice images were increased to 2019. Without data augmentation, the database contains an average of 84% pixels in the background region and 16% pixels in the prostate. The augmented database contains an average of 83% background pixels and 17% of prostate pixels.

- $M \times N$ *MRI Images:* Through data augmentation, the 421 $M \times N$ MRI slice images were increased to a total of 2105. The original set contains an average of 80% background pixels and 20% pixel prostate. The augmented database contains 75% background pixels and 25% prostate pixels.

- *MRI images adjusted to the mask:* The original database contains an average of 55 % background pixels and 45 % prostate pixels. Using data augmentation, the database was increased from 421 samples to 2064 MRI slice images. The augmented database contains 63 % background pixels and 37 % prostate pixels.

(c) Feature extraction and Segmentation: The global U-net architecture was represented in Figure 3.2. This architecture was formed by a contracting path followed by an expansive path. The contracting path, which aims to extract features for classifying each voxel, consists in the repeated application of 3x3 convolutions (padding='same'), followed by dropout layer, a 3x3 convolutions (padding='same'), and a 2x2 max pooling layer for downsampling. The number of feature channels are doubled at each downsampling. On the other hand, the expansive path, which aims to locate regions of interest (ROI) more precisely, consists of a 2x2 transposed convolution layer (i.e. deconvolution) with stride of 2 that halves the number of feature channels, concatenated with the corresponding cropped feature map from the contracting path, followed by one 3x3 convolutions (padding='same'), a dropout layer, and a 3x3 convolutions (padding='same'). The final layer is formed by a 1x1 convolution that is used to map each feature vector to the desired number of classes. ELU activation function was used for every convolution block in U-net.

d) Post-processing: this stage was implemented to improve the detection carried out by the implemented models. First, the predicted image is binarized using a 0.5 threshold. Second, an opening morphology operation with a radius 5 disk structuring element is applied. Finally, small areas are removed and the largest area corresponds to the desired segmentation mask.

Experimental selection in Global U-Net architecture

A total of 6 experiments were carried out for the global U-net architecture. These experiments mainly modified the size of the ROI image and maintained the training parameters and the post-processing. ROI were proved with and without data augmentation. The extension of the files considered were PNG and MAT, of which the MAT
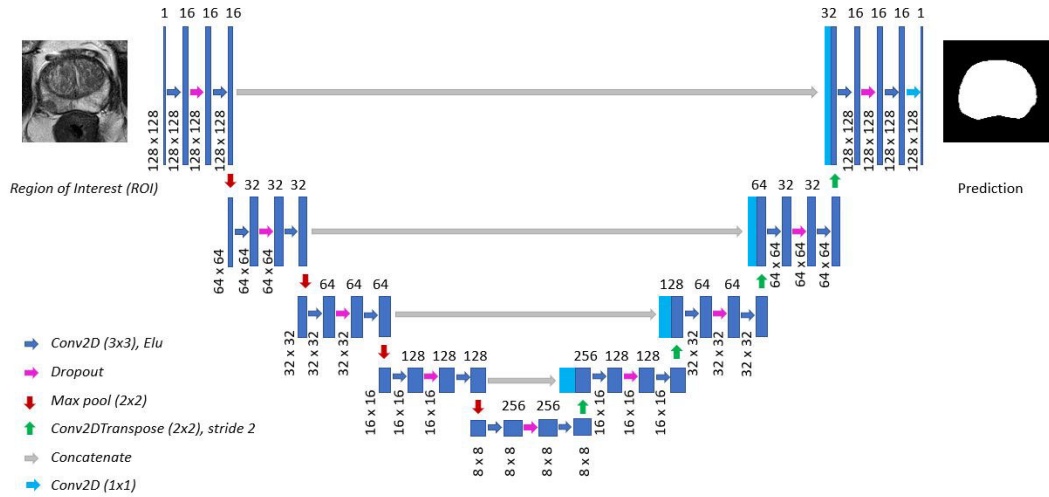
Figure 3.2: Global U-Net architecture

extension was chosen to obtain higher results. The MAT extension saved the data in the range of 0-1 and without re-scaling. The details of each experiment are presented below and in Table 3.4.

- *Experiment 1:* this experiment evaluated the performance of the proposed model when: (a) the ROI had a size of 151×151 for training, evaluation and testing and (b) no data augmentation was used for training, evaluation nor testing.

- *Experiment 2:* this experiment evaluated the performance of the proposed model when: (a) the ROI had a size of $M \times N$, for training, evaluation and testing, and (b) no data augmentation was used for training, evaluation nor testing.

- *Experiment 3:* this experiment evaluated the performance of the proposed model when: (a) for training and evaluation adjusted mask size ROI were used, whereas for testing, a 151×151 ROI was used, and (b) no data augmentation was used for training, evaluation nor testing.

- *Experiment 4:* In this experiment evaluated the performance of the proposed model when: (a) ROI had a size of 151×151, for training, evaluation and testing , and (b) data augmentation was used for training and evaluation. For testing, data augmentation was not used.

- *Experiment 5:* this experiment evaluated the performance of the proposed model when: (a) the ROI of interest had a size of $M \times N$ for training, evaluation and testing, and (b) data augmentation was used for training and evaluation but not for model testing.

- *Experiment 6:* this experiment evaluated the performance of the proposed model when: (a) the ROI used mask adjusted sizes for training and evaluation; While a size of 151×151 was used for testing, and (b) data augmentation was used for training and evaluation, but not for testing.

39

Experiment	ROI	Data Augmentation
1	151×151	No
2	$M \times N$	No
3	Adjust to mask	No
4	151×151	Yes
5	$M \times N$	Yes
6	Adjust to mask	Yes

Table 3.4: Experiments carried out on the global U-Net.

All the experiments were evaluated with the DSC and IoU metrics. The segmentation task evaluated corresponded to segment the whole prostate gland from the background. Table 3.5 presents the training results. Experiments 4, 5 and 6 achieved a higher result compared to experiments 1, 2 and 3 on DSC and IoU. This is due to data augmentation.

Experiment	DSC [%]	IoU [%]
1	91.60 ± 2.42	84.96 ± 3.78
2	92.58 ± 1.20	86.29 ± 2.06
3	93.98 ± 0.94	88.69 ± 1.66
4	96.32 ± 0.39	92.92 ± 0.72
5	96.99 ± 0.22	94.16 ± 0.42
6	95.37 ± 0.43	91.19 ± 0.78

Table 3.5: Training result of the global U-Net.

Table 3.6 presents the results of model validation. Experiments where data augmentation was used present a superior result compared to experiments that did not use data augmentation. Moreover, the result of experiment 6 was the highest result with $93.11\% \pm 0.20\%$ in DSC metric. This result is due to the size of the image input, which has a ROI adjusted to themask; Also, ROI adjusted to the mask decreased background pixels and focuses on the prostate pixels.

Experiment number	DSC [%]	IoU [%]
1	84.46 ± 2.24	73.50 ± 2.92
2	86.09 ± 0.92	75.67 ± 1.38
3	91.24 ± 0.70	83.92 ± 1.18
4	90.42 ± 0.33	82.54 ± 0.54
5	90.61 ± 0.33	82.84 ± 0.55
6	93.11 ± 0.20	87.12 ± 0.34

Table 3.6: Validation result of the global U-Net.

Similar to training and validation, testing results show higher results for experiments with data augmentation than for experiments without data augmentation. Result of experiments was shown in Table 3.7. Also, in DSC metric the experiments 4 and 5 achieved $91.71\% \pm 0.48\%$ and $91.83\% \pm 0.23\%$, respectively.

Comparing the result of experiments 4 and 5 in the validation and testing, we noticed that the results vary by less than 2% in DSC metric. On the other hand, experiments 3

and 6 show difference because training and validation use images adjusted to the size of the mask, whereas 151×151 input size was used for testing; Therefore, the result of experiments 3 and 6 decrease in testing.

Experiment number	DSC [%]	IoU [%]
1	84.94 ± 2.38	74.2 ± 3.15
2	87.15 ± 0.83	77.31 ± 1.31
3	69.22 ± 1.87	53.05 ± 2.21
4	91.71 ± 0.48	84.71 ± 0.81
5	91.83 ± 0.23	84.92 ± 0.38
6	80.52 ± 1.58	67.57 ± 2.17

Table 3.7: Testing result of the global U-Net.

Figure 3.3 presents the results achieved by all the experiments. The metrics considered to compare the results were: DSC and IoU. Also, the standard deviation of each of the results is shown. For DSC metric, the standard deviation is $\leq 2.38\%$, whereas IOU metric is $\leq 3.15\%$. The best results correspond to experiments 4 and 5 with $91.71\% \pm 0.48\%$ and $91.83\% \pm 0.23\%$ in DSC metric, whereas in IoU metric reached $84.71\% \pm 0.81\%$ and $84.92\% \pm 0.38\%$, respectively. Noticed that experiment 4 used a ROI of 151×151, whereas experiment 5 used a ROI of $M \times N$.
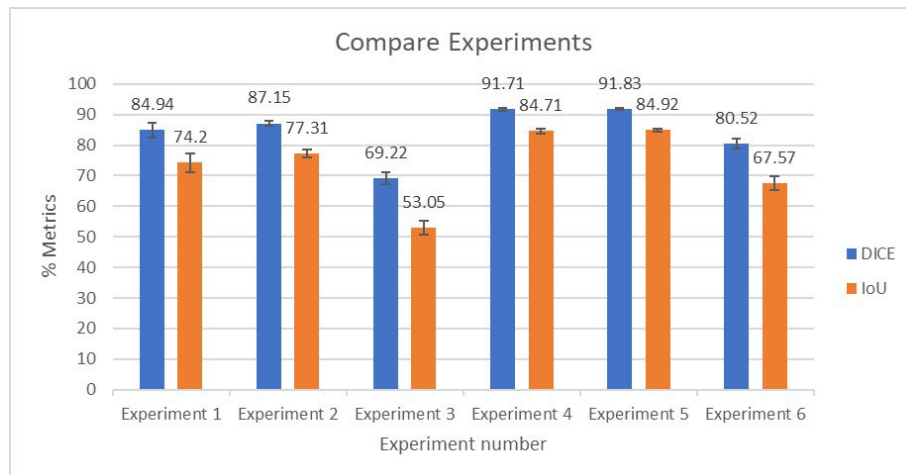

Figure 3.3: Result of the experiments of the global U-Net

Local U-Net architecture

The local U-net architecture predicted the background and Central Gland classes. The steps defined to create the local U-Net architecture were: a) pre-processing, b) data augmentation, c) feature extraction and segmentation, and d) post-processing. The pre-processing stage consists to select of the ROI with the best results achieved in global U-net experimentation. The parameters and architecture used in local U-Net are defined in the feature extraction and segmentation stage. Finally, post-processing suppresses distortions of the resulting image and enhances the predictions.

(a) Pre-processing: The ROI size was chosen according to the best results achieved in Global U-net model testing, detailed in the previous section. Experiments 4 and 5 achieved the best results for the global U-net model, however their results have an overlap. Therefore the highest and lowest standard deviation were considered to discriminate between them. Consequently experiment 5 was selected which has $M \times N$ ROI. Table 3.8 and Figure 3.4(a) were presented ROI pre-processing for local U-net. This process were implemented in Matlab software.

Num.	Process
4.1	The $M \times N$ ROI were selected, and then the Peripheral Zone pixels were eliminated in the volume of masks (IDEAL). As result, the Central Gland pixels in the volume of masks (IDEAL) were obtained (see Sub- figure 3.4(a))
4.2	There are 2 pre-processing alternatives according to the dependence of the 2 U-net networks (see sub-steps 4.2.1 and 4.2.2).
4.2.1	The $M \times N$ input is considered independent when it does not consider the prediction given by the global U-net model. This is represented in Sub-figure 3.4(b).
4.2.2	In this case, the global U-net trained with $M \times N$ images was used to predict a prostate binary mask. The predicted mask was multiplied with the MRI image. The result was an overlay between the predicted mask and the MRI image. Figure 3.4(c) presents this process. Finally, the procedure was repeated in all the samples of the database.
4.4	The MRI images were normalized and slices without the prostate removed. Slices containing the prostate gland were saved in MAT extension.

Table 3.8: ROI pre-processing for local U-net
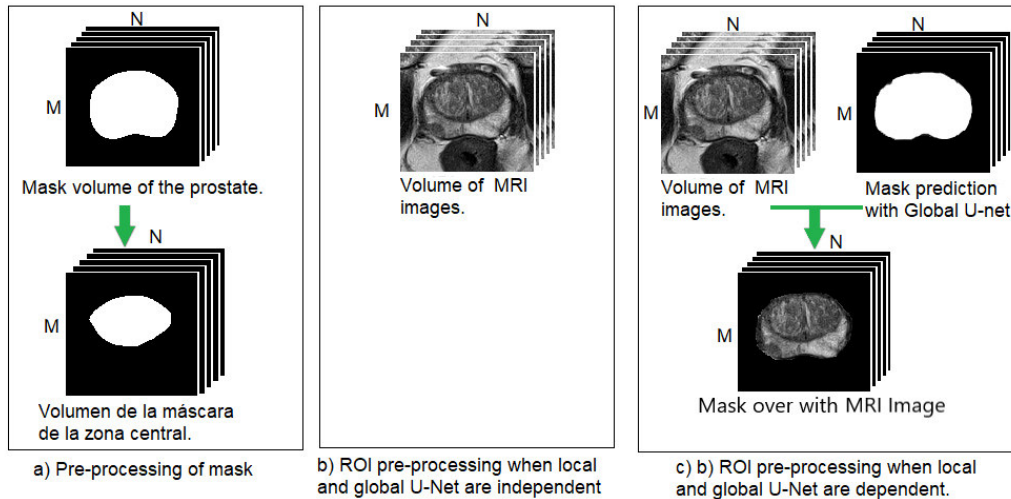

Figure 3.4: ROI pre-processing for local U-net

(b) Data augmentation The original number of MRI images in the database is 578, where 421 MRI slices contain the prostate. By using data augmentation, the database was increased to 2064 slice MRI images. Also, the augmented database contains 80% background pixels and 20% central zone pixels. The following augmentation techniques

were used: flipping, $\pm 45°$ degree random rotation, zoom and a combination between zoom and rotation operations.

(c) Feature extraction and Segmentation: The local U-net architecture was represented in Figure 3.5. This architecture is the same than global U-net architecture (see Figure 3.2) with different from the network inputs and the final layer. The final layer is formed by a 1x1 convolution that is used to map each feature vector to the desired number of classes as Central Gland y background.
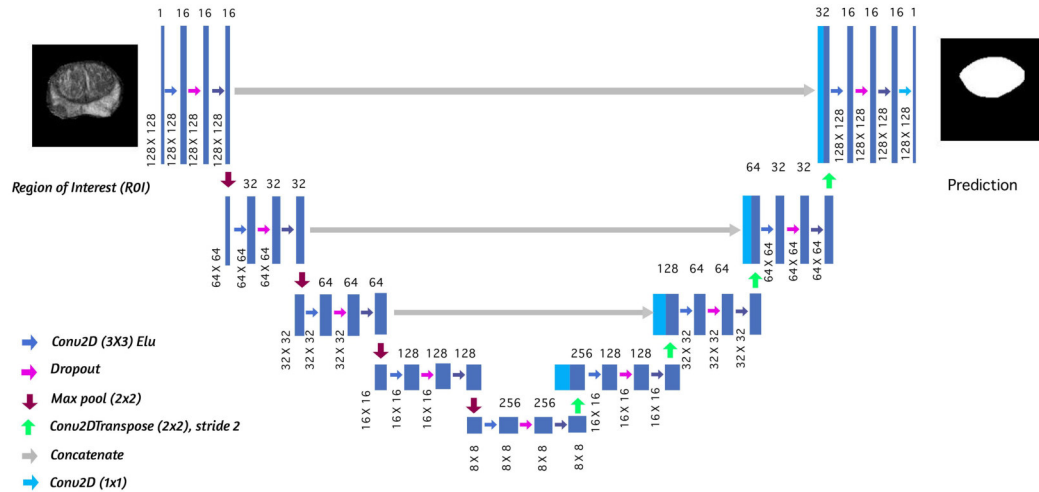

Figure 3.5: Local U-Net architecture

(d) Post-processing: In post-processing were used the same techniques of global U-net to improve the segmentation. An additional step in post-processing was obtained the peripheral zone from the segmentation of the prostate and central central gland, therefore the main zones of prostate were obtained.

Experimental selection in Local U-Net architecture
This section presents the experimental results for the local U-net architecture. This experiment used $M \times N$ ROI image size and data augmentation for training and validation. Also, the dependence or independence between the local U-net and global U-net networks was considered because the segmentation of the entire prostate gland contains parts of the central zone. Details of the experiments for the architecture are detailed below and in Table 3.9:

- *Experiment 5.1:* this experiment evaluated the performance of the proposed model when: (a) the ROI of interest had a size of $M \times N$ for training, evaluation and testing data, (b) data augmentation was used for training and evaluation, and (c) the local U-Net was independent of the global U-Net (i.e. the results of the global U-net were not taken into account to perform the segmentation of the local U-net).

- *Experiment 5.2:* this experiment evaluated the performance of the proposed model when: (a) the ROI of interest had a size of $M \times N$ for training, evaluation and testing data, (b) data augmentation was used for training and evaluation, and (c) the local U-Net was dependent on the global U-Net (i.e. the global U-net's predictions were used and multiplied by the MRI to get the input of the local U-net).

Experiment number	ROI	Increased Data	Dependence on local and global U-Net
5.1	$M \times N$	Yes	No
5.2			Yes

Table 3.9: Experiments carried out on the local U-Net.

The result of the training of the local U-Net is shown in the table 3.10, which contains the number of experiments, the DSC and IoU metrics. The result of the experiments have an overlap, therefore we cannot be said which of them have the best result. The experiments achieved an $mean \geq 91\%$ in the DSC metric and an $mean \geq 84\%$ in the IoU metric.

Experiment number	DSC [%]	IoU [%]
5.1	91.71 ± 0.93	84.79 ± 1.6
5.2	92.76 ± 0.64	86.52 ± 1.08

Table 3.10: Training result of the local U-Net.

The result of validation of the local U-Net is presented in Table 3.11, where the experiment 5.2 got higher results than experiment 5.1. The improvement showed by experiment 5.2 over experiment 5.1 is due to the use of the global U-net prediction. The experiment 5.2 achieved mean of 89.20% and 80.55% in the DSC and IoU metrics, respectively.

Experiment number	DSC [%]	IoU [%]
5.1	85.22 ± 0.53	74.27 ± 0.80
5.2	89.20 ± 0.21	80.55 ± 0.34

Table 3.11: Validation result of the local U-Net.

Table 3.12 is presented the results for the test of the local U-Net. The experiment 5.2 achieved a high performance compare with the experiment 5.1. The experiment 5.2 reached a mean of 93.34% and 87.54% in the DSC and IoU metrics, respectively. Notice that validation and testing results were higher for experiments 5.2. Furthermore, testing results score higher than those of validation.

Experiment number	DSC [%]	IoU [%]
5.1	88.34 ± 0.38	79.17 ± 0.61
5.2	93.34 ± 0.31	87.54 ± 0.54

Table 3.12: Testing result of the local U-Net.

Figure 3.6 is presented the comparation of test results for the local U-net using DSC and IoU metrics. Notice than standard deviations were $\leq 0.38\%$ in DSC metric and $\leq 0.61\%$ in DSC metric. Therefore, the result of experiments do not have overlapping.

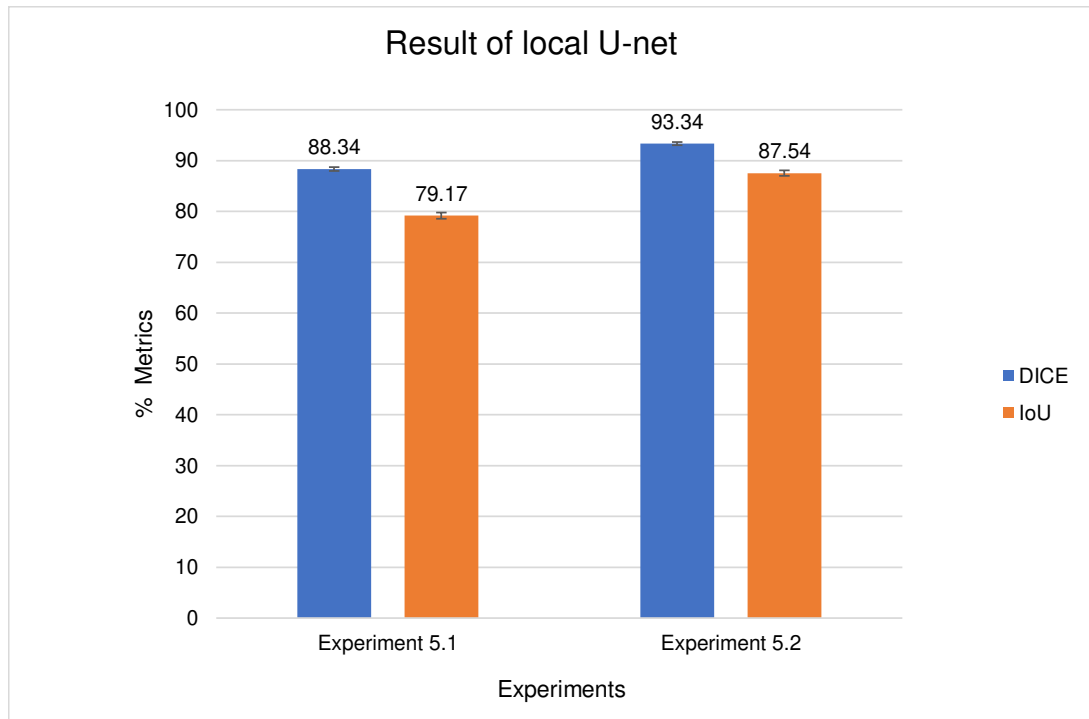Figure 3.6: Result of the experiments of the local U-Net

From, result of the experiments of local U-net were built the architecture for the model-A. The final architecture for the model-A is presented in Figure 3.7, where the Global U-net and the local U-net are dependent to achieve the segmentation of the prostate's main zones.
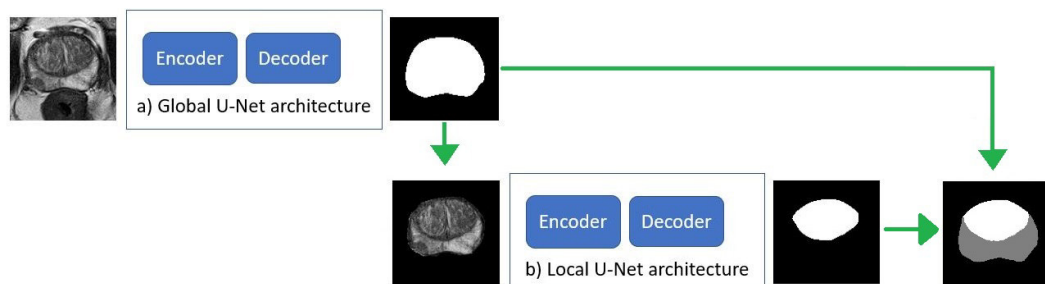

Figure 3.7: Proposed model-A based on the Encoder-Decoder Architecture

4.2.1.2 Model-B: Encoder-Classifier Architecture

The proposed model-B is based on VGG16 pre-training model [51]. VGG16 has an encoder-decoder architecture and usually is used in classification tasks. The proposed model-B is used an sliding-window setup to predict the class label of each pixel by providing a local region (patch). The steps defined to create the model-B were: a) pre-processing,b) data augmentation, c) feature extraction and segmentation, and d) post-processing.

(a) Pre-processing: The pre-processing procedure aimed to obtain MRI image patches (input) according to the classification classes. Therefore, the following sub-steps were followed:

Num.	Process
6.1	From, the $M \times N$ MRI Images specified in pre-processing of the global U-net, the patch extraction process was carried out. There is only exception to the sub-step 2.17 in Table 3.2, which consist in use a opening morphological operation with a radio 30 disk structuring element.
6.2	After, 15x15 sliding window with a stride of 5 was used to obtain slices from each MRI image, as indicated in Figure 3.8.
6.3	The class of the patch was determined according to the central pixel.
6.4	The number of classes is unbalanced, therefore a process is carried out to balance the interest classes. There are more pixels of the background class, therefore 29 background patches are discarded and the background patch number 30 remains, so on for all background patches. The following process is carried out with all ZC patches: 7 ZC patches are discarded and ZC patch number 8 is kept. For all PZ class patches, the process of discarding 1 PZ patches is carried out and PZ patch number 2 is maintained.
6.5	MRI patches were saved in image and matrix formats. This procedure was repeated with all the database (30 volume).

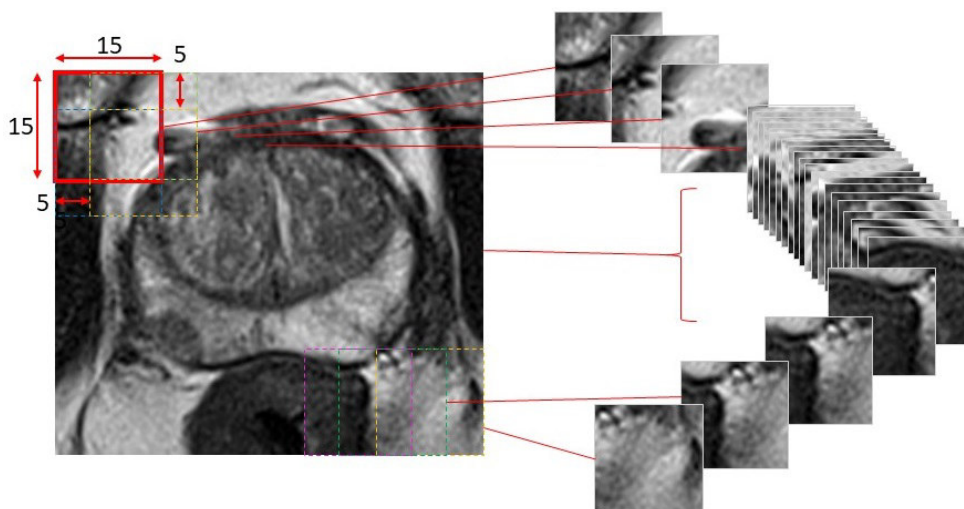Table 3.13: pre-processing for model-B

Figure 3.8: Obtain MRI image patches

(b) Data augmentation: The original set contains 578 slices of MRI images (30 MRI volumen). As indicated before, only 421 slices contain the prostate. Due to data imbalance, the number of patches for each of the classes was manipulated as follows: a) for the central zone, the number of patches was maintained; b) the number of background patches was reduced in 2/3; c) the number of patches for the peripheral zone was increased 3 times by using data augmentation techniques such as: flipping, and rotation in a range of (-45, +45). The total number of patches obtained through the sliding window were 161607 for background, 152042 for central zone, and 155040 for peripheral zone.

(c) Feature extraction and segmentation: The model-B is used the VGG16 pre-trained model. The weights of the 15 first layers were kept, where the weights of the last layer were adjusted for the classification of 3 classes (Central Gland, Peripheral Zone, and background). The process of adjusting the weights of the last layers, whereas keeping the weights values of the previous layers is called fine-tuning. Therefore,only 7211523 parameters were trained, and 7635264 parameters were re-used. Also, the model-B is used an input fixed size of 32 x 32 x 3. The model-B has 5 blocks. The first block contains 2 convolutional Layers with 64 feature maps (the image dimensions change to 32x32x64), followed by maximum pooling (the image dimensions are reduced to 16x16x64). The second block have 2 convolutional layers with 128 feature maps (the image dimensions is modified to 16x16x128), followed by maximum pooling layer (output will be reduced to 8x8x128). The third block include 3 convolutional layers with 256 feature maps (the image dimensions is 8x8x256), followed by maximum pooling layer (which will be reduced to 4x4x256). The fourth block possess 3 convolutional layers (the image dimensions change to 4x4x512), followed by maximum pooling layer (output are reduced to the half, although the feature map remains). The fifth block have of 3 convolutional layers with 512 feature maps (the image dimensions is modified to 2x2x512), followed by maximum pooling layer (output will be reduced to 1x1x512). Finally, there is a softmax output layer with 3 possible values (see Figure 3.9).
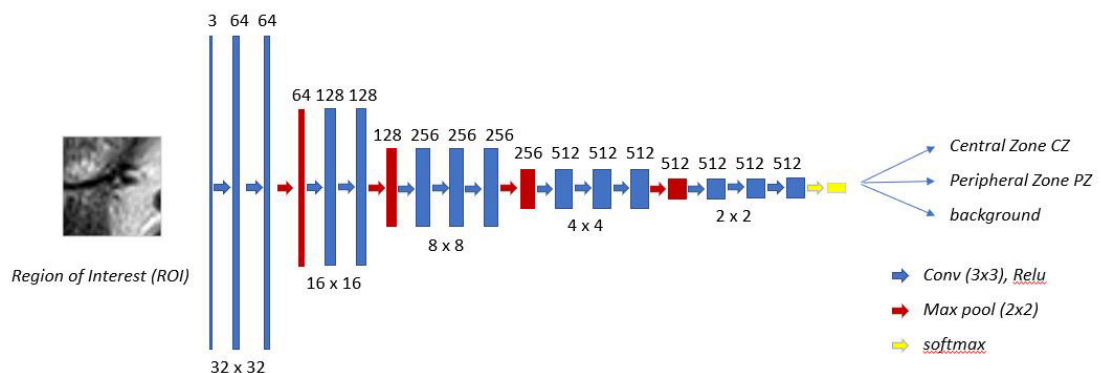

Figure 3.9: Architecture of proposed model-B

(d) Post-processing: The prediction of each pixel of the MRI image is performed

and then small areas are removed and the largest area corresponds to the desired segmentation mask.

3.2.2 Training, validation and testing parameters

Cross-validation was used to evaluate the results of a statistical analysis and ensure that they are independent of the partition between training and test data. Cross-validation consists of repeating and calculating the arithmetic mean obtained from the evaluation measures on different partitions. We established $k - fold = 5$, and total MRI number was $30\ volume$. In this case, $30\ MRI\ volume$ was divided of into 5 parts (4 parts are for training and 1 part for testing), which is repeated 5 times. The 4 parts contains $24\ MRI\ volumen$, where each volume had around $14\ MRI\ images$. Therefore, the 4 parts had a $336\ MRI\ images$, which were used in training and validation of models. Also, using data augmentation, the number of MRI images can be increased 4 more times in training the models. For testing of model is used 1 part which contains $6\ MRI\ volume$, where each volume has around $14\ MRI\ images$. The $6\ MRI\ volume$ have around $84\ MRI\ images$.

4.2.2.1 Model-A: Encoder-Decoder Architecture

The model-A was retrieved from the Tensorflow machine learning framework Also, the global and local U-net networks of the model-A used the following training parameters:

- loss function: binary cross-entropy
- optimizer: ADAM
- batch size: 16
- total number of epochs: 200
- early stopping patience: 100
- the class weight was calculated to balance the different classes of the ROI

4.2.2.2 Model-B: Encoder-Classifier Architecture

The model was retrieved from the Tensorflow machine learning framework. To train the model-B, the following parameter were used:

- loss function: categorical cross-entropy
- optimizer: SGD
- learning rate: 0.001 with decay of 0,000001 and momentum of 0.9
- batch size: 64
- number of epochs: 10
- early stopping patience: 20

Results and Analysis

In this section, the results of the proposed models are presented. In addition, the comparative analysis between the proposed models and with other models proposed in the literature are shown.

4.1 Result

The results of model-A and model-B are presented. To report the results of these models are used DSC and IoU metrics. Additionally, the confusion matrix is shown to obtain the presicio, recall and accuracy. This metrics allow to make the comparison with the model proposed in the literature.

4.1.1 Results of model-A

This section presents the results achieved by model-A (see section 3.2.1 and Figure 3.7) on the segmentation of the main zones of the prostate. The estimation of the mean and standard deviation of the model results consisted of a 10-time repetition of the cross validation process. The results of training, validate and testing were presented in Table 4.1. We can see that the testing results reached 96.79% and 93.79% in the DSC and IoU metrics, respectively. Also, the standard deviation was $\leq 0.15\%$ in the DSC metric, whereas in IoU metric was $\leq 0.29\%$. Therefore, the testing results oscillated in a range of $96.64\% - 96.94\%$ in DSC metric and $93.50\% - 94.08\%$ in IoU metric. This testing result was one of the highest achieved to compare the model-B and proposed model in the literature. In addition, the results show that there was not over-fitting. Notice that the number of MRI volumes of the NCI-ISBI 2013 was 4-times less than the average of images in private databases.

Table 4.1: Result of Model A based on Encoder-Decoder Architecture

	DSC [%]	**IoU** [%]
training	96.59 ± 0.23	93.42 ± 0.43
valid	94.09 ± 0.07	88.86 ± 0.13
test	96.79 ± 0.15	93.79 ± 0.29

According to the proposed study [12] in the literature (see Chapter 2), the precision, recall and accuracy metrics were presented when segmenting the main zones of the prostate. Therefore, in Table 4.2 was shown the confusion matrix and the precision, recall and accuracy metrics. Also, to establish comparison, the model-A was tested using the same volumes of MRI images (volumes 16-30) of study [12], where the number of processed voxels were 29237248. The confusion matrix shown a true positive rate of 95.81% for the background, 0.55% for the PZ, and 3.20% for the CG. Notice that PZ and CG pixels occupied $\leq 4.16\%$ compared to background pixels . The precision values reported, how many predictions were correct: 99.88% for the background, 77.44% for the PZ, and 94.93% for the CG. The recall values show, how many actual positives were captured: 99.95% for the background, 70.76% for the PZ, and 94.86% for the CG. Remind that PZ occupied 0.78% of the total number of pixels, therefore, the result of precision and recall were lower compared to results of precision and recall for the background and CG.

Table 4.2: Confusion matrix of Model A based on Encoder-Decoder Architecture

| | | Actual labels (%) | | | |
		Background	Peripheral zone	Central zone	Precision
Predicted labels (%)	Background	**28013085**	**22445**	**11147**	99.88%
		95.81%	*0.08%*	*0.04%*	*0.12%*
	Peripheral zone	**7004**	**159555**	**39484**	77.44%
		0.02%	*0.55%*	*0.14%*	*22.56%*
	Central zone	**6193**	**43766**	**934569**	94.93%
		0.02%	*0.15%*	*3.20%*	*5.07%*
	Recall	99.95%	70.67%	94.86%	99.56%
		0.05%	*29.33%*	*5.14%*	*0.44%*

4.1.2 Results of Model-B

The results achieved by model-B (see section 3.2.1 and Figure 3.9) were presented in Table 4.3. This table shown the results of training, validate and testing using DSC and IoU metric. The estimation of the mean and standard deviation of the model results consisted of a 10-time repetition of the cross validation process. The testing results of model-B obtained 92.50 ± 1.19 and 86.13 ± 2.02 in the DSC and IoU metrics, respectively. Also, the standard deviation was $\leq 1.19\%$ in the DSC metric, whereas in IoU metric was $\leq 10.02\%$. Therefore, the testing results of model-B fluctuated in a range of $91.31\% - 93.69\%$ in DSC metric and $84.11\% - 88.15\%$ in IoU metric. The testing result of model-B was lower than model-A with 4.26% in DSC metric and 7.66% in IoU metric. Notice, that the result of model-B did not take adjacent pixels into account. In addition, the results show that there was not over-fitting.

To compare the result of model-B with the study [12], the precision, recall and accuracy

Table 4.3: Result of Model B based on Encoder-Classifier Architecture

	DSC [%]	**IoU** [%]
training	94.27 ± 1.33	89.27 ± 2.355
valid	76.48 ± 0.36	62.25 ± 0.47
test	92.50 ± 1.19	86.13 ± 2.02

metrics were used. Therefore, in Table 4.4 was shown the confusion matrix and the precision, recall and accuracy metrics. Also, the model-B was tested using the same volumes of MRI images (volumes 16-30) of study [12], where the number of processed voxels were 29237248. The confusion matrix shown a true positive rate of 95.93% for the background, 0.33% for the PZ, and 2.27% for the CG. The model-B reached a precision de 99.05% for the background, 62.00% for the PZ, and $87,02\%$ for the CG. Also, recall values obtained 99.63% for the background, 48.80% for the PZ, and 75.05% for the CG. Despite, the classes was balanced in the training, it is noted that the precision and recall of the central gland and the peripheral area are less than background class in the testing. Notice that PZ and CG patches occupied $\leq 3.71\%$ compared to background patches.

Table 4.4: Confusion matrix of Model based on Encoder-Classifier Architecture

		Actual labels (%)			
		Background	Peripheral zone	Central zone	Precision
Predicted labels (%)	Background	**28047402** *95.93%*	**69069** *0.24%*	**199736** *0.68%*	99.05% *0.95%*
	Peripheral zone	**38330** *0.13%*	**97275** *0.33%*	**21284** *0.07%*	62.00% *38.00%*
	Central zone	**66231** 0.23%	**32994** 0.11%	**664927** 2.27%	87.02% 12.98%
	Recall	99.63% *0.37%*	48.80% *51.20%*	75.05% *24.95%*	98.54% *1.46%*

4.2 Analysis

In this section, the analysis and comparison of the results between the models-A and models-B were performed. Additionally, another analysis and comparison were performed between the models proposed and the study [12] proposed in the scientific literature.

4.2.1 comparison of the proposed models

The comparison of model-A and model-B consisted in consider test results (see the Tables 4.1 and 4.3). This comparison was shown in Figure 4.1. The results in DSC

metric obtained that model-A reached $96.79\% \pm 0.15\%$ compared to $92.5\% \pm 1.19\%$ of model-B. The results of models A and B did not show overlap in the DSC metric, therefore we concluded that model-A was better with 4.29% compared to model-B. Note that the standard deviation for model-A was less than 0.15%, whereas for model-B was less than 1.19% with a difference of 1.04% in DSC metric. Moreover, the results in IoU metric obtained that model-A reached $93.79\% \pm 0.29\%$ compared to $86.13\% \pm 2.102\%$ of model-B. In addition, the results of models A and B do not show overlap in the IOU metric, therefore we concluded that model-A was better with 7.66% compared to model B. Also note that the standard deviation of model A is less than 0.29%, whereas the model is less than 2.02% with a difference of 1.73% in IoU.
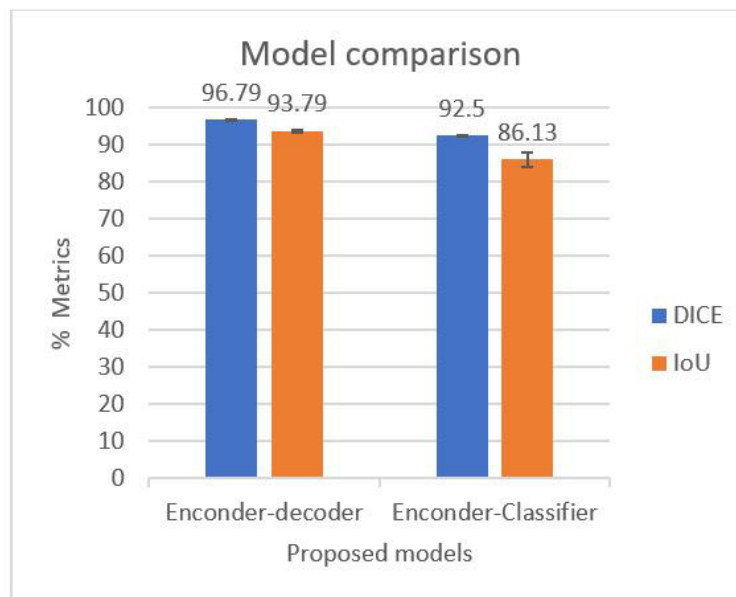

Figure 4.1: Comparison of the proposed models

4.2.2 Compare with other model

Table 4.5 shows the comparison of the results of models-A and model-B (see Tables 4.2 and 4.4) with respect to the proposed model [12] in literature. To compare the model were used the same volumes of MRI images (volumes 16-30) and after we got precision, recall, and accuracy result. Notice that model-A achieved the high results for precision, and recall with 94.9256% and 94.8608% for the central gland. Also, model-A precision for the central gland was higher with 7.9082% compared to model-B, and with 22.9081% outperformed to proposed model [12]. Whereas, model-A recall for the central gland was higher with 19.8096% compared to model-B, and with 38.0682% outperformed to proposed model [12]. On the other hand, the model-A in peripheral zone reached 77.4377%, and 70.7627% for precision, and recall, respectively. Also, the model-A precision for the peripheral zone was higher with 15.4362% compared to model-B, and with 1.1153% outperformed to proposed model [12]. Also, model-A recall for the peripheral zone was higher with 21.8682% compared to model-B, and

with 7.6330% outperformed to proposed model [12]. Moreover, the model-B achieved the results of 87.0174% and 75.0512% for precision and recall for the central gland. The model-B precision for the central gland was higher with 14.9999% to proposed model [12]. Whereas, model-B recall for the central gland was higher with 18.2586% to proposed model [12]. Additionally, the model-B reached the results of 62.0015% and 48.8045% for precision and recall for the peripheral zone. The model-B precision for the peripheral zone was lower with 14.3209% to proposed model [12]. Whereas, model-B recall for the peripheral zone was lower with 14.2352% to proposed model [12].

Table 4.5: Comparison of models in precision, recall and accuracy metrics

Metric	Model A	Model B	Other model [12]
Precision (Predicted labels | Actual labels)			
*Prostate gland	Prostate gland*	98.8915	88.6480	77.7071
*Central gland	Central gland*	94.9256	87.0174	72.0175
*Peripheral zone	Peripheral zone*	77.4377	62.0015	76.3224
Recall (Actual labels | Predicted labels)			
Prostate gland	Prostate gland	97.2260	75.2310	62.0712
Central gland	Central gland	94.8608	75.0512	56.7926
Peripheral zone	Peripheral zone	70.6727	48.8045	63.0397
Accuracy	99.56	98.54	96.84

In Figure 4.2 was compared the ground truth with segmentation of 4 slice MRI images slices by model-A, model-B, and the model [12]. The peripheral zone was depicted in green, whereas the central gland was presented in red. We can see that model-A got similar result a ground truth. After, the result of model-B occupied second place in the segmentation of the main areas.

From result of Table 4.5, and Figure 4.2, we concluded that model-A outperformed to model-B and model [12]. The result of model-A used the $M \times N$ ROI, where $M \times N$ ROI reduced 52% background pixels. Besides, the architecture of model-A considered the adjacent pixels of MRI image to segment. On the other hand, the model-B occupied second place. Model-B same as model-A used the $M \times N$ ROI. However, the architecture of Model-B did not consider the adjacent patches to segment. Therefore, model-B detected spurious objects of Central Gland or Peripheral Zone. Notice that Peripheral Zone occupied small area of pixels in the MRI image, therefore this result can rise.

Finally, the model [12] proposed in the literature, which obtained lower results compared to model-A and model-B. The segmented area of model [12] contained fewer true positives in Figure 4.2. Therefore, the model [12] detected more spurious objects of Central Gland or Peripheral Zone.
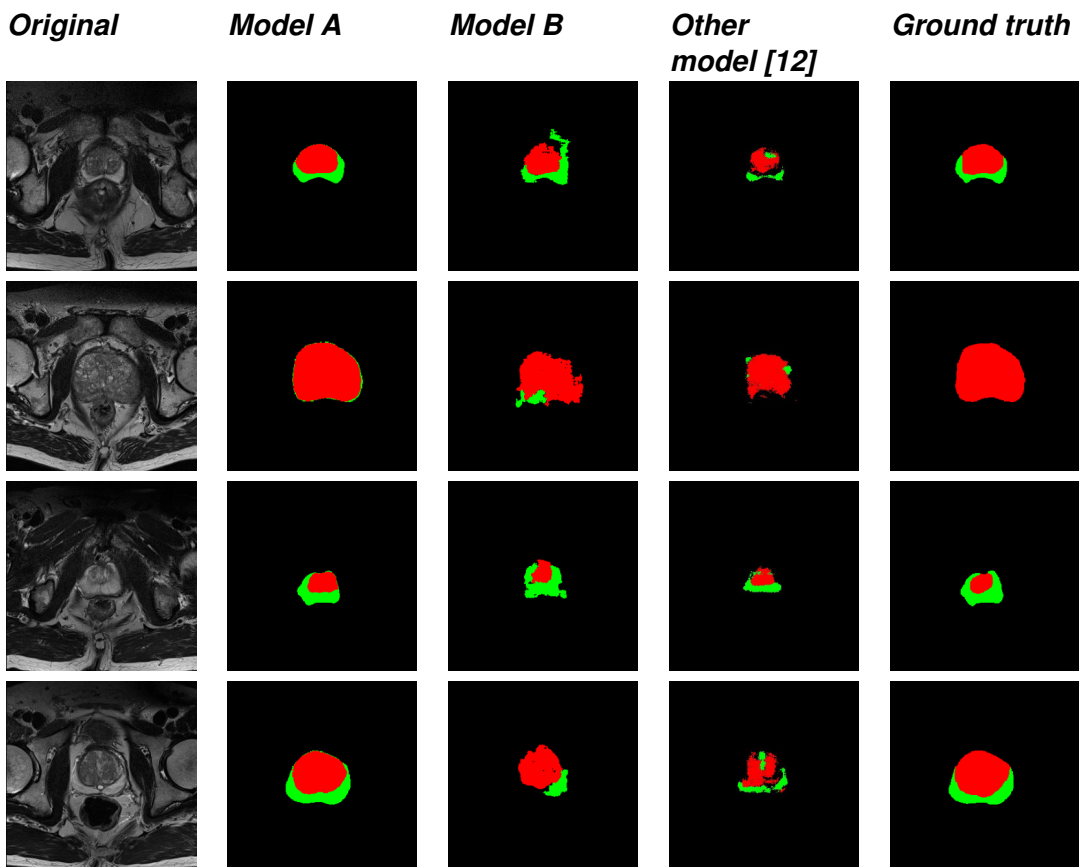
| Original | Model A | Model B | Other model [12] | Ground truth |

Figure 4.2: Examples of original, results of model based on encoder-decoder,results of model based on encoder-classifier,results of other model proposed in SLR, and ground truth slices. Green = peripheral zone; Red = central gland

Conclusions and Recommendations

5.1 Conclusions

- To accomplish the general objective, two of 2D prostate segmentation models of the main areas were created. These models encompass a different segmentation approach, and achieved DSC results of 96.78% and 92.50% for models-A and models-B, respectively. A very important part, when creating a segmentation model is the area that the prostate occupies. The prostate area is around 1/4 and this area varies depending on the patient, therefore it was necessary to eliminate the background pixels that are not necessary through a ROI region of interest, which allowed us to obtain established results. In addition, because few research regarding the segmentation of the main areas of the prostate can use other standard architectures, and test their result.

- The SLR presented in Chapter 2, addressed a research question related to the common metrics used in prostate segmentation. The most used metric was the DSC. However, the DSC metric is a variation to IoU metric. Therefore, results can be misleading because a high result on the DSC metric can be linked to a lower result on the IoU metric.

- To address the first objective is proposed the mode-A with an encoder-decoder architecture. This model is based on U-net. Model-A is composed by a global and local U-net, as indicated in Figure 3.7. The global U-net segments the whole prostate. Whereas the local U-net segments the central gland. The global U-net input uses 2D MRI images with ROI of $M \times N$, where $M \leq 151$ and $N \leq 151$. The ROI of $M \times N$ reduced the background pixels by 52%. The ROI of $M \times N$ was obtained from comparison experiments (see Table 3.7 and Figure 3.3). On the other hand, the U-net local uses 2D MRI image with ROI of $M \times N$ together with the segmentation result of the global U-net. Therefore, the inputs of the local U-net depend on the result of the global U-net (see Table 3.12 and Figure

3.6). From the results of the global and local U-net, the segmentation of the Peripheral Zone is obtained. The training parameters of the local and global unet is specified in section 3.2.2. Also, the training, validation and testing of model-A are carried out in CPU INTEL 2680 of MODEMAT. Model-A is tested in the NCI-ISBI 2013 database, and achieved a performance of $DSC = 96.79\% \pm 0.15\%$ and $IoU = 93.79\% \pm 0.29\%$.

- To address the second objective is present the Model- B with an encoder-classifier architecture (see Figure 3.9). This model is based on VGGnet, which is used in classification tasks. Model-B inputs use patches obtained from 2D MRI images of $M \times N$. According to central pixel, the patches are labeled as central gland, peripheral zone, and background. A very important part for this model is the balance of the classes, which were obtained 161607 patches for the background, 152042 patches for the central gland and 155040 patches for the peripheral zone. Also, Model-B use fine-tuning to train the model and classify in the aforementioned classes. The classification of all the patches of an 2D MRI image segment the main zone of the prostate. The training parameters of model-B is specified in section 3.2.2. Also, the training, validation and testing of model-B are carried out in CPU INTEL 2680 of MODEMAT. Model-B is tested in the NCI-ISBI 2013 database, and achieved a performance of $DSC = 92.50\% \pm 1.19\%$ and $IoU = 86.13\% \pm 2.02\%$.

- Regarding the third objective, the comparison between the results of Model-A and Model-B (Section 4.2.1) suggest that Model A achieves the high result with 4.29% compare to model-B for the DSC metric. Also, Model-A achieves the high result with 7.66% compare to model-B for the IoU metric. Notice that Model-A takes into account the adjacent pixels, In addition, a visual comparison of the models is shown in the Figure 4.2, in which confirmed that the segmentation results of model-A are similar to the ground truth. Whereas the model-B detects spurious objects of central gland or peripheral zone. Furthermore, the model-B does not consider the adjacent patches, therefore model-B tends to erroneously predict the peripheral zone class, because peripheral zone occupies a smaller area $\leq 3.71\%$. Notice that the number of MRI volumes of the NCI-ISBI 2013 are small compared to the number of MRIs in private databases which have an average of 120 MRI volumes.

- The comparison the model-A, model-B and model [12] are shown in Table 4.5. The precision, recall and accuracy metrics and the same MRI volumes were considered to compare the models. The results for the central gland were: $precision$ reached $94.92\%, 87.01\%$, and 72.01% for model A, model B and model [12], respectively. Whereas the $recall$ reached $94.86\%, 75.05\%$, and 56.79% for model A, model B and model [12], respectively. On the other hand, the results for the peripheral zone were: $precision$ reached $77.43\%, 62.00\%$, and 76.32% for model A, model B and model [12], whereas the $recall$ reached $70.76\%, 48.80\%$, and

63.03%, for model A, model B, and model [12]. Note that both the $precision$ and $recall$ metric for the peripheral zone have a lower result in the three models. This is because the peripheral zone has fewer pixels compared to the background and central gland.

- Additionally, a comparative visual analysis was performed (see Figure 4.2) and we concluded that models A and B have superior results than model [12]. The model [12] uses Feed-Forward Artificial Neural Network of three layers, and this model to extract the characteristics manually to enter the neural network and predicts the class. Although, the model-B has a similar approach to the model [12], the model-B uses a convolutional neural network and patches of the ROI of $M \times N$. Therefore model-B work better in comparison with the model. Also, the architecture of models A and model [12] are different, where model-A use encoder-decoder architecture.

5.2 Recommendations

- This project has verified that the number of pixels in the area of the prostate in the MRI slice image is small in the NCI-ISBI 2013 [47]. It could be suggested that smaller ROI (crop of a sub-volume) are generated to improve the segmentation results of the main zones of the prostate. To obtain smaller ROI, the pre-procesing technique specified in Figure 2.4, literal a, can be considered.

- Despite extracting smaller ROI from the NCI-ISBI 2013 [47] database, the number of pixels for background is greater than that of the other main areas of the prostate. Therefore, it is recommended to use balancing techniques to improve the segmentation of the main zones of the prostate gland.

- One of the challenges in prostate segmentation is the limited number of medical images. Therefore, it is recommended to use data augmentation techniques (see Figure 2.3) that allow to increase the number of data. However, the augmented data contains the same information.

- The implementation of the proposed models used the Python programming language. This programming language allows the implementation of models using CNN. Furthermore Python has many open source libraries and a lot of code to reuse. Therefore, it is recommended to use Python programming language.

- The pre-processing and data augmentation of the images of the proposed models was implemented in Matlab, because of it s powerful image processing. Hence, it is recommended to use Matlab for image preprocessing.

Appendix *A*

Studies for review

	Year	Segment		MRI Images			Database	
		Type	Areas	sequences	Plane	acquisition	type	Public
p1	2014	2D	ZC, ZP	T2W	Axial	non-ERC	Public	NCI-ISBI 2013
p2	2016	3D	Gland	T2W	Axial	Both	Public	PROMISE12
p3	2016	3D	Gland	T2W	Axial	Both	Public	PROMISE12
p4	2017	3D	Gland	T2W	Axial	ERC	Public	PROMISE12
p5	2017	2D	Gland	T2W	Axial		Private	
p6	2017	2D	Gland	T2W, CED	Axial	ERC	Private	
p7	2017	2D	Gland, ZT	T2W, DWI, ADC, other	Axial		Private	
p8	2017	2D	Gland		Axial	ERC	Private	
p9	2018	3D	Gland	T2W	three plane	non-ERC	Public	PROSTATEx
p10	2018	3D	Gland	T2W	Axial		Private	
p11	2018	3D	Gland	T2W	Axial	Both	Public	NCI-ISBI 2013
p12	2018	2D	Gland	T2W	Axial	ERC	Public	PROMISE12
p13	2018	2D	Gland	T2W	Axial	ERC	Public	PROMISE12
p14	2018	3D	Gland, ZT, ZP	T2W	Axial		Private	
p15	2018	2D	Gland	T2W	Axial	ERC	Public	PROMISE12
p16	2018	2D	Gland	T2W	Axial	ERC	Public	PROMISE12
p17	2018	3D	Gland	T2W, DWI	Axial	ERC	Private	
p18	2018	3D	Gland	T2W	Axial	ERC	Private	
p19	2018	2D	Gland	T2W	Axial	Both	Public	PROMISE12
p20	2018	3D	Gland	T2W	Axial	Both	Public	PROMISE12
p21	2018	2D	Gland	T2W	Axial	Both	Public	PROMISE12
p22	2018	3D	Gland	T2W	Axial	ERC	Public	PROMISE12
p23	2018	2D	Gland	T2W	Axial	ERC	Both	PROMISE12
p24	2019	2D	Gland	T2W	Axial	Both	Public	PROMISE12
p25	2019	2D	Gland	T2W	Axial	ERC	Public	PROMISE12
p26	2019	3D	Gland	T2W	Axial	ERC	Public	PROMISE12

Appendix B

segmentation challenges
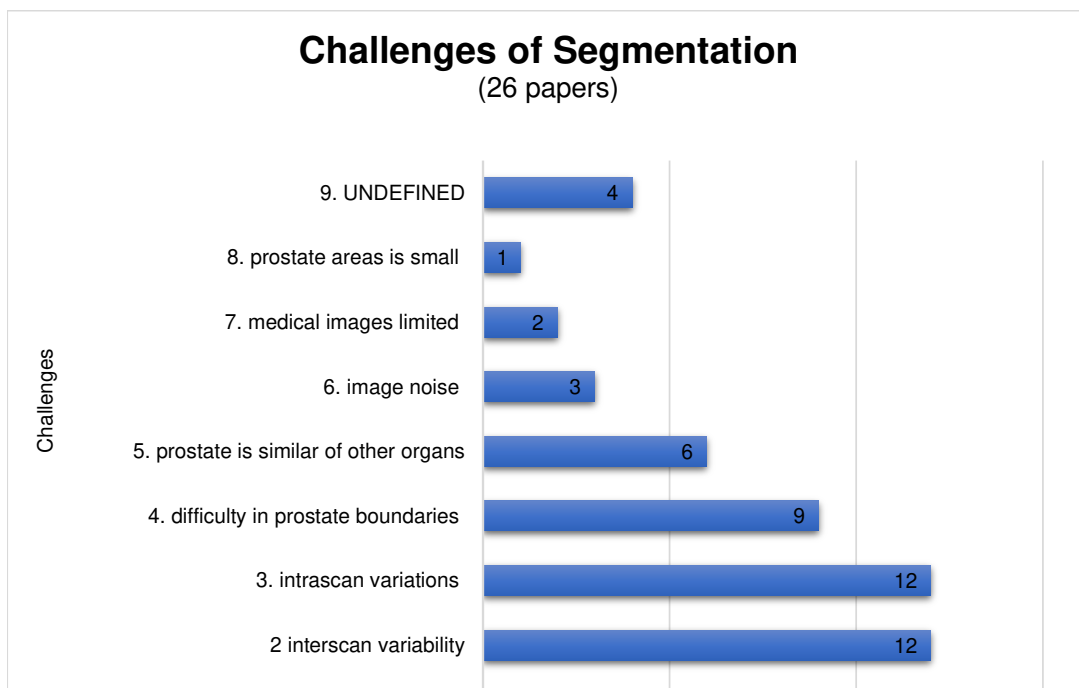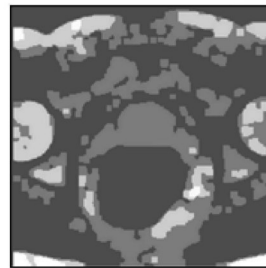

Challenges of Segmentation
(26 papers)

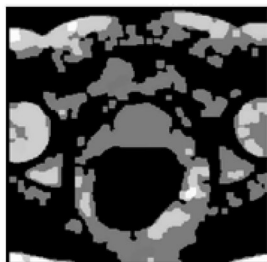Figure B.1: segmentation challenges
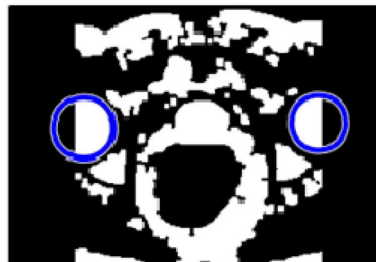
Appendix C

151 x 151 ROI Pre-processing
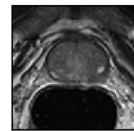

a) Original MRI imagen

b) Mathematical morphology

c) Fill Holes and regions

d) Pattern search (pelvic bone)

e) ROI of 151x151

Figure C.1: 151×151 ROI pre-processing: a) the original 320×320 pixels image; b) mathematical morphology was used to highlight patterns that locate the pelvic bones; c) hole filling process; d) ROI X, Y coordinates were found and the ROI image was increased by adding 75 pixels to the ends; e) 151×151 ROI
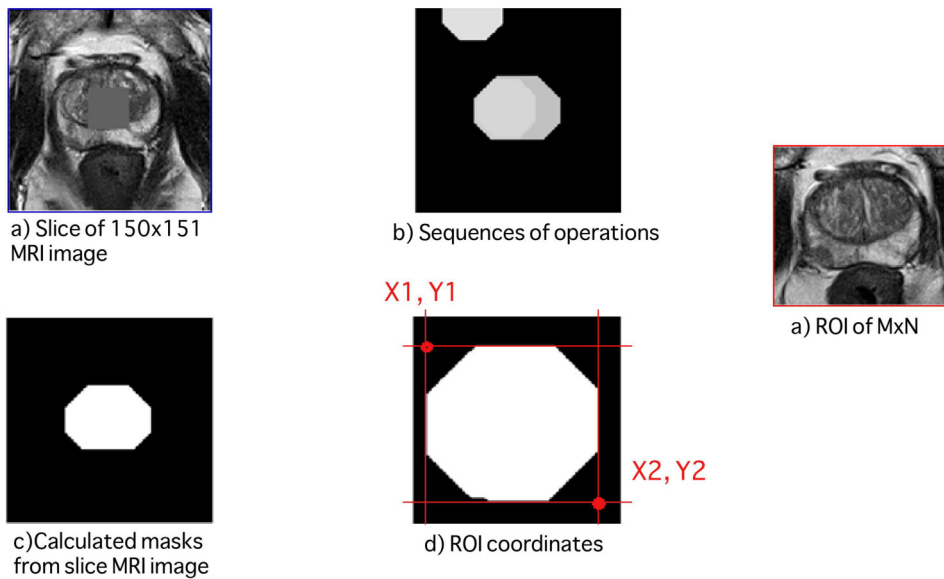
Appendix D

M x N ROI Pre-processing

a) Slice of 150x151 MRI image

b) Sequences of operations

a) ROI of MxN

c) Calculated masks from slice MRI image

X1, Y1

X2, Y2

d) ROI coordinates

Figure D.1: $M \times N$ ROI pre-processing: a) is shown an MRI image of size 151×151 pixels, which has central area of 30×30 with the average value pixels. b) presents the result of the sequence of operations to highlight areas of interest. c) display the calculation of the mask. d) shows the expansion of the mask calculated, to obtain the coordinates to trim the ROI. e) shows the ROI results of $M \times N$.
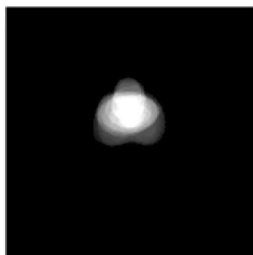
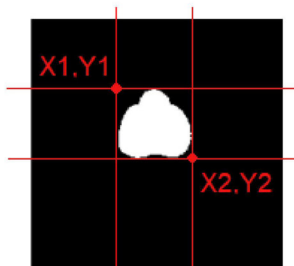# Pre-processing of the ROI adjusted to the mask
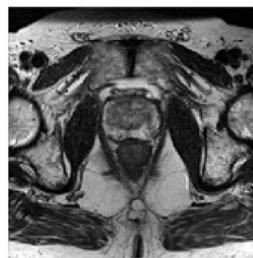
a) Slice of 230x320 MRI image

a) Mask of 230x320 MRI image

c) Sum of all masks

d) ROI coordinates

a) ROI adjusted to the masks

Figure E.1: Pre-processing of the ROI adjusted to the mask

Bibliography

[1] F. B. J. F. I. S. L. S. L. A. T. A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancersin 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, pp. 394–424, 2018.

[2] K. H. J. M. T. Stamey, "Ultrasound guided transrectal core biopsies of the palpably abnormal prostate," *The Journal of Urology*, vol. 142, no. 1, pp. 66–70, 1989.

[3] R. W. M. N. J. Weigel, "Correlation of ultrasound guided and digitally directed transrectal biopsies of palpable prostatic abnormalities," *The Journal of Urology*, vol. 145, no. 3, pp. 516–518, 1991.

[4] C. A. L. S. A. d. l. T. Guillaume Ploussard; Nathalie Nicolaiew; Charles Marchand, Stephane Terry; Francis Vacherot; Dimitri Vordos; Yves Allory, "Prospective evaluation of an extended 21-core biopsy scheme as initial prostate cancer diagnostic strategy," *European Urology*, vol. 65, no. 1, pp. 154–161, 2014.

[5] A. D. la Taille; Patrick Antiphon; Laurent Salomon; Maguy Cherfan; Raphael Porcher; Andras Hoznek; Fabien Saint; Dimitri Vordos; Anthony Cicco; Renee Yiou; Elie Zafrani; Dominique Chopin; Claude Abbou, "Prospective evaluation of a 21-sample needle biopsy procedure designed to improve the prostate cancer detection rate," *Urology*, vol. 61, no. 6, p. 1181–1186, 2003.

[6] D. Lomas and H. Ahmed, "All change in the prostate cancer diagnostic pathway," *Nature Reviews Clinical Oncology*, vol. 17, pp. 1–10, 02 2020.

[7] A. Villers, F. Marliere, A. Ouzzane, P. Puech, and L. Lemaître, "Mri in addition to or as a substitute for prostate biopsy: The clinician's point of view," *Diagnostic and interventional imaging*, vol. 93, no. 4, pp. 262–267, 03 2012.

[8] C. Cardenas, J. Yang, B. Anderson, L. Court, and K. Brock, "Advances in auto-segmentation," *Seminars in Radiation Oncology*, vol. 29, pp. 185–197, 07 2019.

[9] A. El-Shater Bosaily, C. Parker, L. Brown, R. Gabe, R. Hindley, R. Kaplan, M. Emberton, and H. Ahmed, "Promis — prostate mr imaging study: A paired validating cohort study evaluating the role of multi-parametric

mri in men with clinical suspicion of prostate cancer," *Contemporary Clinical Trials*, vol. 42, pp. 26–40, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1551714415000415

[10] K. Yan, X. Wang, J. Kim, M. Khadra, M. Fulham, and D. D. F. Feng, "A propagation-dnn: Deep combination learning of multi-level features for mr prostate segmentation," *Computer Methods and Programs in Biomedicine*, vol. 170, pp. 11–21, 03 2019.

[11] W. Zhang, R. li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, 01 2015.

[12] M. B. M. B. V. Ballarin, "Automatic design of window operators for the segmentation of the prostate gland in magnetic resonance images," *IFMBE Proceedings*, p. 5417–420, 2015.

[13] M. Benalcázar, M. Brun, and V. Ballarin, "Automatic design of window operators for the segmentation of the prostate gland in magnetic resonance images," in *IFMBE Proceedings*, vol. 49. Springer, 10 2014, pp. 417–420.

[14] D. Patil and S. Deore, "Medical image segmentation: A review," *International Journal of Computer Science and Mobile Computing*, vol. 2, pp. 22–27, 01 2013.

[15] D. Kaur and Y. Kaur, "Various image segmentation techniques: A review," *International Journal of Computer Science and Mobile Computing*, vol. 3, p. 809 – 814, 5 2014. [Online]. Available: https://ijcsmc.com/docs/papers/May2014/V3I5201499a84.pdf

[16] S. Shirly and K. Ramesh, "Review on 2d and 3d mri image segmentation techniques," *Current medical imaging reviews*, vol. 15, no. 2, p. 150—160, 2019.

[17] G. Garg and M. Juneja, "A survey of prostate segmentation techniques in different imaging modalities," *Current Medical Imaging Reviews*, vol. 14, no. 1, pp. 19–46, 02 2018.

[18] S. Ghose, A. Oliver, R. Martí, X. Lladó, J. C. Vilanova, J. Freixenet, J. Mitra, D. Sidibé, and F. Meriaudeau, "A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 1, pp. 262–287, 2012.

[19] B. Kitchenham and S. Charters, *Guidelines for performing Systematic Literature Reviews in Software Engineering*, 2007.

[20] C. for Reviews and D. (CRD), "Systematic reviews. crd's guidance for undertaking reviews in health care," *University of York*, 2009.

[21] K. Yan, C. Li, X. Wang, A. Li, Y. Yuan, D. D. F. Feng, M. Khadra, and J. Kim, "Automatic prostate segmentation on mr images with deep network and graph model," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 08 2016, pp. 635–638.

[22] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Fourth International Conference on 3D Vision (3DV)*. IEEE, 10 2016, pp. 565–571.

[23] J. Mun, w.-D. Jang, D. Sung, and C.-S. Kim, "Comparison of objective functions in cnn-based prostate magnetic resonance image segmentation," in *IEEE International Conference on Image Processing (ICIP)*, 09 2017, pp. 3859–3863.

[24] J. Sun, Y. Shi, Y. Gao, and D. Shen, "A point says a lot: An interactive segmentation method for mr prostate via one-point labeling," in *International Workshop on Machine Learning in Medical Imaging*, vol. 10541, 09 2017, pp. 220–228.

[25] R. Cheng, H. Roth, N. Lay, L. Lu, B. Turkbey, W. Gandler, E. McCreedy, T. Pohida, P. Pinto, P. Choyke, M. Mcauliffe, and R. Summers, "Automatic magnetic resonance prostate segmentation by deep learning with holistically nested networks," *Journal of Medical Imaging*, vol. 4, p. 1, 08 2017.

[26] T. Clark, J. Zhang, S. Baig, A. Wong, M. Haider, and F. Khalvati, "Fully automated segmentation of prostate whole gland and transition zone in diffusion-weighted mri using convolutional neural networks," *Journal of Medical Imaging*, vol. 4, p. 1, 10 2017.

[27] Q. Zhu, B. Du, B. Turkbey, P. Choyke, and P. Yan, "Deeply-supervised cnn for prostate segmentation," in *International Joint Conference on Neural Networks (IJCNN)*, 05 2017, pp. 178–184.

[28] A. Meyer, A. Mehrtash, M. Rak, D. Schindele, M. Schostak, C. Tempany, T. Kapur, P. Abolmaesumi, A. Fedorov, and C. Hansen, "Automatic high resolution segmetation of the prostate from multi-planar mri," in *IEEE International Symposium on Biomedical Imaging*, 01 2018.

[29] E. Gibson, Y. Hu, N. Ghavami, H. U. Ahmed, C. Moore, M. Emberton, H. Huisman, and D. Barratt, "Inter-site variability in prostate segmentation accuracy using deep learning," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, vol. 11073, 01 2018.

[30] H. Jia, Y. Song, D. Zhang, H. Huang, D. Feng, M. Fulham, Y. Xia, and W. Cai, "3d global convolutional adversarial network for prostate mr volume segmentation," 2018.

[31] S. Hossain, A. P. Paplinski, and J. M. Betts, "Residual semantic segmentation of the prostate from magnetic resonance images," in *Neural Information Processing*,

L. Cheng, A. C. S. Leung, and S. Ozawa, Eds.   Springer International Publishing, 2018, pp. 510–521.

[32] D. Ji, J. Yu, T. Kurihara, L. Xu, and S. Zhan, "Automatic prostate segmentation on mr images with deeply supervised network," in *5th International Conference on Control, Decision and Information Technologies (CoDIT)*, 04 2018, pp. 309–314.

[33] A. de Gelder and H. Huisman, "Autoencoders for multi-label prostate mr segmentation," 2018.

[34] W. Chen, Y. Zhang, J. He, Y. Qiao, Y. Chen, H. Shi, E. Wu, and X. Tang, "Prostate segmentation using 2d bridged u-net," in *International Joint Conference on Neural Networks (IJCNN)*, 07 2019, pp. 1–7.

[35] D. Jit, J. Qian, J. Yu, T. Kurihara, and S. Zhan, "Automatic prostate segmentation on mr images using enhanced holistically-nested networks," in *24th International Conference on Pattern Recognition (ICPR)*, 08 2018, pp. 3820–3825.

[36] M. To, Q. Vu, B. Turkbey, P. Choyke, and J. T. Kwak, "Deep dense multi-path neural network for prostate segmentation in magnetic resonance imaging," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, pp. 1687–1696, 08 2018.

[37] Q. Zhu, B. Du, J. Wu, and P. Yan, "A deep learning health data analysis approach: Automatic 3d prostate mr segmentation with densely-connected volumetric convnets," in *International Joint Conference on Neural Networks (IJCNN)*, 07 2018, pp. 1–6.

[38] B. He, D. Xiao, Q. Hu, and F. Jia, "Automatic magnetic resonance image prostate segmentation based on adaptive feature learning probability boosting tree initialization and cnn-asm refinement," *IEEE Access*, vol. PP, pp. 1–1, 12 2017.

[39] T. Brosch, J. Peters, A. Groth, T. Stehle, and J. Weese, "Deep learning-based boundary detection for model-based segmentation with application to mr prostate segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*.   Springer International Publishing, 09 2018, pp. 515–522.

[40] H. Jia, Y. Xia, Y. Song, W. Cai, M. Fulham, and D. D. F. Feng, "Atlas registration and ensemble deep convolutional neural network-based prostate segmentation using magnetic resonance imaging," *Neurocomputing*, vol. 275, pp. 1358–1369, 10 2017.

[41] Z. Tian, L. Liu, Z. Zhang, and B. Fei, "Psnet: Prostate segmentation on mri based on a convolutional neural network," *Journal of Medical Imaging*, vol. 5, p. 1, 01 2018.

[42] Y. Zhang, J. Wu, W. Chen, Y. Chen, and X. Tang, "Prostate segmentation using z-net," in *International Symposium on Biomedical Imaging (ISBI 2019)*.   IEEE, 2019, pp. 11–14.

[43] D. Karimi, G. Samei, Y. Shao, and S. Salcudean, "A deep learning-based method for prostate segmentation in t2-weighted magnetic resonance imaging," 2019.

[44] W. B. Y. L. S. T. T. W. Y. L. P. P. A. J. H. M. W. C. T. L. X. Yang, "Deeply supervised 3d fcn with group dilated convolution for automatic mri prostate segmentation," *Medical Physics*, vol. 46, 01 2019.

[45] M. F. H. A. J. L. A. Larach, "Utilidad de la resonancia magnÉtica con endocoil en la etapificaciÓn y planeamiento del tratamiento del cancer rectal," *Revista Chilena de Radiología*, vol. 13, no. 2, pp. 57–62, 2007.

[46] T. Mussi, R. Baroni, R. Zagoria, and A. Westphalen, "Prostate magnetic resonance imaging technique," *Abdominal Radiology*, vol. 45, 07 2019.

[47] "Nci-isbi 2013 challenge: Automated segmentation of prostate structures," Cancer Imaging Archive, 2013. [Online]. Available: https://wiki.cancerimagingarchive.net/display/Public/NCI-ISBI+2013+Challenge+-+Automated+Segmentation+of+Prostate+Structures

[48] SPIE-AAPM-NCI, "Spie-aapm-nci prostatex challenges," *Cancer Imaging Archive*, 2016.

[49] PROMISE12, "Prostate mr image segmentation - challenge 2012 (promise12)," *MICCAI Grand Challenge*, 2012.

[50] O. R. P. F. T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *University of Freiburg, Germany*, 2015.

[51] K. S. A. Zisserman, "Very deep convolutional networks for large-scake image recognition," *Department of Engineering Science, University of Oxford*, 2015.