

**Politecnico di Torino**

---



Master of Science in *Biomedical Engineering*

***Prostate Cancer Detection  
based on Bi-parametric MRI***

Supervisors

*Prof. Gabriella Balestra*

*Prof. Valentina Giannini*

*Prof. Samanta Rosati*

Candidate

*Giulia Maria Mattia*

Academic Year

2018-2019

# Acknowledgments

I would like to thank my supervisors, Prof. Gabriella Balestra, Prof. Valentina Giannini and Prof. Samanta Rosati, for letting me pursue this thesis project and helping me in achieving this goal.

To my mother Angela and my grandparents Rosanna and Pino. Thank you for your unconditional love and encouragement, and for always believing in me. Without you, I could not have accomplished all of this.

To my soul mate Alessandro. Thank you for your endless patience and precious advices, for being always by my side, and for all the love and joy you fill my heart with.

To Sergio and Emilia. Thank you for your constant wonderful presence and for loving me as a daughter.

To Silvia. Thank you for your overwhelming *simpatia* and for being an amazing sister.

To my uncles Paola and Massimo and my cousins Marco and Alessia. Thank you because I can always count on you and I know how much you care about me.

To Annagrazia, Miriam and Letizia. Thank you for letting me know the value of friendship, for all the ice-creams we ate, the coffee we drank and the obstacles we endured.

To Desiré and Valentina. Thank you for accompanying me through these years, sustaining me and making me laugh.

To Antonio. Thank you for finding the time to listen and give me your support.

Thank you to all the people I have not mentioned, who have been part of this challenging and satisfying journey.

These are no words to express all the gratitude and the love I feel.

Giulia

*"Without crisis there are no challenges, without challenges life is a routine, a slow agony. Without crisis there is no merit. In crisis emerges the best of each, because without crisis all winds are only mild breezes. [...] Instead, we work hard. Let us stop, once and for all with the only dangerous crisis, which is the tragedy of not being willing to overcome."*  
Albert Einstein

# Contents

## Contents

<b>Summary</b>	<b>1</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Rationale . . . . .	3
1.2 Prostate Cancer . . . . .	4
1.3 MRI for Prostate Cancer . . . . .	7
<b>2 Prostate Segmentation</b>	<b>10</b>
2.1 Patients . . . . .	10
2.2 MR Image Acquisition . . . . .	11
2.3 Algorithm . . . . .	13
2.4 Validation . . . . .	21
<b>3 Machine Learning for Tumour Detection</b>	<b>27</b>
3.1 Patients . . . . .	28
3.2 Feature Extraction . . . . .	29
3.3 Classifier Construction . . . . .	31
3.4 Classification Optimisation . . . . .	46
3.5 Overall Comparison . . . . .	84
<b>Conclusions and future work</b>	<b>89</b>
<b>Appendix</b>	<b>91</b>
<b>List of Abbreviations</b>	<b>129</b>
<b>List of Figures</b>	<b>131</b>
<b>List of Tables</b>	<b>136</b>
<b>Bibliography</b>	<b>138</b>

# Summary

The aim of this study is to determine the potential of the informative content enclosed in the two image modalities proper of bp-MRI (i.e. T2-weighted and diffusion weighted) in order to automatically detect malignant lesions on behalf of the prostate. All the elaborations will be performed separately for the two image modalities, as initial exploratory trial.

Firstly, segmentation is necessary to isolate the prostate gland. The algorithm designed for this purpose is completely automatic and based on the location of the prostate in each image. It receives as input the ADC map, created from DW sequences, whilst the outcoming segmentation is utilised for the T2-weighted images as well.

Once the prostate is delineated, feature extraction of first order statistics and texture features (i.e. GLCM and GLRLM) is performed, considering non-overlapping regions of interests from the obtained masks for both image modalities.

Characterised by a significant disproportion concerning healthy and cancerous areas, the labelled data set is then used to build a balanced training set embedding the two classes by means of decision trees.

A genetic algorithm is employed to simultaneously select the most relevant and instructive features and optimise the parameters of a support vector machine. This approach led to comparable performances in terms of sensitivity and specificity (around 0.7), which are not enough considering the consistent difference in test set size between malignant and benign samples. What immediately emerges is the need of a greater representativeness of healthy zones, proved by the relatively low specificity, which would compromise the correctly identified cancerous areas.

To enhance the representation of benign tissue, another training set is conceived beginning from decision trees outcome and applying dendrogram clustering on the most numerous benign nodes. The corresponding clusters will constitute diverse benign classes, randomly extracting from each a number of elements equal to the number of malignant ROIs comprised in the decision tree based training set.

A multiclass version of SVM is trained with dendrogram based training set and the optimised parameters chosen by the genetic algorithm, effectively revealing improved identification of non cancerous regions.

K-nearest neighbours classifier is also tested but without any amelioration.

Due to insufficient identification of benign tissue, neural networks are implemented to seek high specificity along with a good tumour recognition.

Feedforward NNs are investigated analysing their outcome according to several hidden layers structures using both aforesaid training sets. Although benign classes from dendrogram based training set ensure high specificity, tumour identification results worsened.

To further exploit the ability of the NNs to recognise a single label, consecutive networks are implemented in a cascade and the class featured with lowest misclassified number of benign elements as malignant is not included in the next NN training set.

Since only small changes are found, the final attempt tries to increment the presence of unseen benign samples, by performing dendrogram clustering on the healthy portion of test set belonging to the network with smallest number of false positives to construct new classes for healthy tissue, thus enforcing their representation. This conclusive method actually achieved a better performance in the recognition of benign samples, at the expenses of a weakened tumour detection.

This study outlines the complexity in designing a system enabling reliable tumour detection, accompanied by appropriate identification of healthy tissue.

As demonstrated by all these trials, the key is to create a training set which encompasses as much knowledge as possible, compensating for data diversity and difference in sample size relative to the two classes.

In conclusion, there exists a basis for bp-MRI protocol to be employed in prostate cancer detection, hoping this work will pave the way for finding an efficient strategy.

# 1 Introduction

## 1.1 Rationale

Prostate cancer is one of the most widespread malignant male pathologies [1]. In Europe, the standardised rates referring to 2012 concerning new diagnosis and deaths amount respectively to 96 and 18 cases per 100000 subjects [2]. Due to the continuous increase of the population and their ageing, the incidence rate associated with prostate cancer is set to increment in the next decades, along with a considerable growth of the costs related to treatment and diagnosis, even considering indolent cancer [3].

Nowadays, PCa diagnosis establishes that men beyond 50 years of age undergo PSA blood test and digital rectal examination during the urological visit. If some anomaly arises from either of them, the patient is usually subjected to random systematic transrectal ultrasound biopsy. Furthermore, many international research groups have proved the accuracy of multi-parametric resonance imaging to identify clinically significant PCa with 82% as positive predictive value and over 95% as negative predictive value to rule out high risk PCa [4,5]. Nevertheless, up to now the use of mp-MR in the clinical practice to diagnose subjects with PCa suspicion is limited by the elevated economic expenses and the prolonged time needed to perform the exam.

In this context, the research activity carried on by Candiolo Cancer Institute proposes the introduction of an alternative examination reserved to a biopsy-naïve population who may be affected by PCa. This new clinical path includes the so-called bi-parametric magnetic resonance, which differs from the mp-MR in the number of sequences (one less with respect to mp-MR), and the unemployment of both the endorectal coil and the intra-venous injection of a contrast agent. These characteristics result in:

- reduced exam duration (about 30 minutes less);
- decreased probability of complications and preparation time for the patients, who would be more prone to perform the exam;
- more rapid and less demanding readings for the radiologists, because of restricted number of images to analyse;
- diminished cost of the exam.

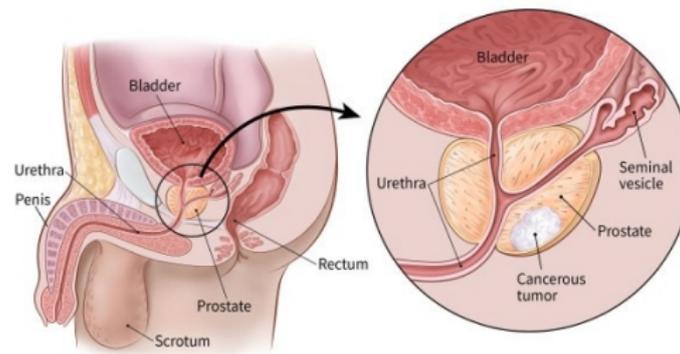
In addition, a Computer Aided Diagnosis system could be developed to automatically detect the presence of prostate cancer from bp-MR images. As a fast and user-independent software tool, it would provide a great aid to the expert radiologists by speeding up the diagnostic process and overcome intra and inter-subject variability.

In order to assess the efficacy of bp-MR in the diagnosis of PCa, a clinical study is performed, enrolling patients declared eligible after urological visit, who are randomised in two groups: in the former the diagnosis is established with bp-MR, while in the latter mp-MR is employed. All the participants are subjected to either random or fusion TRUS biopsy.

For additional information about the aforementioned study please refer to the clinical trial in [6].

## 1.2 Prostate Cancer

The *prostate* (also called *prostate gland*) is an organ belonging to the male reproductive system. It is located underneath the urinary bladder inside the pelvic body cavity and in front of the rectum. Its size is similar to the one of a walnut, but it tends to enlarge with ageing. Posteriorly to the prostate, there are the seminal vesicles (i.e. glands which produce a consistent part of the fluid composing semen), while the urethra passes throughout the prostate, as shown in figure 1.1 [7].



**Figure 1.1:** Representation of the prostate along with surrounding organs [7]

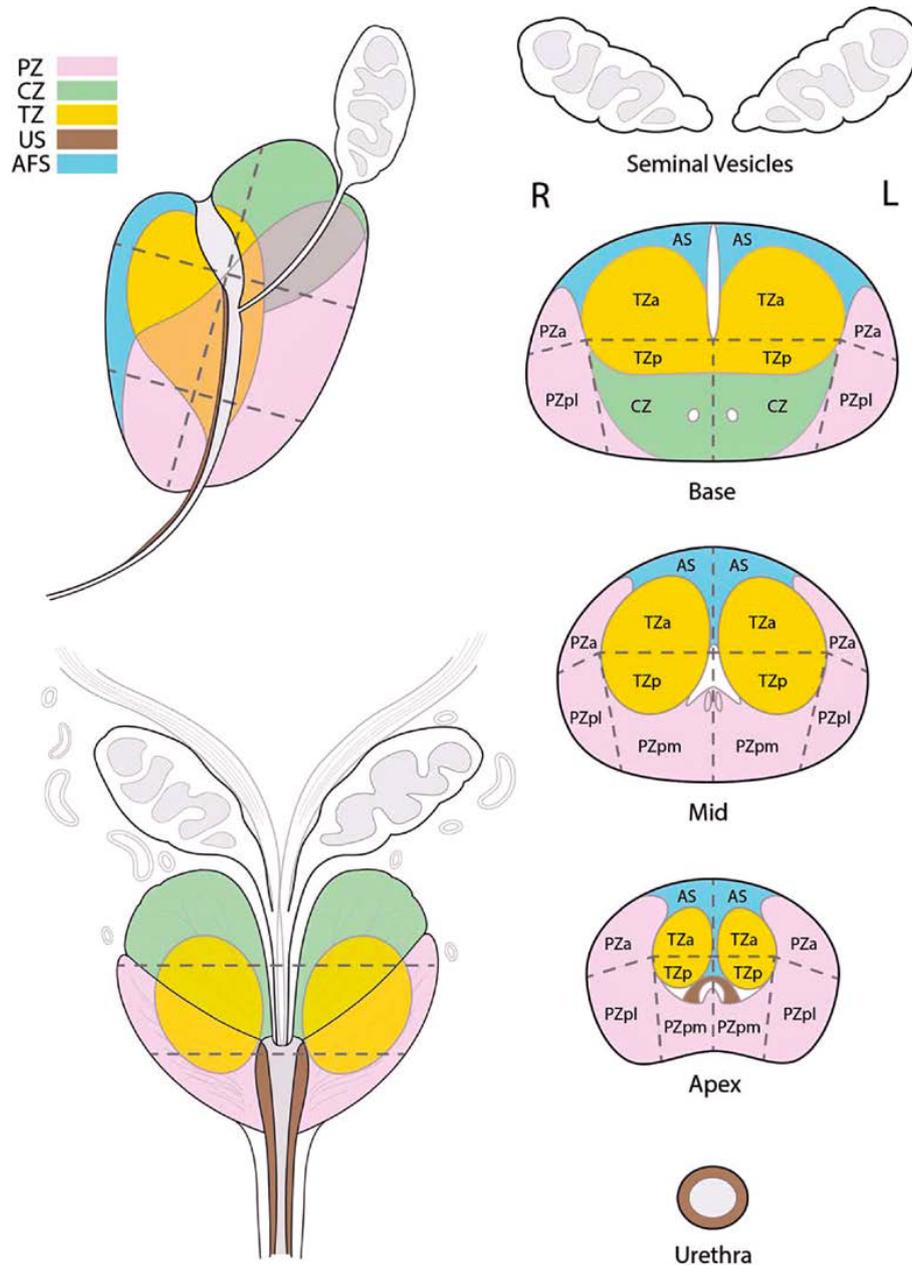
Referring to McNeal's nomenclature of the prostate [8], the following areas can be distinguished [9]:

- *peripheral zone*, comprising more than 70% of the glandular prostate, it is the most common site affected by cancer. This region contains the proximal urethral segment of the prostate, which lies between the verumontanum (i.e. portion in which the ejaculatory ducts enter the urethra) and the urinary bladder. In this tract, the preprostate sphincter is present;
- *central zone*, including about 25% of the glandular prostate, its structure is different than the peripheral one, in terms of size and shape of acini and ducts (much larger and with an uneven contour). It is less predisposed to diseases;
- *transitional zone*, located in the preprostate region, it amounts to only 5% of the total volume of the prostate, but its expansion can cause the urethra to be compressed and the bladder outlet to be obstructed;

- *anterior fibromuscular stroma*, which completely covers the anterior surface of the prostate, presenting a nonglandular composition.

Furthermore, three regions can be delineated considering the coronal plane: base, mid-land and apex.

A schematic representation of the aforesaid zones is given in figure 1.2.



**Figure 1.2:** Detailed illustration of prostate areas (PZ: peripheral zones, TZ: transition zones, CZ: central zone, AS: anterior fibromuscular stroma, SV: seminal vesicles, US: urethral sphincter, a: anterior, mp: medial posterior, lp: lateral posterior). The prostate is divided laterally by a vertical dashed line and into anterior/posterior by a horizontal dashed line passing through the centre of the gland [10]

Among benign pathologies affecting prostate, there are focal prostate atrophy, transition zone nodules, prostatitis, benign prostatic hyperplasia nodules concerning peripheral zone, and calcifications [11]. Instead, prostate cancer represents the most common malignancy in men.

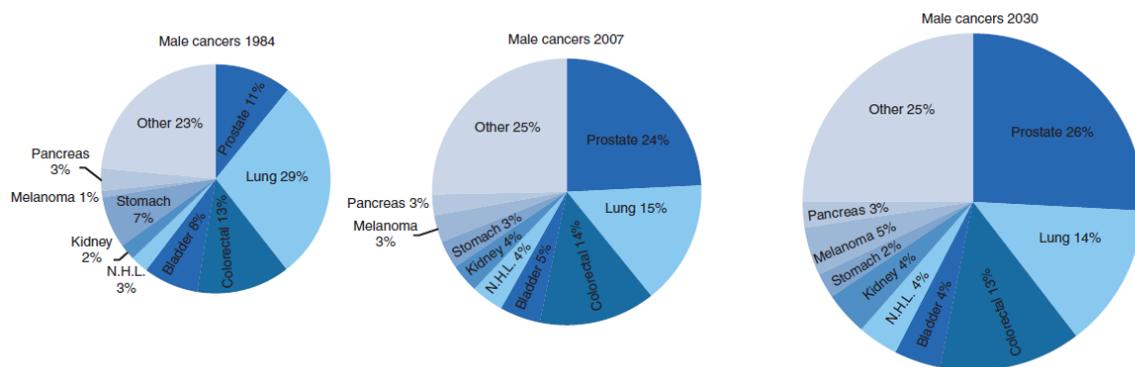
PCa is only second to skin cancer as most common cancer in American male population [12]. The diagnosis of prostate cancer is estimated to be about 1 man in 9, with higher probability in men aged 65 or older. As cause of death, it follows lung cancer with about 1 male subject in 41 who will die of PCa [12]. In Italy, about 1 man in 8 could be diagnosed with prostate cancer during his lifetime [13].

Among risk factors of this disease, the following can be mentioned [13]:

- *age*: very low probability before 40 years, while it increases after 50 years (about 2 tumours in 3 are found in men over 65 years);
- *genetic heritage*: the risk is doubled if a family member has had prostate cancer with respect to a subject with no family cases;
- *high hormonal levels*, such as testosterone and insulin-like growth factor (IGF)-1;
- *lifestyle related causes*, for instance obesity and lack of physical activity.

Race-ethnicity is also encountered as a possible risk factor [14].

Figure 1.3 displays the increment in the incidence of prostate cancer related to Great Britain, as an emblematic example [3].



**Figure 1.3:** Most widespread cancers in men considering 1984, 2007 and a prediction for 2030, showed using pie charts with areas proportional to cases number [3]

Prostate cancer advancement is described mainly by two variants [15]:

- *slow*: these tumours have a gradual development and they remain enclosed in the prostate. They represent around up to 85% of all prostate cancers [16]. Active surveillance can replace treatment in this condition;

- *fast*: their progression is accelerated and they soon reach other organs, creating metastases especially in the bones [17].

Being these two courses so diverse, it is pivotal to discriminate aggressive from slow-growing prostate cancer.

In order to determine the aggressiveness of PCa, a bioptic sample of the gland must be analysed. This enables to assess the *Gleason Score*, utilised to classify prostate cancer based on the quantity of tumour cells, on a scale from 1, lowest grade indicating that malignant tissue is very similar to healthy areas, to 5, when cancerous cells present abnormal aspect and growth. If there is GS bigger than 7, it designates the simultaneous presence of different levels of aggressiveness and it is given by the sum of the two grades, where the first grade refers to the predominant portion of the prostate [18].

Furthermore, PCa tends to evolve in specific areas of the prostate [19–21]:

- about 70 – 80% in the peripheral zone;
- 10 – 20% in the transition zone;
- around 5% in the central zone.

### 1.3 MRI for Prostate Cancer

Magnetic resonance imaging is utilised in the assessment of diseases concerning prostate gland, in light of the high-resolution images it offers [22]. In this section, insights about the MRI protocol employed in common clinical practice will be provided together with a viable alternative.

#### 1.3.1 Multi-parametric MRI

Multi-parametric magnetic resonance imaging has proven to be effective in the identification of malignant and benign pathologies of the prostate, due to the combined information retrieved by the multiple adopted modalities.

MR sequences generally used for prostate imaging are listed in the following:

- *T2-weighted*, yielding anatomical details, in conjunction with pelvic phased array coil and endorectal coil has been extensively exploited for local staging and diagnostic check-ups of PCa [23,24];
- *dynamic contrast enhanced*, created to quantify perfusion parameters to distinguish cancerous from healthy tissue [25,26], revealing tumour vascular architecture;
- *diffusion weighted*, able to examine diffusion of water molecules, which results lessened in dense cellular environments such as cancers [27].

*Magnetic resonance spectroscopy imaging* is another technique which investigates the level of certain metabolites (e.g. creatine, citrate, choline) and polyamines (e.g. spermidine, spermine and putrescine). According to these levels, cancerous areas can be recognised as confirmed by earlier studies [27,28].

Encouraging results infer the efficacy of multimodal MRI to detect PCa, combining morphological information from T2-w with functional modalities, such as DCE-MRI, DWI and/or MRSI. The improvements include enhancement of cancer localisation [29, 30] and volume assessment [31] in the peripheral zone.

The previously mentioned MR sequences can be used together in diverse combinations [32, 33], even though T2-w is often included to preserve morphological references.

Some research groups tried to differentiate low-grade from high-grade PCa starting from mp-MR featuring T2-w and DCE modalities together with MRSI, obtaining promising findings [34]. Applications of mp-MRI comprise also prostate cancer radiotherapy, to delineate healthy and unhealthy tissue and assess staging of the disease [35].

With the aim of making the interpretation of prostate MRI less biased by radiologists' subjectivity, the prostate imaging reporting and data system (PI-RADS) has been conceived to favour standardisation and reporting of MRI [10, 36].

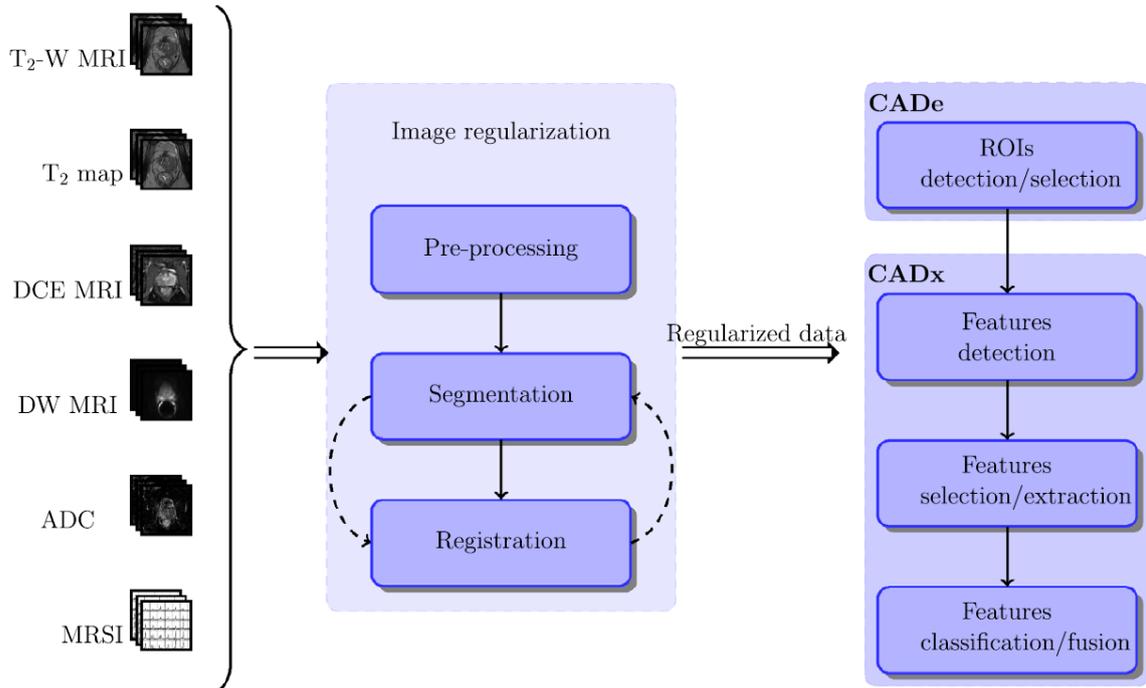
In spite of the aforementioned qualities, interpreting prostate MRI can be very challenging and often confounding also for experienced readers [11, 37]. This sanctioned the construction of software tools for supporting the clinicians in the diagnostic process, strengthened by the significant informative content retrievable from this imaging modalities.

To further consolidate the necessity of standardisation and agreement among expert radiologists, an advancement in this field is represented by the development of computer-aided diagnosis systems, able to provide quantitative and user-independent information about prostate cancer localisation and even staging. Among the main advantages, there are reproducibility and reduction of the time required to analyse MR scanning. A representative example of the architecture of a CAD system is given in figure 1.4.

For a systematic review on CAD systems the reader is referred to [38].

The major downside associated to mp-MRI is the use of the ERC coil and/or the injection of the intra-venous contrast agent, which could cause discomfort in the subjects, making them less willing to perform the exam. There exists also the chance some subjects are allergic to the contrast agent.

To overcome this issue, a non-invasive protocol called *bi-parametric MRI* has been designed and it is currently examined to understand whether it is suited as a diagnostic tool.



*Figure 1.4:* Detection of prostate cancer by means of a computer aided diagnosis system [38]

### 1.3.2 Bi-parametric MRI

Nowadays, bi-parametric magnetic resonance imaging is revealing to be appropriate for the detection of prostate cancer. It comprises non invasive MR sequences such as T<sub>2</sub>-weighted and DW modalities. Some preliminary studies show it provides similar performances among radiologists compared to multi-parametric MR [39–41].

In a recent work of Di Campli et al. [42], the diagnostic accuracy of a bp-MR protocol with respect to the traditional mp-MR was evaluated regarding identification of clinically significant cancer. Specifically, they proved there was no considerable disagreement within the performances of readers with different years of experience (7, 3 and 1 years) using the two sets of images. This result is promising as it underlines the validity of the bp-MR protocol, even in the case of clinicians with a minor level of expertise.

The research conducted in [43] utilised bi-parametric MR protocol in conjunction with PSA-density, discovering the greater performance of the former in detecting prostate cancer. They acquired images with a 3 T scanner from a biopsy-naïve population with clinical suspicion of PCa. They reinforced the potential of bp-MRI, especially before any previous biopsy, to play a crucial role in the diagnosis of prostate cancer, for it may spare invasive procedures and unneeded treatments.

The study of Niu et al. [44] examined texture features extracted from bi-parametric MRI for distinguishing high-grade PCa, proving its feasibility.

These preliminary findings suggest bp-MR protocol may encompass the knowledge necessary to differentiate benign from malignant tissue belonging to the prostate.

## 2 Prostate Segmentation

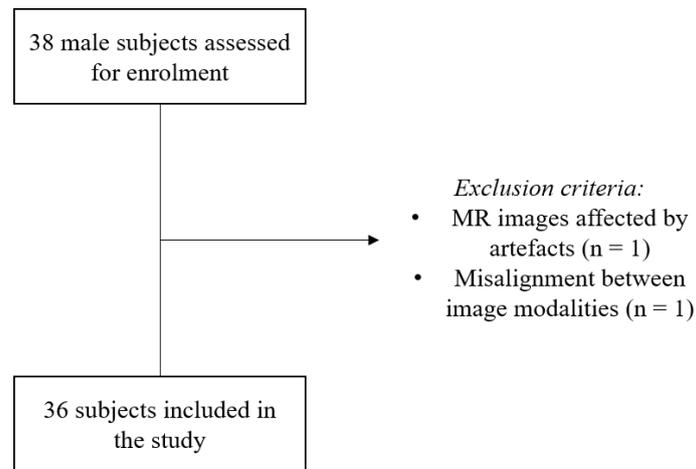
### 2.1 Patients

In order to be eligible for enrolment, male participants complied with these requisites:

- age equal or inferior to 75 years;
- no previous prostate biopsy (i.e. biopsy naïve);
- negative findings in urological visit;
- PSA level smaller than 15 *ng/ml* (found to increase in at least two successive tests).

All patients waived written informed consent to enter this clinical trial.

A schematic representation illustrating the study population for prostate segmentation is presented in figure 2.1.



**Figure 2.1:** Schematic representation of the study population for prostate segmentation

Table 2.1 details the main characteristics of the included subjects.

<b>Patients</b>	<b>All (n = 36)</b>
Age	68.4 (52.9 – 79.1)
PSA ( <i>ng/ml</i> )	6.2 (3.2 – 16)

**Table 2.1:** Characteristics of the study population. Measurements are given as *median (range)*

## 2.2 MR Image Acquisition

Images were acquired using a 1.5 T scanner (OptimaMR450w, GE Healthcare, Milwaukee, Illinois, USA), adopting the setting in table 2.2.

Sequence	ST (mm)	FOV (cm)	NEX	AM (pixels)	RM (pixels)	Additional information
T2-w (SE)	3	16x16	16	256 × 192	512 × 512	
DW (EPI, SE)	3	20x20	1	128 × 100	256 × 256	b-values: 0 – 1000 s/mm <sup>2</sup>

**Table 2.2:** MRI protocol for T2-weighted and diffusion weighted modalities (ST: slice thickness, FOV: field of view, NEX: number of excitations, AM: acquisition matrix, RM: reconstruction matrix, SE: spin echo, EPI: echo planar imaging)

Imaging parameters were in accordance with the scanning requirements for prostate imaging [45].

Since the dimensions of T2-weighted and diffusion weighted images differ, DW images are subjected to up-sampling taking as reference T2-w modality. By doing so, T2-w and DW images present matching dimensions (512 × 512 pixels per 24 slices) and they will be employed for all the succeeding elaborations.

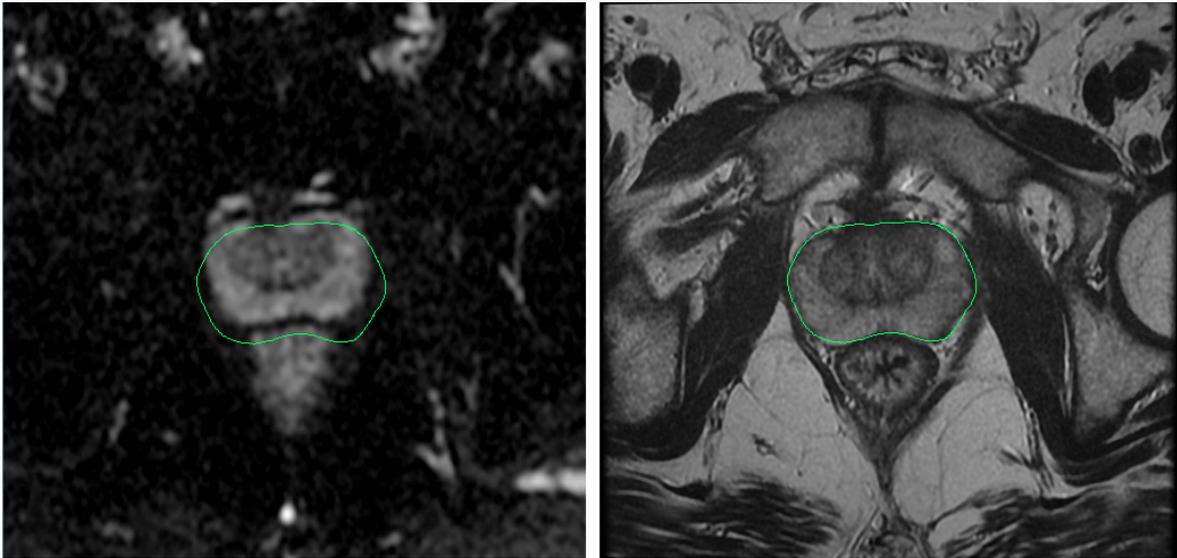
The apparent diffusion coefficient map is calculated pixel-wise by means of the monoexponential model [46, 47] in equation 2.1:

$$ADC_i = \frac{1}{b} \times \ln\left(\frac{S_0}{S_b}\right) \quad (2.1)$$

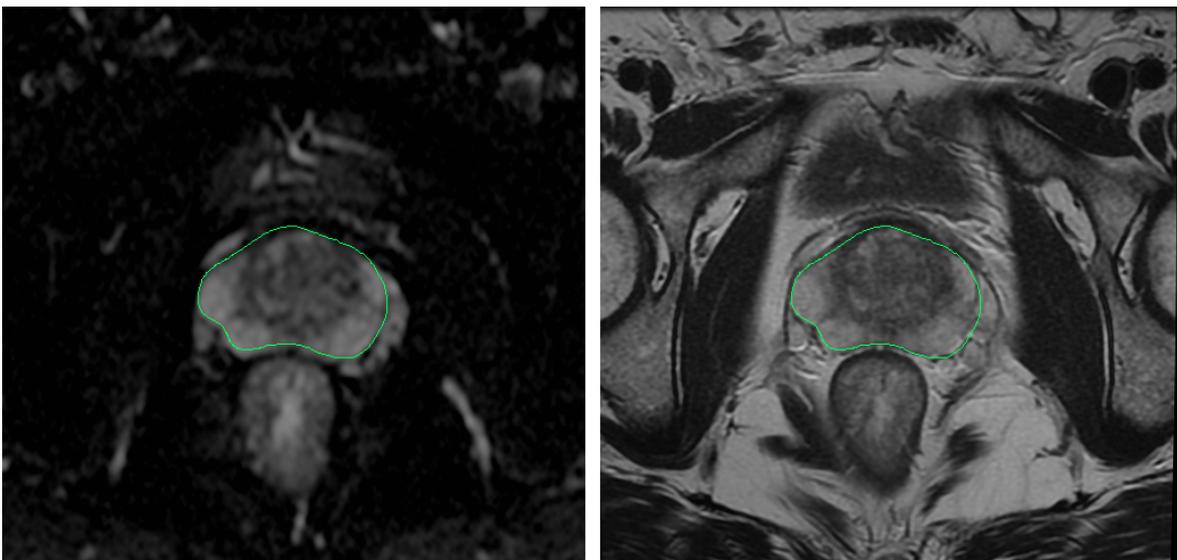
where  $b = 1000 \text{ s/mm}^2$ ,  $S_0$  is the pixel value from DW image with  $b = 0 \text{ s/mm}^2$ ,  $S_b$  is the pixel value from DW image with  $b = 1000 \text{ s/mm}^2$  and  $\ln$  indicates the natural logarithm.

To guarantee the alignment between the two image modalities, every patient's set of images was individually evaluated by segmenting the prostate in one sequence and projecting it to the other.

Referring to the study population, no considerable misalignment existed, leading to the un-employment of a registration procedure. Thanks to the absence of the endorectal coil, included in the mp-MR protocol, there is no substantial deformation, also enforced by the fast image acquisition. In fact, only one patient has been excluded due to misalignment after the re-sampling procedure, very likely caused by motion artefacts during the acquisition (see figure 2.2).



(a) Patient presenting misalignment



(b) Patient with no misalignment

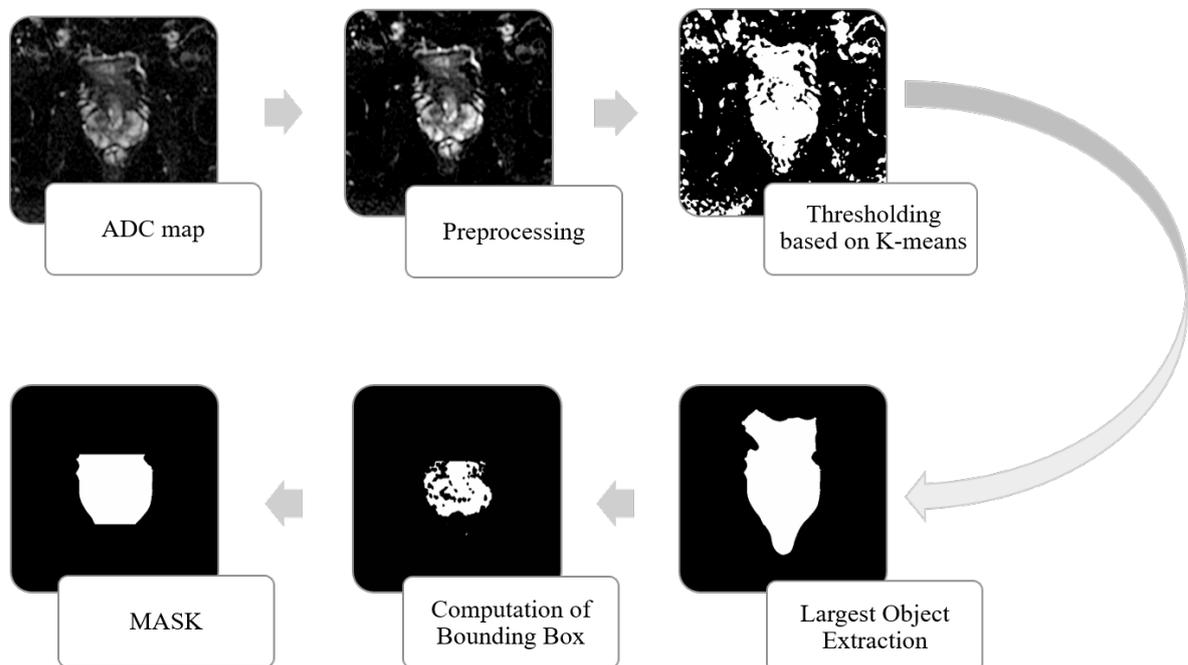
**Figure 2.2:** Representative examples to visualise the problem of misalignment between image modalities

## 2.3 Algorithm

Prostate segmentation is a mandatory step to focus the analysis on the region of interest, i.e. the prostate.

The segmentation algorithm proposed in this work has the aim of automatically isolating the prostate gland from the other biological structures present in the images. Essentially, it is based on the location of the prostate known to be approximately in the centre of each image. The algorithm has been developed using in-house C++ software along with ITK open source libraries [48].

A schematic representation of the procedure here employed is displayed in figure 2.3.



*Figure 2.3:* Phases of segmentation algorithm

### Preprocessing

The input given to the algorithm is constituted by the ADC map (computed as in equation 2.1), which is featured with more homogeneous background and less tissue diversity compared to T2-weighted modality.

Due to the re-sampling procedure, each image in the two sequences presents the same size ( $512 \times 512$  pixels per slice), hence it is equivalent referring to the horizontal dimension ( $x$ ) or vertical dimension ( $y$ ).

The preprocessing phase consists of a contrast adjustment to enhance the most numerous grey tones in the image, by windowing between the 50th and the 99th percentile of pixel intensities (see figure 2.4). Consequently, the biological structures are preserved, whereas the background becomes more uniform.

Furthermore, grey tones are rescaled between 0 and 255.

The main advantage of such steps is to create a common ground to begin with for all images, trying to lessen the variety of characteristics proper of each single patient.

### Thresholding based on K-means

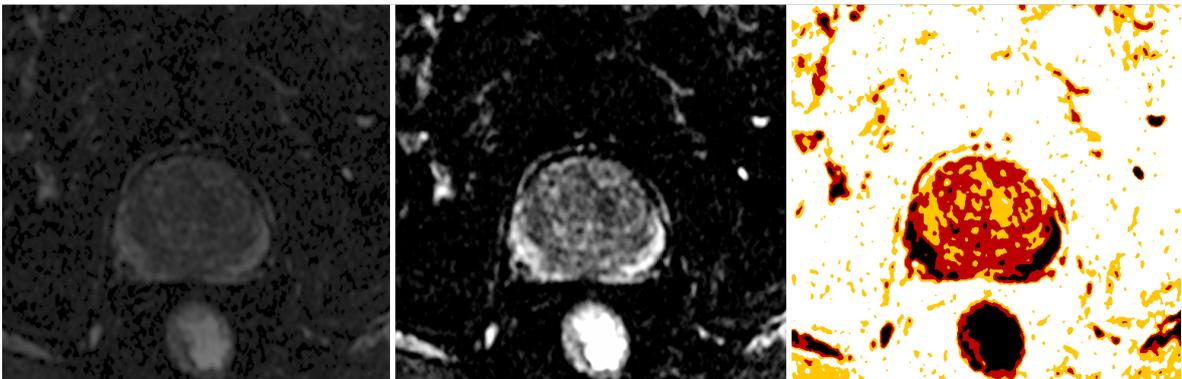
Referring to [49] as an insight, they used Fuzzy-C-Means to smooth inhomogeneities of the MR field. For the purpose of this study, *k-means algorithm* appears suitable to distinguish the prostate, by clustering pixels on the basis of their intensity value.

Choosing  $K = 4$  as number of clusters revealed to be experimentally the most appropriate choice in view of next steps. In particular, the following areas can be differentiated:

- background;
- edges of present structures;
- inner part of present structures;
- outer part of present structures (excluding edges).

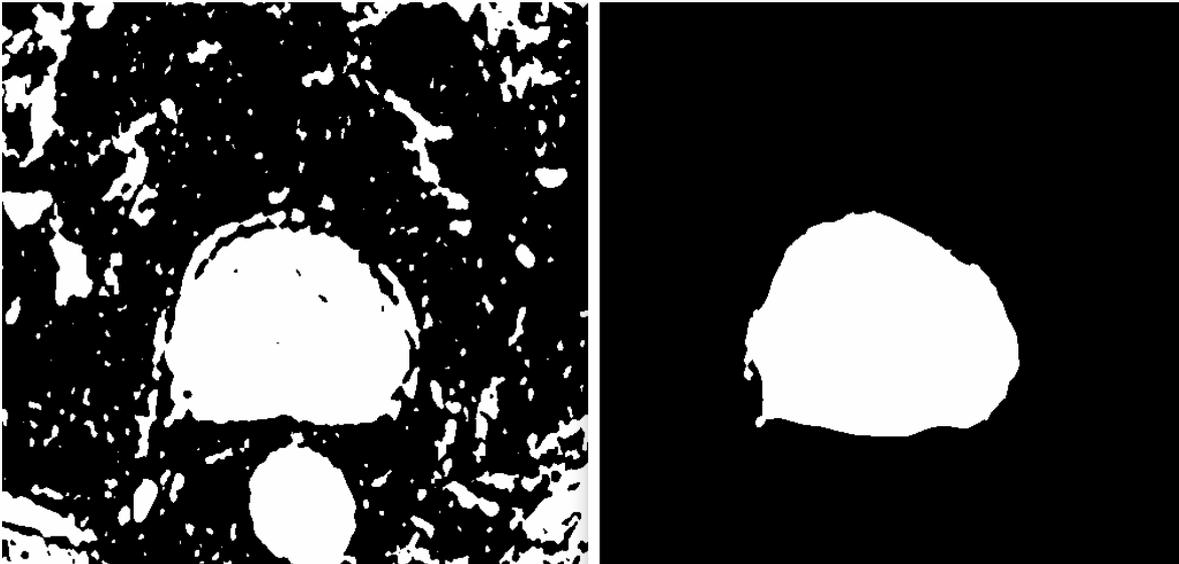
The attribution between each cluster and the aforementioned zones aids to identify macroscopic references, in the attempt to simplify the more variegate reality.

Figure 2.4 shows the images obtained with the preprocessing and the output of k-means algorithm.



**Figure 2.4:** From left to right: ADC map given as input, image resulting from preprocessing, image representing k-means pixel clustering

The outcome of the thresholding applied on the k-means pixel clustering is presented in figure 2.5, so that the background is not considered.



*Figure 2.5:* On the left: result of thresholding applied on output of k-means algorithm. On the right: image obtained after largest object extraction

### Largest Object Extraction

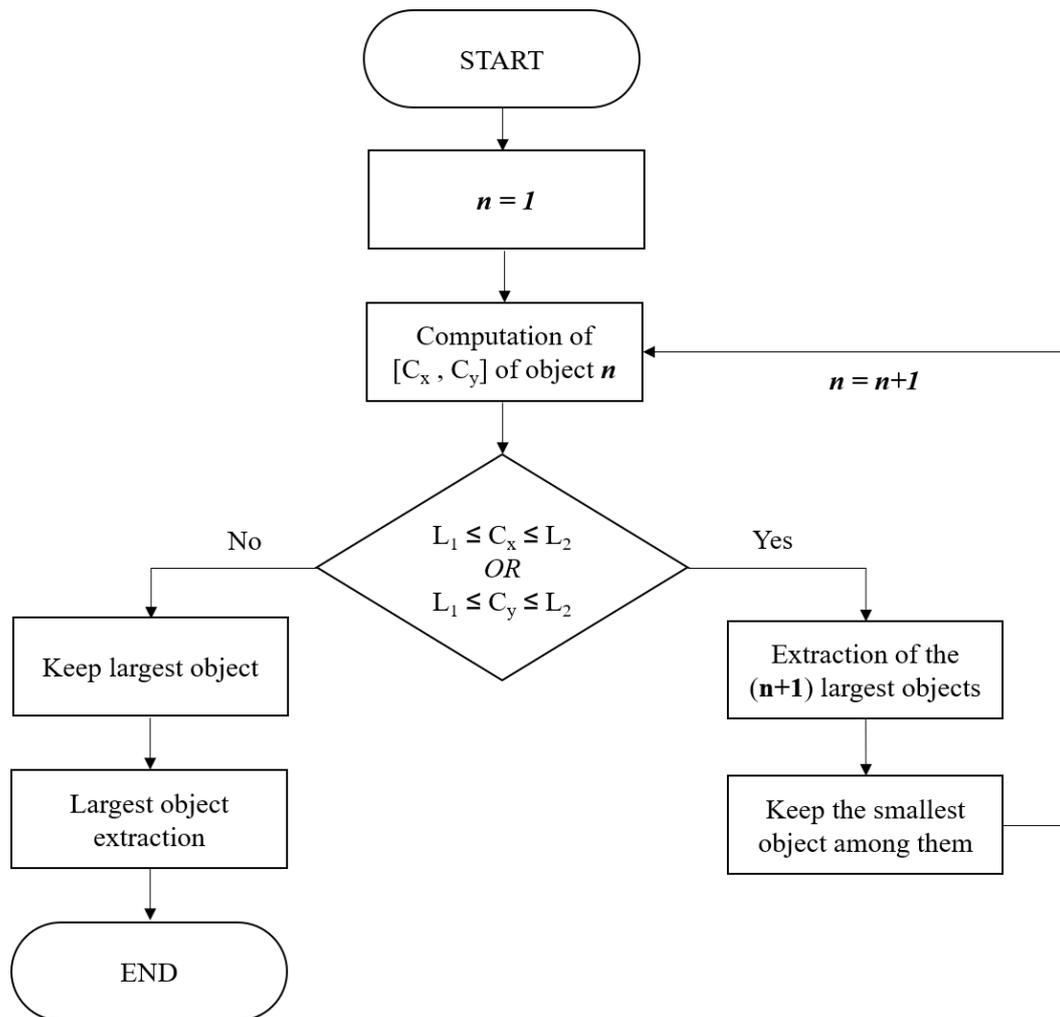
The largest object in each slice is extracted according to the procedure in figure 2.6 based on centroid coordinates. An *object* is intended as a group of joint pixels, creating an area of a certain extension.

After computing centroid coordinates of the first largest object, they are checked to be in the centre of the image, i.e. between the 25% and 75% of the slice  $x$  and  $y$  dimensions. This range is taken into account because of the variability of prostate size among patients and the possibility of having part of bladder or rectum contained in the area, thus altering the coordinates.

If the first object does not satisfy the condition, then the coordinates of the second largest object are tested and so on until a maximum of 4 iterations. This can be explained by the presence of objects at the corners of the image, which compete in terms of number of pixels with the central object, especially in the final slices of the volume (due to the restriction of prostate diameter).

In order to avoid the presence of holes inside the objects, a hole filling operation with a disk of radius 7 pixels on both  $x$  and  $y$  dimension and width of 1 pixel along the  $z$  dimension (intended as direction to scan the slices) is performed.

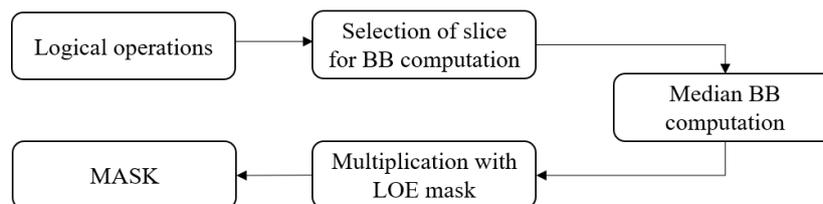
An example of the outcome of the largest object extraction procedure is displayed in figure 2.5, and from now on it will be addressed to as *LOE mask*.



**Figure 2.6:** Flowchart for retrieving largest object in each slice based on centroid coordinates ( $[C_x, C_y]$ : centroid coordinates respectively along  $x$  and  $y$ ,  $L_1$ : 25% of horizontal image dimension,  $L_2$ : 75% of horizontal image dimension,  $n$ : objects counter)

### Computation of Bounding Box

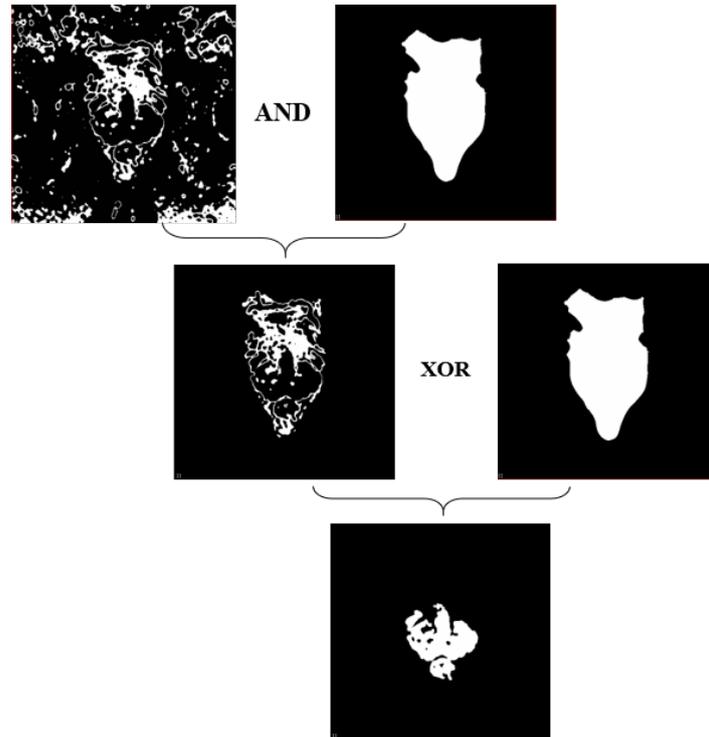
It is fundamental to underline that the *LOE mask* by itself cannot ensure the limitation of biological structures other than the prostate. To solve this issue, a bounding box (abbreviated as BB) enclosing the prostate is computed with the schematic representation provided in figure 2.7.



**Figure 2.7:** Phases of bounding box computation

Primarily, the undermentioned *logical operations* are performed (see figure 2.8):

- *AND*, between the *LOE mask* and the cluster from k-means algorithm containing only the edges of present structures;
- *XOR*, between the result of the previous operation and the *LOE mask*.



**Figure 2.8:** Logical operations performed beginning from k-means cluster comprising the edges of present structures (on the upper left) and *LOE mask* (on the upper right)

Ultimately, the largest object in each slice is extracted to obtain the final image, called *XOR mask*, used to designate the slice to be taken as reference for computing bounding box coordinates.

Bounding box coordinates will be indicated as in equation 2.2:

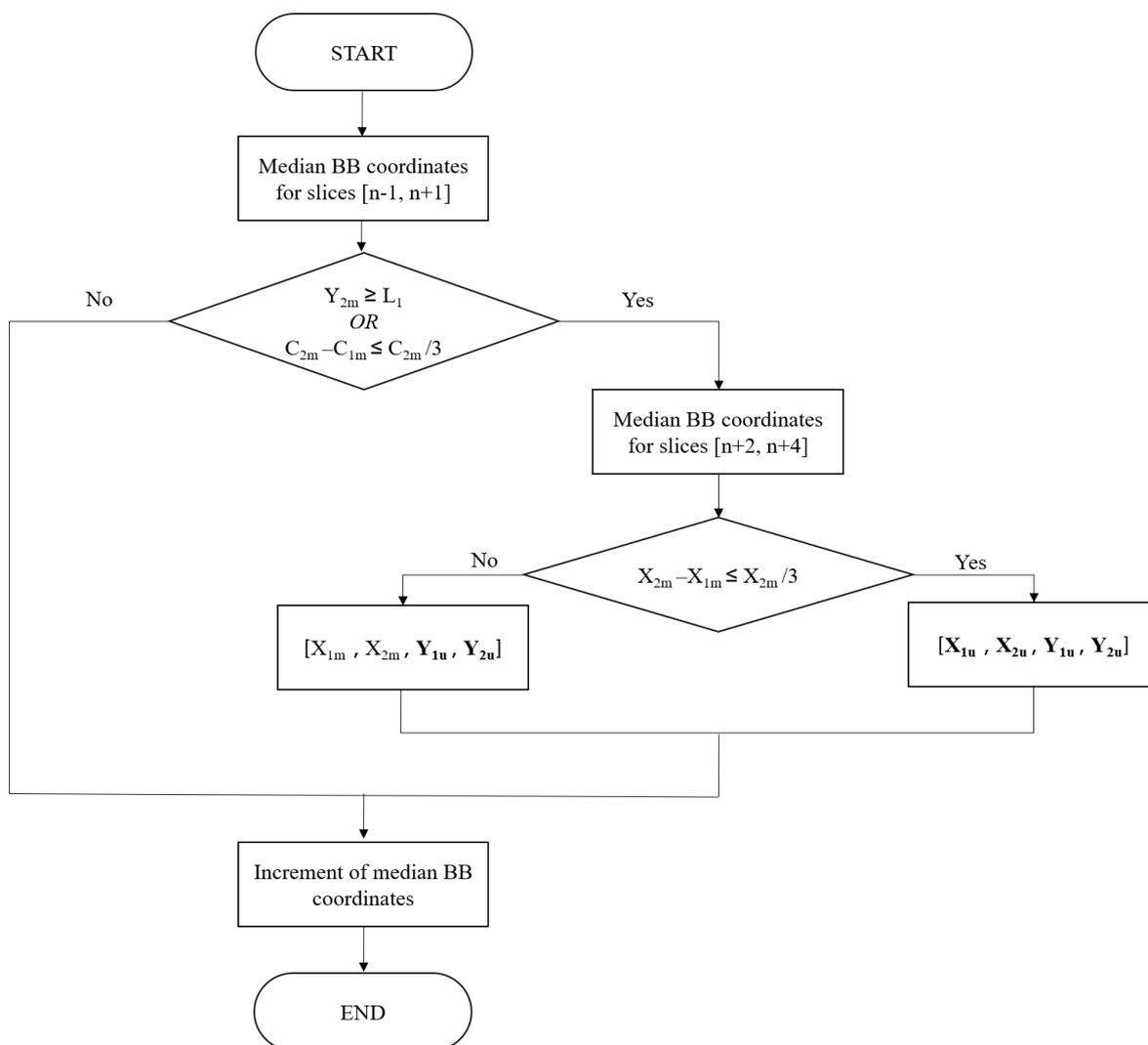
$$[X_1, X_2, Y_1, Y_2] \quad (2.2)$$

where  $X_1$  and  $X_2$  are respectively minimum and maximum  $x$  coordinates, while  $Y_1$  and  $Y_2$  are respectively minimum and maximum  $y$  coordinates.

The *selection of the slice eligible for bounding box computation* consists of ascertaining that the minimum  $y$  coordinate of the current BB is beyond 25% of image dimension ( $L_1$ ):

$$Y_1 \geq L_1 \quad (2.3)$$

The first slice to be tested is the central slice in the volume, since it certainly includes the prostate. If the condition in equation 2.3 is not satisfied, then next slice is checked until the acceptable one is found.



**Figure 2.9:** Flowchart for retrieving median bounding box coordinates ( $L_1$ : 70% of vertical image dimension,  $C$  can be either  $X$  or  $Y$ , subscript  $m$  indicates median coordinates considering slices  $[n - 1, n + 1]$ , while subscript  $u$  indicates median coordinates considering slices  $[n + 2, n + 4]$ )

With the goal of finding the bounding box which most faithfully contains the prostate, the method in figure 2.9 is implemented.

At the beginning, median coordinates considering slices in the range  $[n - 1, n + 1]$  are computed, where  $n$  corresponds to the number of the slice eligible for bounding box computation. Thereupon, the following two conditions are tested:

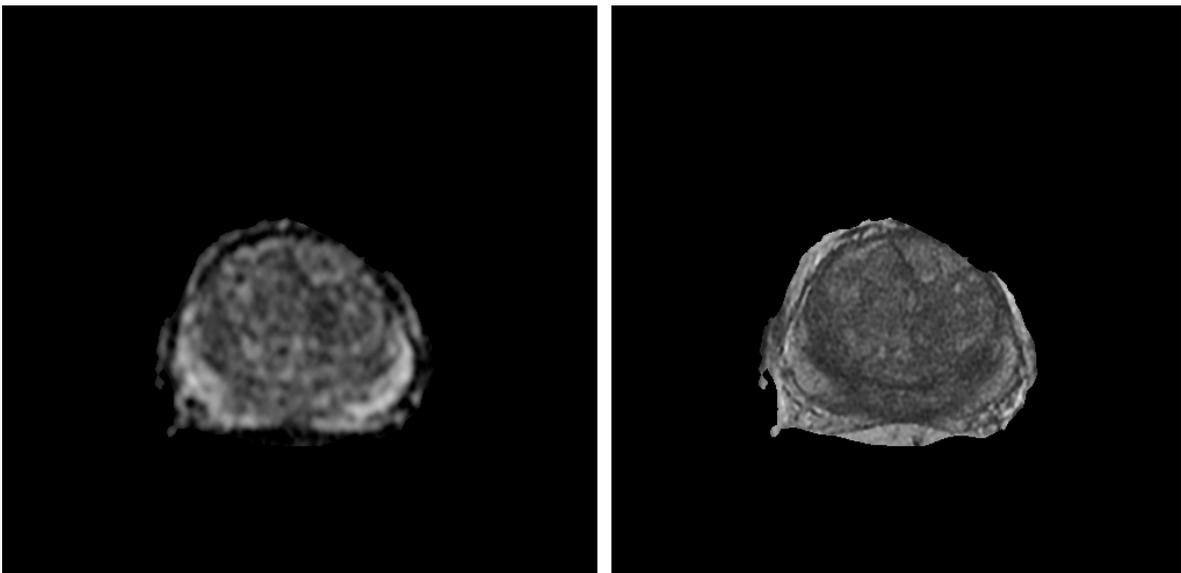
- minimum  $y$  coordinate beyond 70% of vertical image dimension, to ensure the rectum is not included;
- difference between maximum and minimum  $x$  coordinate being smaller than one third of the maximum  $x$  coordinate, to ascertain prostate diameter is within the expected range (same condition is applied for  $y$  coordinate).

If either one of the previous statements is verified, slices in the range  $[n + 2, n + 4]$  are considered to calculate median BB coordinates. As a further control point, the condition regarding the difference between maximum and minimum  $x$  coordinate being smaller than one third of the maximum  $x$  coordinate is tested: if it is true, all four coordinates are updated taking into account median coordinates of current slices, otherwise only minimum and maximum coordinates along  $y$  are modified. In such manner, the diameter of the prostate along the horizontal dimension is preserved, whether no anomaly arises from former checks. In fact, moving toward the final slices, the diameter is restricted as the whole prostate is scanned.

An increment of 40% and 30% respectively along  $x$  and  $y$  coordinate with respect to bounding box area is added to the obtained BB. This is a preventive measure, not to miss any pixel belonging to the prostate, although there is the chance to segment tissue outside the prostate itself.

The ultimate mask is given by the multiplication of the bounding box image (i.e. mask with pixel equal to 1 in BB area) with the *LOE mask*.

An example of the outcome resulting from the segmentation algorithm is reported in figure 2.10.



*Figure 2.10:* From left to right: final mask on ADC map and T2-w image

### 2.3.1 Reference Standard

Manual segmentation of the prostate has been provided slice by slice on T2-weighted and ADC map for each patient by an expert radiologist and will be used to validate the automatic masks.

### 2.3.2 Statistical Analysis

The statistical analysis to estimate the performance of the segmentation algorithm is focused on the following metrics, obtained from the confusion matrix in table 2.3:

- *recall* (also called *sensitivity*), establishing the quantity of pixels belonging to the prostate actually segmented [50]:

$$Recall = \frac{TP}{TP + FN} \quad (2.4)$$

- *precision*, assessing the proportion of the automatic segmentation which is comprehended in the reference mask [51]:

$$Precision = \frac{TP}{TP + FP} \quad (2.5)$$

- *Dice Coefficient*, measuring the proportion of the overlapping region considering reference and automatic segmentation [52, 53]:

$$DC = \frac{2TP}{2TP + FP + FN} \quad (2.6)$$

These metrics are all unitless, ranging from 0, meaning worst-case scenario (no pixel in common) to 1, i.e. ideal segmentation (corresponding to the ground truth).

		<i>True class</i>	
		<b>T</b>	<b>P</b>
<i>Predicted class</i>	<b>T</b>	TN	FN
	<b>P</b>	FP	TP

**Table 2.3:** Confusion matrix used to compute metrics (P: label for prostate, T: label for tissue other than prostate, TP: true positive, FP: false positive, TN: true negative, FN: false negative)

Every metric is computed slice by slice for both T2-w and ADC masks.

## 2.4 Validation

The segmentation algorithm proposed in this study shows high sensitivity across all slices belonging to each patient. This is very crucial, since for the purpose of classification it is essential not to leave out any pixels proper of cancerous tissue.

Median values and range for the computed statistics are provided in table 2.4.

<b>Image type</b>	<b>DC</b>	<b>Recall</b>	<b>Precision</b>
<i>ADC map</i>	0.74 (0.32 – 0.94)	1.00 (0.94 -1.00)	0.59 (0.19 – 0.88)
<i>T2-w</i>	0.72 (0.27 – 0.93)	1.00 (0.97 -1.00)	0.56 (0.16 - 0.88)

**Table 2.4:** Metrics to evaluate segmentation performances considering the entire set of patients. Measurements are given as *median (range)*

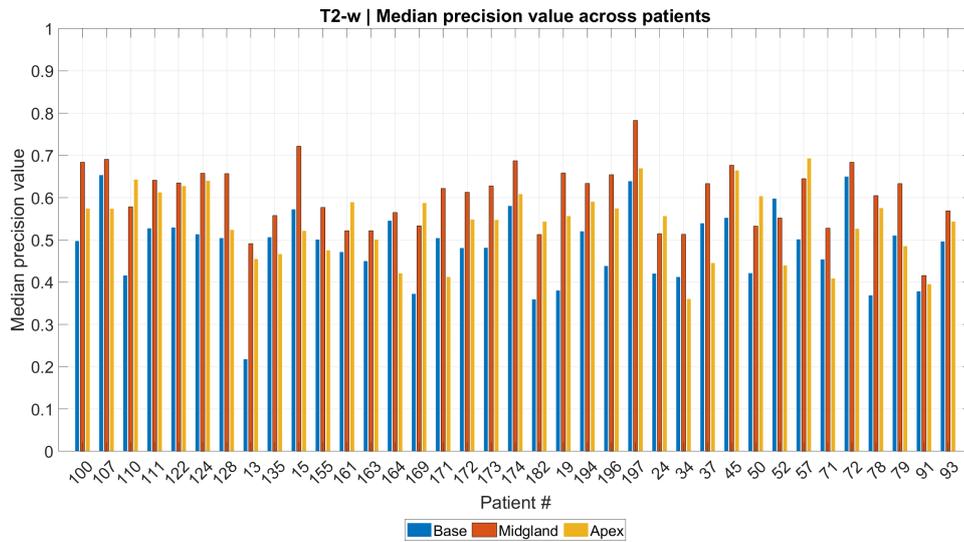
Whilst the range of sensitivity is limited to a very small interval, precision and Dice coefficient reach a much broader series of values. Moreover, the median precision value yields the lowest performance.

Dice coefficient values retain a median performance across all slices around 0.7 for both T2-w and ADC map.

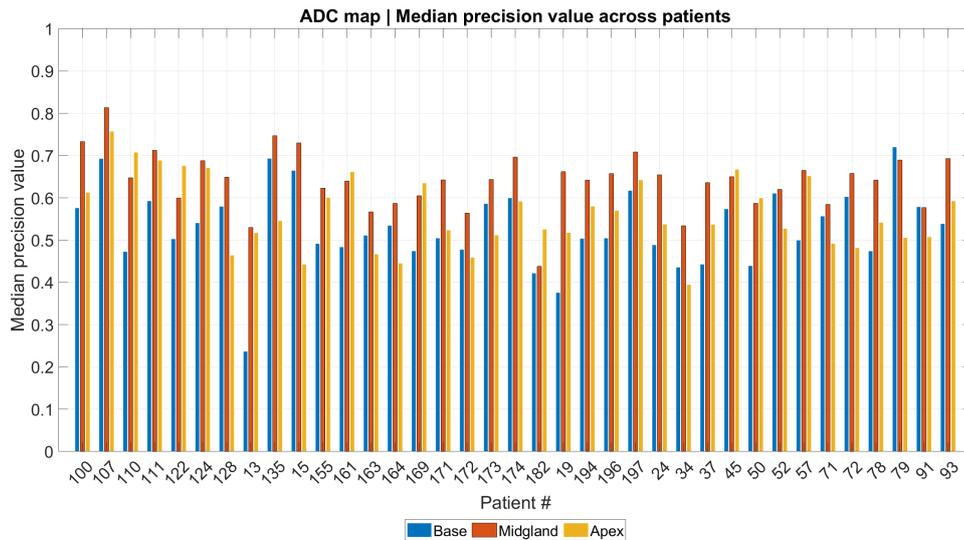
In order to better comprehend the reason behind poor performance regarding precision, figure 2.11 provides the median precision value for each patient, whose slices have been divided in base, midgland and apex.

As observable from figure 2.11, the majority of subjects presents the highest value of precision, mostly above 0.6, in the slices corresponding to the midgland. As a matter of fact, the computation of the bounding box takes as reference one of the slices surely belonging to the central part of the prostate, thus achieving better results.

Furthermore, the increment of the bounding box area contributes to diminish the precision, especially considering the extremities of the volume.



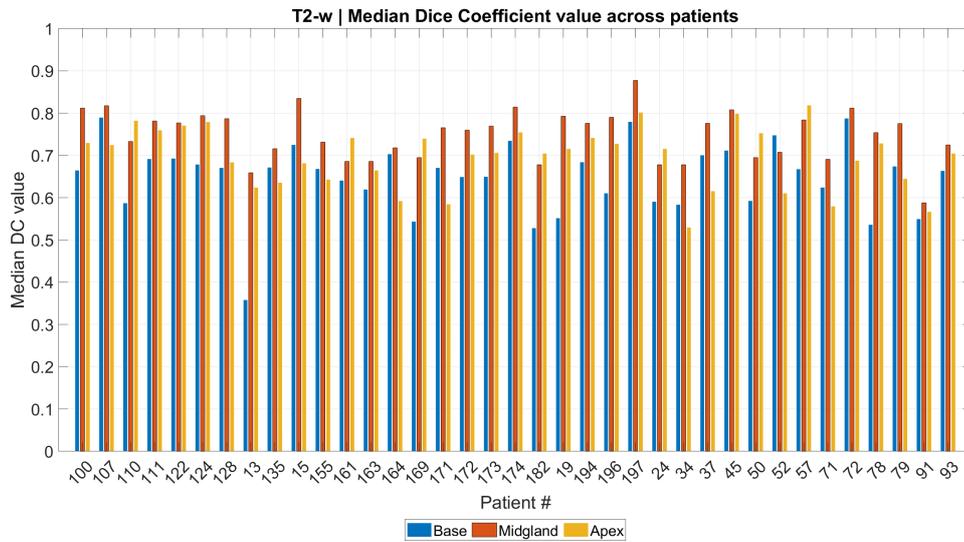
(a) T2-weighted



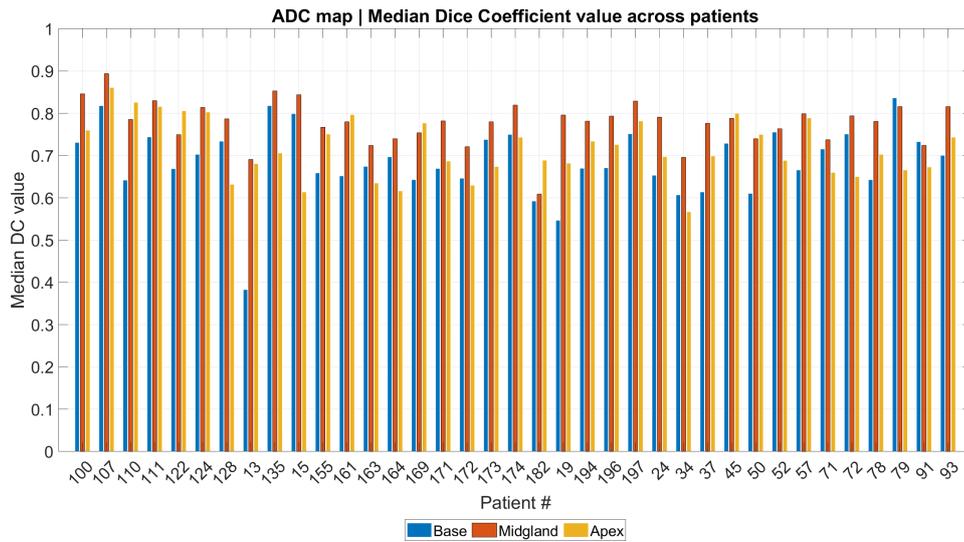
(b) ADC map

**Figure 2.11:** Median value of precision considering base, midland and apex for each patient

The separation of the three zones of the prostate presents better performances also considering median Dice coefficient in the midland, as visible from figure 2.12.



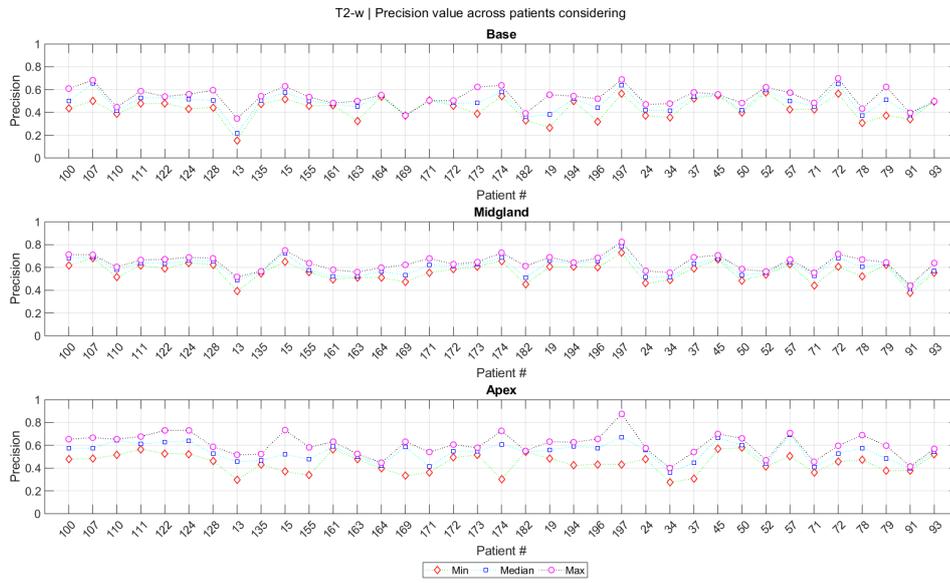
(a) T2-weighed



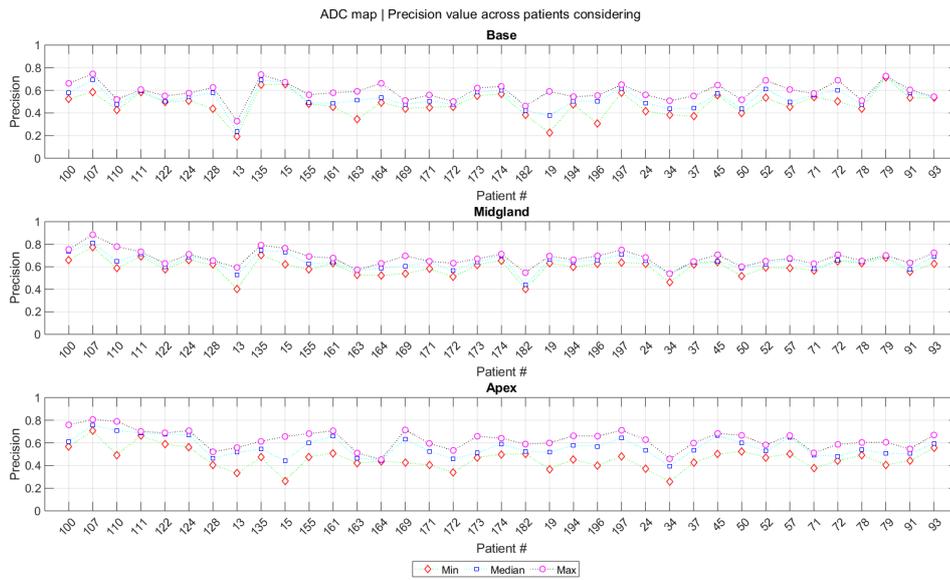
(b) ADC map

**Figure 2.12:** Median value of Dice coefficient considering base, midglad and apex for each patient

As evident from figures 2.13 and 2.14, both Dice coefficient and precision are characterised by a wider range in the area proper of the apex. This could be explained by the fixed dimension of the bounding box, which is much greater with respect to the dimension of the prostate in the final slices of the volume.

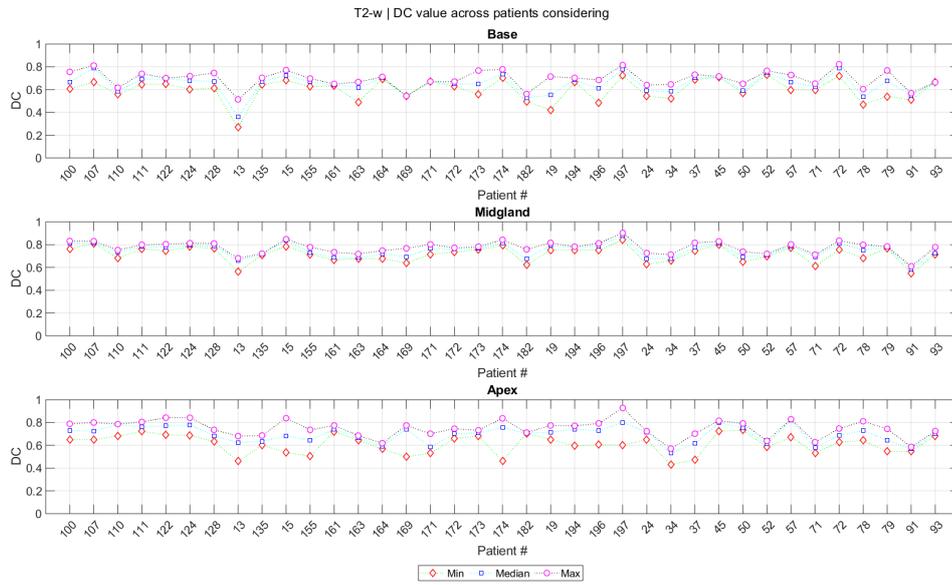


(a) T2-weighed

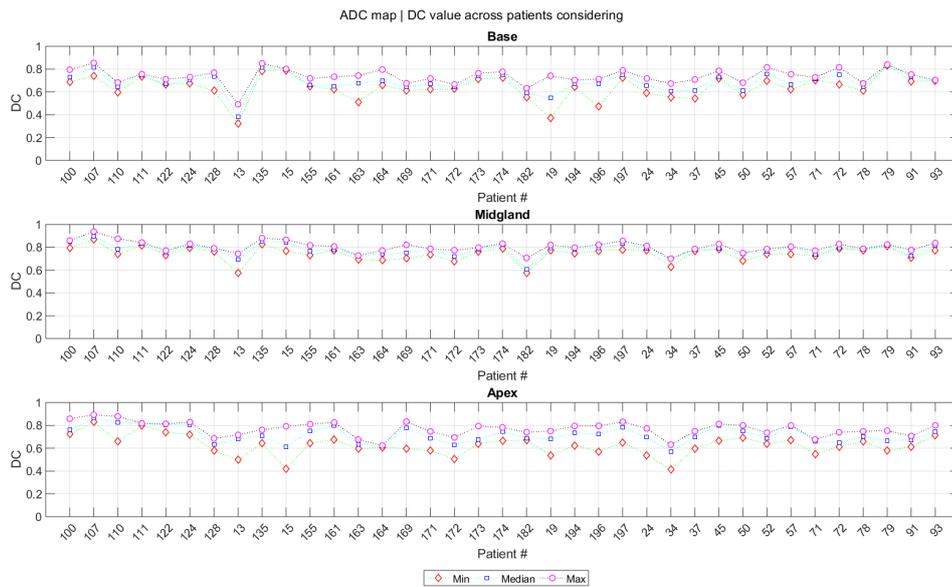


(b) ADC map

**Figure 2.13:** Minimum, median and maximum value of precision considering base, midglad and apex for each patient



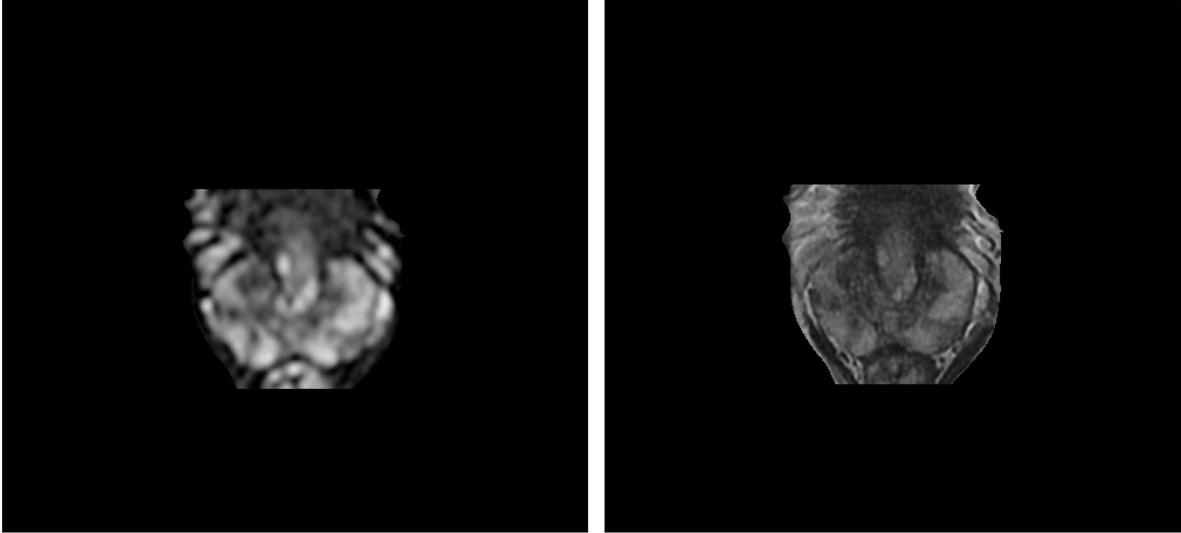
(a) T2-weighed



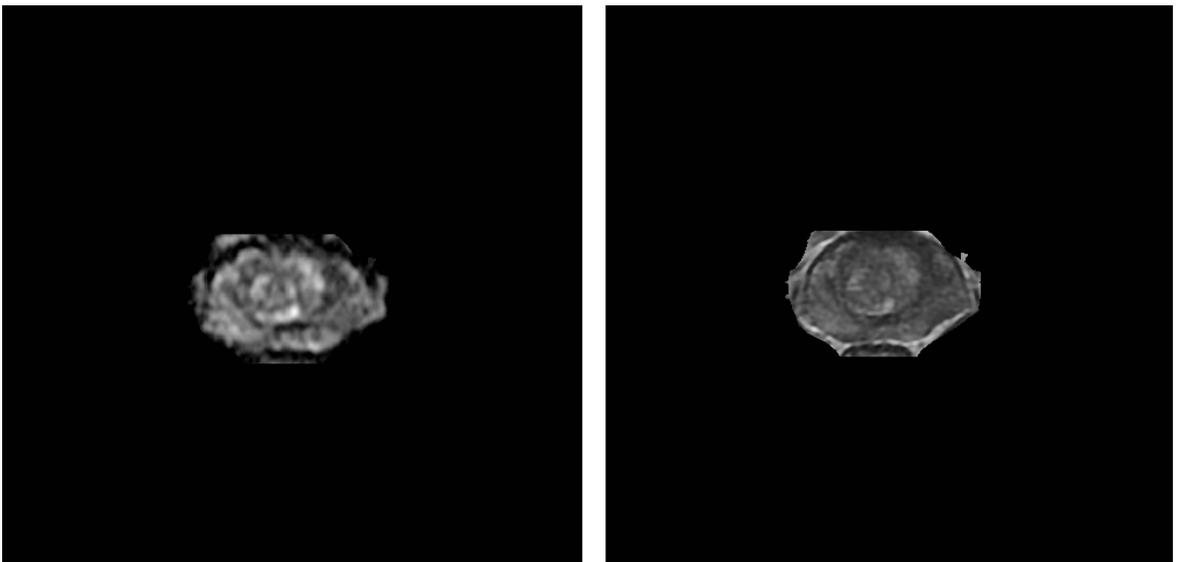
(b) ADC map

**Figure 2.14:** Minimum, median and maximum value of Dice coefficient considering base, midgland and apex for each patient

Figure 2.15 exhibits segmentation performances on two subjects, showing a case of more and less pronounced oversegmentation.



(a) Example of considerable oversegmentation



(b) Example of slight oversegmentation

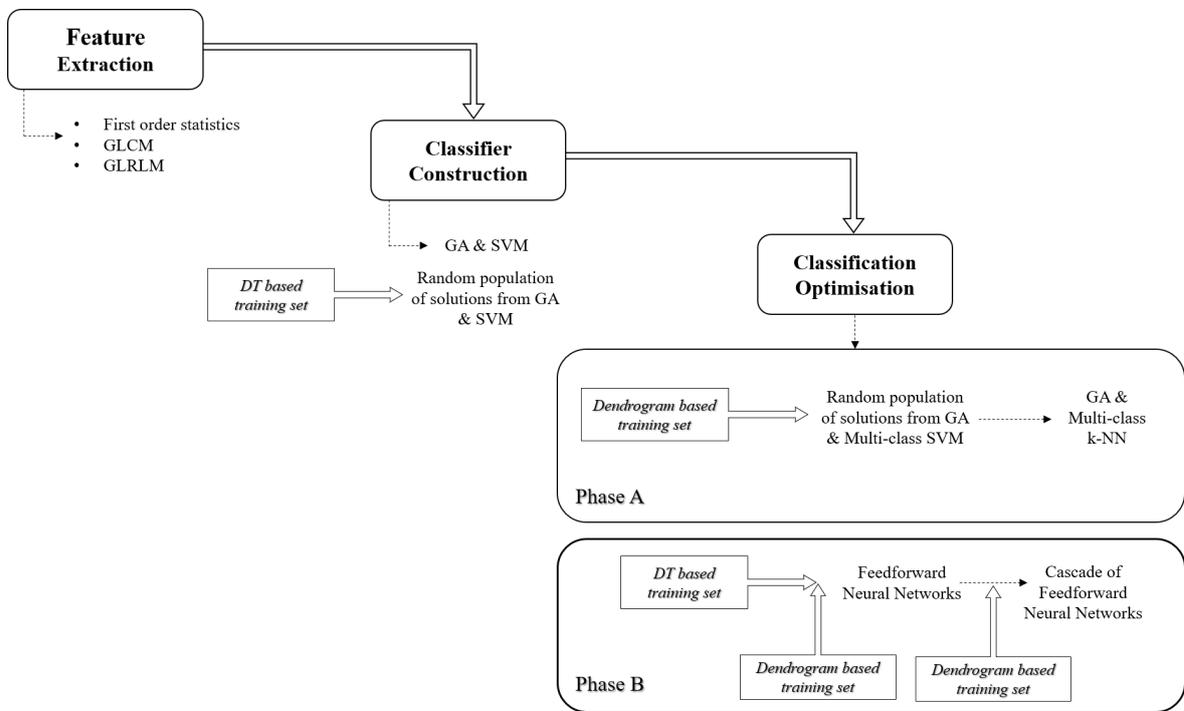
**Figure 2.15:** Outcome of the segmentation algorithm. On the left, prostate mask applied on ADC map. On the right, prostate mask applied on T2-w

In spite of the presence of pixels beyond the prostate, especially in the slices outside the central part of the gland, the automatic masks guarantee segmentation of the whole prostate, assuring complete inclusion of healthy and cancerous tissue, thus being suitable to be employed in the next elaborations.

# 3 Machine Learning for Tumour Detection

Tumour detection is known to be one of the most complex and challenging tasks in biomedical engineering. This study proposes to investigate diverse classification techniques to locate malignant lesions on behalf of the prostate.

Figure 3.1 provides a concise scheme to illustrate the elaboration process, which is to be delineated in details throughout this chapter.



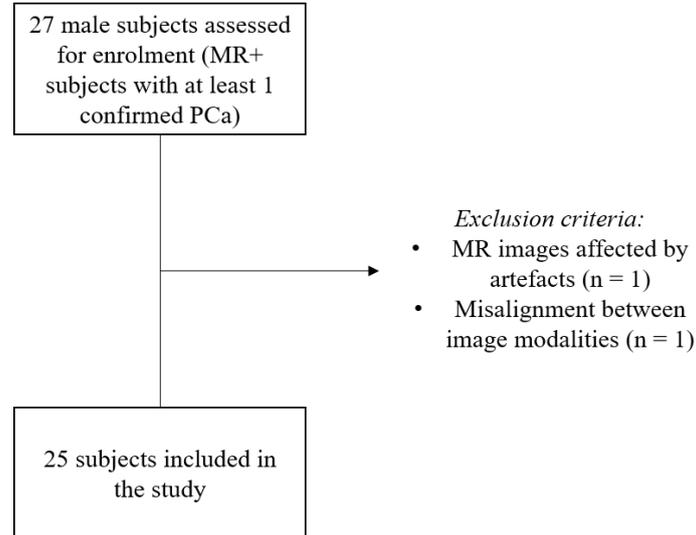
**Figure 3.1:** Schematic representation highlighting every step implemented for tumour detection

All the passages here described are applied separately to the two image types (ADC map and T2-w).

The elaborations for the tumour detection phase have been implemented with MATLAB Software, Release 2018b (The MathWorks, Inc., Natick, Massachusetts, United States).

### 3.1 Patients

The study population involved in the tumour detection phase is shown in figure 3.2. They represent the subset of the subjects in section 2.1, who resulted positive according to MR.



**Figure 3.2:** Schematic representation of the study population for tumour detection

Details about patients and lesions can be found in table 3.1.

<b>Patients</b>	<b>MR+ subjects (n = 25)</b>			
<i>Age (years)</i>	70.2 (52.9-79.1)*			
<i>PSA (ng/ml)</i>	6.2 (4.1-16)*			
<b>Lesions</b>	<b>All (n = 36)</b>	<b>PZ (n = 34)</b>	<b>TZ (n = 2)</b>	
<i>No. of lesions with GS</i>				
	3+3	10	8	2
	3+4	10	10	0
	4+3	3	3	0
	4+4	8	8	0
	4+5	3	3	0
	5+5	2	2	0
<i>Size (mm)</i>				
	≤7	18	17	1
	8-9	5	5	0
	≥10	15	14	1

**Table 3.1:** Characteristics of patients and lesions. \*Measurements are given as *median (range)*

#### 3.1.1 Reference Standard

Manual segmentation of malignant lesions was performed by an expert radiologist on T2-weighted and ADC map.

### 3.2 Feature Extraction

Feature extraction aims at gaining relevant information from the original set of data (e.g. an image) by creating an ensemble of variables, thus diminishing the number of involved parameters without losing any prime content [54].

In this study, the features listed in table 3.2 are extracted from T2-w and ADC masks, comprising first and higher order statistics (i.e. texture analysis by means of GLCM and GLRLM). Details about these features and their abbreviations can be found in section 3.5 of the appendix.

FE has been performed by means of C++ in-house software and ITK open source libraries [48].

<b>First order statistics</b>	<b>GLCM</b>	<b>GLRLM</b>
Mean	Contrast	HGRE
SD	Correlation	LGRE
Skewness	Energy	GLNU
Kurtosis	Entropy	RLNU
10 <sup>th</sup> percentile	Cluster prominence	
25 <sup>th</sup> percentile	Cluster shade	
50 <sup>th</sup> percentile	Haralick correlation	
75 <sup>th</sup> percentile	Homogeneity	
	Variance	
	Sum average	
	Sum entropy	
	Sum variance	
	Difference variance	
	Difference entropy	
	Information correlation 1	
	Information correlation 2	
	Dissimilarity	
	Max GLCM	

**Table 3.2:** List of extracted features (SD: standard deviation, GLCM: Gray Level Co-occurrence Matrix, GLRLM: Gray Level Run Length Matrix, HGRE: High Gray-Level Run Emphasis, LGRE: Low Gray-Level Run Emphasis, GLNU: Gray-Level Non-uniformity, RLNU: Run Length Non-uniformity )

Each feature is computed considering non-overlapping regions of interest composed by  $5 \times 5$  pixels, in which every slice is divided taking into account only the prostate mask resulting from the algorithm in section 2.3.

Many research groups [55–58] proved the efficacy of ADC map and T2-w signal intensities to distinguish healthy from unhealthy tissue. To preserve and exploit this knowledge, first order statistics are included in the extracted variables. It is important to underline that this type of analysis is not concerned with the positioning of grey levels in the image and their relationships, as it considers the single grey tone by itself.

Texture analysis was first introduced by Haralick [59], who intended to acquire information

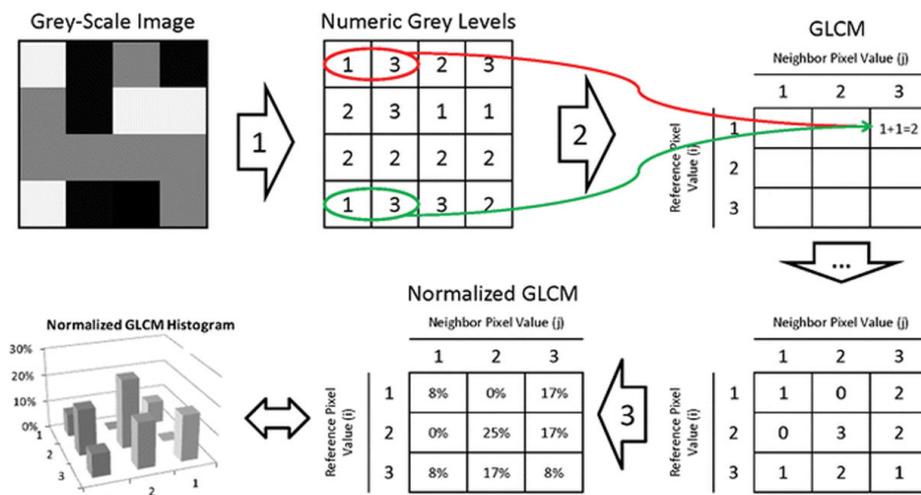
about the spatial distribution of grey intensities, creating a set of features able to describe an image patch by associating it to a visual perception (i.e. rippled, coarse, irregular, smoothed). His idea was to transform into mathematical expressions what the human eye can perceive and naturally describe.

In general, a gray-level co-occurrence matrix can be pictured as a bi-dimensional histogram of the intensities present in an image.

Referring to an image of  $N$  grey levels, a gray level co-occurrence matrix is a  $N \times N$  matrix  $T_v$ , where  $v$  is the displacement vector indicating the difference in spatial locations of  $n$  pixels. Each entry of the matrix  $T_v$  of coordinates  $(x, y)$  represents the number of times the pixel pair with grey levels  $x$  and  $y$  at distance equal to  $v$  occurs.

Posteriorly to normalisation with respect to total number of occurrences, the GLCM contains the probability of occurrence of each pixel pair characterised by intensities  $x$  and  $y$ , which are distant  $v$  [60, 61].

Figure 3.3 shows a simplified example to understand GLCM computation.

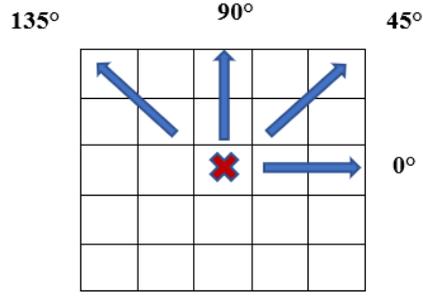


**Figure 3.3:** Simplified example about GLCM calculation. 1) An image of  $4 \times 4$  grey tones is taken into account and transformed into a numeric grid. 2) Considering a pixel with intensity 1 and its immediate neighbour to the right (i.e. 3), there are two occurrences in the image, thus the entry (1, 3) of the GLCM matrix is filled with 2. Following the same criterion, the entire GLCM is filled. 3) The gray-level co-occurrence matrix is thus normalised so that each entry holds the probability of every pixel pair. Adapted from [62]

A gray level run length can be defined as a contiguous set of pixels presenting the same intensity [63].

Gray level run length matrix is a texture representation which counts the presence of runs of pixels of each grey tone present in an image. As reported in [64], the entry  $(x, y)$  of the GLRLM matrix identifies the occurrence of pixels of grey level  $x$  with run length  $y$ .

While GLCMs are square matrices because the quantity of bins must match and the co-occurrence couples must be symmetric, GLRLMs may have an unequal number of rows and columns. In fact, it is crucial to preserve the information about the symmetrical couples for GLCM, whereas in the case of the GLRLM this would only be cause of redundancy.



**Figure 3.4:** Representation of the four directions along with texture features are computed. The red cross designates the center of the  $5 \times 5$  ROI

Regarding the calculation of GLCM derived features performed in this study, two contiguous neighbouring pixels are considered as displacement vector in four possible directions ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ). The same directions are evaluated for GLRLM (see figure 3.4). The number of gray levels for GLCM and GLRLM matrices is set to 64 bins.

Every ROI has been assigned malignant class (denoted by 1) only if the pixels belonging to the ROI are fully contained in the tumour mask, otherwise the ROI is attributed benign class (denoted by 0).

Prior to further elaborations, the entire set of features was scaled according to *min-max scaling* in equation 3.1, where  $f_i$  represents the  $i$ -th value of the current feature, while  $\max(f)$  and  $\min(f)$  are respectively its maximum and minimum value.

$$f_i = \frac{f_i - \min(f)}{\max(f) - \min(f)} \quad (3.1)$$

From now on, the terms *sample*, *element* and *ROI* denote indiscriminately the same entity, i.e. the vector of features corresponding to each ROI. The malignant class will be also denominated *positive* class, while the benign class corresponds to the *negative* class.

### 3.3 Classifier Construction

The classification problem is part of the computational intelligence field and it can be conceded as a critical issue in biomedical engineering, especially concerning the diagnostic process [65].

Throughout this study, *supervised classification* will be exploited in order to detect malignant lesions belonging to the prostate gland. Generally speaking, a supervised classifier can be viewed as a machine which is responsible for finding the relationships between the class label and the structure of the data. To do so, a *training set* composed of samples with their corresponding class must be given as input to such a machine.

The construction of the training set plays a vital role, as it influences the ability of the classifier to correctly recognise unexplored data. Depending on the level of knowledge enclosed in the training samples, the classification performance will be affected as more possible scenarios are comprised in the training samples.

Concerning the medical domain, it is particularly challenging to collect a vast and exhaustive quantity of data because of [66]:

1. huge amount of time needed for image acquisitions;
2. difficulty in gathering data in the case of low prevalence diseases (e.g. cancer);
3. variability related to the same disease.

The here proposed study is as well characterised by an imbalanced data set made of benign and malignant samples, therefore facing all the correlated issues (e.g. finite sample size, large data dimensionality). The main concern is then to create a training set which encounters for the heterogeneity of available data, regardless of the prevalence of one class compared to the other.

### 3.3.1 Decision tree based training set

The first method applied to create the training set is the *decision tree*, a supervised classifier that recursively partitions the instance space. Useful as an exploratory technique, a DT is composed by:

- a *root* node, first existing node with no incoming branches;
- *test* nodes, with outgoing branches;
- *leaf* nodes, terminal nodes holding the class label.

In the most common and plain case, each internal node divides the instance space according to the values of a single attribute [67]. In order to determine the best splitting attribute, it is necessary to choose a heuristic able to divide the labelled data set into single classes.

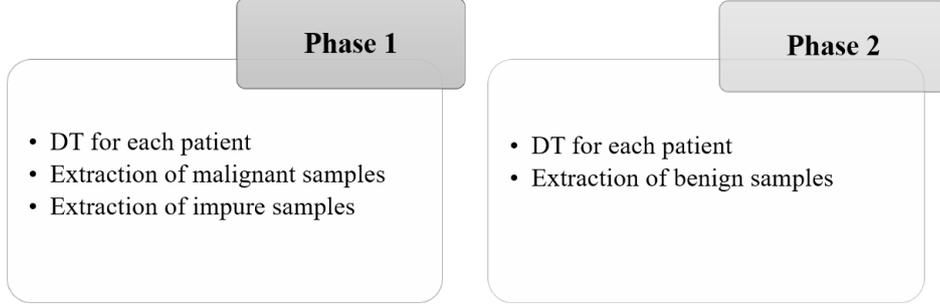
One of the most used attribute selection measures is the *Gini index*, which establishes the impurity of the data partition by means of the probability that an arbitrary sample belongs to each class.

In the creation of training and test sets, only those slices embedding cancerous lesions are considered. Table 3.3 shows the number of available samples separated per class and image type.

Image type	Malignant ROIs	Benign ROIs
ADC map	1136	58413
T2-w	1090	60235

**Table 3.3:** Total number of ROIs separated per class and image type

The approach conceived to build the training set starting from decision trees is constituted by two phases, summarised in figure 3.5.



**Figure 3.5:** Phases of decision tree based approach for training set construction

In the first phase of figure 3.5, the decision tree algorithm is applied to each patient separately, randomly extracting elements belonging to the nodes as follows:

1. find all nodes containing at least one malignant sample;
2. extract a number of samples ( $n_e$ ) equal to half of the elements held by the node;
3. check whether the node contains samples of the benign class and in the positive case extract benign elements ( $n$ ) according to equation 3.2:

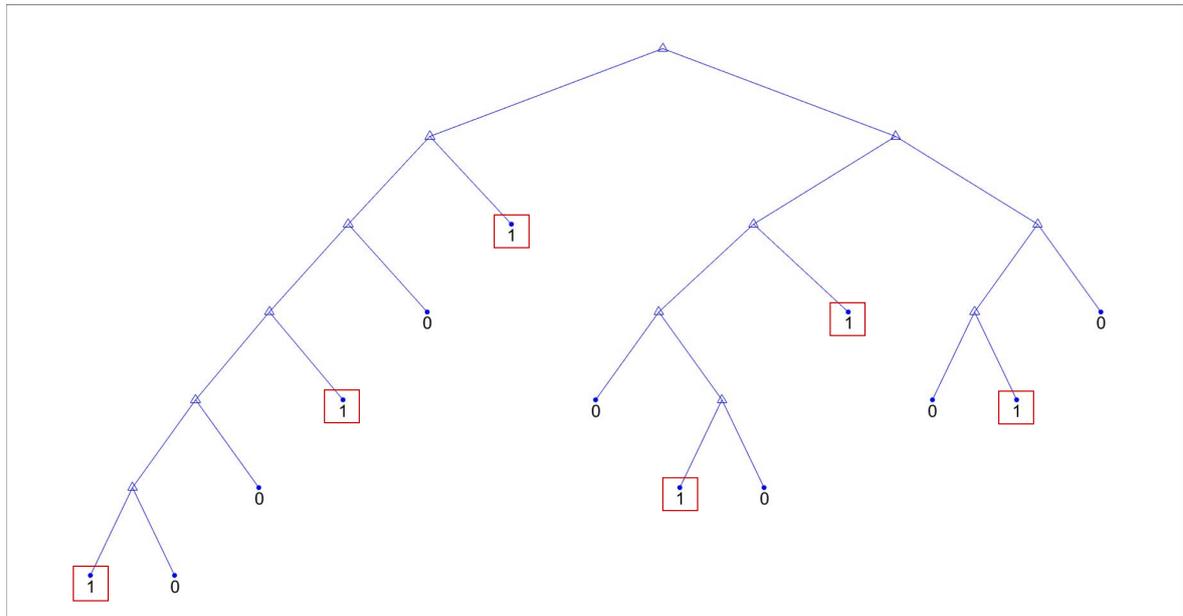
$$n = \min\left(n_m, \frac{n_b}{2}\right) \quad (3.2)$$

where  $n_m$  indicates the number of malignant elements in the considered node, and  $n_b$  is the number of benign elements contained in the same node.

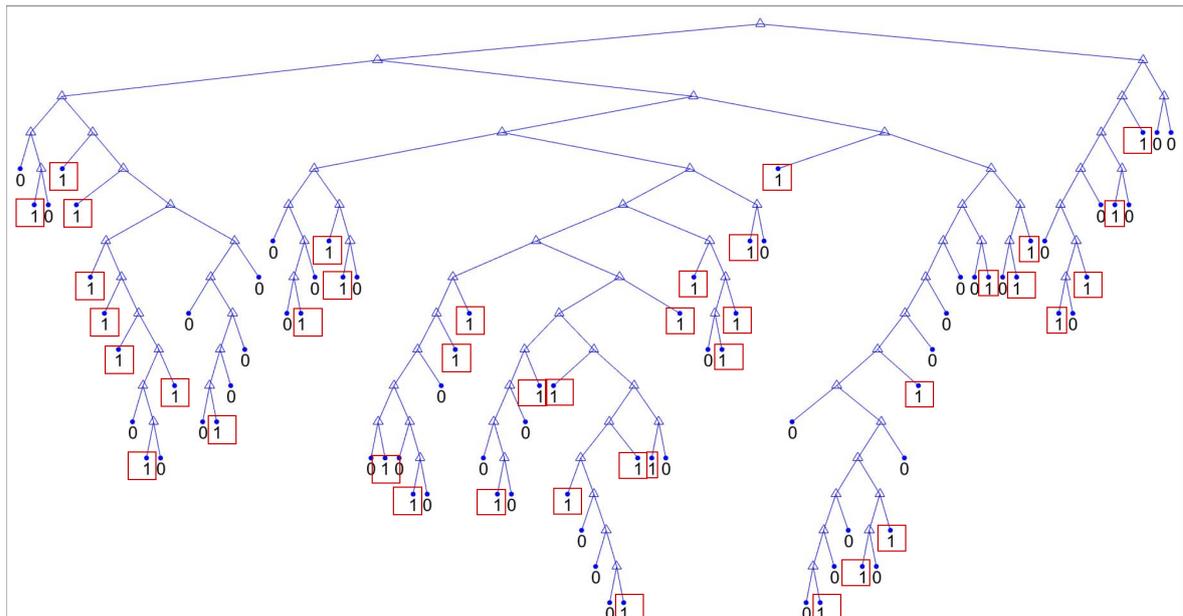
Equation 3.2 has been adopted in order to include those elements which are more resembling to malignant samples, thus ensuring a good representation of data diversity without compromising the total number of benign elements.

It is worth underlining that every decision tree presented a considerable inhomogeneity taking into account malignant ROIs, which led to have nodes containing few malignant elements together with a limited amount of benign elements in the majority of cases. Those peculiar nodes are here denominated *impure* nodes, due to the simultaneous presence of both classes, and the benign elements extracted from them are called *impure elements*. The possibility of having few malignant samples in a node assigned benign label, characterised by a consistent number of benign elements, can also occur. Nevertheless, this is taken care of during the process of picking impure elements.

The more malignant ROIs are present, the more branches are created in the trees, as seen in figure 3.6, therefore suggesting the unexpected inhomogeneity proper of cancerous lesions.



(a) DT of a patient with few malignant ROIs



(b) DT of a patient with many malignant ROIs

**Figure 3.6:** Emblematic examples of decision trees for patients with different number of malignant samples and a comparable number of benign samples (superior to 2500). Red boxes highlight leaf nodes assigned malignant label

Once the total number of malignant ROIs ( $M$ ) and impure elements ( $I$ ) is known, the second phase in figure 3.5 can begin with the computation of each patient's DT, followed by next steps:

1. find all nodes containing at least one benign sample (regardless of the presence of malignant elements);

2. consider only the nodes containing a number of elements at least equal to 5% of the total amount of benign ROIs present in the subject's data set ( $n_{tot}$ );
3. extract elements randomly ( $n_B$ ) in accordance with equation 3.3:

$$n_B = n_b \times \frac{\max(n_1, |M - I|)}{n_{tot}} \quad (3.3)$$

where  $n_b$  is the number of benign elements belonging to the node, and  $n_1$  corresponds to 1% of  $n_{tot}$ .

The limit imposed at 5% of the total amount of benign ROIs of each patient derives from the fact that there exists a restricted number of nodes containing nearly all benign elements. This value enables to include the most represented benign ROIs in the training set. Indeed, equation 3.3 allows to pick a reasonable number of benign samples, in view of the selected malignant and impure elements.

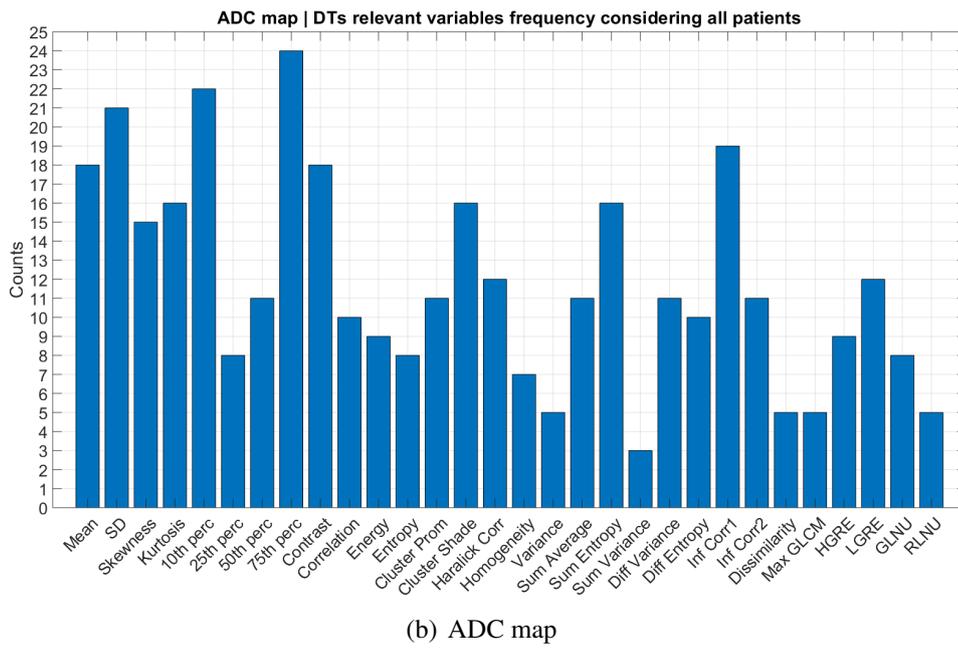
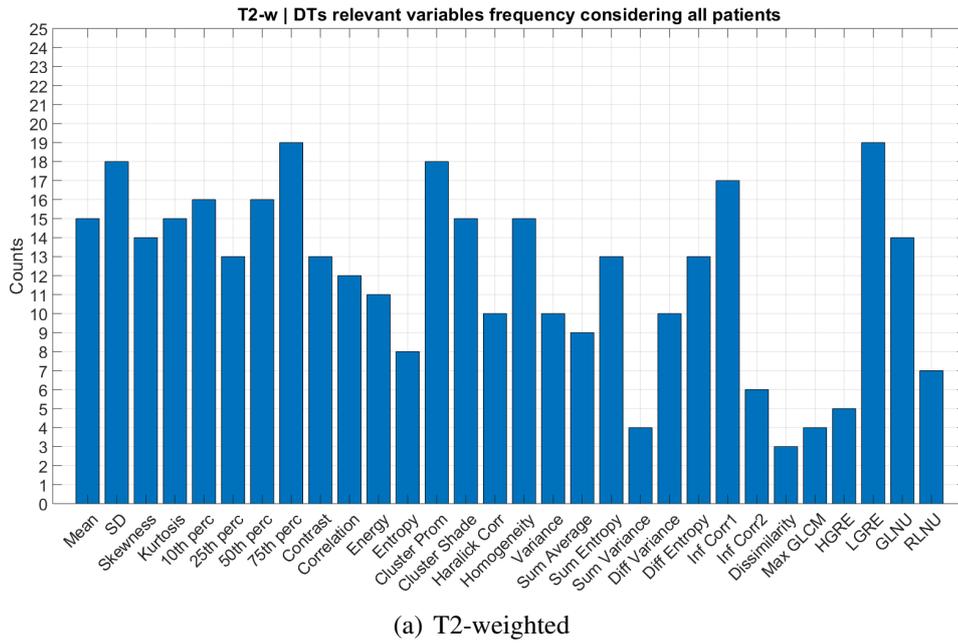
As noticeable from table 3.4, the number of positive and negative samples comprised in the training set slightly differs. However, being the difference so small, this should not affect the classifier performance and the two classes can be retained balanced.

<b>Image type</b>	<b>Malignant ROIs</b>	<b>Benign ROIs</b>
ADC map	780	886
T2-w	768	925

**Table 3.4:** Number of ROIs separated per class and per image type constituting decision tree based training set

Decision trees are also featured with the benefit of establishing the most relevant attributes useful to discriminate the classes. The two image modalities present frequently selected variables and others rarely appearing, hence enforcing the great variability proper of the involved set of data.

Figure 3.7 exhibits bar diagrams representing the frequency of each feature for all decision trees: it is evident that no variable is selected considering the entire study population.



**Figure 3.7:** Occurrence of selected attributes across all decision trees

### 3.3.2 Feature Selection with Genetic Algorithms

The purpose of feature selection is to identify a subset of variables able to [68]:

- preserve relevant informative content;
- decrease dimensionality;
- eliminate superfluous and redundant data;
- ameliorate classification performance;
- enhance result understandability.

Two principal strategies can be adopted for FS: *filter methods*, characterised by recognition of the most relevant features without considering the learning task (regardless of subsets comprising high discriminative variables), and *wrapper methods*, which select features along with evaluating the goodness of the learning algorithm [69]. Although wrapper methods may present a heavy computational burden, they contribute to superior classification performance and parameters optimisation for classifier or learning algorithm.

When implementing a classifier, it is pivotal to tune its parameters as well as try to find a subset of features which may reach the best classification outcome [70]. Among the various methods such as grid or random search, a heuristic search can be utilised to simultaneously perform FS and tuning classifier parameters.

In this study, a *genetic algorithm* is applied, which has been already used to solve optimisation problems [71] or find tuning parameters [70]. Up to now, few applications are recovered concerning these two aspects at the same time (one of them can be found in [72]).

Belonging to the field of evolutionary computing, GAs are optimisation methods inspired by evolutionary processes (e.g. natural selection, reproduction) to create computer-based problem solving methods. They can be used to prune the set of involved variables, by discovering those which reduce generalisation errors [65].

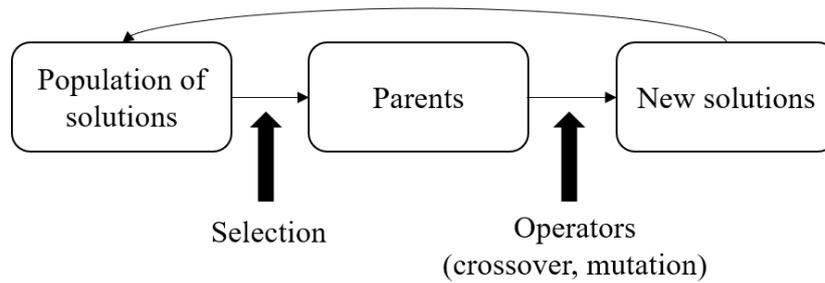
Three main operators can be encountered:

- *selection*, to represent the survival of the fittest individuals;
- *crossover*, to model reproduction;
- *mutation*, to introduce diversification through the generations.

Every solution of a genetic algorithm is codified as a binary vector of fixed dimension, called *chromosome*, whereas each bit is referred as a *gene*.

GAs perform a stochastic population-based research, thus the first step is to produce an initial population of candidate solutions, by assigning a random value from the permitted domain to each gene. They may need to satisfy an admissibility condition.

Moreover, it is cardinal to determine population dimension and number of iterations, for they influence the exploration capacity of the algorithm. The goodness of each individual is established by means of an objective function, called *fitness function*.



**Figure 3.8:** Schematic representation of an evolutionary algorithm

Once the best solutions are identified, mutation and crossover operators are applied to generate the offspring, respectively by combining parts of existing solutions and randomly changing the genes. Afterwards, the selection operator is in charge of preserving the fittest individuals of each generation, thus yielding the population for the next generation. This operator is particularly important to assure the survival of good solutions and avoid stopping in a local optimum.

A genetic algorithm ceases when a limit is reached in terms of executed number of generations or fitness function evaluations or when a convergence criterion is met.

### 3.3.3 Classifier

A suitable classifier to be used in combination with a genetic algorithm is the *support vector machine* [65]. It is a binary supervised learning classifier, which maps the non-linear input vectors into a high dimensional feature space in which they become linearly separable [73]. Training a SVM consists in finding the hyperplane, able to best split the two classes, characterised by furthest distance from the nearest training samples (so called *support vectors*).

In order to build a SVM, the following must be set:

- *kernel function*, non linear function which projects samples from the input space to the output space;
- *parameter C*, numerical value to establish the number of misclassified samples: a strict margin results from a high value of C, even if the points are linearly separated, whereas a small C leads to a large margin with misclassification errors.

The learning task expressed as an optimisation problem with linear constraints aims at balancing the maximisation of the margin with the error penalty.

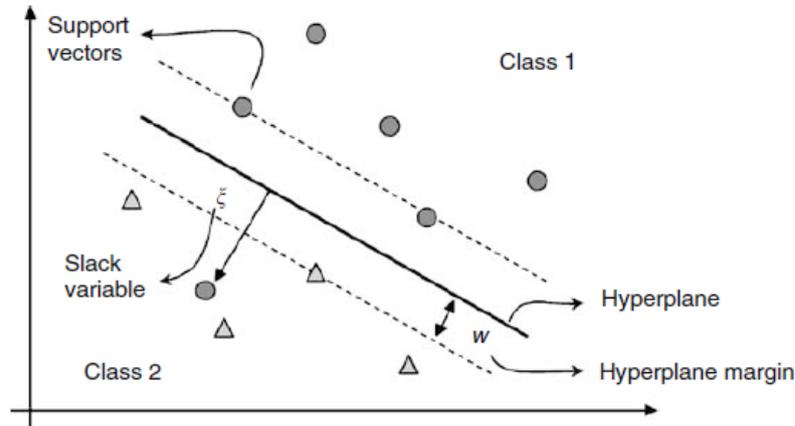


Figure 3.9: Soft-margin support vector machine in a 2-D example [65]

### Parameters setting

The fitness function implemented in the GA is reported in equation 3.4 [72]. It solves a minimisation problem, thus the smaller its value (ideally 0) the better the solution is.

$$fitness_i = 1 - \frac{spec + sens}{2} + 0.3 \times (|spec - sens|) \quad (3.4)$$

In equation 3.4, *sens* and *spec* refers respectively to sensitivity and specificity of the classifier with parameters and selected variables chosen by *i*-th solution, considering classification performance on the test set (in table 3.3).

Sensitivity and specificity (respectively in equations 3.6 and 3.7) are computed from the confusion matrix in table 3.7.

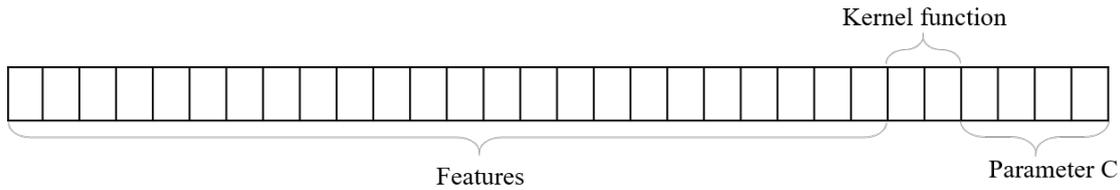
This fitness function attempts to compensate for the imbalanced classes present in the test set, by using the so called *averaged accuracy* (i.e. the second term in equation 3.4) in order to evenly weigh the two metrics.

Experimental trials conducted in this study found that the genetic algorithm did not evolve finding in the initial population of random individuals the solution with the best fitness. The setting used in these trials is here detailed:

- number of individuals: 200;
- number of iterations: 5000;
- number of parents:  $0.8 \times \text{number of individuals}$ ;
- number of repetitions: 1;
- number of genes: 36, codified as:
  - first 30 bits corresponding to the features;
  - next 2 bits for SVM kernel function (*polynomial of order 2, polynomial of order 3, linear, radial basis function*);

- last 4 bits for parameter C of SVM according to equation 3.5, where  $C_{bin}$  represents the decimal value of the bits;

$$\begin{cases} C = 0.5, & \text{if } C_{bin} = 0 \\ C = 1, & \text{if } C_{bin} = 1 \\ C = 10 \times (C_{bin} - 1), & \text{if } C_{bin} > 1 \end{cases} \quad (3.5)$$



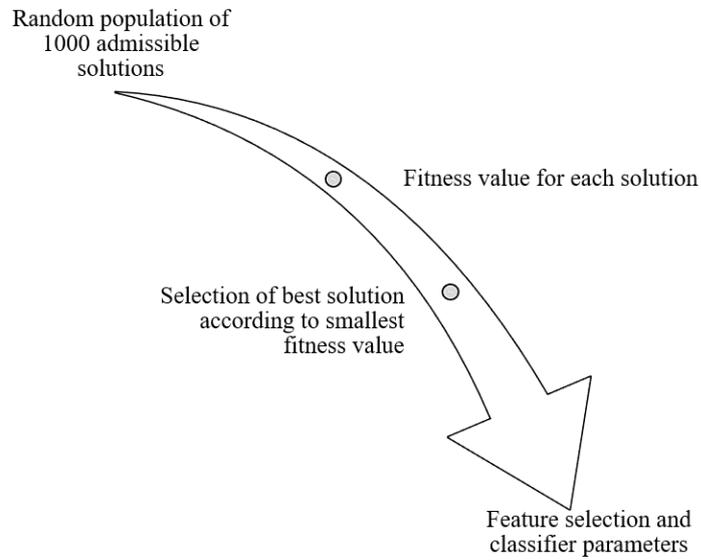
**Figure 3.10:** Schematic representation of GA solution encoding with SVM parameters. Each square represents a single bit

- crossover probability: 1;
- crossover cutting points: 2;
- mutation probability: 0.2, decreased to 0.15 after 200 iterations and 0.1 after 400 iterations.

With regard to the bits representing the features, they are codified according to the *Yes/No variables* encoding, i.e. the bit equals 1 if the variable is selected, 0 otherwise.

A solution is considered admissible if at least 2 features are equal to 1.

Due to the premature identification of the best solution, the genetic algorithm in figure 3.11 is implemented (the encoding for each solution remains unchanged): generating a random population of 1000 individuals, the fitness function is evaluated for each of them and the one with the smallest value is elected as best. By doing so, the GA serves to determine the classifier parameters and examine the behaviour of the classifier according to the selected features.



**Figure 3.11:** Implementation of the genetic algorithm to set classifier parameters and selected features

## Results

Parameters established by the genetic algorithm along with number of selected variables and fitness value of the best solutions are reported in table 3.5.

<b>Image type</b>	<b>Best fitness value</b>	<b>No. of selected variables</b>	<b>Kernel function</b>	<b>Parameter C</b>
<b>T2</b>	0.28	22	Radial basis function	100
<b>ADC</b>	0.27	20	Radial basis function	140

**Table 3.5:** SVM. Parameters selected by the genetic algorithm and best fitness values

Details about the selected features can be found in figure 3.6, with a binary representation indicating presence of the variable (cell in white) and absence of the variable (cell in black).



Evaluation metrics derived from the confusion matrices (reference in figure 3.7) are detailed below:

- sensitivity indicates the number of ROIs classified as tumours, actually belonging to the malignant class;

$$Sens = \frac{TP}{TP + FN} \quad (3.6)$$

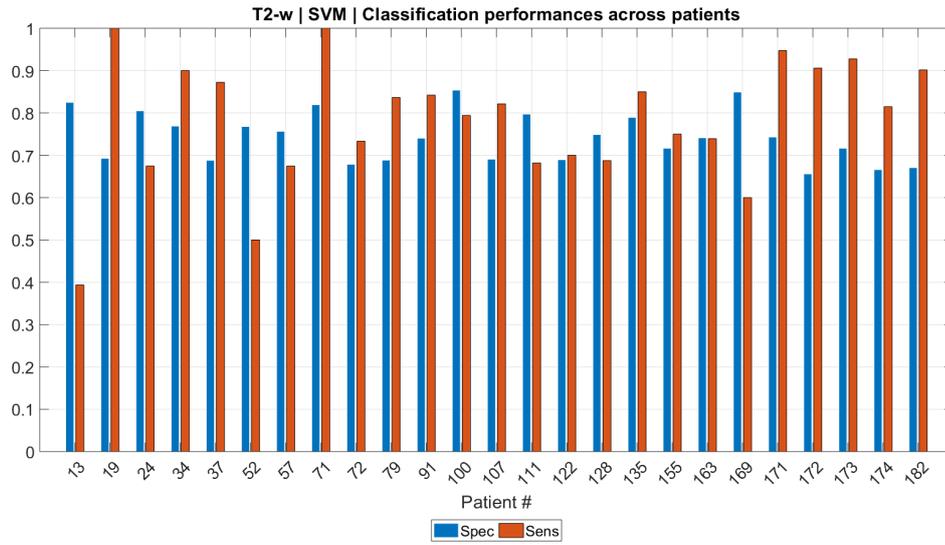
- specificity estimates the amount of ROIs classified as healthy, actually belonging to the negative class.

$$Spec = \frac{TN}{TN + FP} \quad (3.7)$$

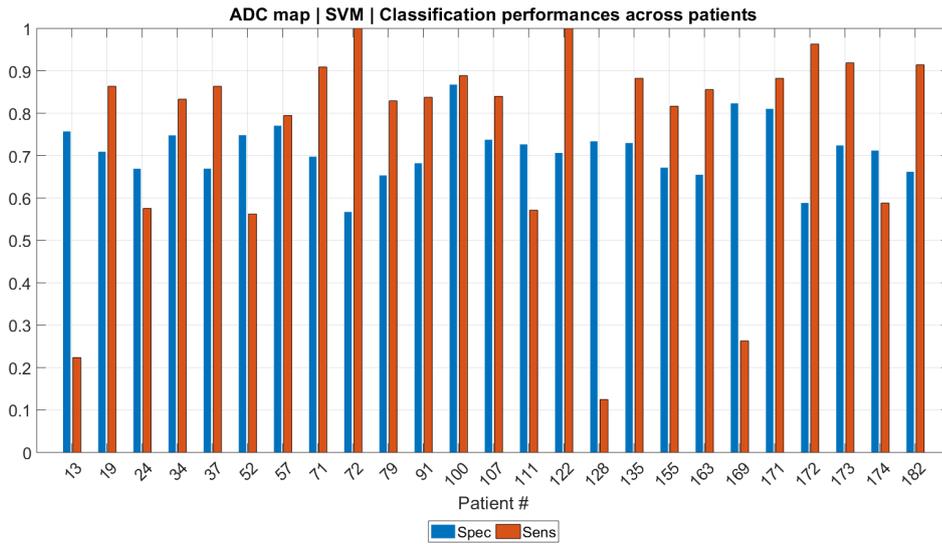
Whereas the classifier seems to recognise malignant samples in an acceptable amount (both in training and test sets), the benign class is characterised by a lower identification, especially taking into account the much greater presence of negative samples.

Sensitivity and specificity obtained for each patient are represented in figure 3.12.

It can be noticed that the metrics concerning T2-w are less variable among patients with respect to ADC map. However, while the specificity stays around 0.6 or above in the two image sets, sensitivity reveals to have a value below 0.5 for more than one patient.



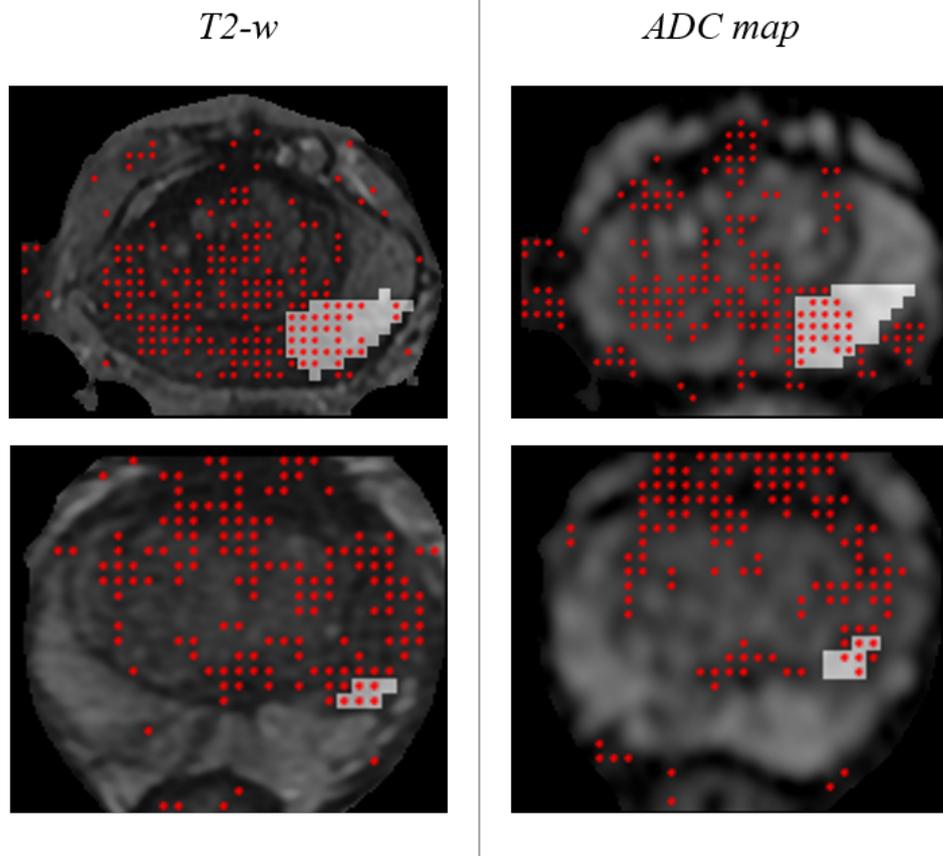
(a) T2-weighted



(b) ADC map

Figure 3.12: SVM. Sensitivity and specificity across patients

In figure 3.13, the resulting outcome of SVM tumour detection is displayed for two patients, showing the difference between a large and a small tumour. False positives (i.e. benign samples classified as malignant) are particularly present in the T2-w mask, even though the cancerous lesion is recognised at least partially.



**Figure 3.13:** SVM. Examples of tumour detection. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue

### 3.4 Classification Optimisation

Due to the relatively low recognition of the benign class characterising the former approach, next attempt proposes to improve the representation of healthy tissue by means of enlarging the training set with more than a single negative class.

#### 3.4.1 Dendrogram based Training Set

The dendrogram based approach arises from the necessity to increase the representation of the benign class. As a matter of fact, the latter is affected by a huge diversification inter and intra-patients. To solve this issue, benign elements are grouped through *dendrogram clustering*, a technique which merges samples on the basis of a similarity measure. Keeping unaltered the malignant ROIs extracted with the decision tree based training set (see section 3.3.1), benign ROIs are gathered as follows starting from DT of each patient:

1. find all nodes containing at least one benign sample (regardless of the presence of malignant elements);
2. sort these nodes in descending order given number of elements held by each node;
3. consider the first three nodes (i.e. the ones with the largest number of elements).

At this point, dendrogram clustering is applied on all benign samples enclosed by the aforementioned nodes. The classes are obtained by cutting each dendrogram (one per image modality) and extracting randomly from every cluster a number of samples equal to:

$$\begin{cases} M & \text{if } n_d \geq M \\ n_d & \text{otherwise} \end{cases} \quad (3.8)$$

where  $M$  is the total number of malignant elements present in the DT based training set and  $n_d$  corresponds to the number of elements present in the current cluster.

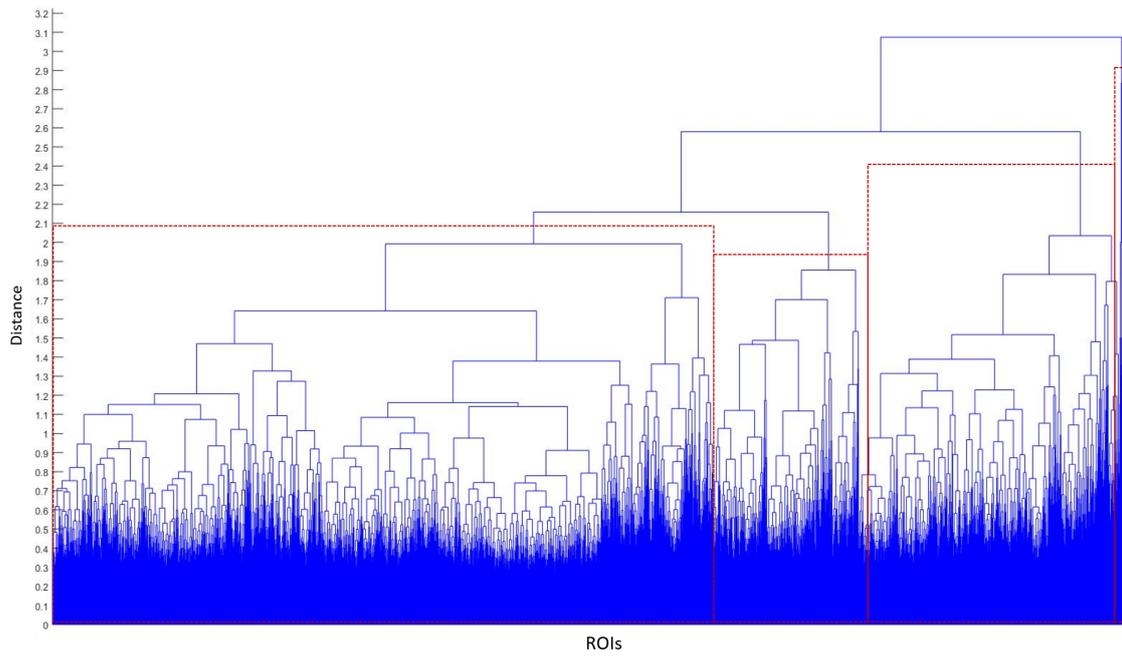
The cutting level has been chosen to create groups separated by a certain distance and rather homogeneous in terms of inner variability and number of elements.

Figure 3.14 shows the clusters obtained from the dendrograms.

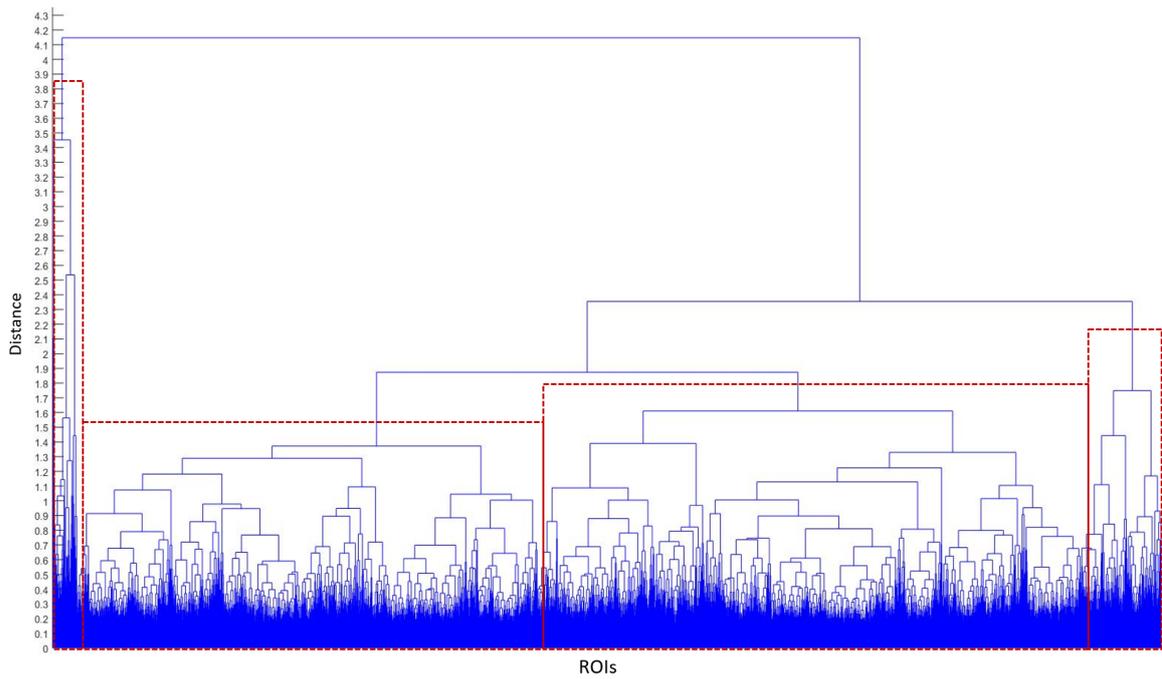
In table 3.8, the number of elements belonging to each class is reported per image modality.

Image type	Malignant ROIs	Benign ROIs			
		Class A	Class B	Class C	Class D
ADC map	780	780	780	780	780
T2-w	768	768	768	768	607

**Table 3.8:** Number of ROIs reported per class and per image type constituting dendrogram based training set



(a) T2-weighted



(b) ADC map

**Figure 3.14:** Dendrograms with overlapping red boxes representing clusters

### 3.4.2 Classifiers

#### Multiclass SVM

When more than two classes are present, it is possible to exploit a *multiclass* version of the SVM, which can work with different strategies [74]. For further details on the implementation of the two-class support vector machine, the reader is referred to section 3.3.3.

In the context of this study, the one-versus-one coding design is adopted, consisting in training  $L(L - 1)/2$  binary classifiers, where a class is considered positive and the other negative, and  $L$  is the total number of classes.

Even with the dendrogram based training set, the genetic algorithm mentioned in section 3.3.3 showed no developments from the initial population, so the approach in figure 3.11 is implemented in this case as well.

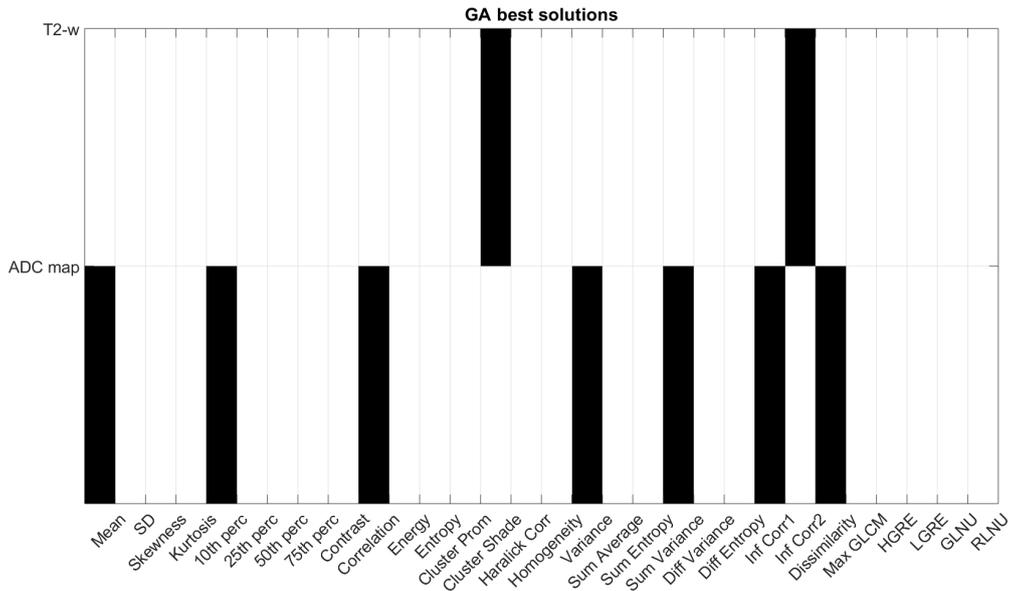
The fitness function in equation 3.4 will consider sensitivity and specificity relative to multi-class SVM performance.

Table 3.5 reports classifier parameters set by the genetic algorithm along with number of selected variables and fitness value of the best solutions.

Image type	Best fitness value	No. of selected variables	Kernel function	Parameter C
T2	0.25	28	Polynomial of order 3	140
ADC	0.27	23	Polynomial of order 3	140

**Table 3.9:** *Multiclass SVM.* Parameters selected by the genetic algorithm and best fitness values

Details about the selected features can be found in figure 3.10, with a binary representation indicating presence of the variable (cell in white) and absence of the variable (cell in black).



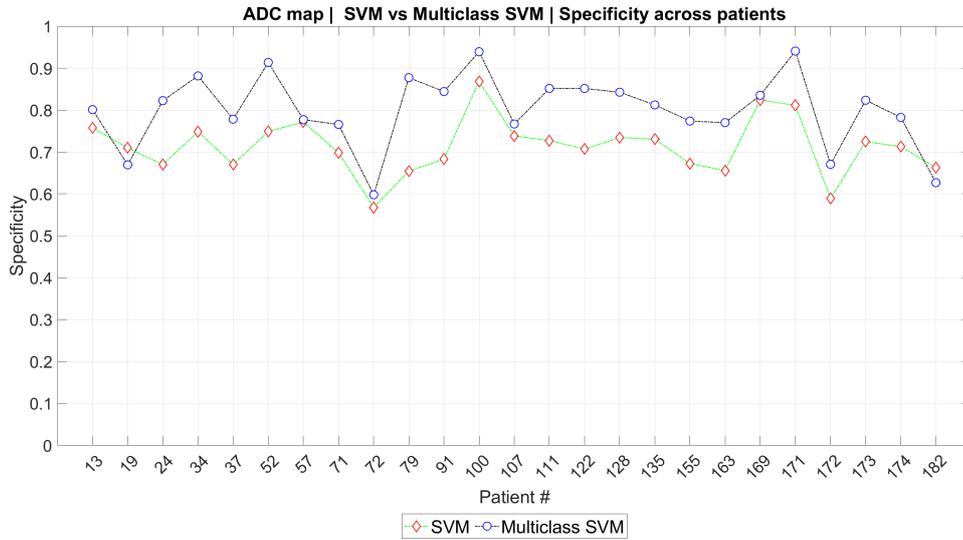
**Table 3.10:** Multiclass SVM. Binary representation of FS performed by the genetic algorithms. White cells correspond to selected variables, black ones indicate those features that are not considered

Final performances obtained with this classifier are shown in form of normalised confusion matrices in table 3.11. The superior recognition of benign samples found in the training set is not retrieved in the test set, characterised by pretty balanced sensitivity and specificity (approximately equal to 0.7).

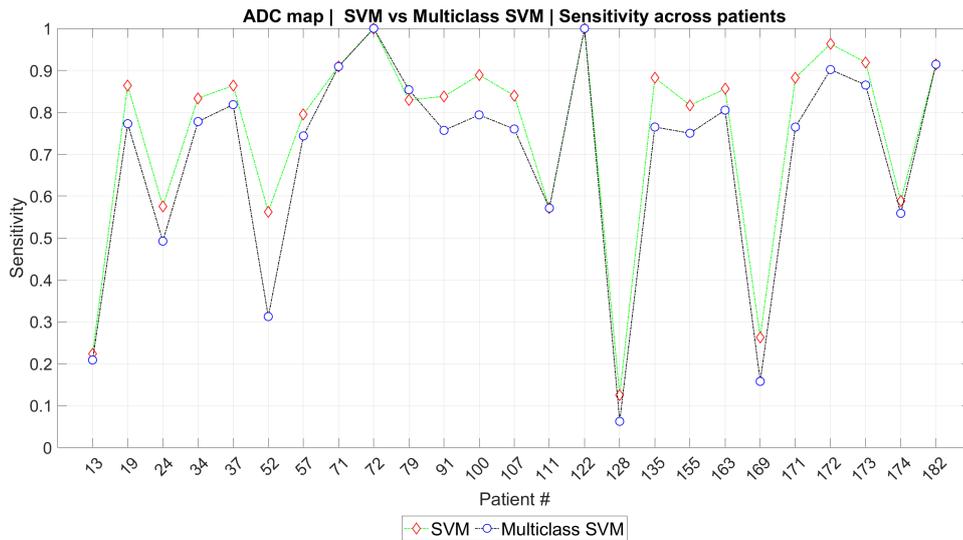
		T2-w					
		TRAINING SET		TEST SET			
		True class		True class		True class	
		0	1	0	1	0	1
Predicted class	0	0.92	0.26	0.76	0.27		
	1	0.08	0.74	0.24	0.73		
		ADC map					
		TRAINING SET		TEST SET			
		True class		True class		True class	
		0	1	0	1	0	1
Predicted class	0	0.92	0.29	0.80	0.29		
	1	0.08	0.71	0.20	0.71		
		0	1	TN	FN	FP	TP
Predicted class	0						
Predicted class	1						

**Table 3.11:** Multiclass SVM. Classification performances considering training and test sets per image modality. The reference confusion matrix is reported on the right (1: malignant label, 0: benign label)

Concerning the ADC map, the results in figure 3.15 show an overall improvement in terms of specificity comparing SVM with multiclass SVM for the study population. Regarding sensitivity values instead, more than one half of the subjects presents their malignant samples better recognised by the SVM.



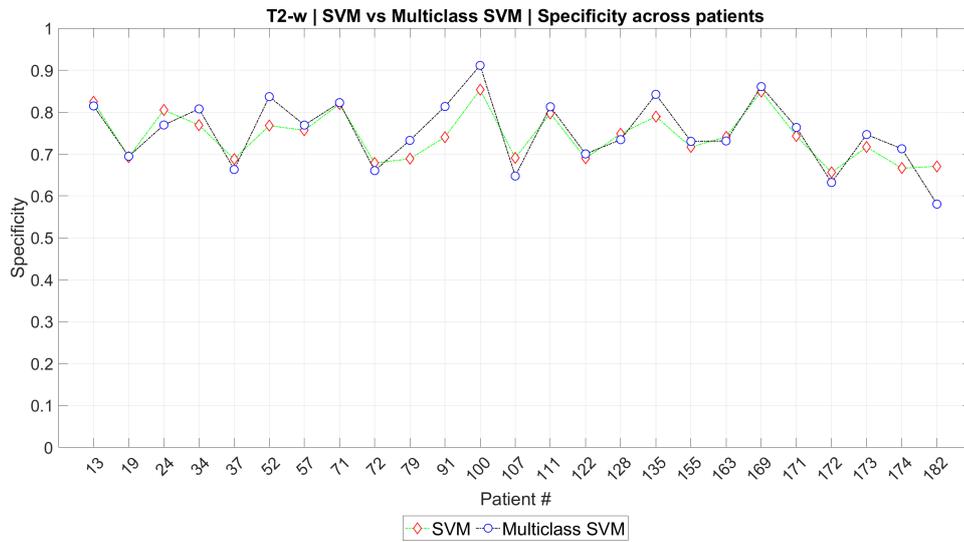
(a) Specificity



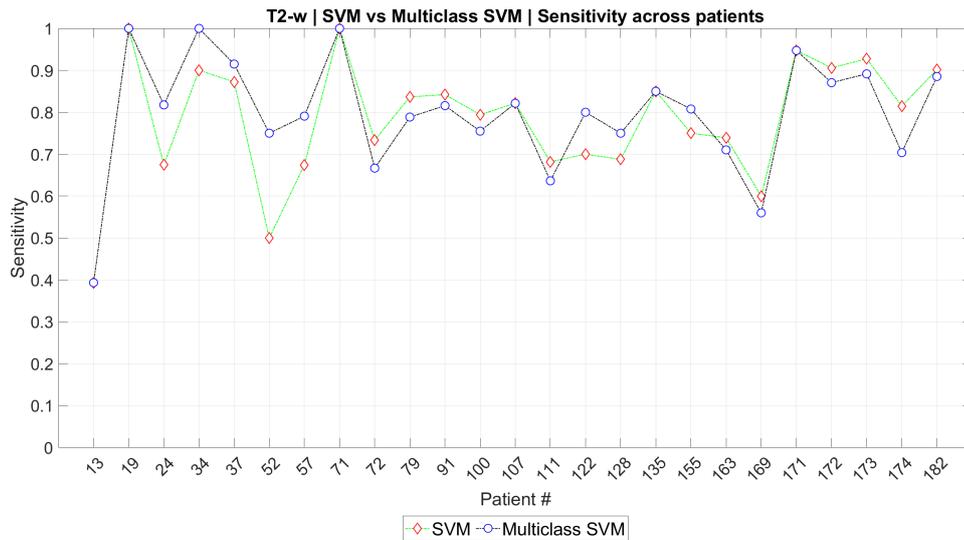
(b) Sensitivity

Figure 3.15: SVM vs Multiclass SVM. Classification performances across patients on ADC map

A similar trend can be observed for T2-w considering specificity, whereas less than half of the subjects shows higher sensitivity with SVM classifier.



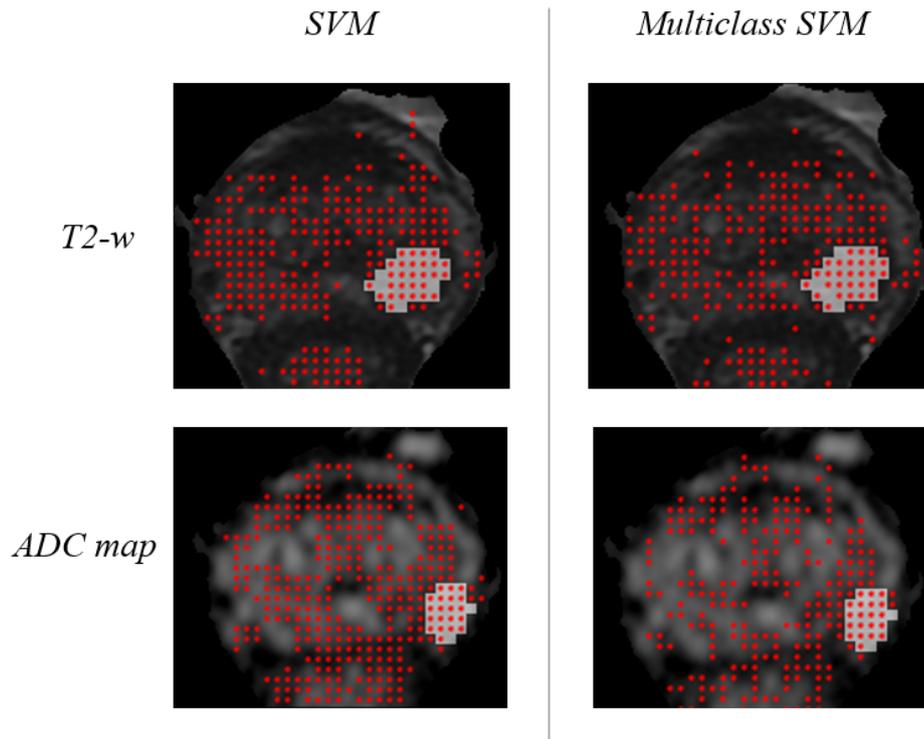
(a) Specificity



(b) Sensitivity

**Figure 3.16:** SVM vs Multiclass SVM. Classification performances across patients on T2-weighted modality

A comparison concerning a subject with many misclassified elements is provided in figure 3.17, in which no considerable improvement is found for T2-w, whilst the outcome of ADC map for multiclass SVM shows less false positives.



**Figure 3.17:** *SVM vs Multiclass SVM.* Examples of tumour detection for a patient with several false positives. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue

Dendrogram based training set demonstrates its greater efficacy in detecting benign elements, although a slight reduction in tumour recognition can be inferred.

Next attempt is to assess the performance of another classifier to seek general improvements and examine the differences which will arise.

### **k-Nearest Neighbours**

*k-Nearest Neighbours* performs classification based on the nearest distance unlabelled examples present with respect to the training set.

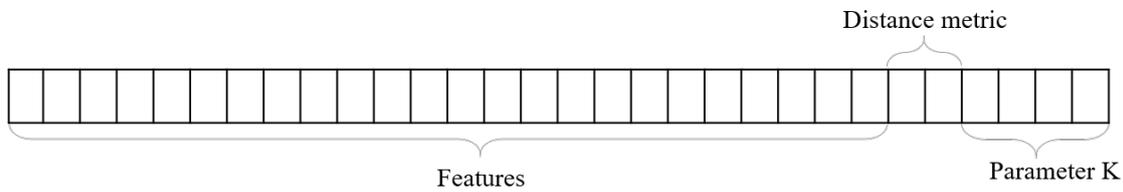
Considering a number  $k$  of neighbours, the distance of the analysed element is computed with each sample of the training data, according to a precise metric. These distances are sorted in descending order and the tested element is assigned to the most represented class among the neighbours. This is designated as majority voting procedure [75].

As preliminary trial with k-NN classifier, a genetic algorithm with the undermentioned parameters is implemented:

- number of individuals: 50;
- number of iterations: 20;
- number of parents:  $0.8 \times \text{number of individuals}$ ;

- number of repetitions: 1;
- number of genes: 36, codified as:
  - first 30 bits corresponding to the features;
  - next 2 bits for distance metric (*euclidean, chebychev, minkowsky, cityblock*);
  - last 4 bits for parameter  $K$  (i.e. number of neighbours) according to equation 3.9, where  $K_{bin}$  represents the decimal value of the bits;

$$K = 16 + K_{bin} \quad (3.9)$$



**Figure 3.18:** Schematic representation of GA solution encoding with k-NN parameters. Each square represents a single bit

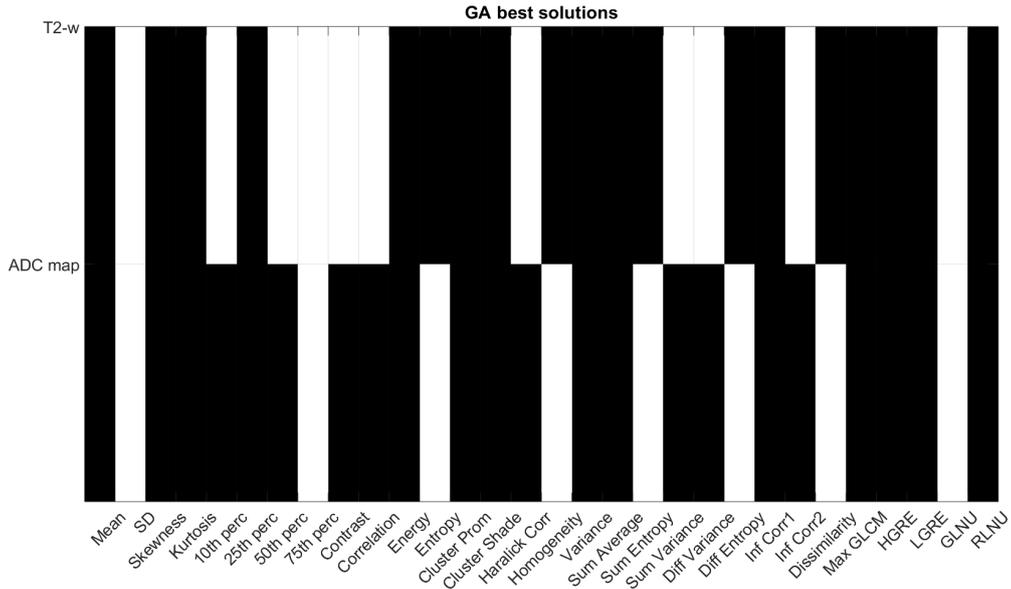
- crossover probability: 1;
- crossover cutting points: 4;
- mutation probability: 0.2.

In contrast to the version with SVM classifiers (see sections 3.3.3 and 3.4.2), the here performed GA evolved from the initial population in its research for the best solution.

The outcome of this genetic algorithm is exhibited in table 3.12, while the selected features can be viewed in figure 3.13.

Image type	Best fitness value	No. of selected variables	Distance metric	Parameter K
T2	0.28	11	Cityblock	16
ADC	0.26	8	Chebychev	16

**Table 3.12:** Multiclass k-NN. Parameters selected by the genetic algorithm and best fitness values



**Table 3.13:** Multiclass *k*-NN. Binary representation of FS performed by the genetic algorithm. White cells correspond to selected variables, black ones indicate those features that are not considered

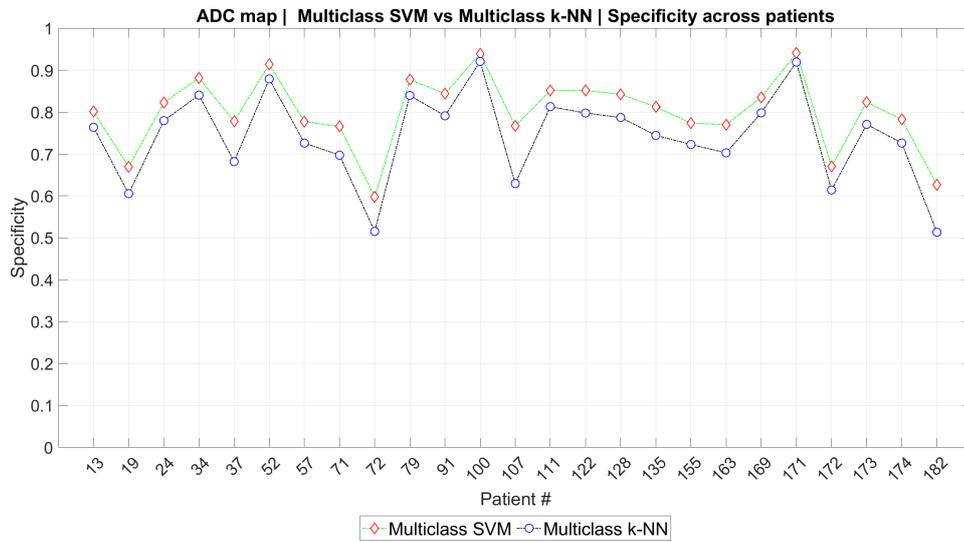
Final performances reached by this classifier are shown in form of normalised confusion matrices in table 3.11.

		T2-w					
		TRAINING SET		TEST SET			
		True class		True class		True class	
		0	1	0	1	0	1
Predicted class	0	0.90	0.25	0.72	0.27		
	1	0.10	0.75	0.28	0.73		
		ADC map					
		TRAINING SET		TEST SET			
		True class		True class		True class	
		0	1	0	1	0	1
Predicted class	0	0.90	0.26	0.74	0.25	TN	FN
	1	0.10	0.74	0.26	0.75	FP	TP

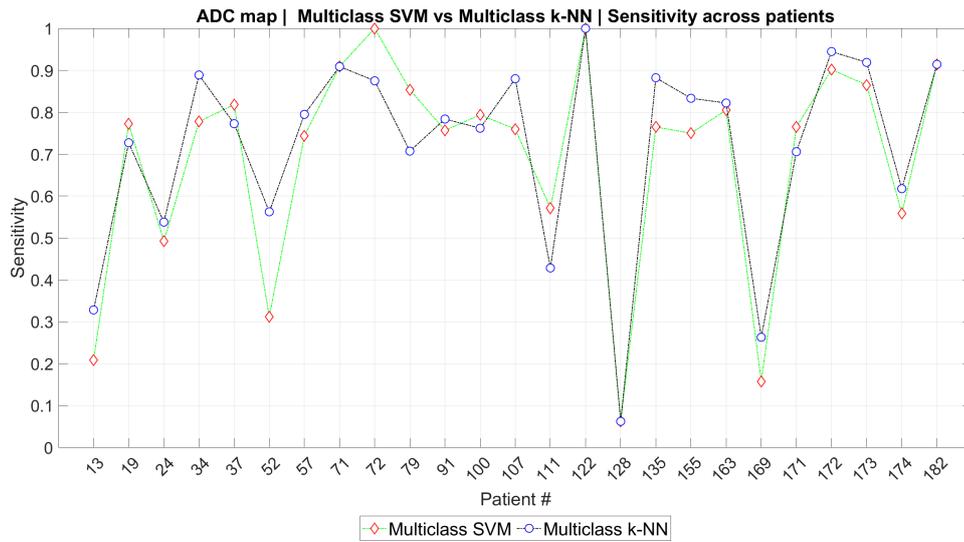
**Table 3.14:** Multiclass *k*-NN. Classification performances considering training and test sets per image modality. The reference confusion matrix is reported on the right (1: malignant label, 0: benign label)

As clear in figure 3.19, multiclass SVM outperforms *k*-NN in terms of specificity on ADC map. Sensitivity across patients differs in the two classifiers, but *k*-NN reveals better performance on malignant samples in more than half subjects. Comparable performances are obtained on the test set for T2-weighted modality, related to those proper of multiclass SVM (in table 3.11).

Both classifiers present a high recognition of benign samples on the training set, which however cannot be retrieved on the test set.



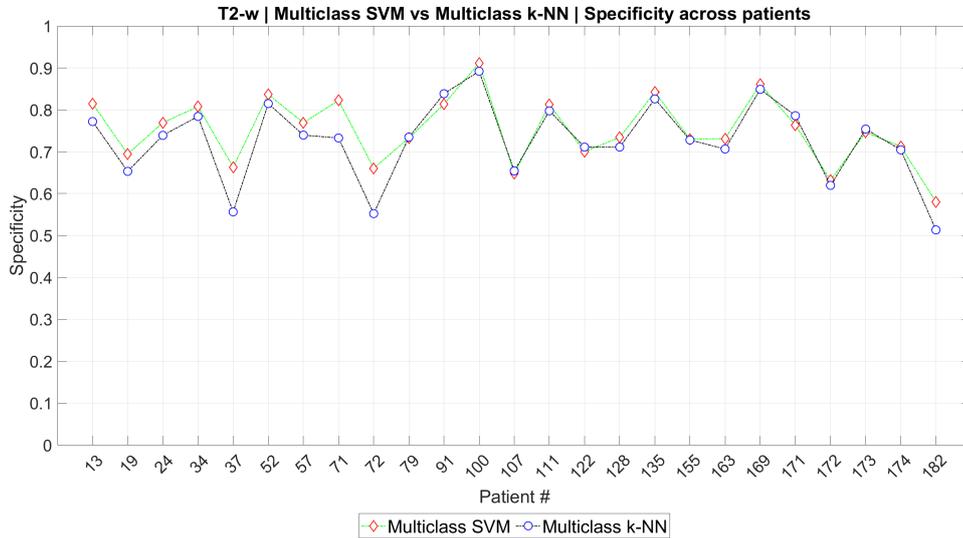
(a) Specificity



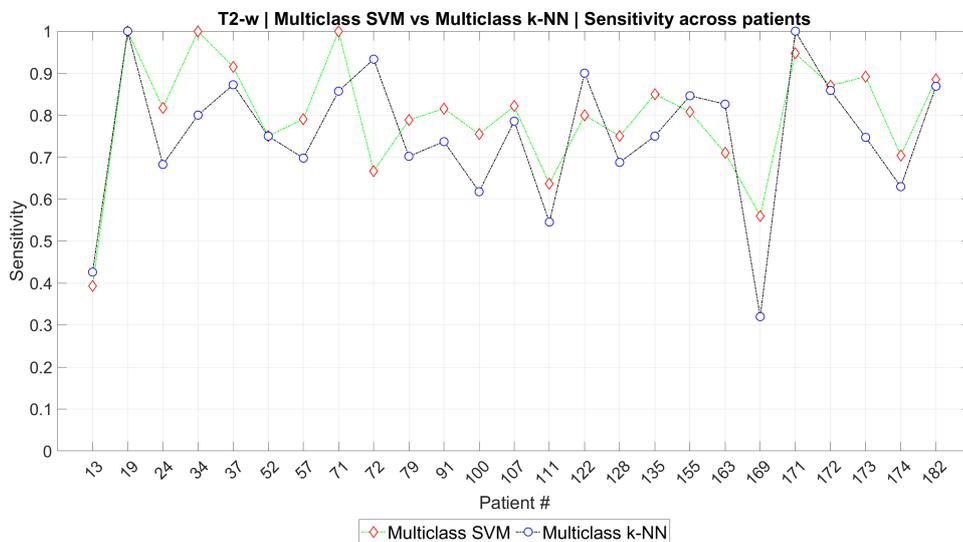
(b) Sensitivity

**Figure 3.19:** Multiclass SVM vs Multiclass k-NN. Classification performances across patients on ADC map

Referring to the outcome regarding T2-w, a pretty similar trend arises for the specificity, whereas the sensitivity is higher for the majority of patients in the case of multiclass SVM.



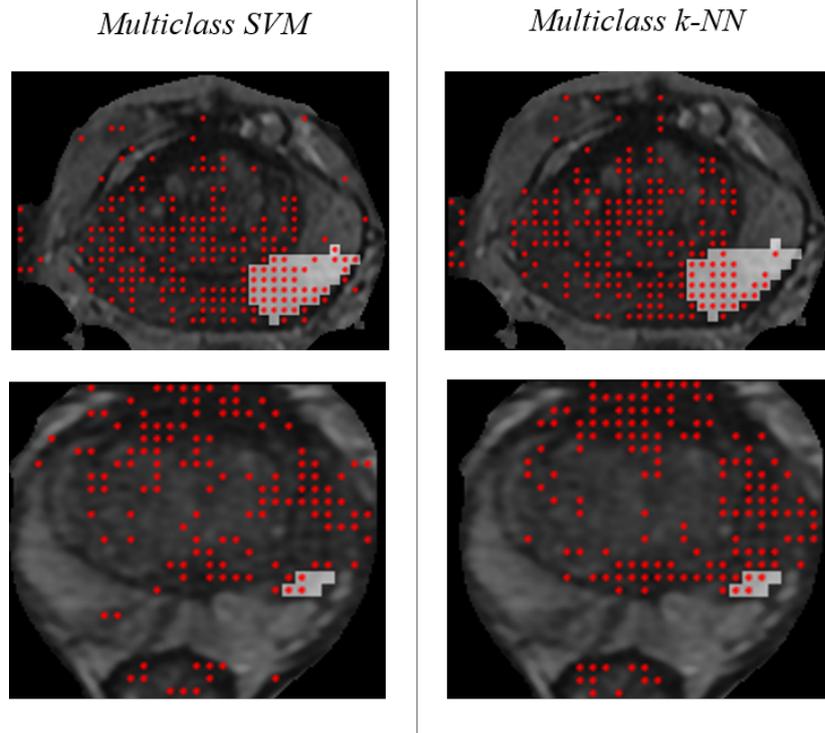
(a) Specificity



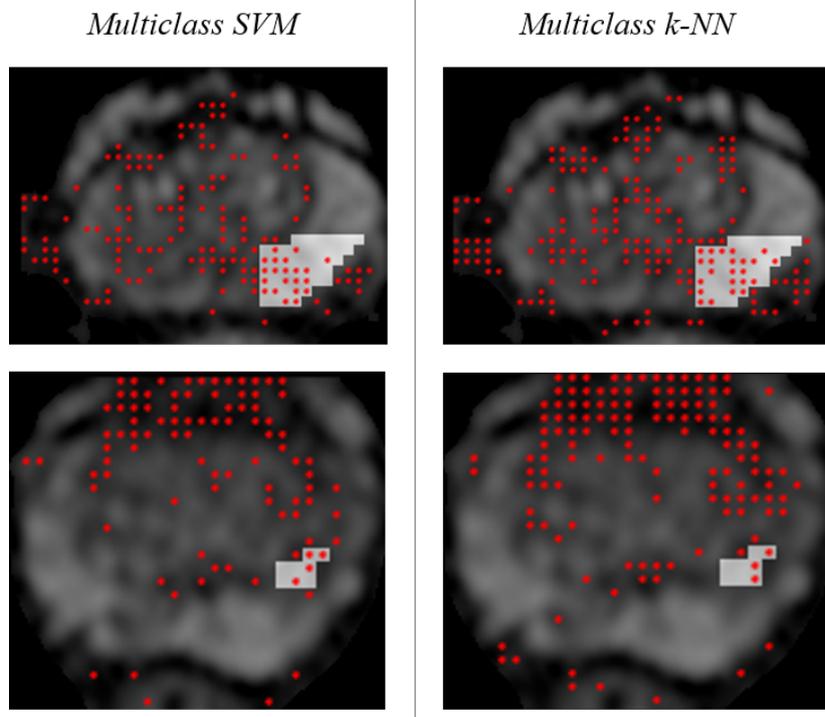
(b) Sensitivity

**Figure 3.20:** Multiclass SVM vs Multiclass k-NN. Classification performances across patients on T2-weighted modality

Observing figures 3.21 and 3.22, multiclass SVM performance is slightly superior in terms of specificity, concerning both image modalities, with respect to multiclass k-NN. Sensitivity is comparable between the two classifiers.



**Figure 3.21:** SVM vs Multiclass SVM. Examples of tumour detection referred to T2-w. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous areas



**Figure 3.22:** SVM vs Multiclass SVM. Examples of tumour detection referred to ADC map. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous areas

Since the recognition of benign elements is still not satisfying, a completely diverse approach will be tested to investigate how this aspect can be improved without jeopardising tumour detection.

It is worth pointing out that decision trees involved in the construction of the training sets did not exclude any available feature (see figure 3.7). For this reason, the complete ensemble of variables will be taken into account in the next phase.

### Feedforward Neural Networks

*Artificial neural networks* are information-processing models, which mimic the functioning of the biological nervous system [65].

The main actor is the *neuron*, i.e. the fundamental core of a neural network responsible for the learning process achievable with massive inter-synaptic connections.

As a biological neuron is made up by a nucleus, enclosed in the cell body receiving incoming signals, and a bunch of dendrites which encounter the information from neighbouring neurons (i.e. the data fed as input to the system), the artificial neuron is characterised by [65]:

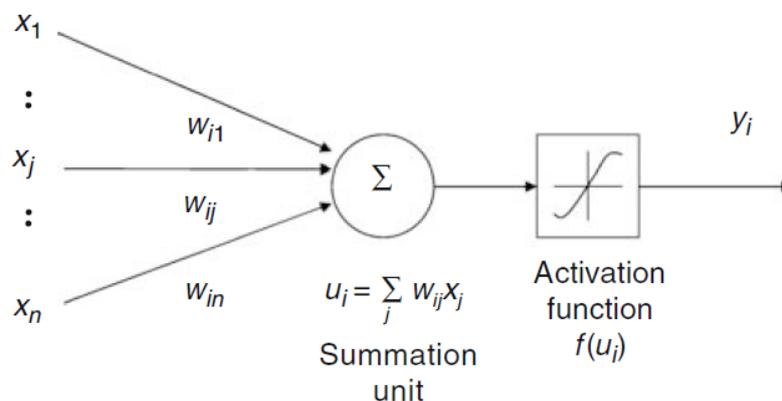
- *summation and activation function*, part of the processing centre to decode and devise the input responses. Mathematically, it embeds an activation function  $f$ , which elaborates the inputs gathered at the summation *node*;
- *input*  $x_i$  is modelled by a scalar weight  $w_i$  multiplied with itself. It can be associated with the electrical input to the biological neuron:

$$y_i = f\left(\sum w_k x_k\right) \quad (3.10)$$

There may be multiple inputs;

- *output*  $y_i$ , which signifies the strength of the electrical pulse moving along the axon;
- *feedback*, it can be provided in certain types of NNs and gives the capability to become adaptive by feeding the output back to the input.

A schematic representation of an artificial neuron, also called *perceptron*, is displayed in figure 3.23.



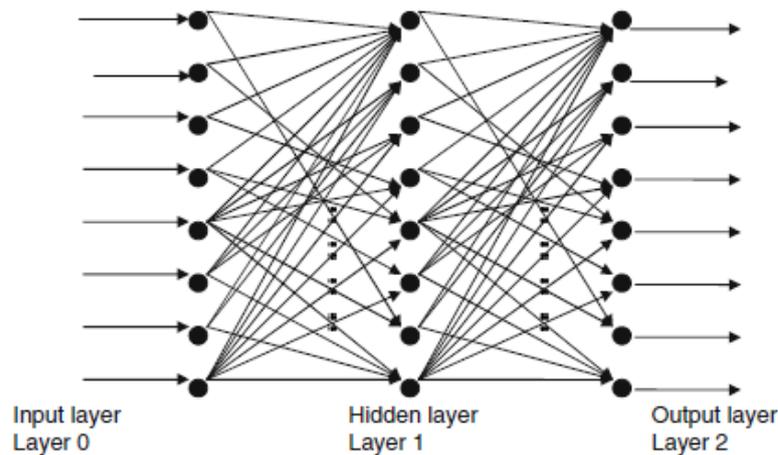
**Figure 3.23:** Representation of an artificial neuron [65]

The learning process requires the modification of neuron weights, in order to make the network output match the true label, based on the error between the desired and computed output. Training is usually performed with general optimisation algorithms, to be effortlessly implemented on a computer.

Connecting more than one perceptron into a network leads to the construction of a *multi-layer perceptron* featured with the *backpropagation algorithm* (see figure 3.24). The latter is constituted by two steps:

1. computation of the derivatives belonging to the network training error with respect to the weights;
2. adjustment of the weights through a gradient descent method, exploiting the error derivatives to reduce the output errors.

Obviously, the backpropagation algorithm supposes the neurons to possess differentiable activation functions.



**Figure 3.24:** Example of a two-layer MLP [65]

Particularly when dealing with non-linear dependencies among the variables, the presence of hidden layers is advisable to increase the connections in the network, thus enhancing its learning ability.

Neural networks have been already exploited to reduce false positives in the case of a computer-aided diagnosis system based on mp-MRI [76]. Seeking the same goal, different architectures for neural networks and training sets are tested to understand the consequent impact on performances.

**Experimental trials** With the aim of discovering the behaviour of feedforward neural networks in the context of this study, the first attempt has been made using decision tree based training set (see section 3.3.1).

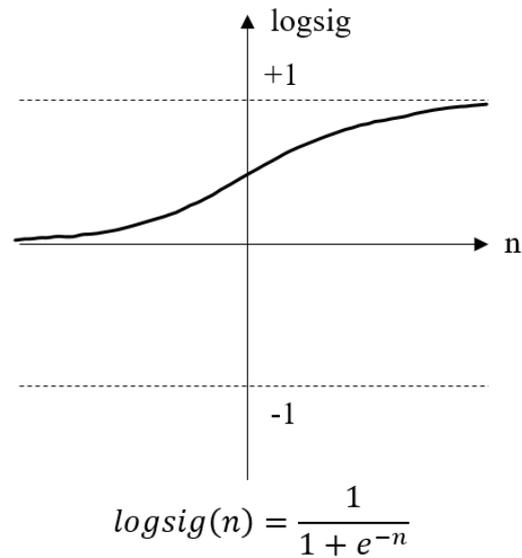
The established setting is the following:

- number of input neurons: 30;
- number of output neurons: 1;
- hidden layers architecture: the structures listed in table 3.15 are tested to explore various scenarios, ranging from 1 to maximum 3 hidden layers;

No. of hidden layers		
<i>1</i>	<i>2</i>	<i>3</i>
5	[15 30]	[5 15 30]
10	[30 15]	[5 30 15]
15	[30 60]	[15 5 30]
20	[60 30]	[15 30 5]
30	[20 30]	[30 5 15]
60	[30 20]	[30 15 5]
	[10 30]	[15 30 60]
	[30 10]	[15 60 30]
	[10 20]	[30 15 60]
	[20 10]	[30 60 15]
	[10 15]	[60 15 30]
	[15 10]	[60 30 15]
	[20 60]	[10 20 30]
	[60 20]	[10 30 20]
		[20 10 30]
		[20 30 10]
		[30 10 20]
		[30 20 10]

**Table 3.15:** List of tested hidden layers (number of neurons in each layer is provided)

- input neurons activation function: *log-sigmoid* in figure 3.25;



**Figure 3.25:** Input neurons log-sigmoid activation function

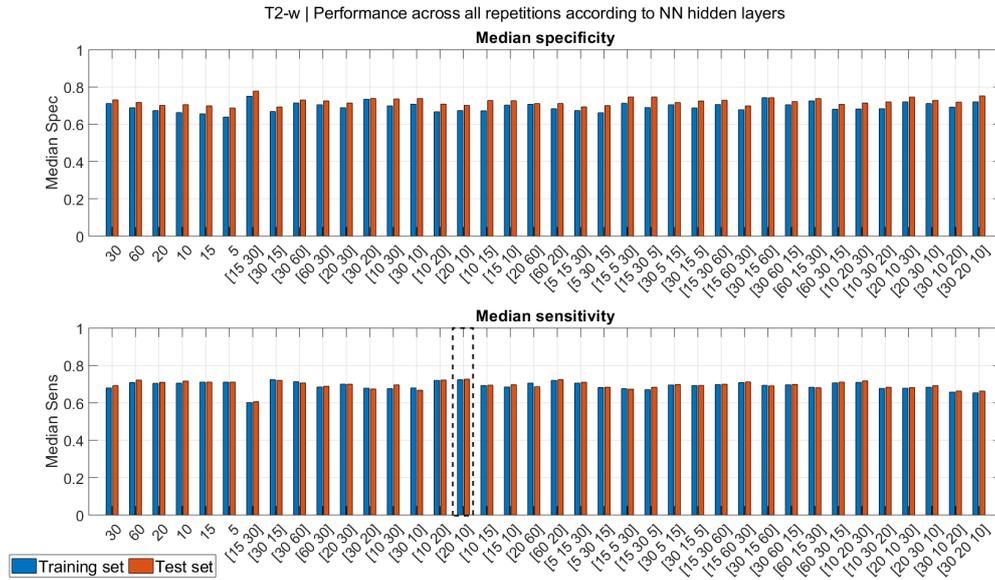
- output neurons encoding function: *linear*;
- backpropagation algorithm: *Levenberg-Marquardt optimization* [77];
- number of repetitions: 10.

The resulting performance of each NN is evaluated in terms of specificity and sensitivity on training and test sets per image type.

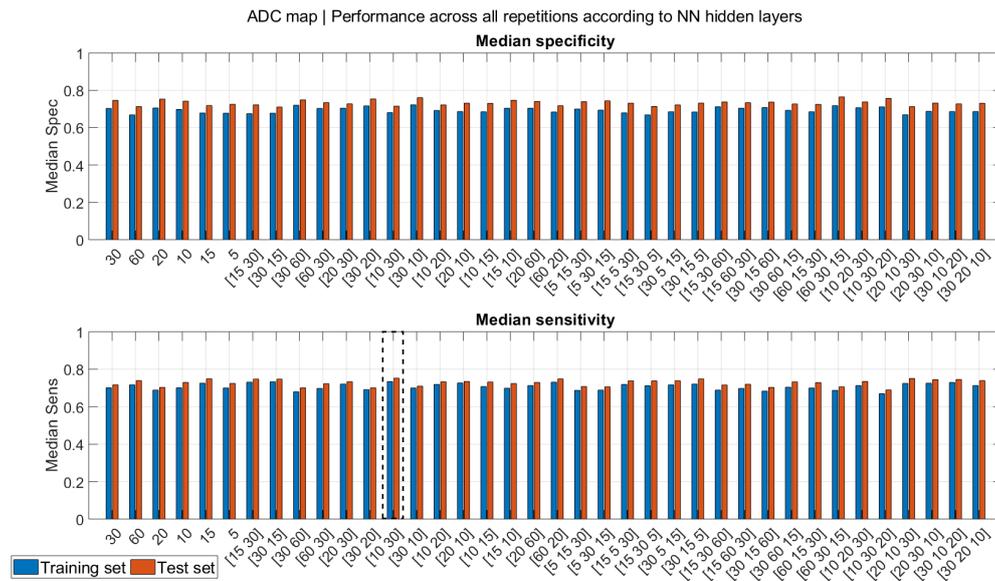
Two criteria are applied to determine the best structures:

1. median sensitivity on test set across all repetitions (here denominated *MaxMedian*);
2. smallest difference between maximum and minimum value of sensitivity on test set across all repetitions (here denominated *MinDiff*).

For both image modalities, median sensitivity and specificity stays above 0.6 for training and test sets, as retrievable from figure 3.28.



(a) T2-weighted



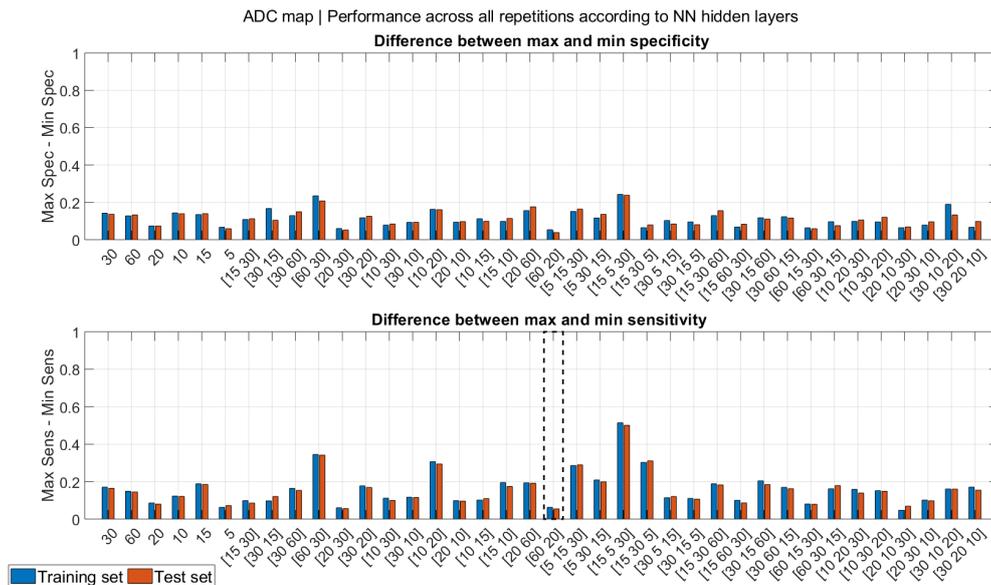
(b) ADC map

**Figure 3.26:** DT based training set - MaxMedian criterion. Performance proper of each NN structure in terms of specificity and sensitivity evaluated on training and test sets. Dashed boxes indicate the structure elected as best

More than a single repetition is performed for each structure: this is justified by the fact that the output can change every time, thus it is important to identify the most reliable configuration. Those NNs satisfying *MinDiff* criterion should fulfil this demand (see figure 3.27).



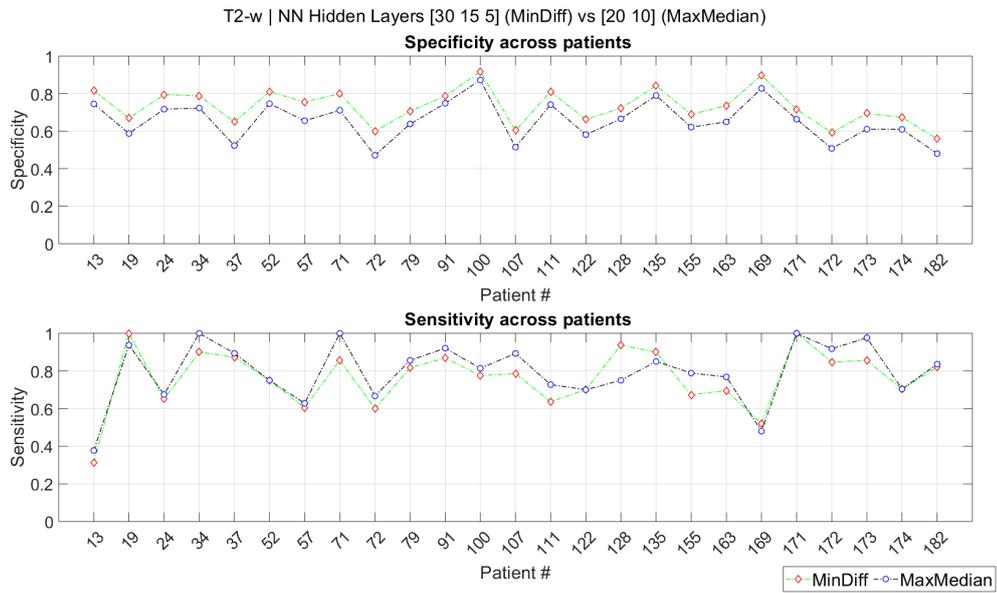
(a) T2-weighted



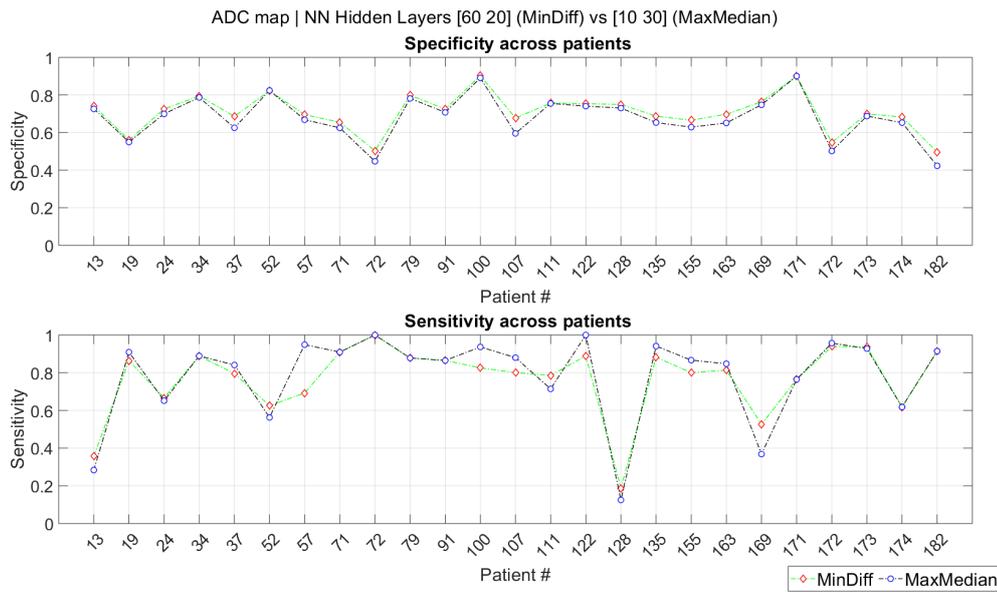
(b) ADC map

**Figure 3.27:** DT based training set - MinDiff criterion. Performance proper of each NN structure in terms of specificity and sensitivity evaluated on training and test sets. Dashed boxes indicate the structure elected as best

In order to give a closer look at subject-wise performance, the outcome of the repetition presenting the greatest sensitivity is examined for both structures, whose results are displayed in figure 3.28.



(a) T2-weighted



(b) ADC map

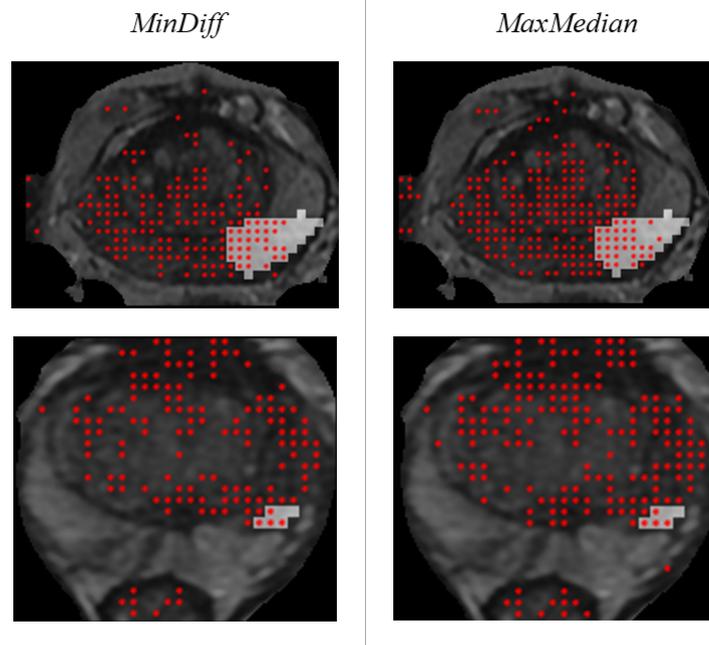
**Figure 3.28:** DT based training set. Sensitivity and specificity across patients considering the best elected structures

While the ability to recognise benign elements seems more stable across patients and greater in the case of the *MinDiff* criterion, malignant samples are detected in a smaller amount in some subjects more than others, considering both selected structures. From the normalised confusion matrices in table 3.16, the *MaxMedian* structure presents higher sensitivity, which is instead pretty balanced with specificity for *MinDiff* structure, concerning the two image modalities.

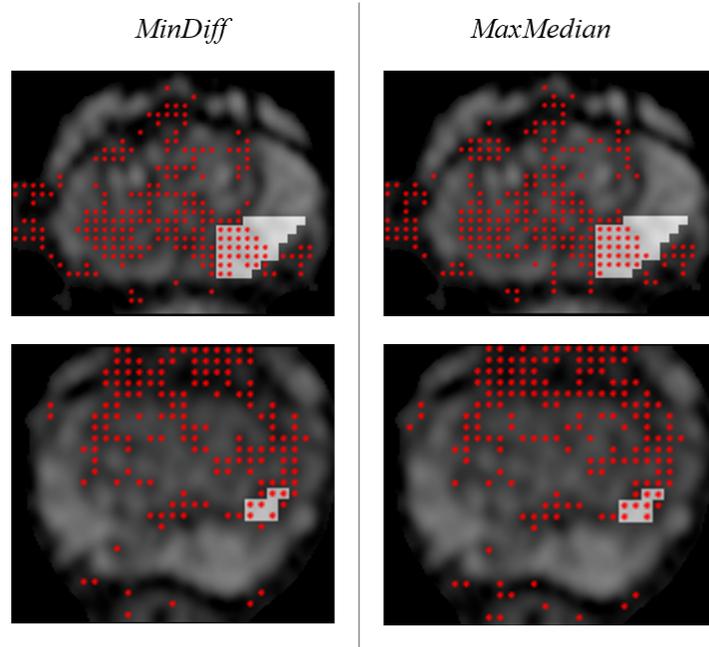
<i>MinDiff</i>							
<b>T2-w</b>							
<i>TRAINING SET</i>				<i>TEST SET</i>			
		<b>True class</b>				<b>True class</b>	
		<b>0</b>		<b>1</b>			
<b>Predicted class</b>	<b>0</b>	<b>0.68</b>	0.26	<b>Predicted class</b>	<b>0</b>	<b>0.73</b>	0.26
	<b>1</b>	0.32	<b>0.74</b>		<b>1</b>	0.27	<b>0.74</b>
						<b>True class</b>	
		<b>0</b>		<b>1</b>			
<b>Predicted class</b>	<b>0</b>	TN	FN	<b>Predicted class</b>	<b>1</b>	FP	TP
	<b>1</b>	FP	TP				
<b>ADC map</b>							
<i>TRAINING SET</i>				<i>TEST SET</i>			
		<b>True class</b>				<b>True class</b>	
		<b>0</b>		<b>1</b>			
<b>Predicted class</b>	<b>0</b>	<b>0.66</b>	0.24	<b>Predicted class</b>	<b>0</b>	<b>0.70</b>	0.22
	<b>1</b>	0.34	<b>0.76</b>		<b>1</b>	0.30	<b>0.78</b>
<b>T2-w</b>							
<i>TRAINING SET</i>				<i>TEST SET</i>			
		<b>True class</b>				<b>True class</b>	
		<b>0</b>		<b>1</b>			
<b>Predicted class</b>	<b>0</b>	<b>0.61</b>	0.22	<b>Predicted class</b>	<b>0</b>	<b>0.65</b>	0.21
	<b>1</b>	0.39	<b>0.78</b>		<b>1</b>	0.35	<b>0.79</b>
<b>ADC map</b>							
<i>TRAINING SET</i>				<i>TEST SET</i>			
		<b>True class</b>				<b>True class</b>	
		<b>0</b>		<b>1</b>			
<b>Predicted class</b>	<b>0</b>	<b>0.64</b>	0.21	<b>Predicted class</b>	<b>0</b>	<b>0.68</b>	0.20
	<b>1</b>	0.36	<b>0.79</b>		<b>1</b>	0.32	<b>0.80</b>

**Table 3.16:** DT based training set. Classification performances obtained considering training and test sets per image modality and best elected NN structures. The reference confusion matrix is reported on the right (1: malignant label, 0: benign label)

Figure 3.29 shows the *MinDiff* criterion to be lightly more effective in limiting the number of false positives for T2-w compared to *MaxMedian*, whereas for ADC map in figure 3.30 there is no apparent difference between performances of NN structures.

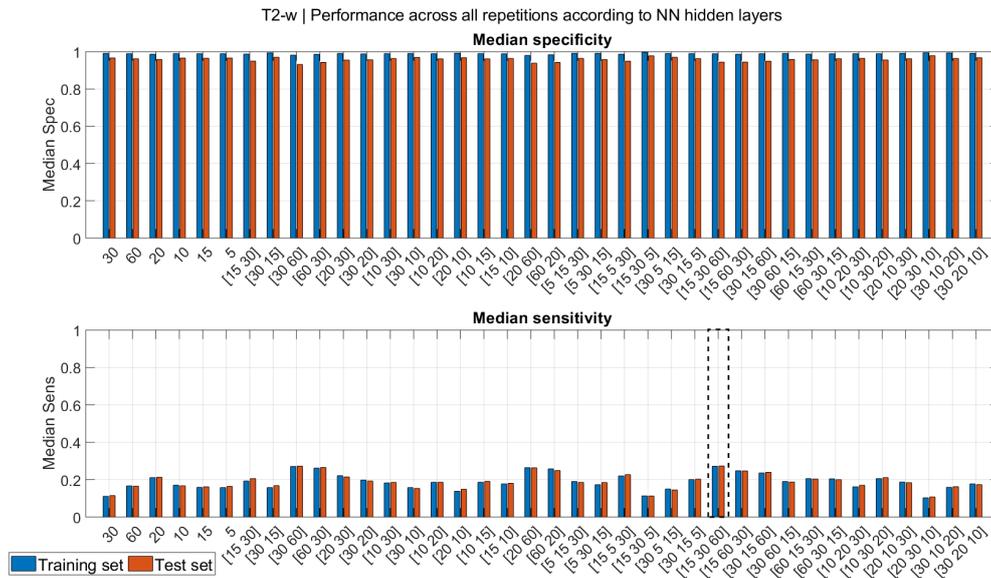


**Figure 3.29:** *DT based training set - T2-w.* Examples of tumour detection according to best elected NN structures. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue

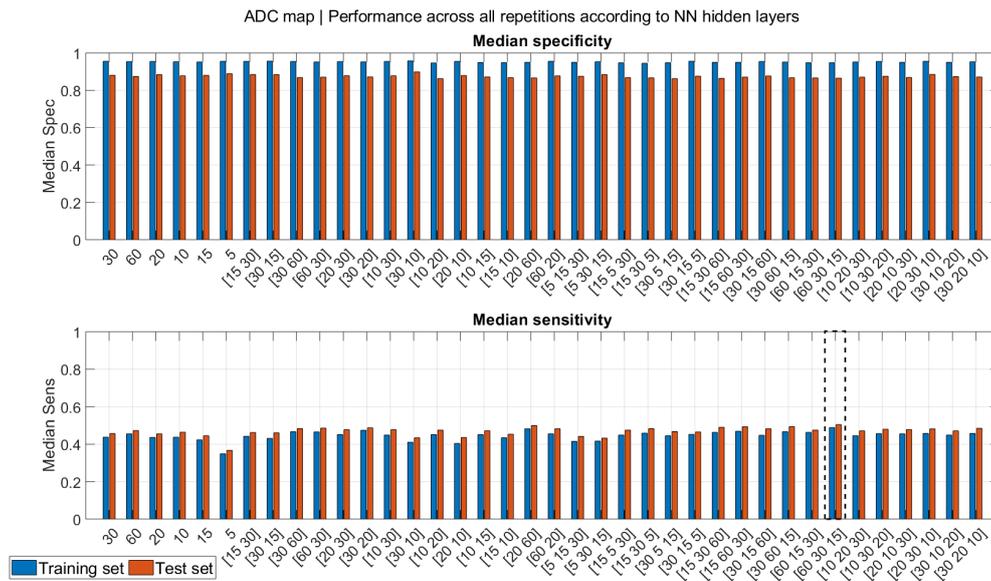


**Figure 3.30:** *DT based training set - ADC map.* Examples of tumour detection according to best elected NN structures. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue

Without any modification to the parameters described earlier in this section, the same structures have been tested using dendrogram based training set (see section 3.4.1), to analyse how the NNs work when fed with a more diversified data set.



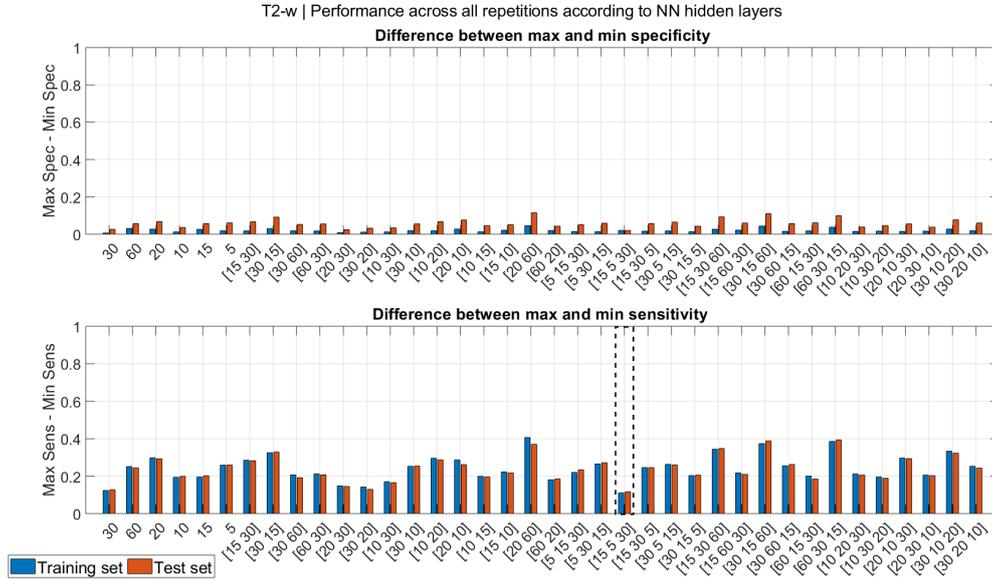
(a) T2-weighted



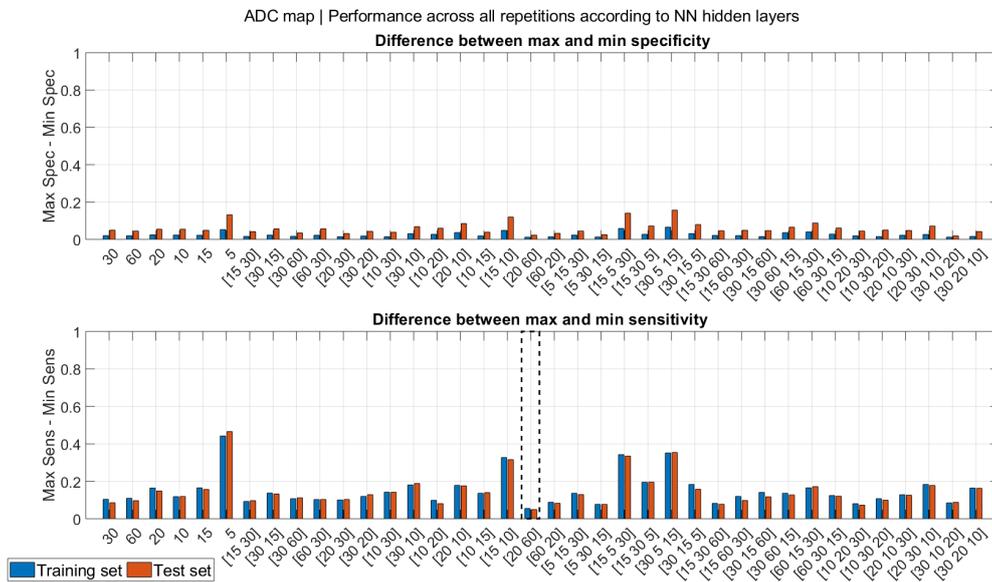
(b) ADC map

**Figure 3.31:** Dendrogram based training set - MaxMedian criterion. Performance proper of each NN structure in terms of specificity and sensitivity evaluated on training and test sets. Dashed boxes indicate the structure elected as best

Figure 3.31 highlights a considerable decrease in the median sensitivity per structure, involving both image modalities.



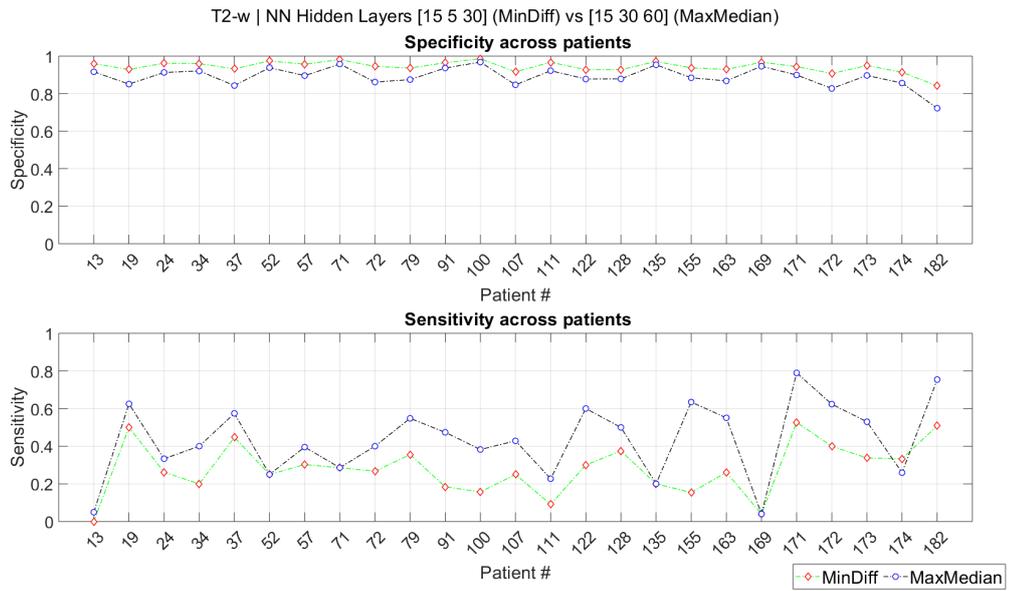
(a) T2-weighted



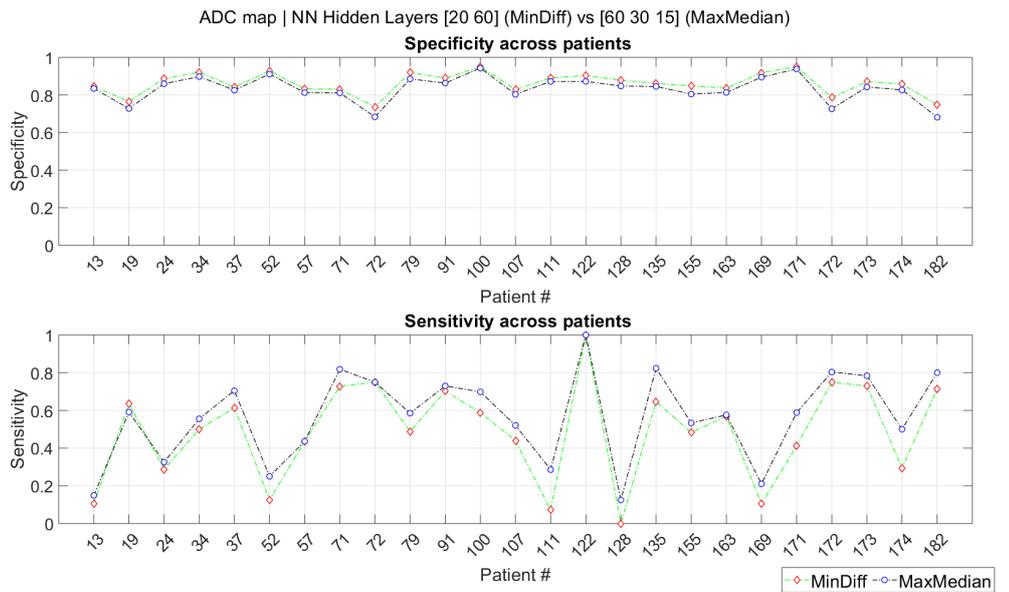
(b) ADC map

**Figure 3.32:** Dendrogram based training set - MinDiff criterion. Performance proper of each NN structure in terms of specificity and sensitivity evaluated on training and test sets. Dashed boxes indicate the structure elected as best

Considering patient-wise performances in figure 3.33, recognition of benign samples is actually ameliorated (it was in fact the desired outcome). Nevertheless, tumour detection appears to be worsened, being very diverse across subjects.



(a) T2-weighted

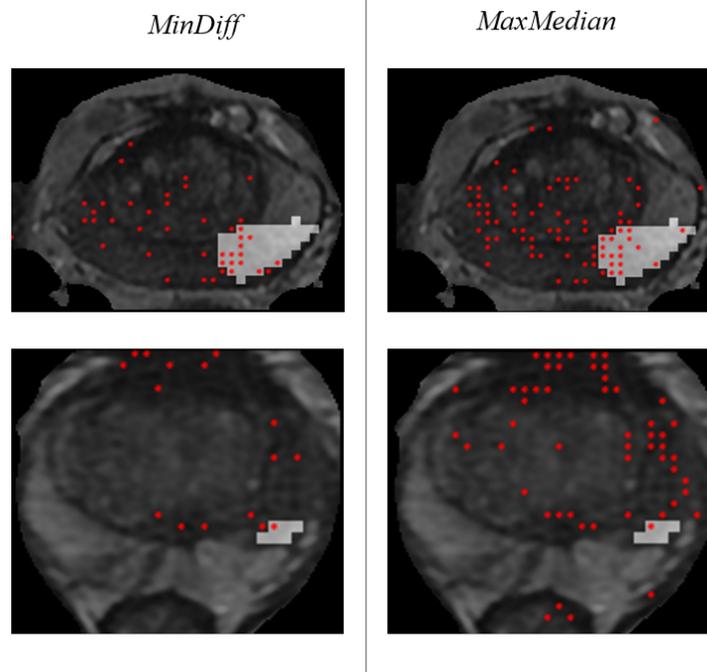


(b) ADC map

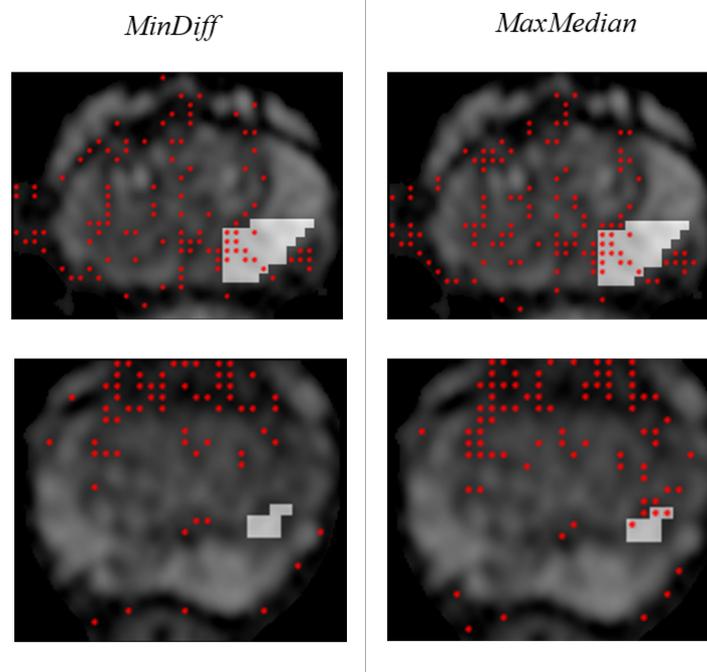
**Figure 3.33:** Dendrogram based training set. Sensitivity and specificity across patients considering the best elected structures

Observing performance on training and test sets given in table 3.17 specific for the elected structures, specificity remains high for both criteria, while sensitivity is considerably lower for *MinDiff*.





**Figure 3.34:** Dendrogram based training set - T2-w. Examples of tumour detection according to best elected NN structures. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue



**Figure 3.35:** Dendrogram based training set - ADC map. Examples of tumour detection according to best elected NN structures. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue

To summarise the results obtained with these last elaborations, table 3.18 provides the maximum or minimum values for specificity and sensitivity considering all structures along with an evaluation of the performance variability among the study population.

		DT based training set		Dendrogram based training set	
		<b>T2</b>	<b>ADC</b>	<b>T2</b>	<b>ADC</b>
<b>Performance on test set*</b>	<i>Sens</i>	$\leq 0.8$	$\leq 0.7$	$\leq 0.4$	$\leq 0.5$
	<i>Spec</i>	$\leq 0.7$	$\leq 0.7$	$\geq 0.95$	$\geq 0.8$
<b>Variability across subjects</b>	<i>Sens</i>	Low	Low	High	High
	<i>Spec</i>	Low	Low	Low	Low

**Table 3.18:** NNs performance comparison between the two employed training sets, given per image modality.

\*The reported number is the upper or lower bound of existing values (regarding sensitivity and specificity)

What emerges from table 3.18 is that using dendrogram based training set causes a significant worsening in the recognition of the positive class, although healthy tissue is better identified.

The ideal situation would be to reach such good performances also when detecting malignant lesions without being deprived of the recognition characterising the benign class.

Next approach is thought to achieve this purpose, even to find out if there exist benign classes which are better identified than others.

### Cascade of Feedforward Neural Networks

A cascade of neural networks is implemented to exploit the ability of an ensemble of NNs to recognise each class one by one.

This is performed by subsequently removing the benign class with highest recognition and lowest number of misclassified elements as malignant.

As a consequence, reducing at each step the classes involved in the training set lowers the variability and simplifies the learning process, hopefully leading to an improved performance.

A multiclass data set is then required, in fact dendrogram based training set is exploited, enabling greater representation of healthy tissue.

The elements correctly classified as benign by each neural network in the cascade will fall in the true negatives, while those wrongly identified as malignant will be considered false positives. Both of them will contribute in the performance of the last neural network in the cascade, presented in form of normalised confusion matrices.

The parameters set for all the networks are the same listed in section 3.4.2, except for the hidden layers structure, which is fixed at 1 hidden layer of 30 neurons, as preliminary trial. In the following tables, label 1 denotes malignant class, while labels from 2 to 5 are associated with the classes of benign tissue created from the clusters in figure 3.14.

*TRAINING SET*

$NN_1$		<i>Predicted Class</i>				
		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<i>True class</i>	<b>1</b>	<b>0.436</b>	0.494	0.065	0.005	0.000
	<b>2</b>	0.186	<b>0.742</b>	0.065	0.005	0.001
	<b>3</b>	0.005	0.379	<b>0.538</b>	0.077	0.000
	<b>4</b>	0.000	0.001	0.090	<b>0.909</b>	0.000
	<b>5</b>	0.000	0.005	0.013	0.046	<b>0.936</b>

$NN_2$		<i>Predicted Class</i>			
		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
<i>True class</i>	<b>1</b>	<b>0.444</b>	0.497	0.058	0.001
	<b>2</b>	0.188	<b>0.781</b>	0.031	0.000
	<b>3</b>	0.009	0.421	<b>0.495</b>	0.076
	<b>4</b>	0.000	0.000	0.097	<b>0.903</b>

$NN_3$		<i>Predicted Class</i>		
		<b>1</b>	<b>2</b>	<b>3</b>
<i>True class</i>	<b>1</b>	<b>0.415</b>	0.551	0.033
	<b>2</b>	0.158	<b>0.827</b>	0.015
	<b>3</b>	0.004	0.478	<b>0.518</b>

$NN_4$		<i>Predicted Class</i>	
		<b>1</b>	<b>2</b>
<i>True class</i>	<b>1</b>	<b>0.847</b>	0.153
	<b>2</b>	0.313	<b>0.687</b>

**Table 3.19:** ADC map - 1<sup>st</sup> NNs Cascade. Normalised confusion matrices belonging to each neural network. Rows enclosed in the bold boxes highlight the class which is not included in the training set of the successive neural network (1: malignant label, benign labels from 2 to 5)

Final performance on test set is reported in table 3.20.

*TEST SET*

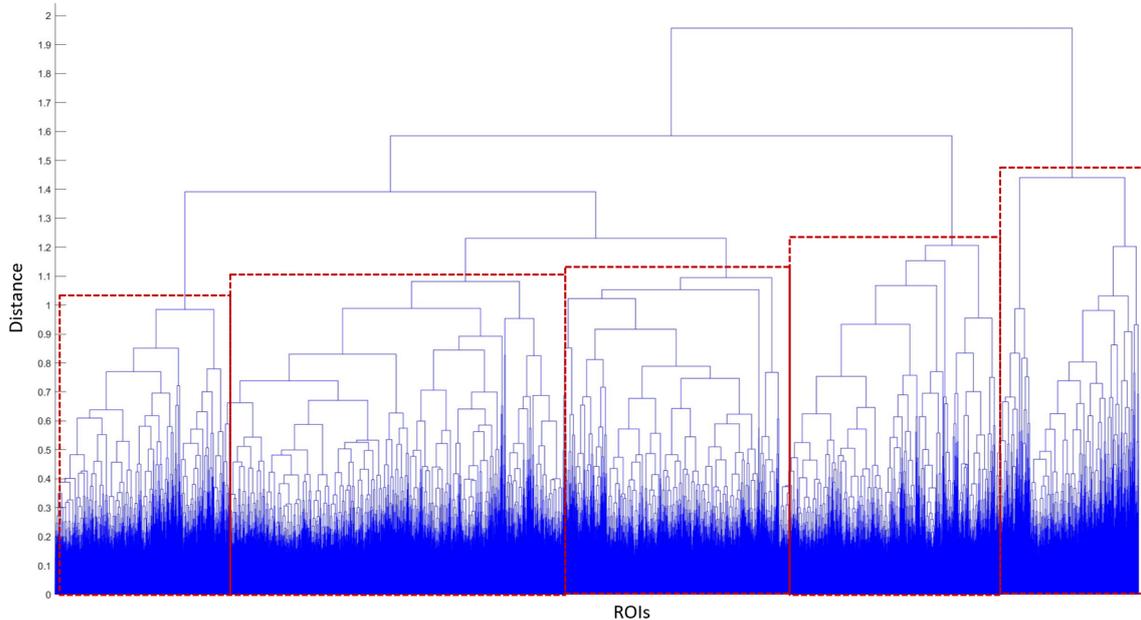
		<i>Predicted Class</i>	
		<b>1</b>	<b>0</b>
<i>True class</i>	<b>1</b>	<b>0.825</b>	0.175
	<b>0</b>	0.387	<b>0.613</b>

**Table 3.20:** *ADC map - 1<sup>st</sup> NNs Cascade.* Normalised confusion matrix belonging to test set (1: malignant class, 0: benign class)

It is evident from table 3.20 there has been no improvement in detecting benign samples considering final performance of the NNs cascade.

Starting from remaining test set used in the last neural network of the cascade (i.e.  $NN_4$ ) while keeping classified elements from previous structures, dendrogram clustering is applied only on benign samples to constitute a new training set, trying to provide a greater representation of the negative class. Malignant samples are the same considered in dendrogram based training set.

Figure 3.36 shows the resulting dendrogram on benign ROIs along with clusters.



**Figure 3.36:** *ADC map.* Dendrogram clustering on benign samples of  $NN_4$  test set. Dashed red boxes enclose clusters

A second cascade is implemented exploiting the classes formed from each cluster in figure 3.36. The new training set for benign samples is built as in equation 3.8, by randomly extracting elements from every cluster.

Table 3.21 specifies performance on training set and highlights the class, excluded from the succeeding training set, based on lowest number of benign elements classified as malignant.

*TRAINING SET*

$NN_5$		<i>Predicted Class</i>					
		1	2	3	4	5	6
<i>True class</i>	<b>1</b>	<b>0.036</b>	0.340	0.313	0.214	0.082	0.015
	2	0.100	<b>0.755</b>	0.140	0.005	0.000	0.000
	3	0.008	0.233	<b>0.494</b>	0.218	0.045	0.003
	4	0.004	0.044	0.396	<b>0.495</b>	0.059	0.003
	5	0.000	0.003	0.108	0.537	<b>0.323</b>	0.029
	6	0.000	0.001	0.037	0.155	0.473	<b>0.333</b>

$NN_6$		<i>Predicted Class</i>				
		1	2	3	4	5
<i>True class</i>	<b>1</b>	<b>0.087</b>	0.418	0.363	0.129	0.003
	2	0.078	<b>0.891</b>	0.031	0.000	0.000
	3	0.005	0.392	<b>0.524</b>	0.078	0.000
	4	0.003	0.046	0.432	<b>0.471</b>	0.049
	5	0.000	0.005	0.315	0.540	<b>0.140</b>

$NN_7$		<i>Predicted Class</i>			
		1	2	3	4
<i>True class</i>	<b>1</b>	<b>0.221</b>	0.553	0.213	0.014
	2	0.059	<b>0.903</b>	0.038	0.000
	3	0.003	0.371	<b>0.608</b>	0.019
	4	0.004	0.094	0.505	<b>0.397</b>

$NN_8$		<i>Predicted Class</i>		
		1	2	3
<i>True class</i>	<b>1</b>	<b>0.445</b>	0.508	0.047
	2	0.021	<b>0.946</b>	0.033
	3	0.001	0.442	<b>0.556</b>

$NN_9$		<i>Predicted Class</i>	
		1	2
<i>True class</i>	<b>1</b>	<b>0.853</b>	0.147
	2	0.762	<b>0.238</b>

**Table 3.21:** ADC map - 2<sup>nd</sup> NNs Cascade. Normalised confusion matrices belonging to each neural network training set. Rows enclosed in the bold boxes highlight the class which is not included in training set of the successive neural network (1: malignant label, benign labels from 2 to 6)

The confusion matrix regarding training set of the last neural network in the second cascade exhibits a considerable decrease in the recognition of benign samples.

*TEST SET*

		<i>Predicted Class</i>	
		<b>1</b>	<b>0</b>
<i>True class</i>	<b>1</b>	<b>0.757</b>	0.243
	<b>0</b>	0.340	<b>0.660</b>

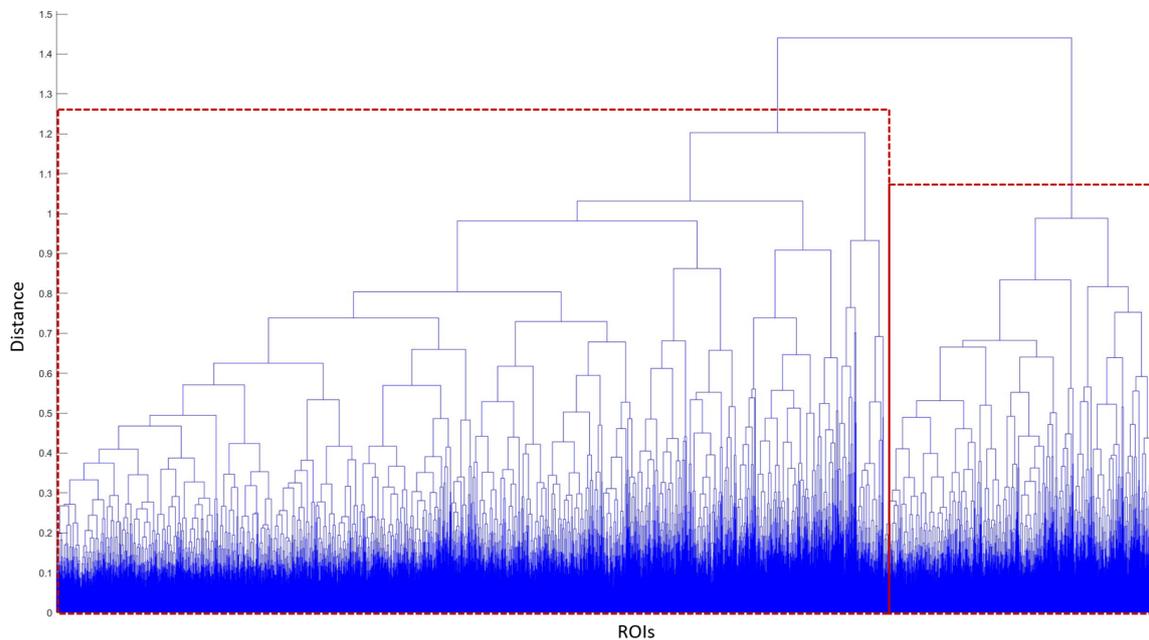
**Table 3.22:** ADC map - 1<sup>st</sup> and 2<sup>nd</sup> NNs Cascades. Normalised confusion matrix considering complete test set (1: malignant class, 0: benign class)

Performances presented in table 3.22 are obtained considering the outcome of  $NN_9$  and those elements already classified by the previous NNs.

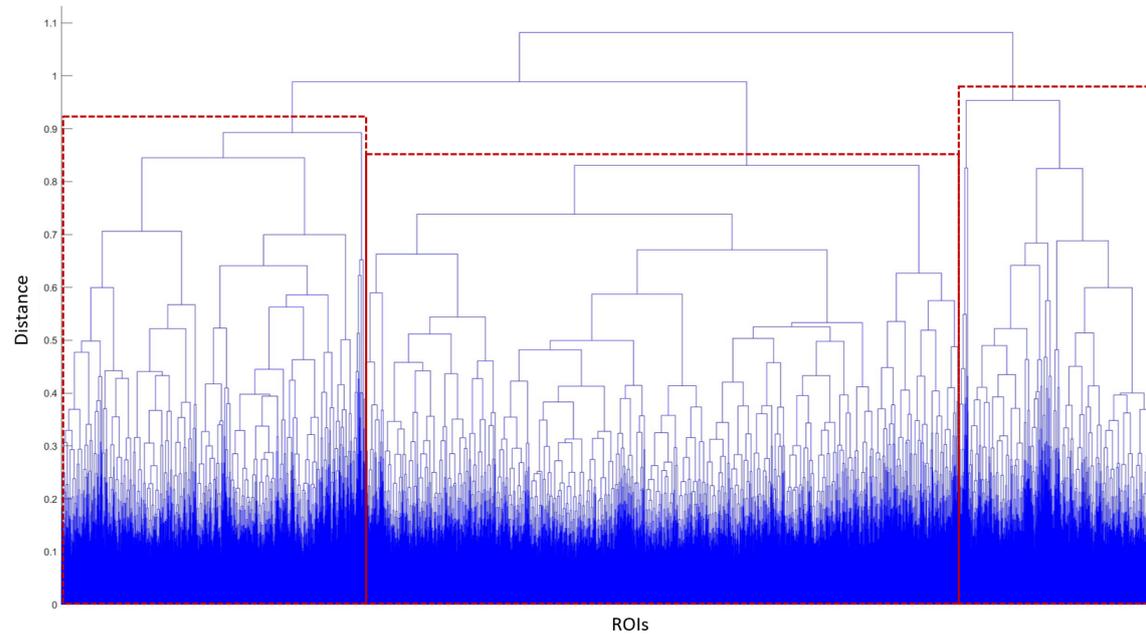
A slight amelioration concerning specificity is found, but it is not sufficient to reach an acceptable outcome.

Given the worsening of the final network in the second cascade, another attempt is made by clustering benign samples of the remaining test set from  $NN_8$ .

Dendrograms belonging to classes 2 and 3 (involved in  $NN_8$ ) are cut in order to extend the presence of these classes in the training set, creating a class from each of the newly formed clusters (see figures 3.37 and 3.38). Samples are randomly extracted from each class according to equation 3.8.



**Figure 3.37:** ADC map. Dendrogram on class 2 from remaining test set of  $NN_4$ . Dashed red boxes enclose clusters



**Figure 3.38:** *ADC map*. Dendrogram on class 3 from remaining test set of  $NN_4$ . Dashed red boxes enclose clusters

Table 3.23 reports the confusion matrices of each neural network in the 3<sup>rd</sup> cascade regarding training, whereas table 3.24 shows final performances on entire test set considering all three cascades.

## TRAINING SET

$NN_{10}$		<i>Predicted Class</i>					
		1	2	3	4	5	6
<i>True class</i>	1	<b>0.312</b>	0.345	0.155	0.078	0.097	0.013
	2	0.010	<b>0.605</b>	0.322	0.051	0.012	0.000
	3	0.015	0.338	<b>0.483</b>	0.128	0.035	0.000
	4	0.001	0.099	0.554	<b>0.328</b>	0.015	0.003
	5	0.000	0.001	0.029	0.208	<b>0.726</b>	0.036
	6	0.000	0.000	0.019	0.119	0.608	<b>0.254</b>

$NN_{11}$		<i>Predicted Class</i>				
		1	2	3	4	5
<i>True class</i>	1	<b>0.309</b>	0.417	0.217	0.049	0.009
	2	0.041	<b>0.555</b>	0.372	0.032	0.000
	3	0.006	0.388	<b>0.559</b>	0.046	0.000
	4	0.000	0.055	0.423	<b>0.506</b>	0.015
	5	0.000	0.006	0.090	0.405	<b>0.499</b>

$NN_{12}$		<i>Predicted Class</i>			
		1	2	3	4
<i>True class</i>	1	<b>0.372</b>	0.454	0.165	0.009
	2	0.015	<b>0.737</b>	0.244	0.004
	3	0.001	0.359	<b>0.605</b>	0.035
	4	0.001	0.071	0.440	<b>0.488</b>

$NN_{13}$		<i>Predicted Class</i>		
		1	2	3
<i>True class</i>	1	<b>0.559</b>	0.417	0.024
	2	0.006	<b>0.965</b>	0.028
	3	0.001	0.599	<b>0.400</b>

$NN_{14}$		<i>Predicted Class</i>	
		1	2
<i>True class</i>	1	<b>0.912</b>	0.088
	2	0.808	<b>0.192</b>

**Table 3.23:** ADC map - 3<sup>rd</sup> NNs Cascade. Normalised confusion matrices belonging to each neural network. Rows enclosed in the bold boxes highlight the class which is removed in the training set of the successive neural network (1: malignant label, benign labels from 2 to 6)

*TEST SET*

		<i>Predicted Class</i>	
		<b>1</b>	<b>0</b>
<i>True class</i>	<b>1</b>	<b>0.812</b>	0.188
	<b>0</b>	0.370	<b>0.630</b>

**Table 3.24:** ADC map - 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> NNs Cascade. Normalised confusion matrix considering complete test set (1: malignant class, 0: benign class)

Although the representation of benign samples has been incremented, the expected improvement about their detection is still not accomplished.

Given these results, a further trial is performed changing the method to extract elements for training set construction.

Starting with test set proper of  $NN_4$  in table 3.19 and considering its dendrogram in figure 3.36, the same clusters are taken into account. Each of them (called *reference cluster*) is divided in 10 subclusters, and a number  $n$  of elements is picked randomly from each using equation 3.11, in which:

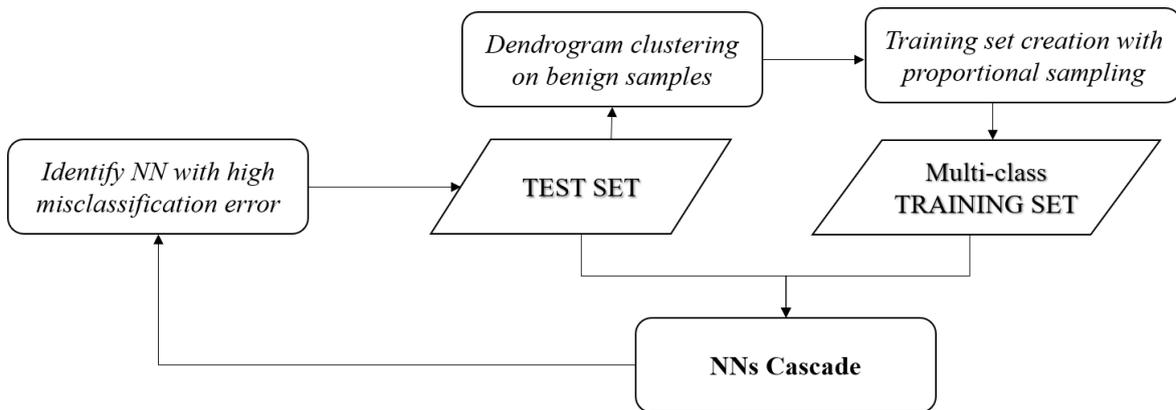
- $n_c$  is number of elements in the current subcluster;
- $M$  is number of malignant samples in dendrogram based training set (see table 3.8);
- $n_{ref}$  is total number of elements in the reference cluster.

$$n = \frac{n_c \times M}{n_{ref}} \quad (3.11)$$

Malignant data for training set are not changed (i.e. equal to the samples from dendrogram based training set) and joined with the new benign classes.

The aforesaid procedure to build the training set will be denominated *proportional sampling*.

A schematic representation of the approach employed in each successive cascade is presented in figure 3.39.



**Figure 3.39:** ADC map - NNs cascade with proportional sampling. Schematic representation of the procedure adopted to implement cascade of NNs

In particular, the first step is to identify the network presenting a low misclassification error regarding benign samples mistaken as malignant. When this error starts to increase with respect to the previous NNs in the cascade, the approach proceeds as follows:

1. benign samples from test set of the elected network undergo dendrogram clustering;
2. cutting level is established;
3. proportional sampling is used to construct training set;
4. NN is fed with this training set and the test set from which it derives.

As a matter of fact, the aim is to limit the presence of false positives, sampling the test set featured with a considerable misidentification.

Accounting for training set performance of each network in the cascade, the class characterised by lowest number of benign elements predicted as malignant is not included in the next NN training set.

Samples classified as benign truly belonging to the benign class will be considered true negatives, whilst those with actual malignant class will be denoted as false positives. All of them will contribute in the performance of the last network of the cascade.

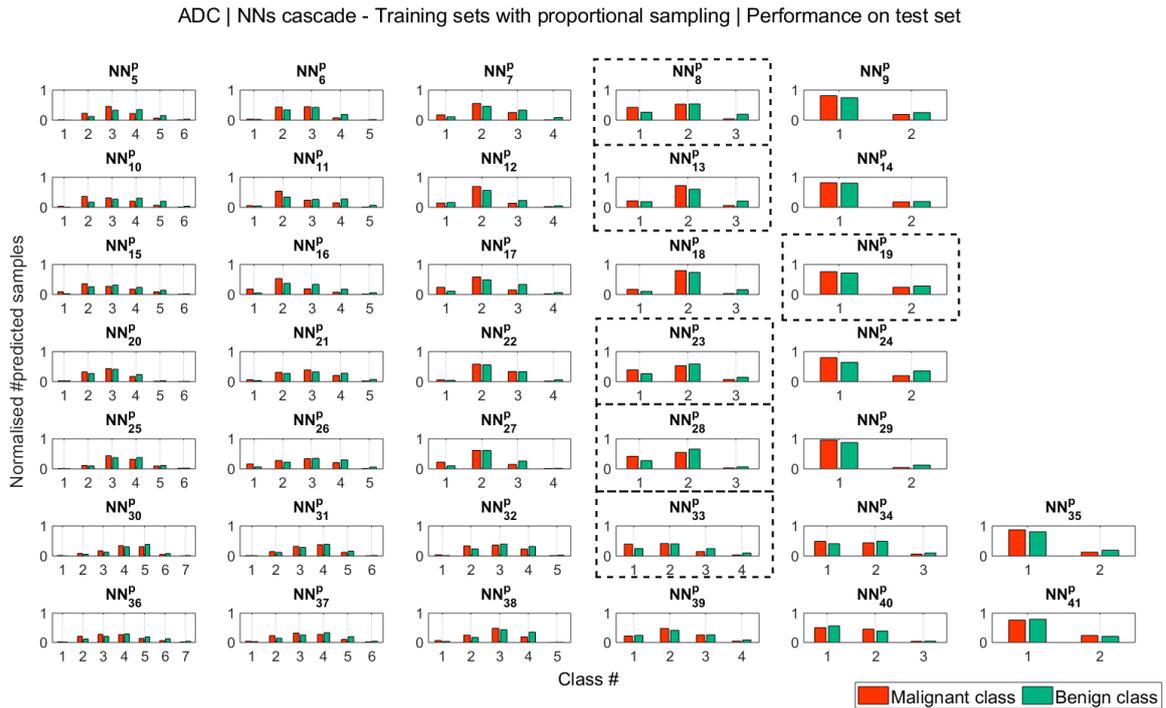
Dendrograms and their clusters proper of this phase are reported in the appendix (section 3.5).

Performances resulting from training set of each neural network can be found in section 3.5 of the appendix.

In order to evaluate misclassification errors, table 3.25 is constructed on test set of every single network giving the normalised number of malignant and benign samples recognised according to NN prediction.

		<i>Predicted Class</i>					
		<b>1</b>	<b>2</b>	<b>3</b>	<b>...</b>	<b>...</b>	<b>n</b>
<i>True class</i>	<b>Malignant</b>						
	<b>Benign</b>						

**Table 3.25:** Template providing normalised number of malignant and benign elements per NN prediction. Class 1 denotes malignant samples, while benign classes are denoted from 2 to  $n$



**Figure 3.40:** ADC map - NNs cascade with proportional sampling. Each row comprises a NNs cascade. The bars represent the normalised number of malignant and benign samples according to NN prediction on test set.

Dashed boxes highlight the test set whose samples will be used in the creation of the training set for the following cascade (1: malignant label, benign labels from 2 to 7)

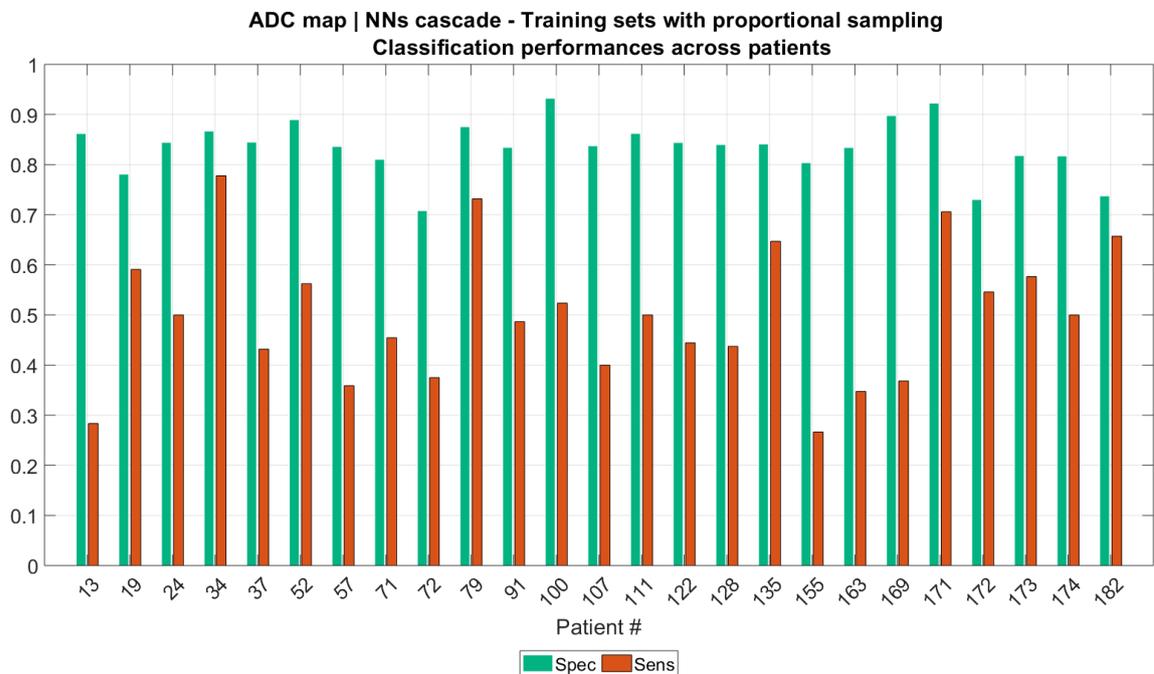
As discernible from table 3.26, there is a significant enhancement in the correct identification of the negative class thanks to the implementation of 8 NNs cascades. Nevertheless, the sensitivity is consistently diminished, resulting around 0.5.

*TEST SET*

		<i>Predicted Class</i>	
		<b>1</b>	<b>0</b>
<i>True class</i>	<b>1</b>	<b>0.49</b>	0.51
	<b>0</b>	0.17	<b>0.83</b>

**Table 3.26:** ADC map - NNs cascade with proportional sampling. Normalised confusion matrix considering complete test set (1: malignant class, 0: benign class)

To closely observe the outcome of this approach in a patient-wise fashion, figure 3.41 provides sensitivity and specificity per subject.



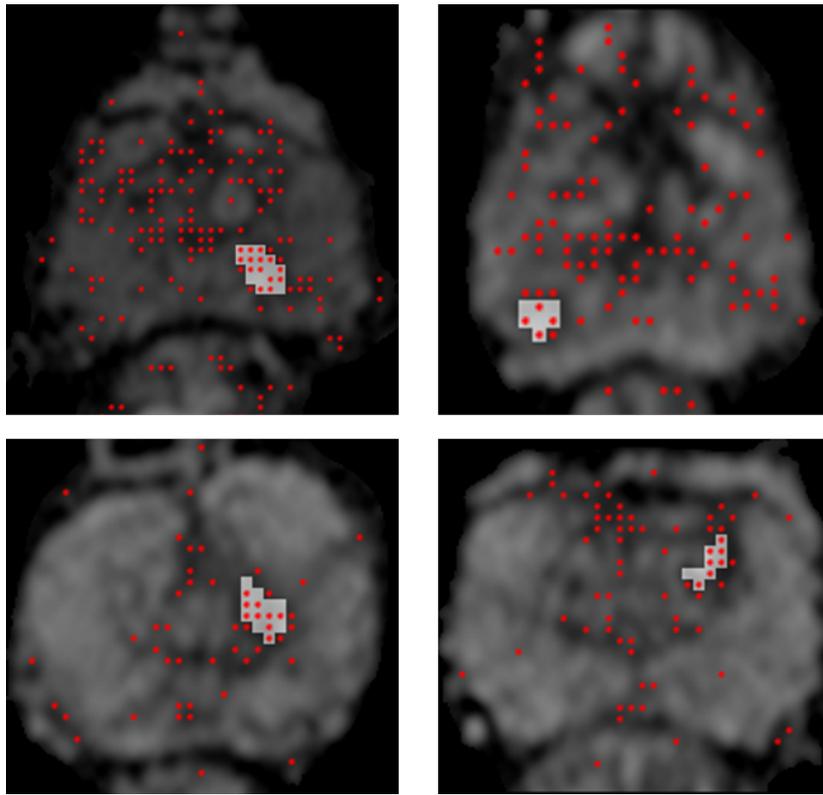
**Figure 3.41:** ADC map - NNs Cascade with Proportional Sampling. Patient-wise sensitivity (*Sens*) and specificity (*Spec*)

Specificity stays above 0.7 for the entire study population, whereas sensitivity varies across patients in a much broader range (roughly from 0.3 to 0.8).

Some examples of tumour detection realised by NNs cascade with proportional sampling are exhibited in figure 3.42.

Several falsely recognised benign samples undermine the correct identification of cancerous zones, as visible in the upper left corner of figure 3.42.

Since the tumour in the upper right corner of figure 3.42 is small, it could be more probably disregarded as a malignant area.

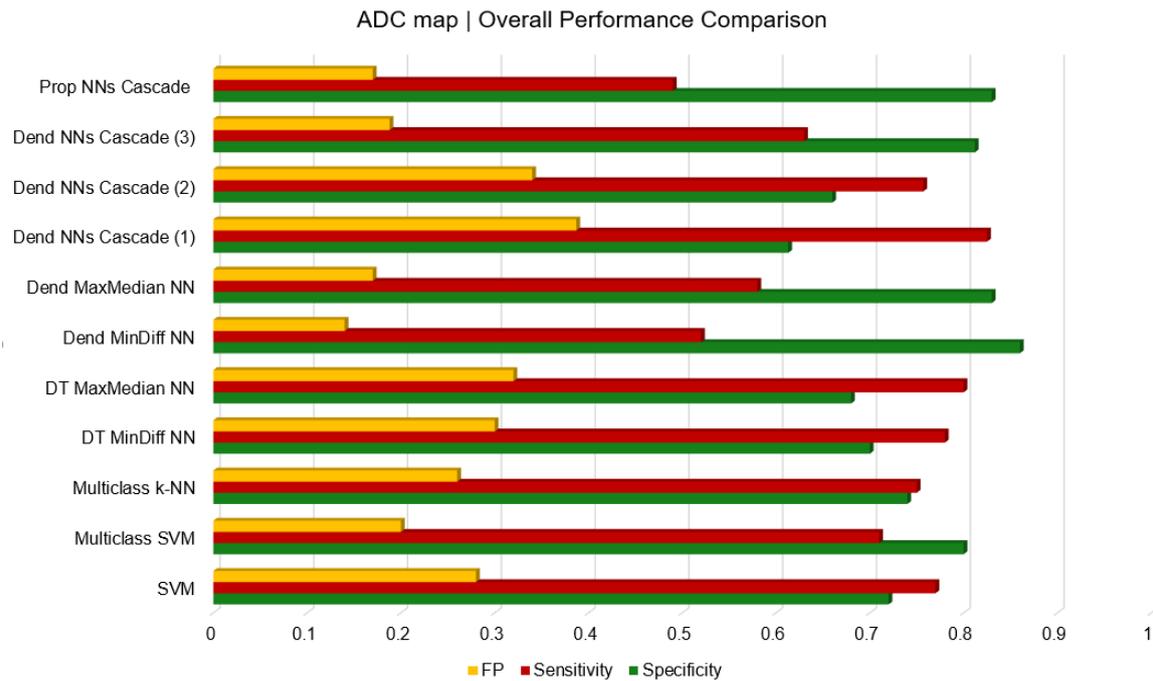


**Figure 3.42:** ADC map - NNs cascade with proportional sampling. Examples of tumour detection. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous areas

### 3.5 Overall Comparison

Throughout this study, several supervised machines have been tested with the aim of achieving a good performance in detecting malignant areas, as well as in limiting the amount of benign samples misidentified as cancerous.

Performances accomplished by each classifier are reported in figure 3.43 for ADC map and in figure 3.44 for T2-weighted, in terms of false positives, sensitivity and specificity.



**Figure 3.43:** ADC map. Comparison in terms of number of false positives, sensitivity and specificity considering all classifiers. (DT: Decision tree based training set, Dend: Dendrogram based training set, Prop: Proportional sampling training set)

Neural networks enable to recognise more correctly benign samples employing dendrogram based training set and in the case of NNs cascade with proportional sampling with ADC map.

On the other hand, multiclass SVM shows higher specificity, which competes with NNs results.

Referring to T2-weighted, all classifiers but dendrogram based training set NNs present a balanced performance between sensitivity and specificity, approximately around 0.7. The best outcome regarding specificity is accomplished by NNs with dendrogram based training set, even though sensitivity is much decreased.

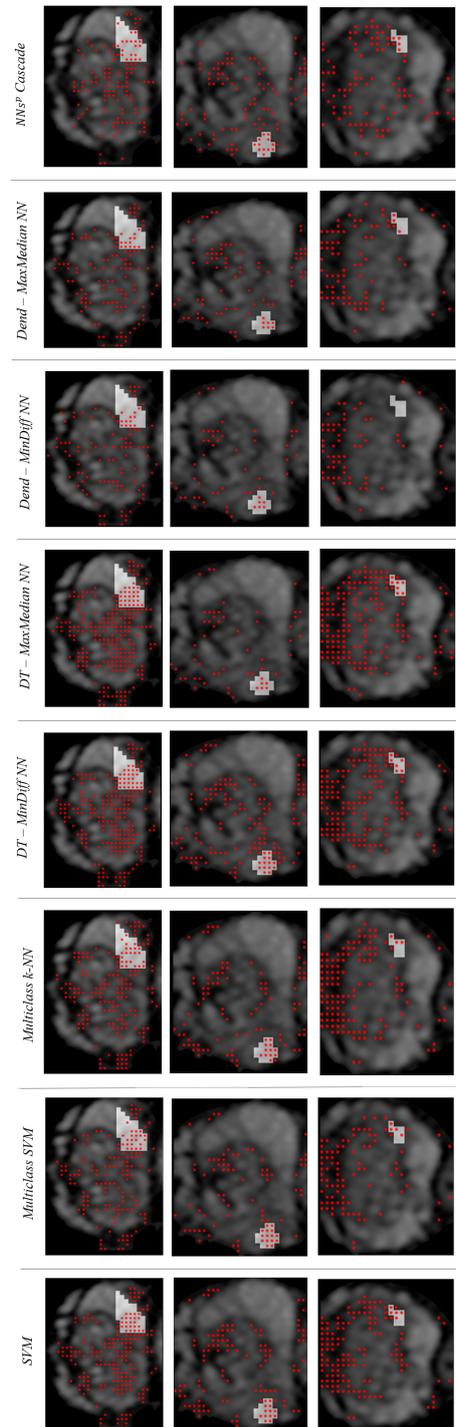


**Figure 3.44:** *T2-weighted.* Comparison in terms of number of false positives, sensitivity and specificity considering all classifiers. (DT: Decision tree based training set, Dend: Dendrogram based training set)

False positives are in general more significant with SVM, multiclass SVM and multiclass k-NN and NNs using decision tree based training set for both image modalities. This could be explained mainly by two reasons:

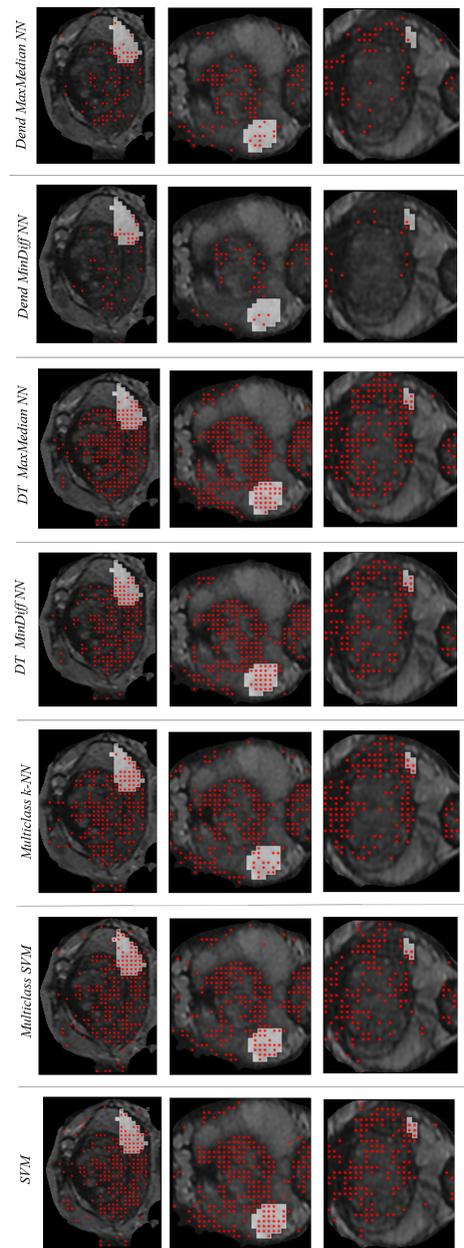
- dendrogram based training set embodies a more complete representation of healthy regions, improving their differentiation;
- neural networks outperform the other classifiers due to their structure and learning process which allow to discover inner relationships among the samples and exploiting them to recognise the true label.

Some representative outcomes of tumour detection per classifier and image type are provided in figure 3.45 and 3.46, analysing subjects with decreasing tumour size.



**Figure 3.45:** *ADC map.* Examples of tumour detection performed by all classifiers, referring to large, medium and small cancerous areas. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue

Regarding ADC map, misclassified ROIs result restricted especially in the case of the medium-sized lesion, although detecting the small tumour becomes more and more difficult.



**Figure 3.46:** *T2-weighted.* Examples of tumour detection performed by all classifiers, referring to large, medium and small cancerous areas. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue

In the case of T2-weighted, false positives considerably worsen image appearance as they are present in a great amount except in the case of dendrogram based training set NNs, where also sensitivity is much lessened.

These are just few examples to visually provide an idea of the performances: it is essential to remember every subject is unique and cannot be interpreted as a generalisable result.

All the trials performed in this study outline the complexity found in constructing a supervised classifier able to identify cancerous lesions.

This complexity can be attributed to:

- imbalanced malignant and benign classes;
- huge heterogeneity inherent benign samples and their similarity with cancerous areas;
- intra and inter subjects variability, including even malignant samples.

To overcome these aspects, the adopted approaches comprised changes in the training set together with the classifiers in the attempt to retrieve a good combination.

The most decisive step has revealed the creation of the training set, as performances of the classifiers radically modified. Of course, feeding each machine with a diverse information content causes the generation of a different model which may turn out to be more or less efficient in classifying than the previous ones.

Certainly, those training sets featured with multiple benign classes (i.e. dendrogram based and proportional sampling) encounter both for data diversity and difference in sample size providing the classifiers with an acceptable knowledge in terms of heterogeneity needed for the detection.

Final results demonstrate there exists still an improvement margin to keep identifying unhealthy regions along with healthy ones, as one cannot be as effective without the other.

Nevertheless, this study can be considered as a starting point to be further analysed and ameliorated with the goal of finding a satisfying set up for prostate cancer detection based on bi-parametric MRI.

## Conclusions and future work

This thesis project represents one of the first attempts to identify prostate cancer using only the non-invasive sequences proper of bp-MR protocol.

Conceived as a completely automatic system, it comprises prostate segmentation and tumour detection in view of the creation of a software tool which could act as support for radiologists in PCa screening.

The algorithm designed to segment the prostate is ensuring with regard to sensitivity, confirmed by robust and satisfying performances across the study population. However, it suffers from over segmenting tissue outside the prostate (e.g. rectum or bladder) due to the adoption of the fixed bounding box dimension in the entire volume.

In order to improve its performance, an adaptive bounding box could be used to follow the actual extension of the prostate in each slice. Another useful advancement could be to independently determine presence or absence of the prostate in each image to select the useful ones. Edge detection may instead be utilised to precisely delineate prostate contour.

As common in the medical domain, the huge difference in sample size between healthy and unhealthy regions characterised this work and led to several trials, with the aim of overcoming this aspect along with the significant data variability which comes with it.

The first approach implemented to detect PCa comprehended simultaneous feature selection and classification using a genetic algorithm in conjunction with support vector machine, exploiting the training set constituted beginning from decision trees proper of each patient.

Although a sensitivity around 70% was acceptable as initial result, it soon emerged the necessity to reach a much higher specificity to make tumour detection valid.

Bearing this concept in mind, all succeeding trials tried to enrich benign samples representation changing the training set. Dendrogram clustering was involved to create more classes concerning healthy areas, then fed to a multiclass version of the SVM and to k-nearest neighbours classifiers. While some ameliorations were brought by the support vector machine, the same did not occur for k-NN.

To discover whether a totally different classification strategy could enhance healthy tissue identification without losing in sensitivity, feedforward neural networks played a crucial role, showing similar performances to previous classifiers when using the two-class training set, but superior recognition in the case of more benign classes.

Given these results, a cascade of neural networks was performed to exploit the ability of each NN to correctly identify a benign class with a negligible error concerning false positives.

Since the desired outcome was still not accomplished, the conclusive attempt included consecutive NNs, working with a training set built by proportionally sampling clusters resulting from dendrogram on test set of the network presenting increasing misclassification error.

At the end, this method achieved a great performance in classifying benign samples with the downside of a worsened tumour detection.

In conclusion, this work demonstrated the difficulty in conceiving a system suitable for prostate cancer detection assuring a good tumour recognition along with an acceptable benign tissue identification. This is pivotal if such a system should be employed as a screening tool for PCa, although always supervised by expert radiologists.

A fundamental encountered issue is the substantial intra and inter subject variability which causes an enormous data heterogeneity concerning both malignant and benign areas.

Being the tumours outnumbered with respect to non cancerous regions and given their size which could be also considerably small, all the supervised classifiers found very difficult to distinguish between the two.

As evident from these preliminary trials, training set construction has proven to be essential in establishing the goodness of the classifiers.

In addition, even though all these elaborations were performed in parallel for the two image modalities, together they may compensate for reciprocal errors and enrich the final performance.

Future advancements may come up with a strategy to increase the informative content enclosed in the data set with the creation of multiple classes for healthy and cancerous tissue as well.

Bi-parametric MRI protocol certainly embeds valuable knowledge to detect prostate cancer: further researches are required to find a robust and reliable strategy to exploit this information.

# Appendix

## GLCM Texture Features

GLCM texture features mentioned in chapter 3.2 are described in the current section [59, 78]. Parameters notation is detailed in table 3.27.

<b>Notation</b>	<b>Meaning</b>
$G$	grey level
$P(i,j)$	entry of the GLCM
$P_x$	marginal probability from the GLCM (sum of the rows)
$P_y$	marginal probability from the GLCM (sum of the columns)
$\mu$	mean value of P
$\sigma$	standard deviation of P

**Table 3.27:** Parameters notation adopted in the computation of GLCM texture features

In particular,  $P(i, j)$  represents the co-occurrence of a couple of grey levels at a specified distance in terms of pixels.

<b>GLCM feature</b>	<b>Abbreviation</b>
Contrast	<i>Contrast</i>
Correlation	<i>Correlation</i>
Energy	<i>Energy</i>
Entropy	<i>Entropy</i>
Cluster prominence	<i>ClusterProm</i>
Cluster shade	<i>ClusterShade</i>
Haralick correlation	<i>HaralickCorr</i>
Homogeneity	<i>Homogeneity</i>
Variance	<i>Variance</i>
Sum average	<i>SumAverage</i>
Sum entropy	<i>SumEntropy</i>
Sum variance	<i>SumVariance</i>
Difference variance	<i>DiffVariance</i>
Difference entropy	<i>DiffEntropy</i>
Information correlation 1	<i>InfCorr1</i>
Information correlation 2	<i>InfCorr2</i>
Dissimilarity	<i>Dissimilarity</i>
Max GLCM	<i>MaxGLCM</i>

**Table 3.28:** Abbreviations for GLCM texture features

In the following equations,  $n = |i - j|$ .

*Contrast* quantifies local variations of intensities appearing in an image, favouring the couples  $(i, j)$  apart from the diagonal (i.e. when  $i \neq j$ ). Its value could be particularly affected by the highest and the lowest value of the considered set of pixels, mostly in the case of displacement vector made of only two pixels. Significant variability of grey tones in the image leads to high values of contrast.

$$Contrast = \sum_{n=0}^{G-1} n^2 \left( \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) \right) \quad (3.12)$$

*Correlation* estimates the linear dependency among grey levels of neighbouring pixels. Regions with similar grey scale intensities will present high values of correlation.

$$Correlation = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{(i - \mu_x)(j - \mu_y)P(i, j)}{\sigma_x \sigma_y} \quad (3.13)$$

*Energy* (also called *angular second moment*) measures the homogeneity of the image: the higher its value, the more homogeneous the image will be, because of the elevated occurrences of the couples  $(i, j)$ .

$$Energy = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j)^2 \quad (3.14)$$

$$Inertia = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - j)^2 P(i, j) \quad (3.15)$$

*Entropy* evaluates the randomness of the intensity distribution. More diverse probabilities result in a high value of *Entropy*, while it decreases for unequal probabilities.

$$Entropy = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) \ln(P(i, j)) \quad (3.16)$$

*ClusterShade* and *ClusterProm* can be considered as 2-D versions respectively of skewness and kurtosis.

$$ClusterShade = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i + j - \mu_x - \mu_y)^3 P(i, j) \quad (3.17)$$

$$ClusterProm = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i + j - \mu_x - \mu_y)^4 P(i, j) \quad (3.18)$$

$$HaralickCorr = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{(i, j)P(i, j) - \mu_x \times \mu_y}{\sigma_x \sigma_y} \quad (3.19)$$

*Homogeneity*, also known as *inverse difference moment*, gives information about local homogeneity. The term  $(1 + (i - j)^2)$  in the denominator of equation 3.19 decreases the contribution of more diversified areas.

$$Homogeneity = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{1 + (i - j)^2} P(i, j)^2 \quad (3.20)$$

*Variance* emphasizes the elements which are more distant from the mean value of  $P(i, j)$ . It can be interpreted as a measure of heterogeneity and it is not correlated to any spatial frequency, unlike *Contrast*.

$$Variance = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - \mu_x)^2 P(i, j) + \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (j - \mu_y)^2 P(i, j) \quad (3.21)$$

In order to define the remaining features, next parameters must be set:

$$P_{x+y} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) \quad (3.22)$$

where  $n = 2, 3, \dots, 2G$  and  $i + j = n$ , while

$$P_{x-y} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) \quad (3.23)$$

where  $n = 0, 1, \dots, G - 1$  and  $|i - j| = n$ ,

$$SumAverage = \sum_{n=2}^{2G} n P_{x+y}(n) \quad (3.24)$$

$$SumVariance = \sum_{n=2}^{2G} (n - SumAverage)^2 P_{x+y}(n) \quad (3.25)$$

$$DiffVariance = \sum_{n=0}^{G-1} (n - \mu_{x-y})^2 P_{x-y}(n) \quad (3.26)$$

In particular,  $\mu_{x-y}$  corresponds to the average value of  $P_{x-y}(n)$ .

$$SumEntropy = - \sum_{i=2}^{2G} P_{x+y}(i) \ln(P_{x+y}(i)) \quad (3.27)$$

$$DiffEntropy = - \sum_{i=2}^{G-1} P_{x-y}(i) \ln(P_{x-y}(i)) \quad (3.28)$$

$$InfCorr1 = \frac{Entropy - HXY1}{\max(HX, HY)} \quad (3.29)$$

$HX$  and  $HY$  represent respectively the entropy of  $P_x$ , indicating sum of GLCM rows, and  $P_y$ , indicating sum of GLCM columns.

$$HXY1 = - \sum_i \sum_j P(i, j) \ln(P_x(i)P_y(j)) \quad (3.30)$$

$$HXY2 = - \sum_i \sum_j P_x(i)P_y(j) \ln(P_x(i)P_y(j)) \quad (3.31)$$

$$InfCorr2 = 1 - e^{-2(HXY2-Entropy)} \quad (3.32)$$

$$MaxGLCM = \max(GLCM) \quad (3.33)$$

$$Dissimilarity = - \sum_i \sum_j |(i+1) - (j+1)| P_{x+y}(i, j) \quad (3.34)$$

### GLRLM Texture Features

GLRLM texture features mentioned in chapter 3.2 are listed in the current section [64, 79]. Notation adopted for the calculation of GLRLM is provided in table 3.29.

Notation	Meaning
$G$	grey level (rows of the GLRLM)
$p(i, j \theta)$	entry of the GLCM
$R$	greatest sequence
$n$	number of pixels in the image
$\mu$	mean value of P
$j$	length of the sequence (columns of the GLRLM)

**Table 3.29:** Parameters notation adopted in the computation of GLRLM texture features

GLRLM feature	Abbreviation
Gray Level Non-uniformity	<i>GLNU</i>
Run Length Non-uniformity	<i>RLNU</i>
High Gray-Level Run Emphasis	<i>HGRE</i>
Low Gray-Level Run Emphasis	<i>LGRE</i>

**Table 3.30:** Abbreviations for GLRLM texture features

*GLNU* weighs more runs of greater lengths and equals smaller values when the analysed sequence has uniformly distributed grey levels.

$$GLNU = \frac{\sum_{i=1}^G (\sum_{j=1}^R p(i, j|\theta))^2}{\sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta)} \quad (3.35)$$

Instead, *RLNU* focuses on the length of the distribution, thus it assumes low values if the sequence is evenly distributed for each run length.

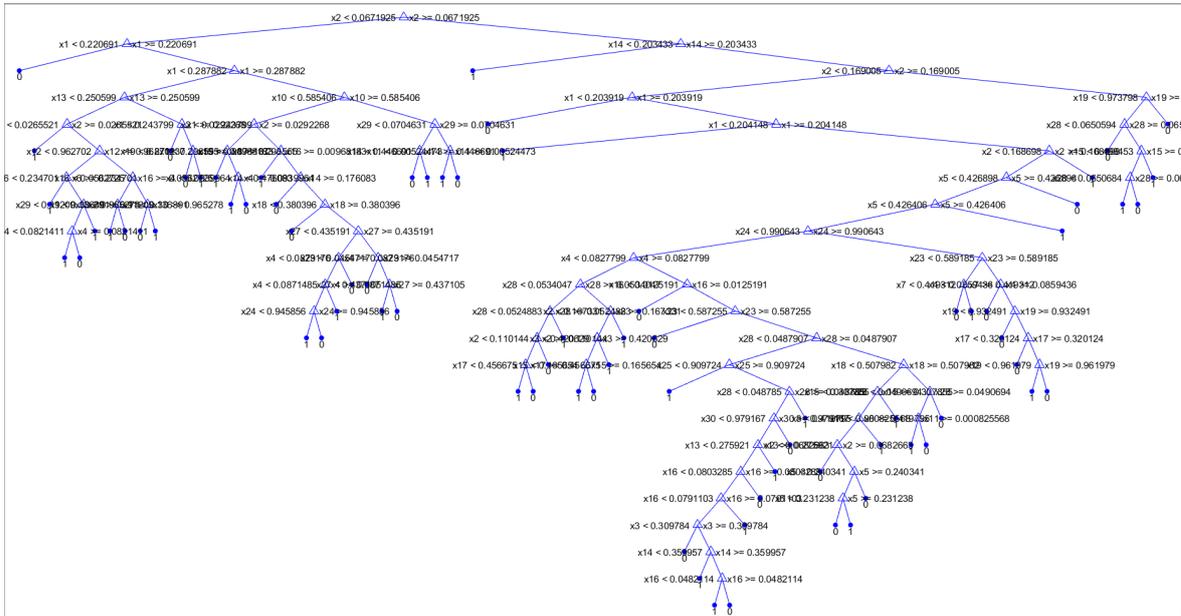
$$RLNU = \frac{\sum_{j=1}^R (\sum_{i=1}^G p(i, j|\theta))^2}{\sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta)} \quad (3.36)$$

*HGRE* and *LGRE* estimates the presence of grey levels in each run lengths.

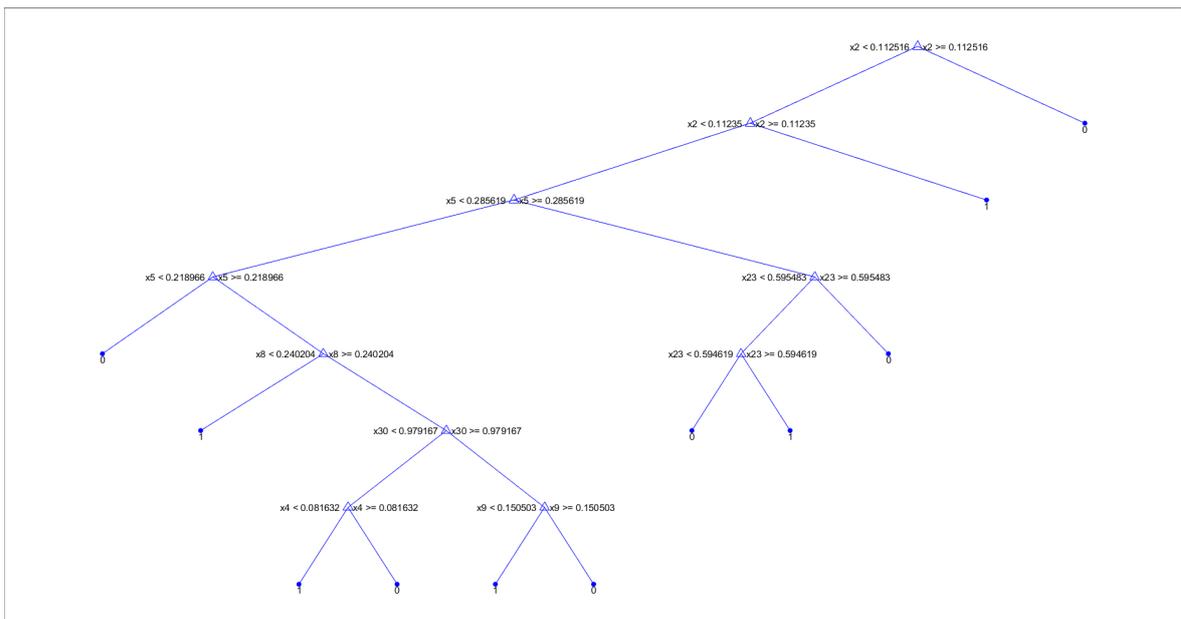
$$HGRE = \frac{\sum_{i=1}^G \sum_{j=1}^R i^2 p(i, j|\theta)}{\sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta)} \quad (3.37)$$

$$LGRE = \frac{\sum_{i=1}^G \sum_{j=1}^R \frac{p(i, j|\theta)}{i^2}}{\sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta)} \quad (3.38)$$

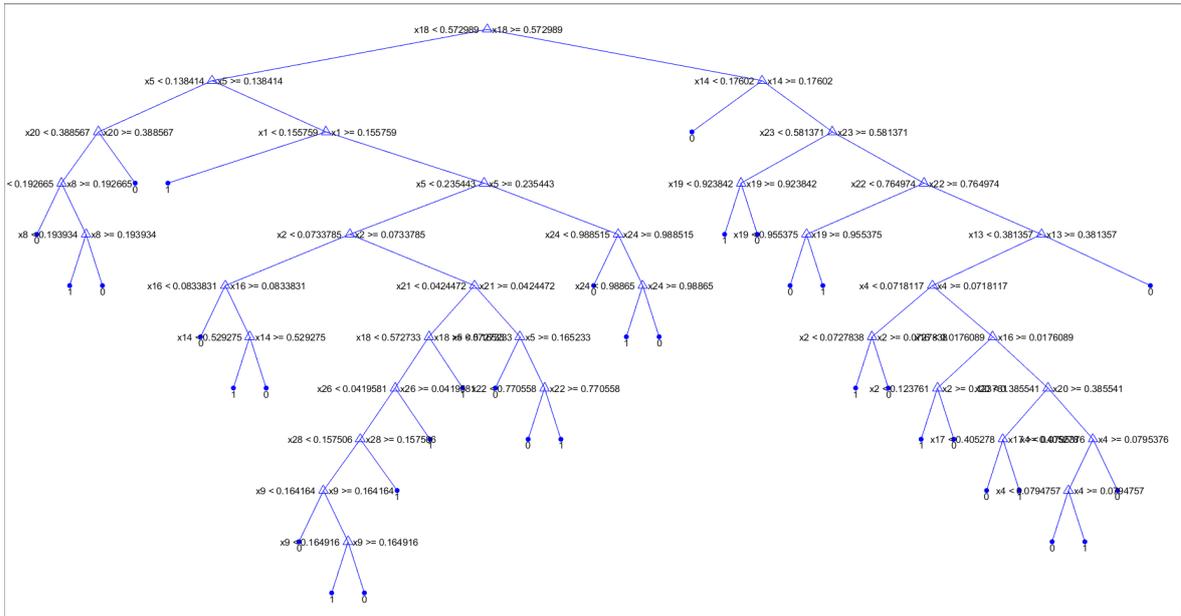




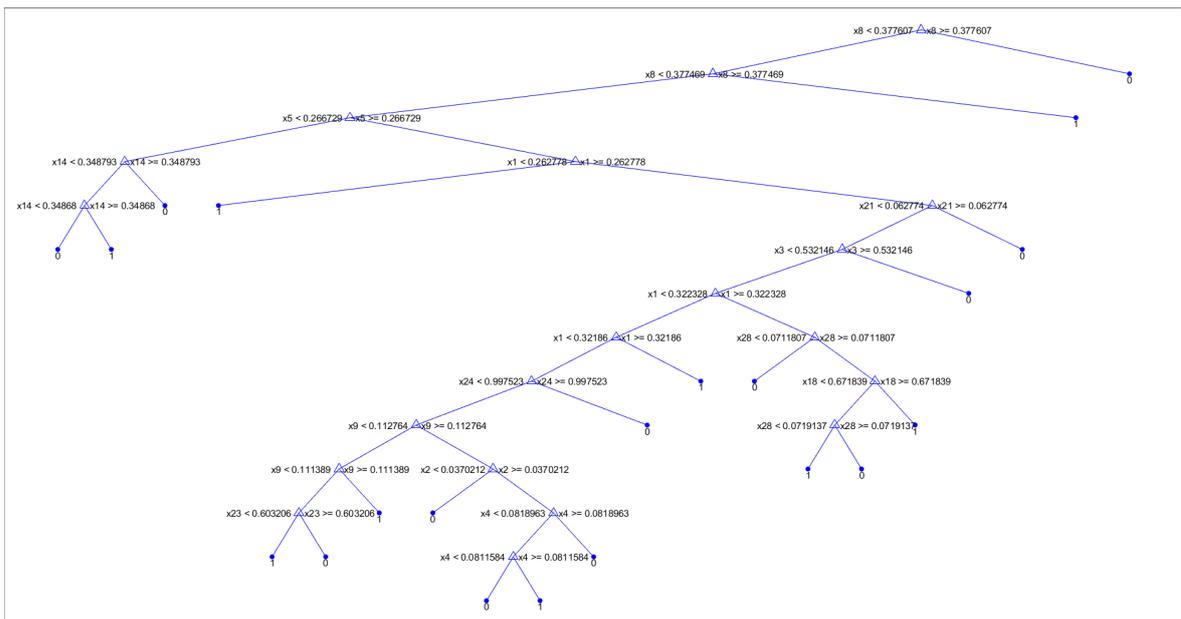
(c) Patient #24



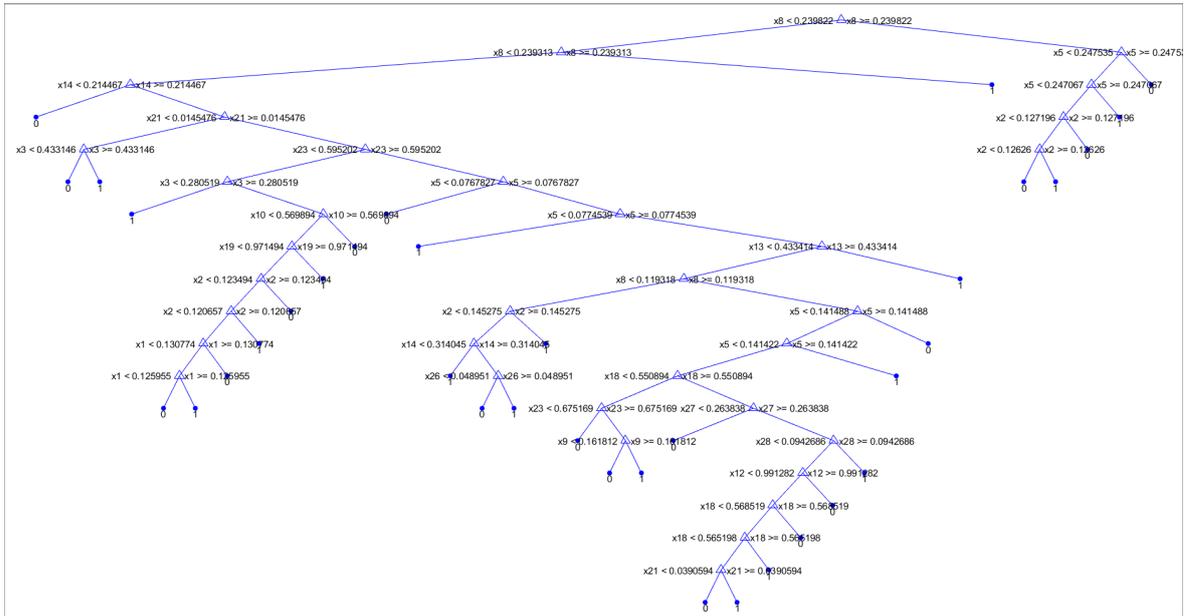
(d) Patient #34



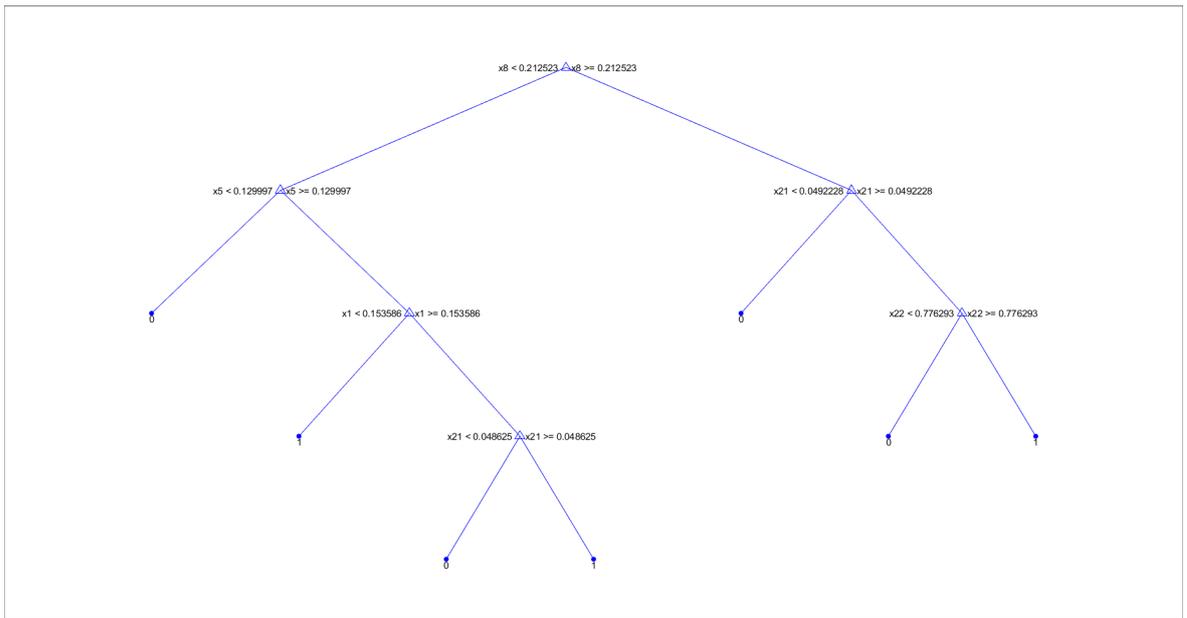
(e) Patient #37



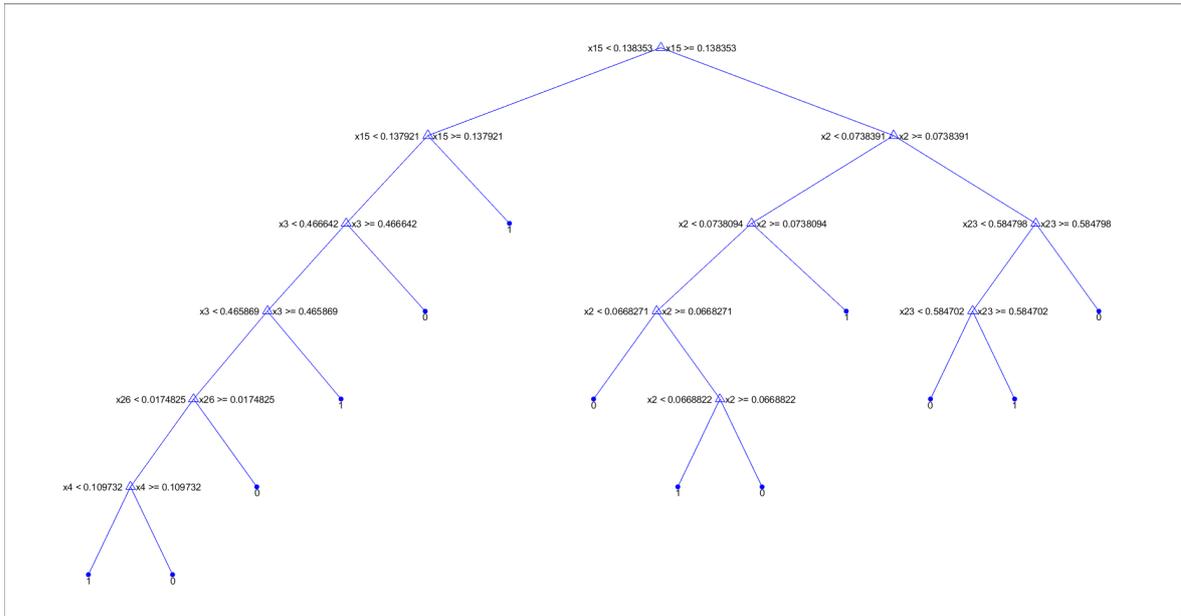
(f) Patient #52



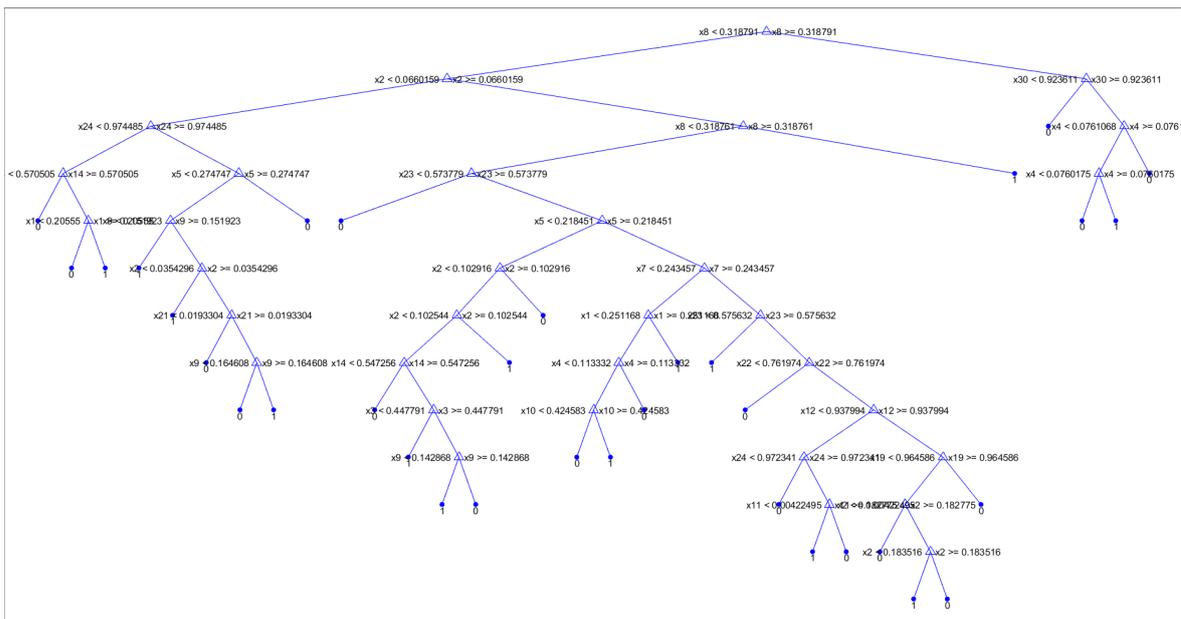
(g) Patient #57



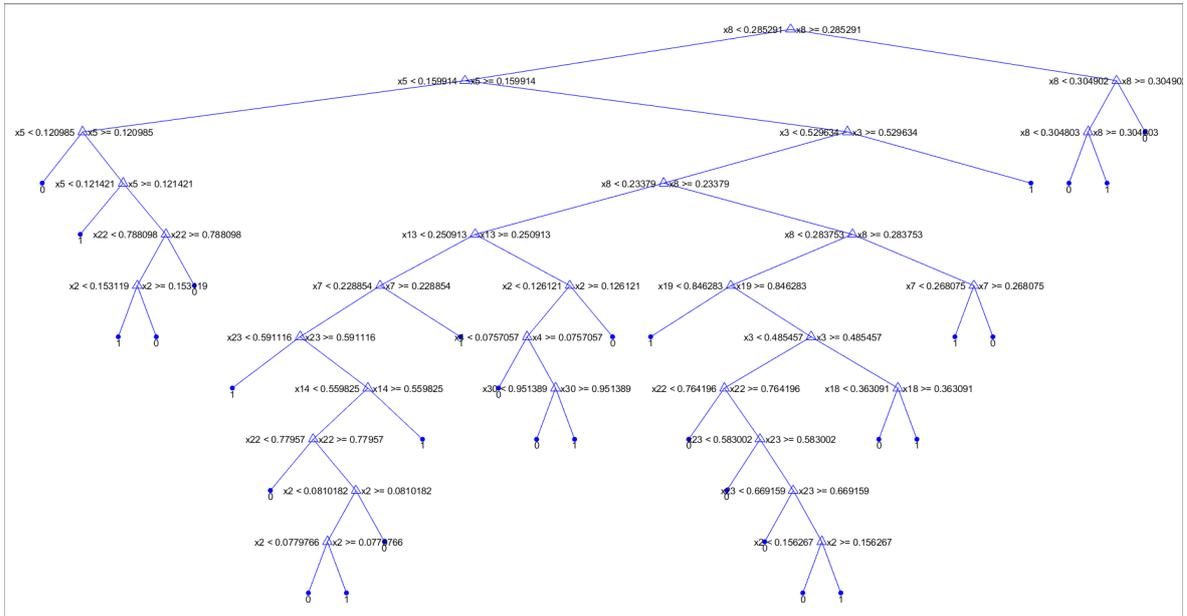
(h) Patient #71



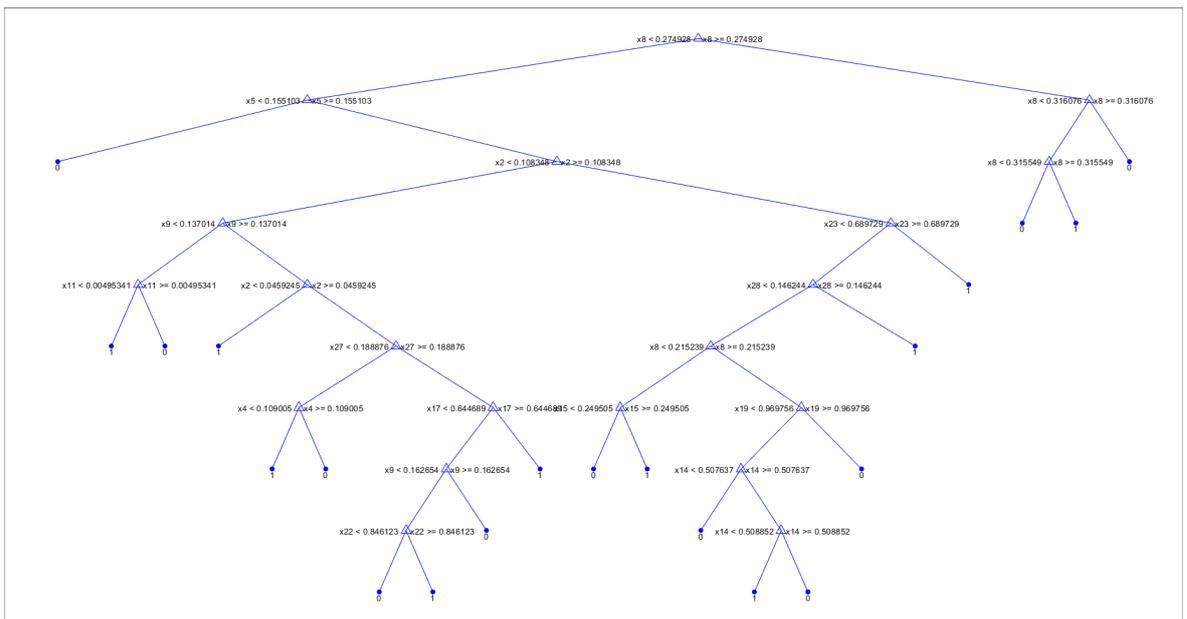
(i) Patient #72



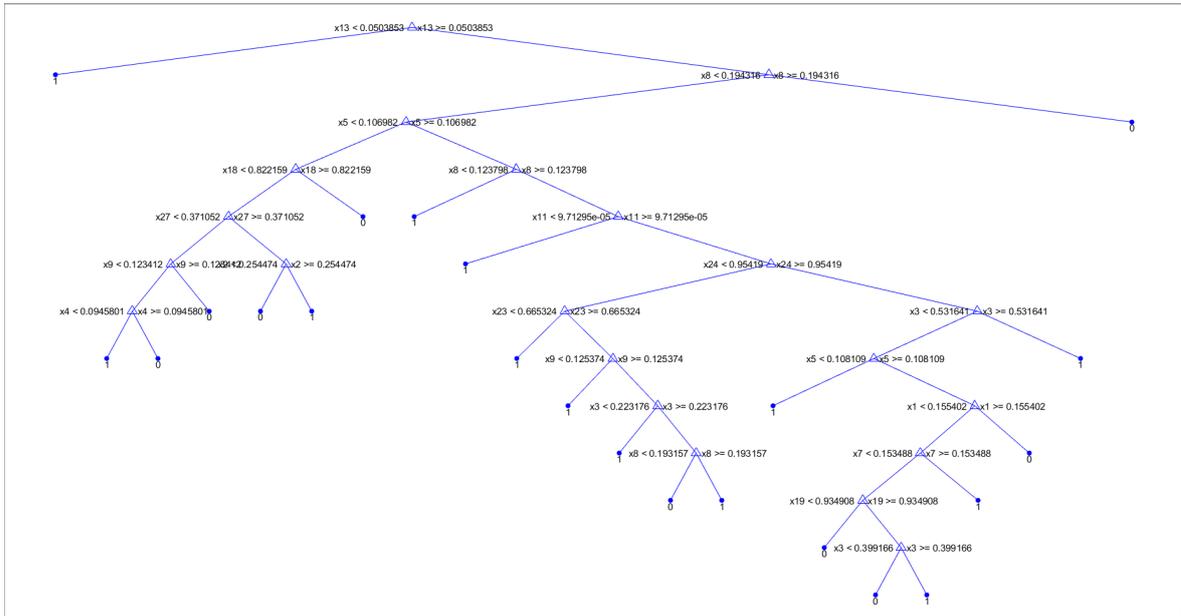
(j) Patient #79



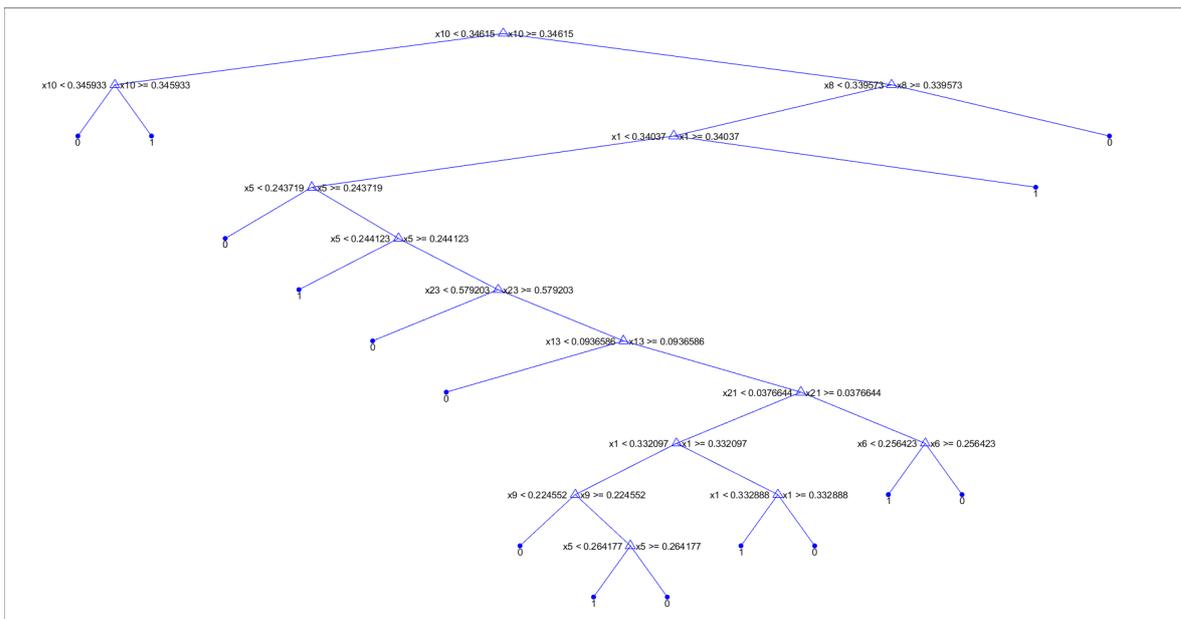
(k) Patient #91



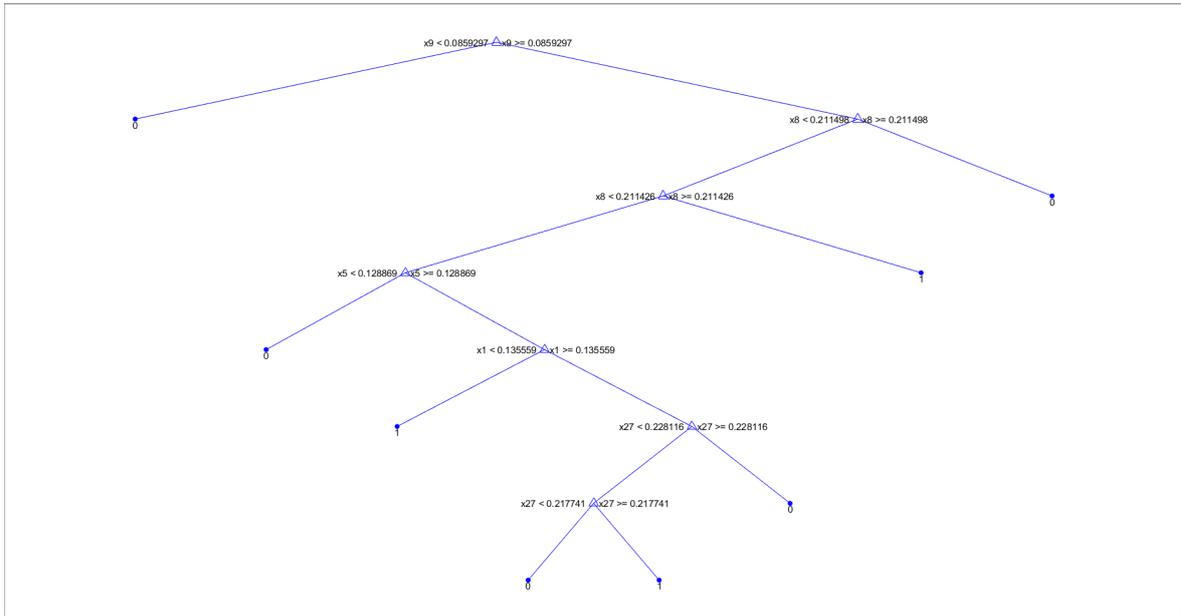
(l) Patient #100



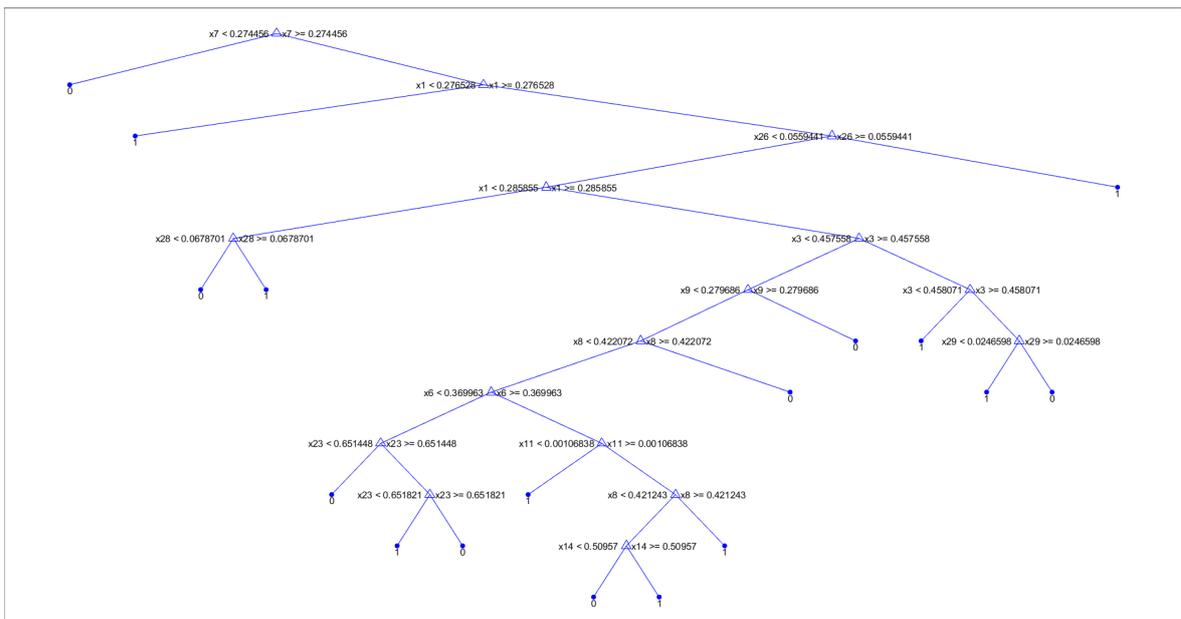
(m) Patient #107



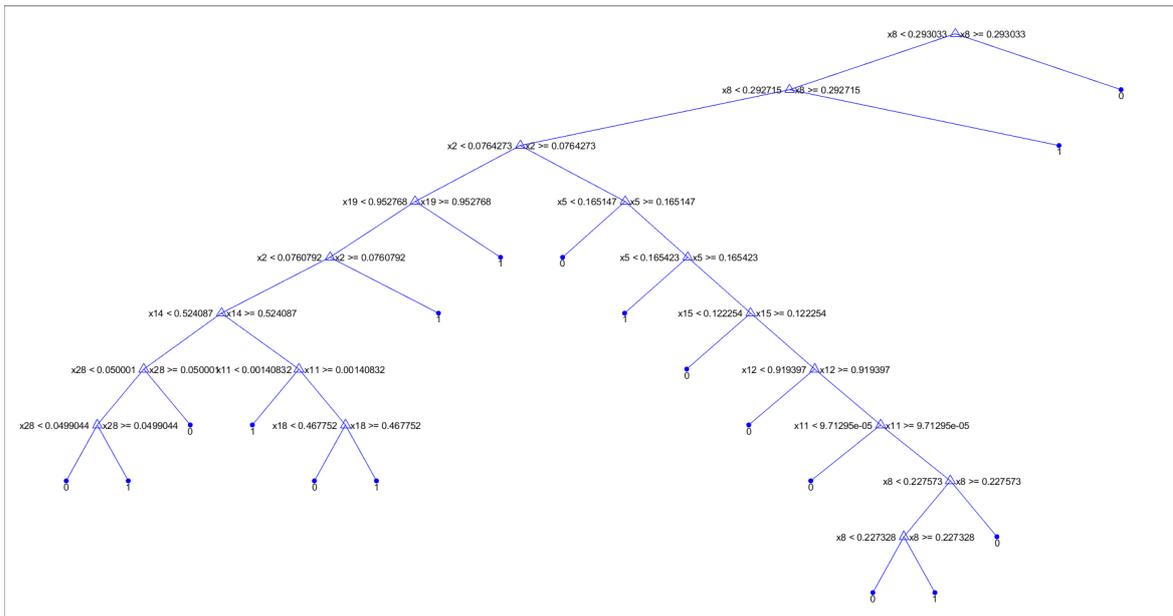
(n) Patient #111



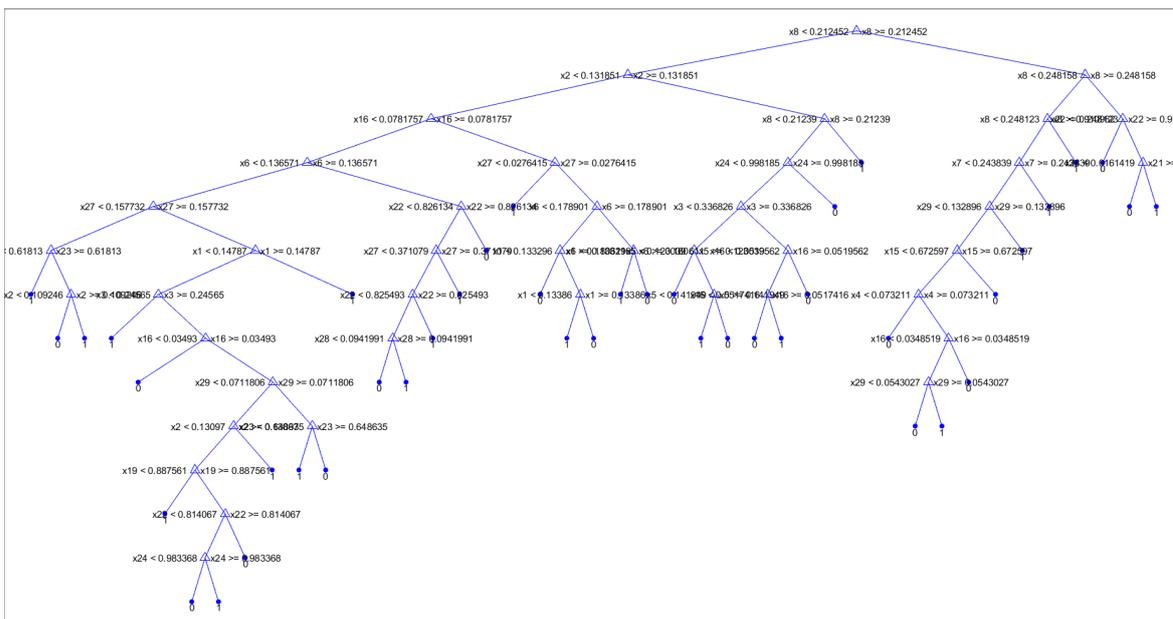
(o) Patient #122



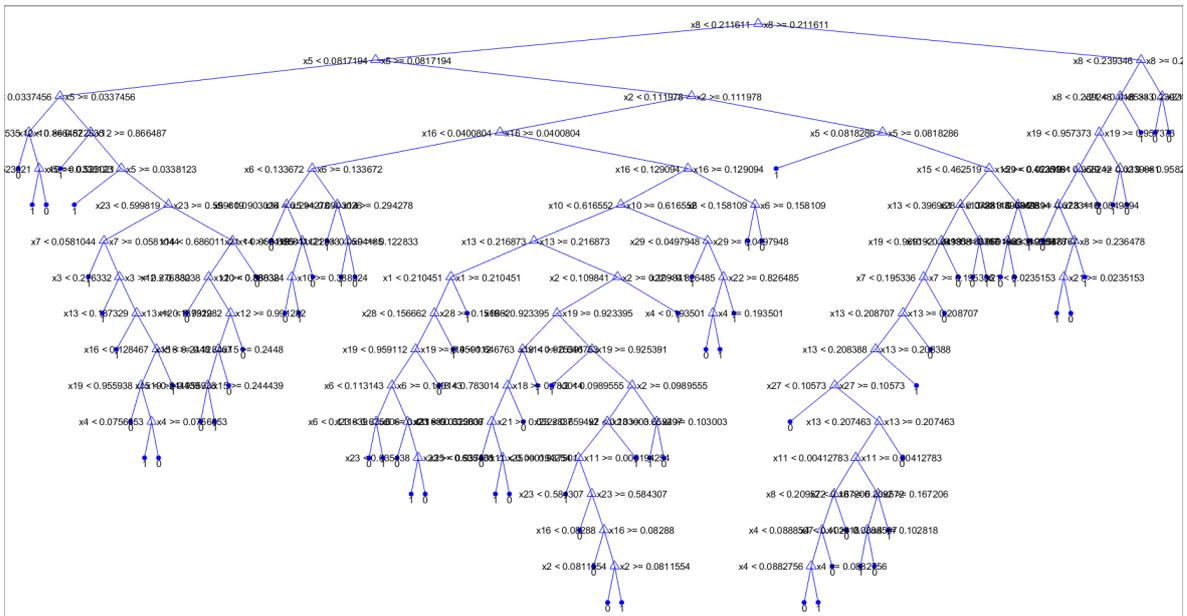
(p) Patient #128



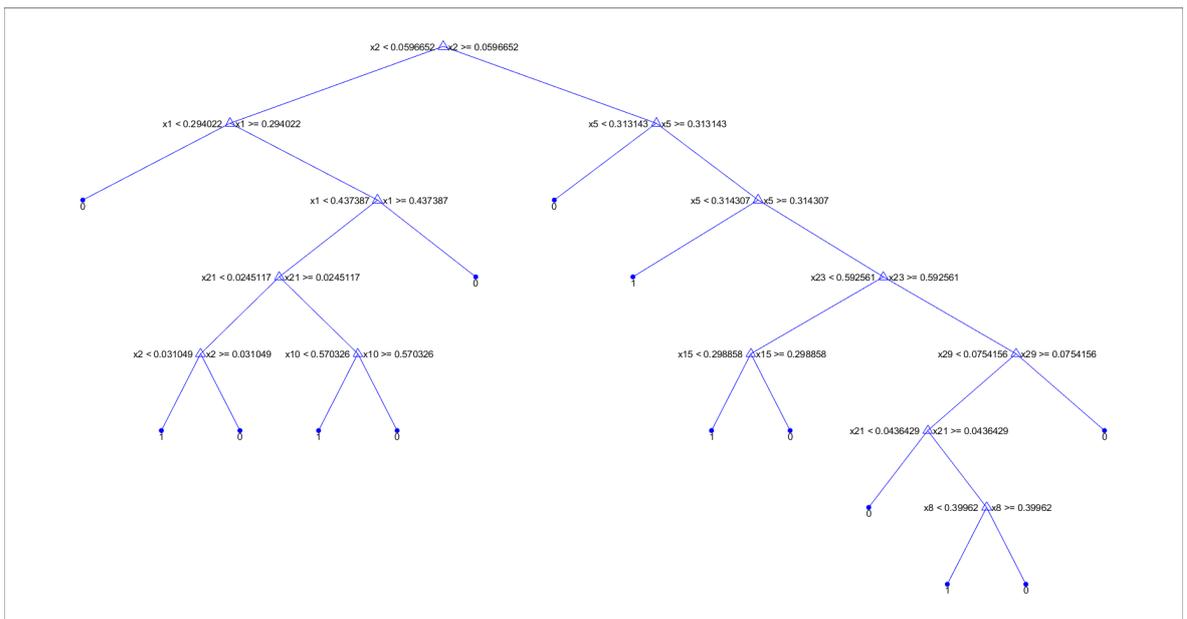
(q) Patient #135



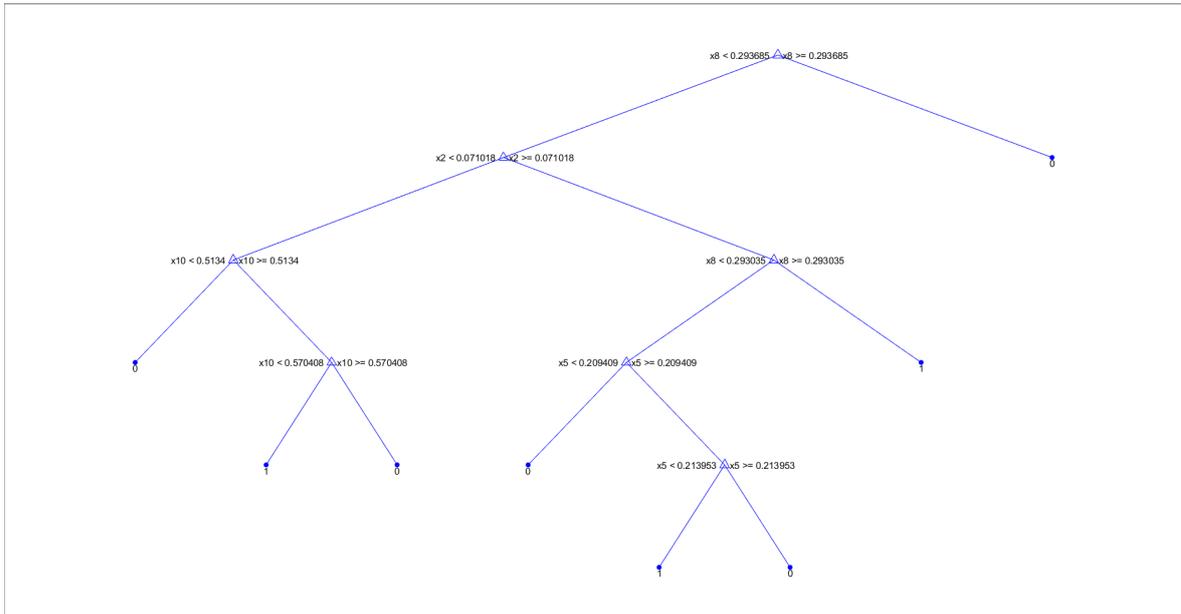
(r) Patient #155



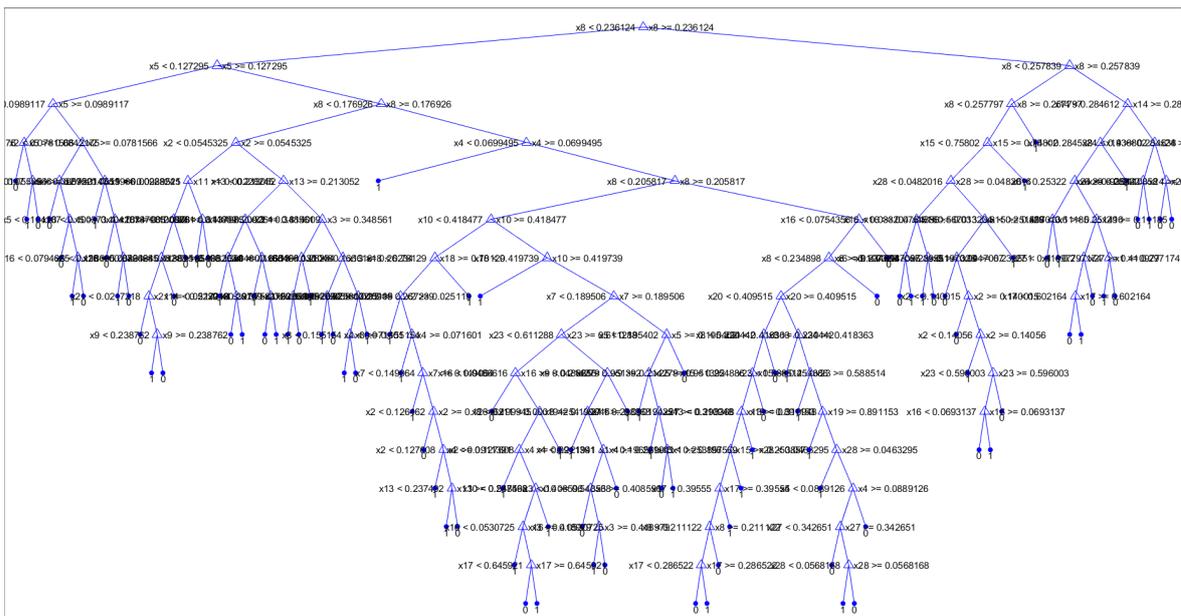
(s) Patient #163



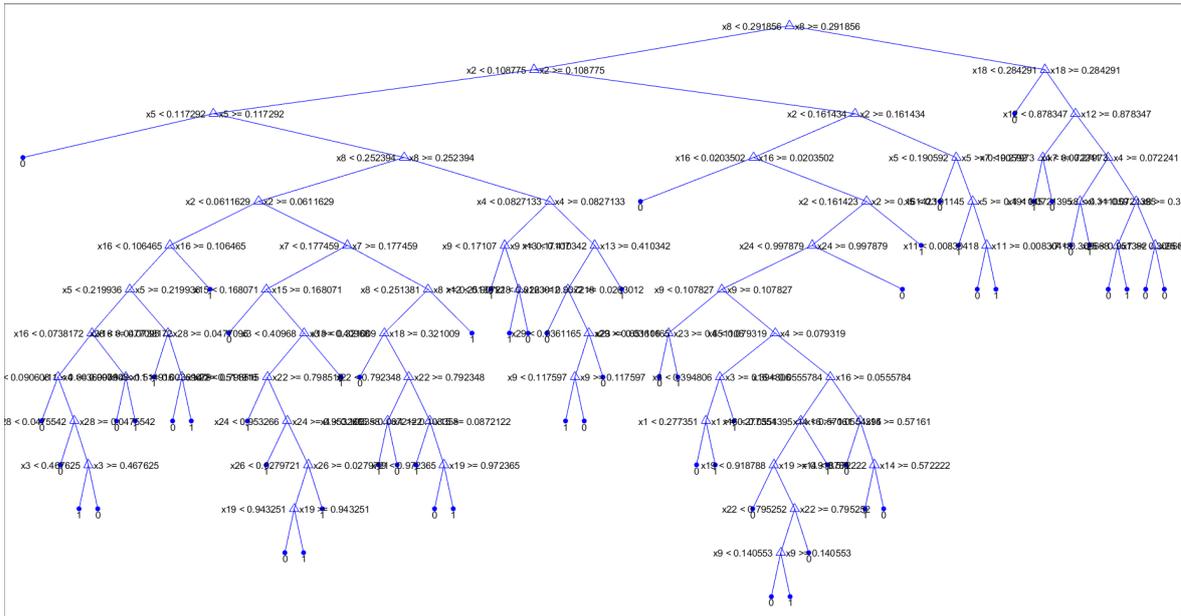
(t) Patient #169



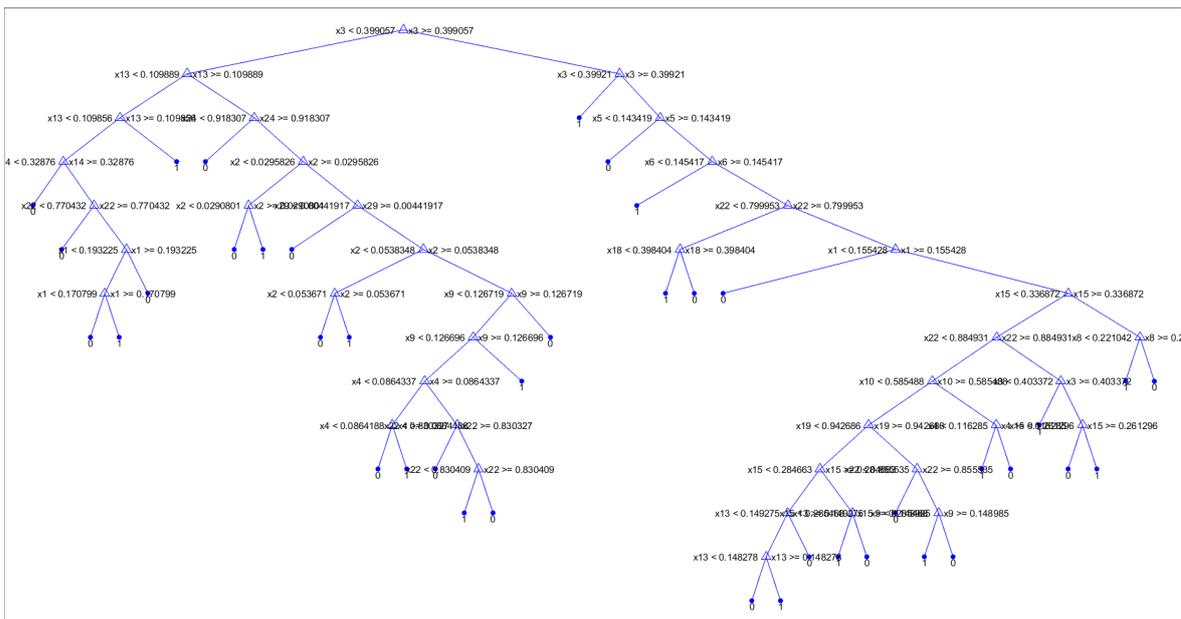
(u) Patient #171



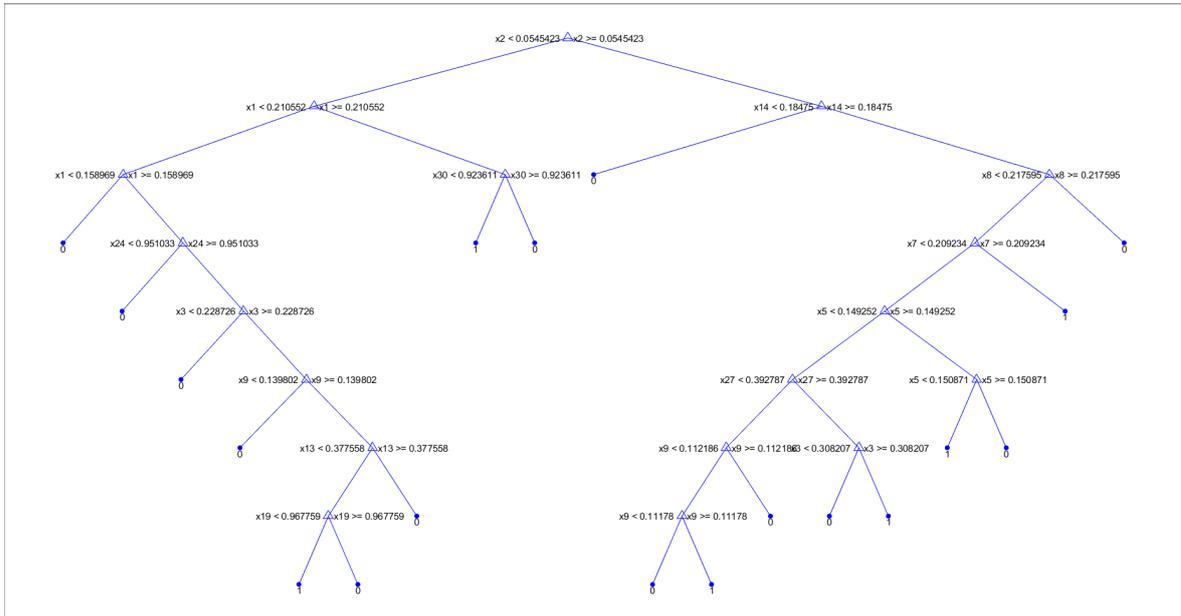
(v) Patient #172



(w) Patient #173

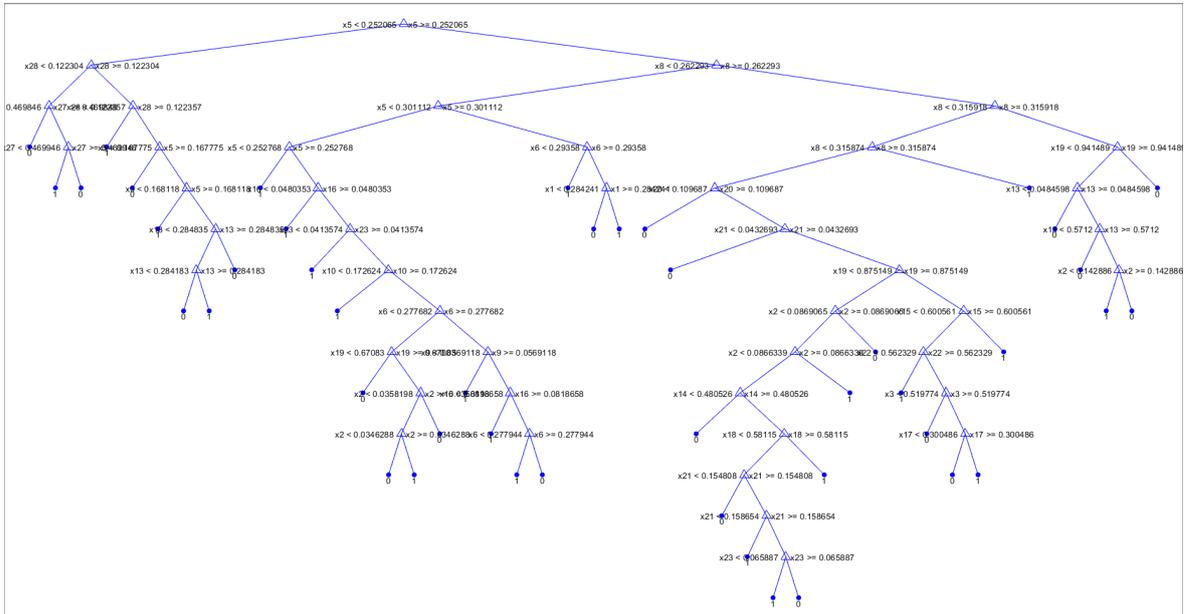


(x) Patient #174

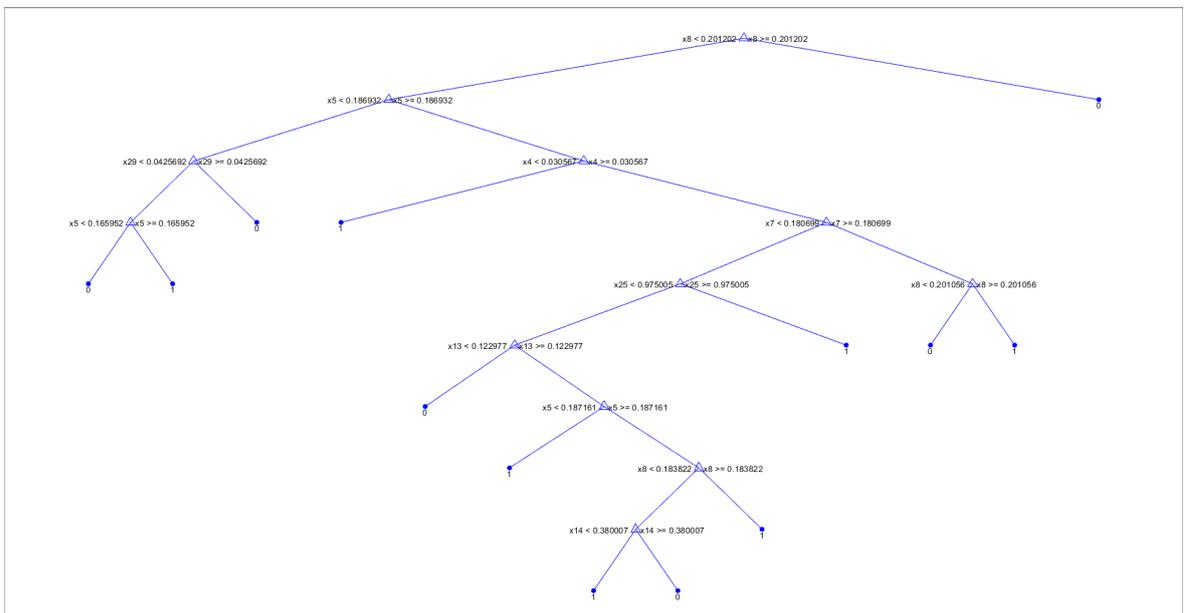


(y) Patient #182

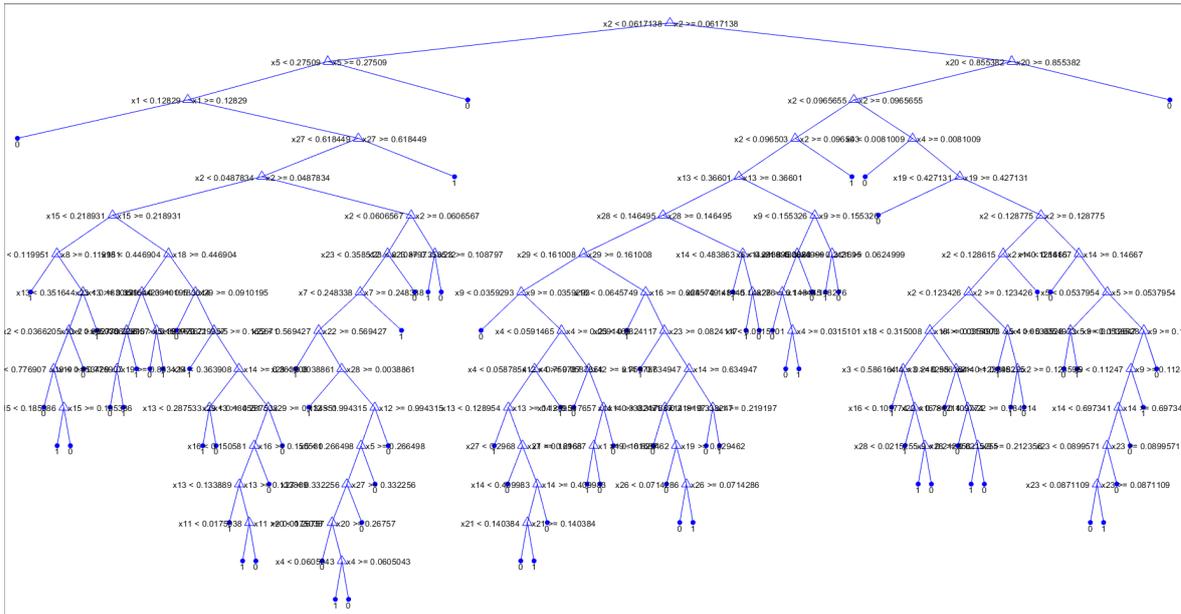
Decision trees computed considering features extracted from ROIs belonging to T2-w used for training sets construction are reported in the following.



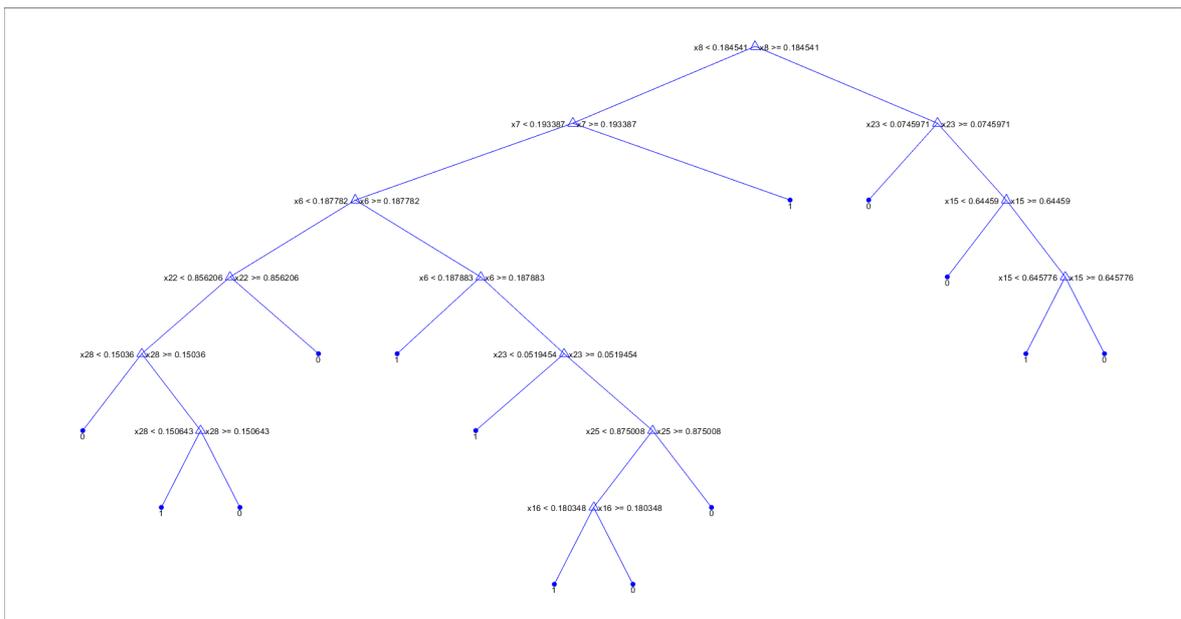
(a) Patient #13



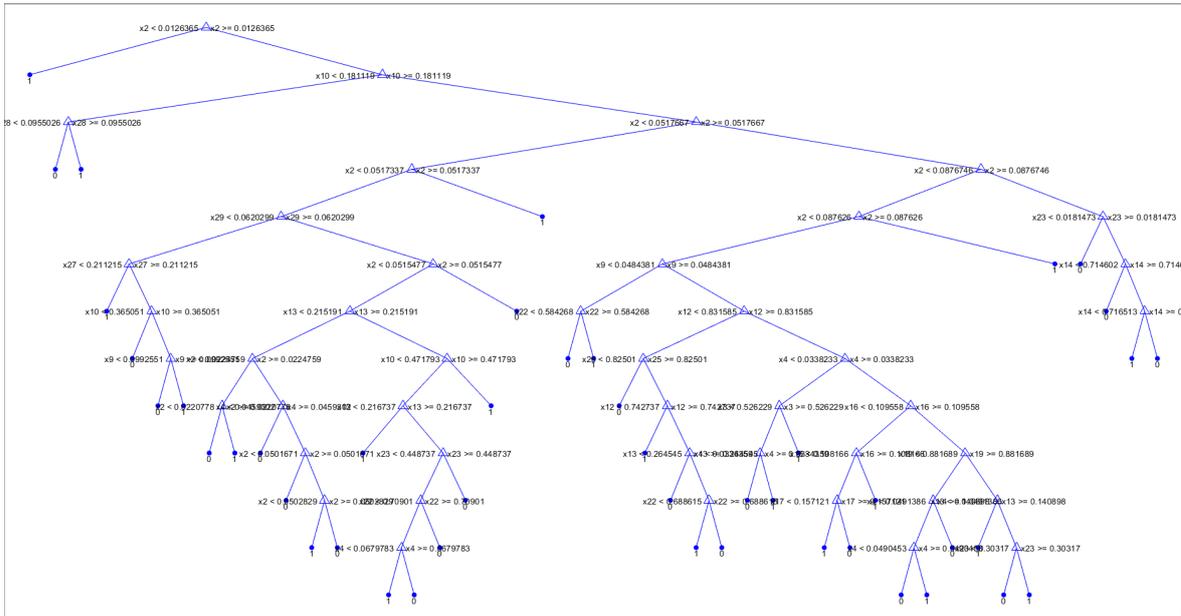
(b) Patient #19



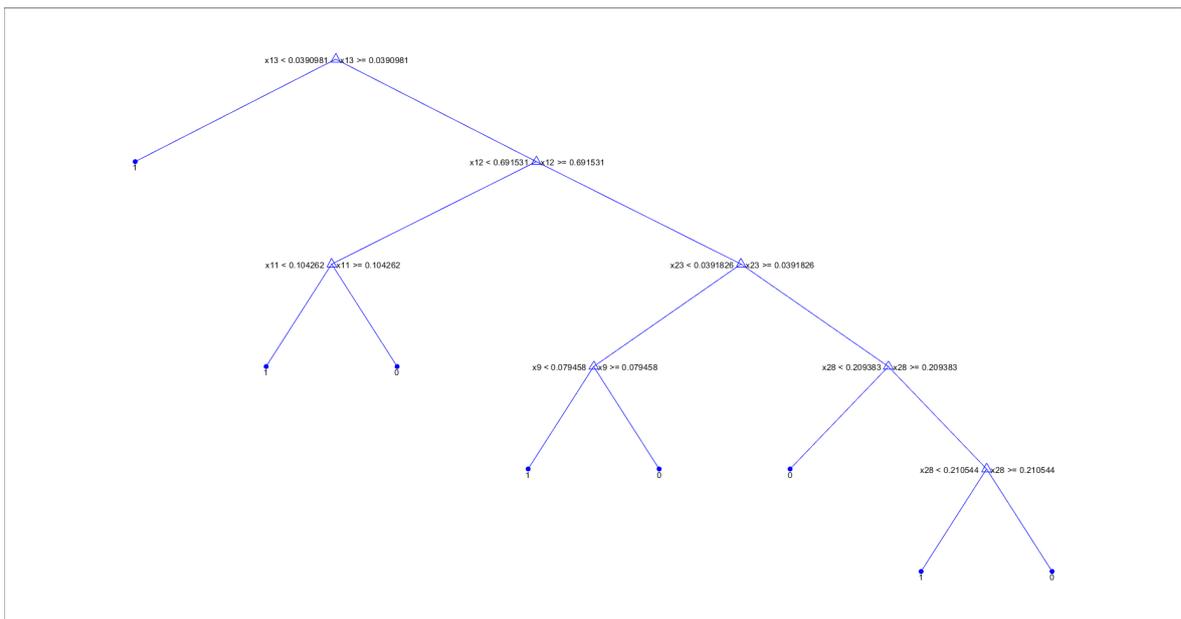
(c) Patient #24



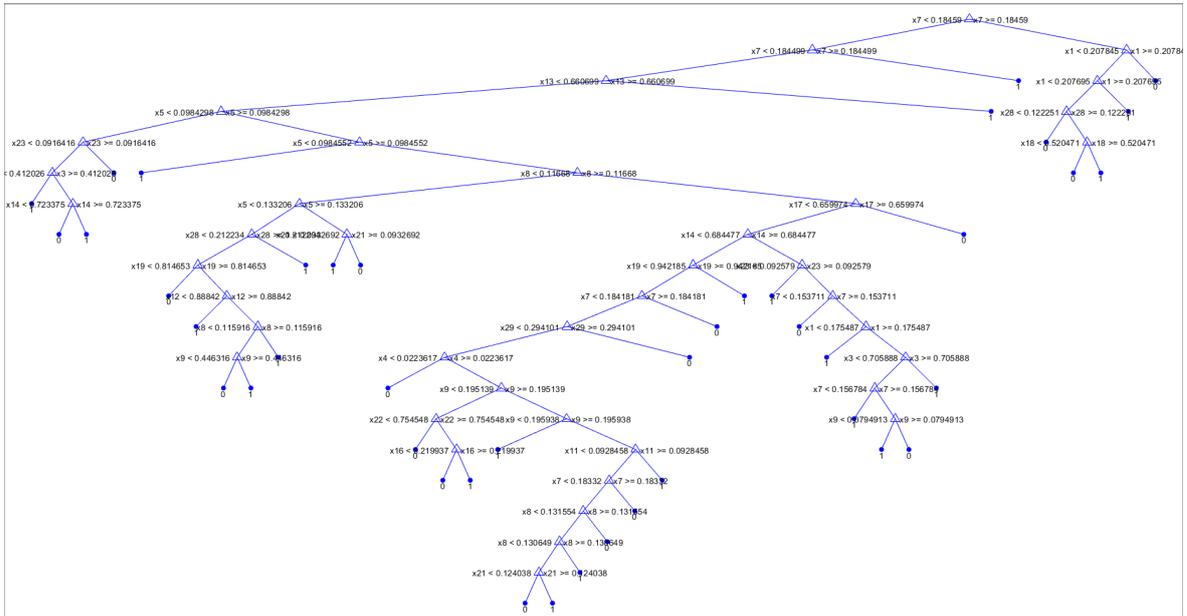
(d) Patient #34



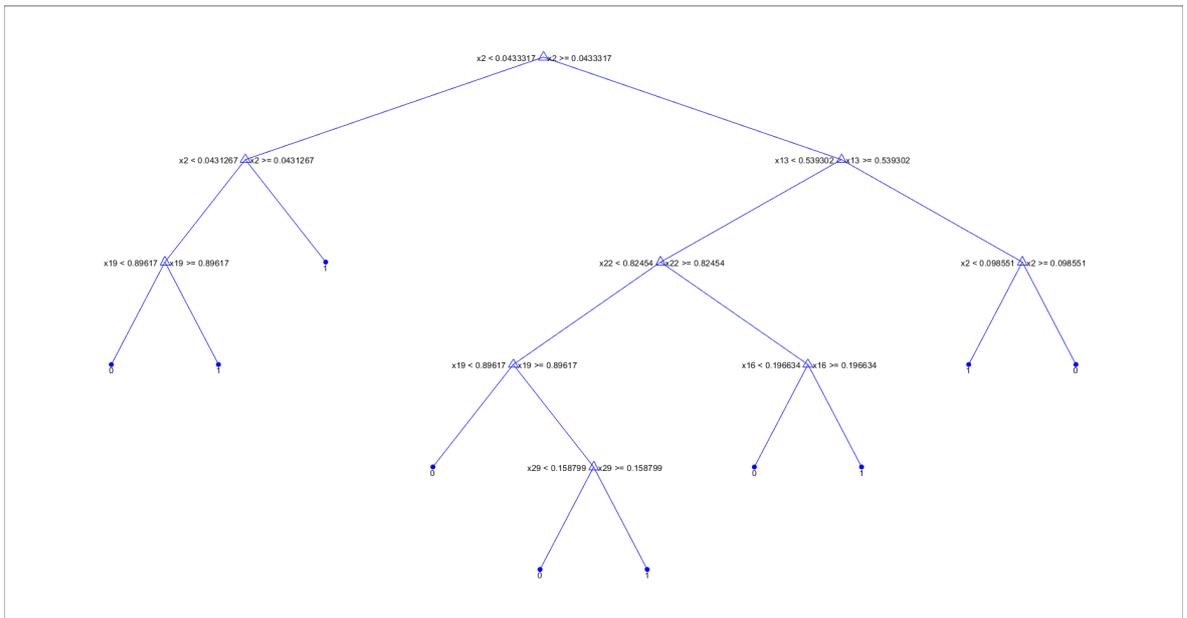
(e) Patient #37



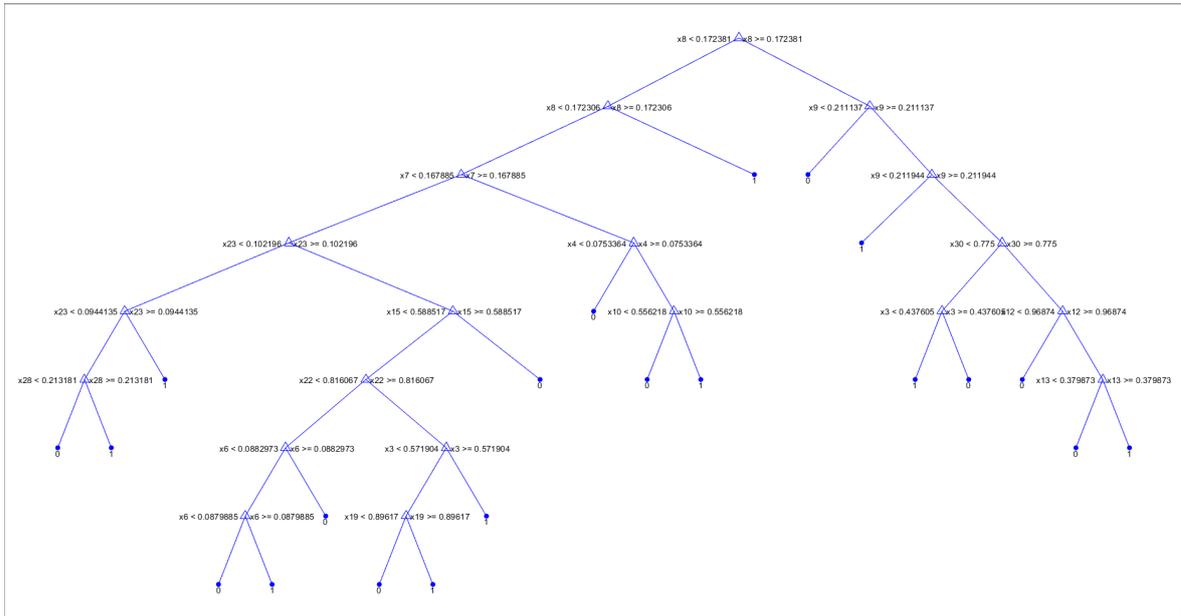
(f) Patient #52



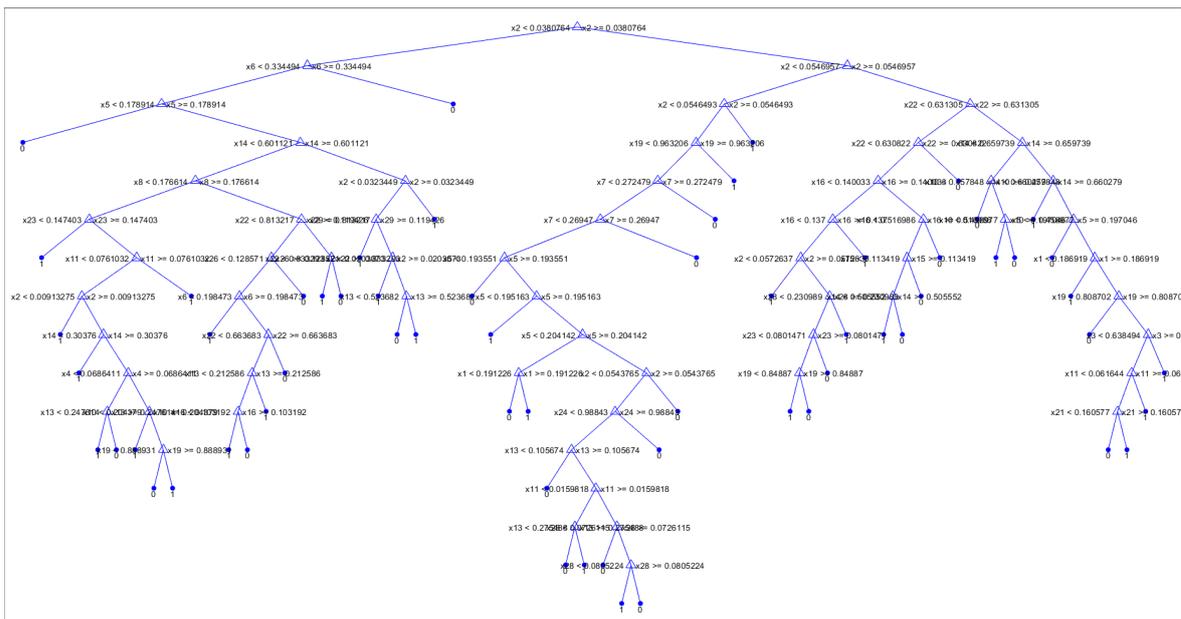
(g) Patient #57



(h) Patient #71



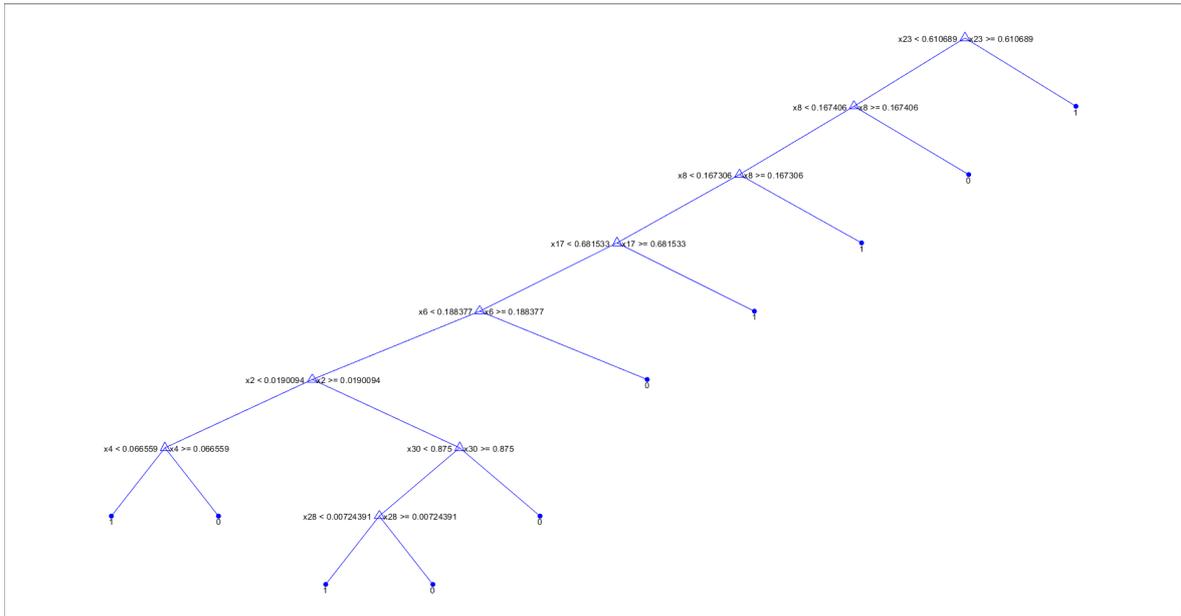
(i) Patient #72



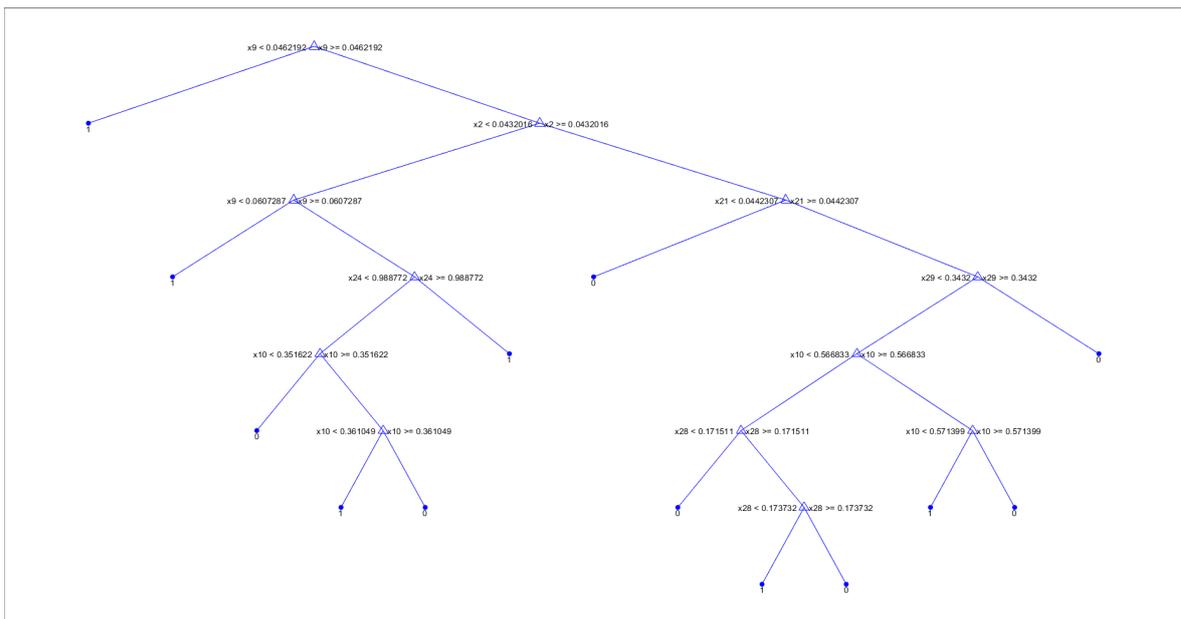
(j) Patient #79



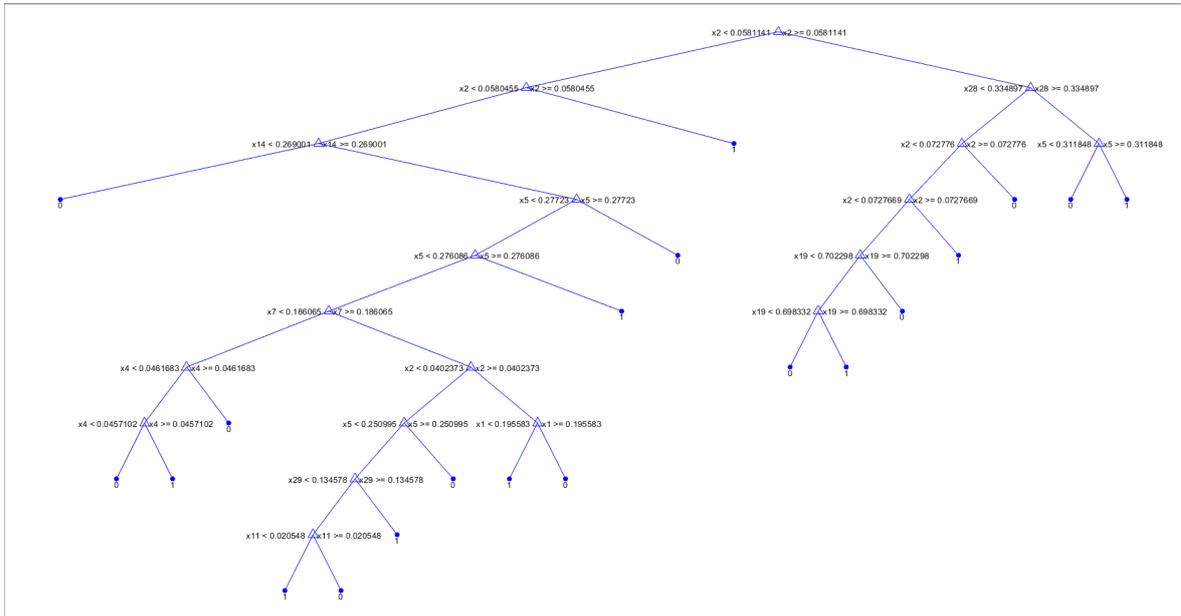




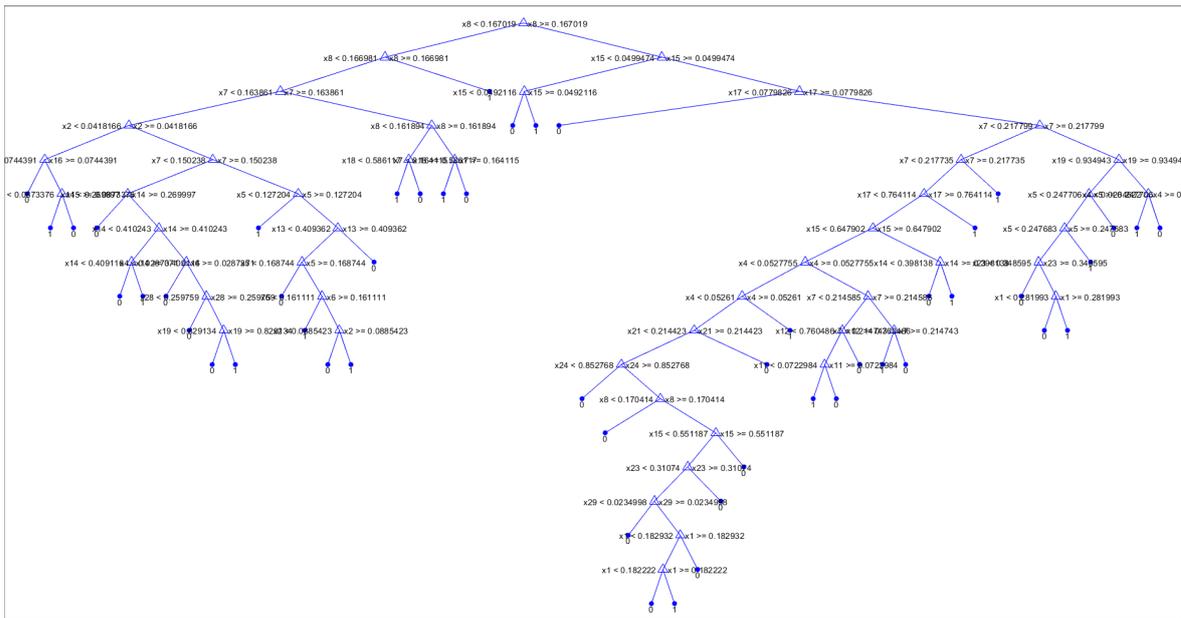
(o) Patient #122



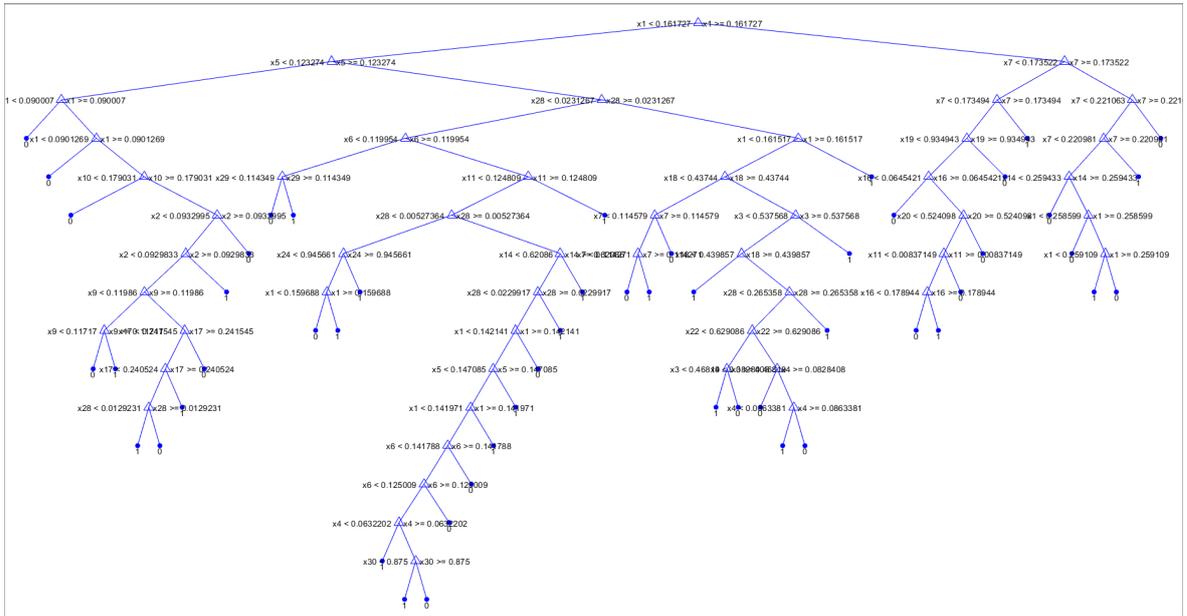
(p) Patient #128



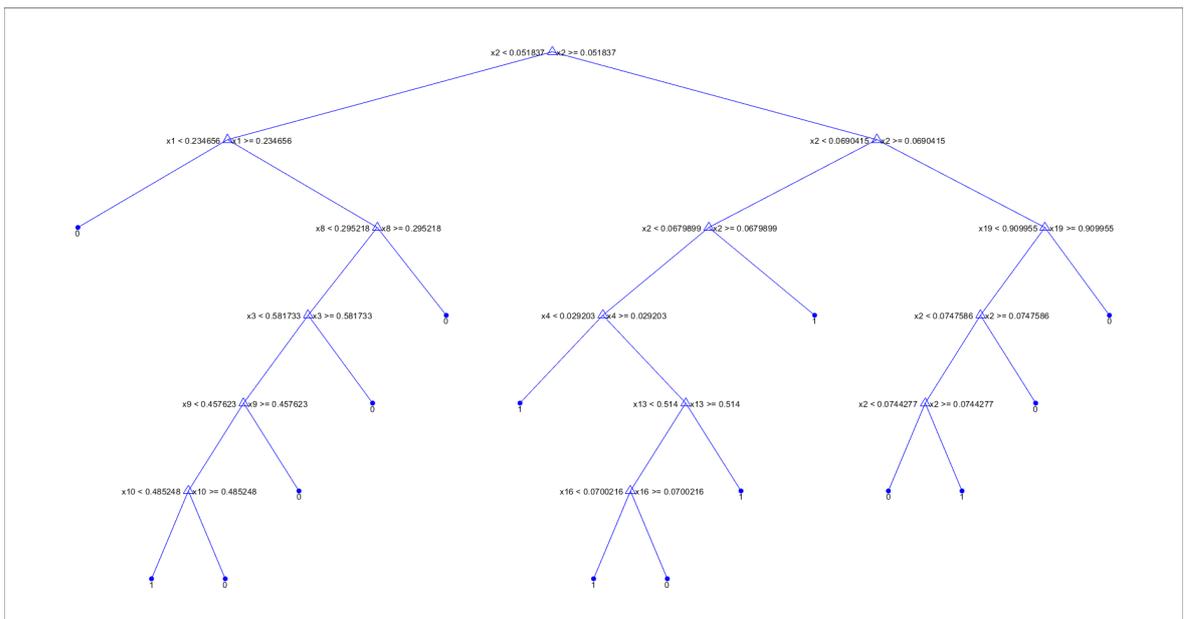
(q) Patient #135



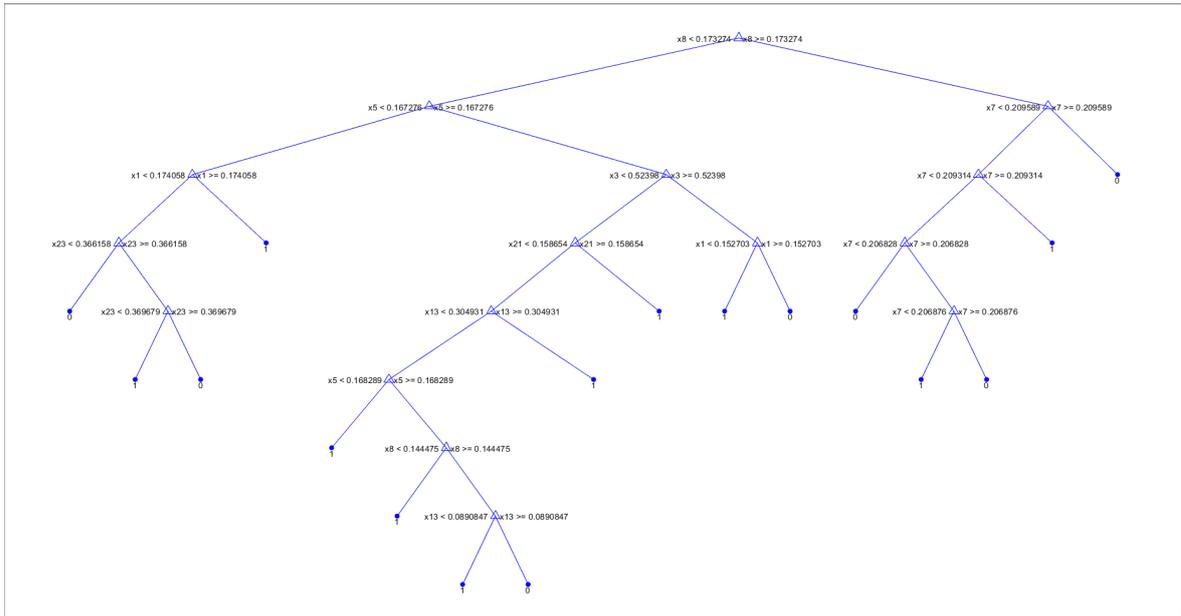
(r) Patient #155



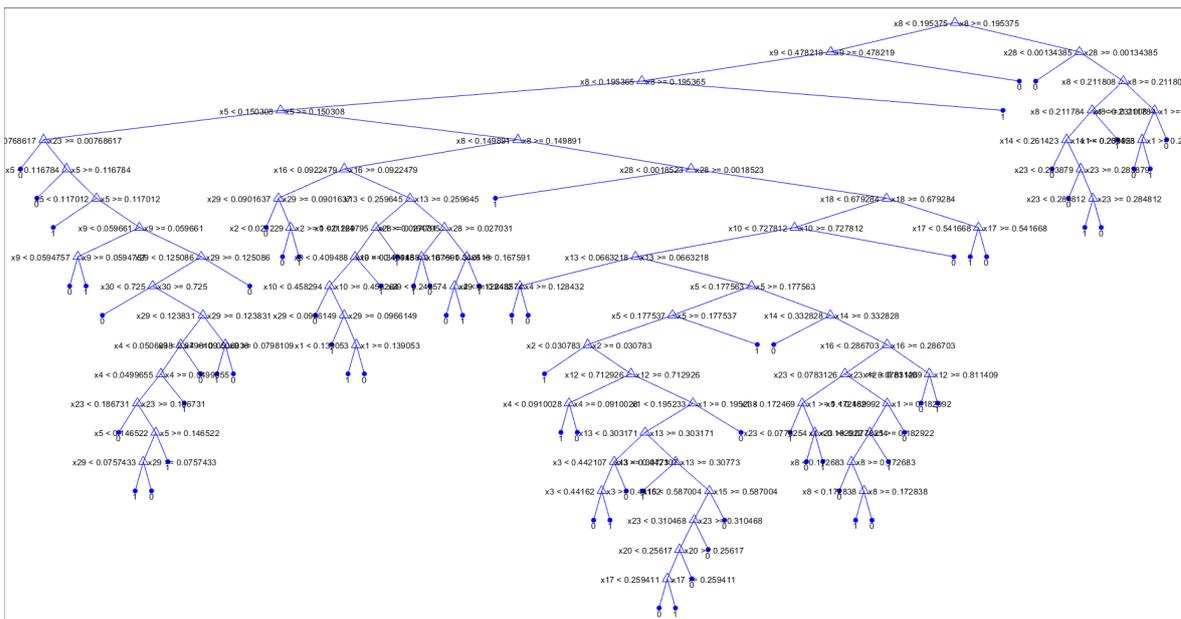
(s) Patient #163



(t) Patient #169

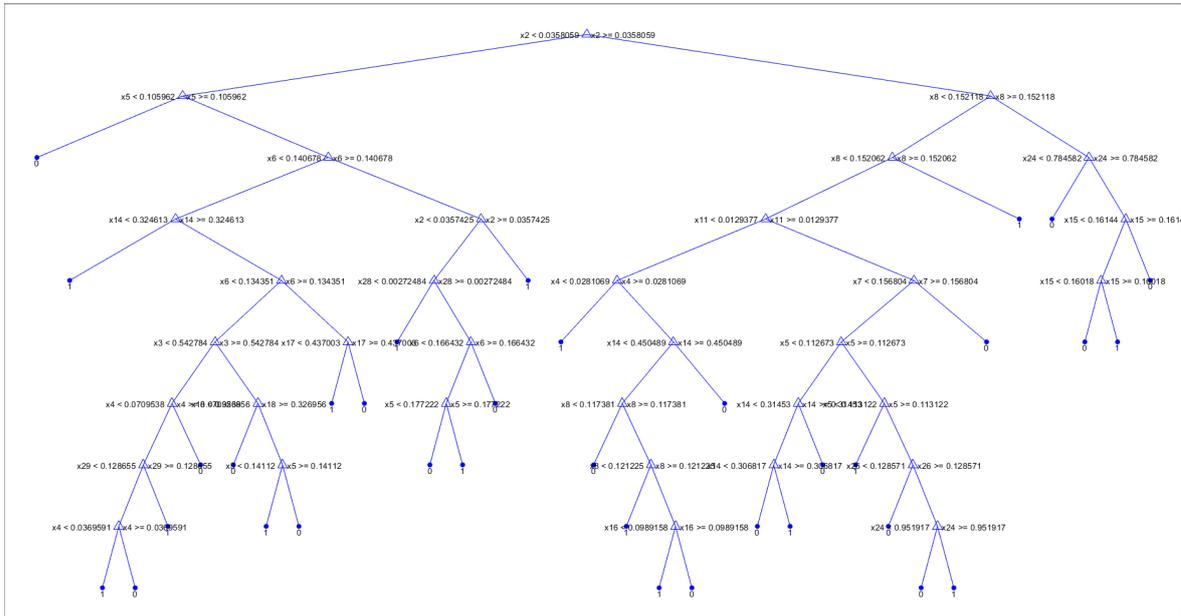


(u) Patient #171



(v) Patient #172





(y) Patient #182

## Performance on training sets of NNs Cascade

Performances obtained with training sets constituted with proportional sampling are provided in form of normalised confusion matrices. The class not included in the next neural network training set corresponds to the row enclosed by bold box in each table, while dashed boxes indicate the NN whose test set undergoes dendrogram clustering.

$NN_5^p$		Predicted Class					
		1	2	3	4	5	6
True class	1	<b>0.015</b>	0.224	0.460	0.223	0.067	0.010
	2	0.023	<b>0.614</b>	0.353	0.010	0.000	0.000
	3	0.001	0.117	<b>0.590</b>	0.264	0.028	0.000
	4	0.000	0.010	0.382	<b>0.551</b>	0.056	0.000
	5	0.000	0.003	0.135	0.628	<b>0.222</b>	0.013
	6	0.000	0.000	0.037	0.282	0.493	<b>0.187</b>

$NN_{10}^p$		Predicted Class					
		1	2	3	4	5	6
True class	1	<b>0.037</b>	0.344	0.315	0.209	0.081	0.014
	2	0.054	<b>0.627</b>	0.269	0.044	0.006	0.000
	3	0.001	0.305	<b>0.458</b>	0.197	0.037	0.001
	4	0.000	0.000	0.232	<b>0.715</b>	0.053	0.000
	5	0.000	0.010	0.117	0.562	<b>0.291</b>	0.020
	6	0.000	0.014	0.087	0.242	0.510	<b>0.146</b>

$NN_6^p$		Predicted Class				
		1	2	3	4	5
True class	1	<b>0.050</b>	0.463	0.413	0.071	0.004
	2	0.049	<b>0.895</b>	0.056	0.000	0.000
	3	0.001	0.354	<b>0.578</b>	0.065	0.001
	4	0.000	0.050	0.486	<b>0.447</b>	0.017
	5	0.000	0.015	0.423	0.469	<b>0.092</b>

$NN_{11}^p$		Predicted Class				
		1	2	3	4	5
True class	1	<b>0.062</b>	0.536	0.246	0.146	0.010
	2	0.032	<b>0.937</b>	0.031	0.000	0.000
	3	0.005	0.394	<b>0.445</b>	0.151	0.005
	4	0.000	0.000	0.190	<b>0.772</b>	0.038
	5	0.000	0.006	0.119	0.621	<b>0.254</b>

$NN_7^p$		Predicted Class			
		1	2	3	4
True class	1	<b>0.190</b>	0.555	0.245	0.010
	2	0.037	<b>0.936</b>	0.027	0.000
	3	0.003	0.432	<b>0.533</b>	0.032
	4	0.003	0.104	0.490	<b>0.404</b>

$NN_{12}^p$		Predicted Class			
		1	2	3	4
True class	1	<b>0.200</b>	0.658	0.128	0.014
	2	0.036	<b>0.954</b>	0.010	0.000
	3	0.006	0.556	<b>0.414</b>	0.023
	4	0.000	0.008	0.254	<b>0.738</b>

$NN_8^p$		Predicted Class		
		1	2	3
True class	1	<b>0.441</b>	0.519	0.040
	2	0.027	<b>0.963</b>	0.010
	3	0.012	0.481	<b>0.508</b>

$NN_{13}^p$		Predicted Class		
		1	2	3
True class	1	<b>0.260</b>	0.681	0.059
	2	0.018	<b>0.976</b>	0.006
	3	0.006	0.542	<b>0.451</b>

$NN_9^p$		Predicted Class	
		1	2
True class	1	<b>0.813</b>	0.187
	2	0.746	<b>0.254</b>

$NN_{14}^p$		Predicted Class	
		1	2
True class	1	<b>0.818</b>	0.182
	2	0.808	<b>0.192</b>

**Figure 3.47:** ADC map - NNs cascade with proportional sampling. Normalised confusion matrices regarding training sets from  $NN_5^p$  to  $NN_{14}^p$ . Rows enclosed in the bold boxes highlight the class which is not included in the training set of the successive neural network. Dashed boxes indicate the NN whose test set undergoes dendrogram clustering (1: malignant label, benign labels from 2 to 6)

$NN_{15}^p$		Predicted Class					
		1	2	3	4	5	6
True class	1	<b>0.082</b>	0.338	0.274	0.204	0.090	0.012
	2	0.080	<b>0.708</b>	0.204	0.008	0.000	0.000
	3	0.005	0.261	<b>0.472</b>	0.223	0.037	0.001
	4	0.000	0.018	0.421	<b>0.488</b>	0.073	0.000
	5	0.000	0.001	0.027	0.423	<b>0.522</b>	0.027
	6	0.000	0.001	0.026	0.260	0.563	<b>0.150</b>

$NN_{20}^p$		Predicted Class					
		1	2	3	4	5	6
True class	1	<b>0.054</b>	0.324	0.419	0.167	0.024	0.012
	2	0.099	<b>0.554</b>	0.292	0.042	0.013	0.000
	3	0.006	0.379	<b>0.553</b>	0.059	0.003	0.000
	4	0.001	0.021	0.476	<b>0.483</b>	0.019	0.000
	5	0.001	0.023	0.362	0.551	<b>0.063</b>	0.000
	6	0.000	0.000	0.034	0.139	0.343	<b>0.484</b>

$NN_{16}^p$		Predicted Class				
		1	2	3	4	5
True class	1	<b>0.177</b>	0.491	0.197	0.118	0.017
	2	0.076	<b>0.758</b>	0.162	0.004	0.000
	3	0.012	0.356	<b>0.537</b>	0.092	0.004
	4	0.000	0.024	0.403	<b>0.506</b>	0.067
	5	0.000	0.004	0.047	0.543	<b>0.406</b>

$NN_{21}^p$		Predicted Class				
		1	2	3	4	5
True class	1	<b>0.104</b>	0.331	0.378	0.171	0.017
	2	0.078	<b>0.713</b>	0.204	0.005	0.000
	3	0.008	0.236	<b>0.631</b>	0.123	0.003
	4	0.000	0.012	0.297	<b>0.632</b>	0.060
	5	0.003	0.012	0.205	0.524	<b>0.256</b>

$NN_{17}^p$		Predicted Class			
		1	2	3	4
True class	1	<b>0.271</b>	0.555	0.145	0.029
	2	0.075	<b>0.806</b>	0.111	0.009
	3	0.028	0.492	<b>0.460</b>	0.021
	4	0.000	0.026	0.447	<b>0.527</b>

$NN_{22}^p$		Predicted Class			
		1	2	3	4
True class	1	<b>0.105</b>	0.582	0.297	0.015
	2	0.050	<b>0.946</b>	0.004	0.000
	3	0.001	0.439	<b>0.554</b>	0.005
	4	0.000	0.033	0.617	<b>0.349</b>

$NN_{18}^p$		Predicted Class		
		1	2	3
True class	1	<b>0.240</b>	0.735	0.026
	2	0.026	<b>0.967</b>	0.008
	3	0.010	0.710	<b>0.280</b>

$NN_{19}^p$		Predicted Class	
		1	2
True class	1	<b>0.759</b>	0.241
	2	0.715	<b>0.285</b>

$NN_{23}^p$		Predicted Class		
		1	2	3
True class	1	<b>0.423</b>	0.504	0.073
	2	0.124	<b>0.862</b>	0.014
	3	0.017	0.565	<b>0.419</b>

$NN_{24}^p$		Predicted Class	
		1	2
True class	1	<b>0.800</b>	0.200
	2	0.644	<b>0.356</b>

**Figure 3.48:** ADC map - NNs cascade with proportional sampling. Normalised confusion matrices regarding training sets from  $NN_{15}^p$  to  $NN_{24}^p$ . Rows enclosed in the bold boxes highlight the class which is not included in the training set of the successive neural network. Dashed boxes indicate the NN whose test set undergoes dendrogram clustering (1: malignant label, benign labels from 2 to 6)

$NN_{25}^p$		Predicted Class					
		1	2	3	4	5	6
True class	1	<b>0.018</b>	0.141	0.436	0.300	0.088	0.017
	2	0.113	<b>0.649</b>	0.233	0.005	0.000	0.000
	3	0.003	0.117	<b>0.595</b>	0.262	0.023	0.001
	4	0.000	0.019	0.342	<b>0.609</b>	0.030	0.000
	5	0.001	0.032	0.323	0.544	<b>0.099</b>	0.001
	6	0.001	0.000	0.068	0.362	0.452	<b>0.117</b>

$NN_{26}^p$		Predicted Class				
		1	2	3	4	5
True class	1	<b>0.154</b>	0.312	0.335	0.185	0.015
	2	0.051	<b>0.667</b>	0.276	0.006	0.000
	3	0.013	0.285	<b>0.568</b>	0.129	0.005
	4	0.000	0.013	0.285	<b>0.650</b>	0.053
	5	0.000	0.015	0.244	0.599	<b>0.142</b>

$NN_{27}^p$		Predicted Class			
		1	2	3	4
True class	1	<b>0.227</b>	0.600	0.155	0.018
	2	0.019	<b>0.867</b>	0.114	0.000
	3	0.005	0.509	<b>0.485</b>	0.001
	4	0.002	0.015	0.403	<b>0.580</b>

$NN_{28}^p$		Predicted Class		
		1	2	3
True class	1	<b>0.440</b>	0.531	0.029
	2	0.005	<b>0.994</b>	0.001
	3	0.027	0.768	<b>0.205</b>

$NN_{29}^p$		Predicted Class	
		1	2
True class	1	<b>0.955</b>	0.045
	2	0.876	<b>0.124</b>

$NN_{30}^p$		Predicted Class						
		1	2	3	4	5	6	7
True class	1	<b>0.026</b>	0.104	0.168	0.341	0.287	0.058	0.017
	2	0.023	<b>0.327</b>	0.489	0.143	0.018	0.000	0.000
	3	0.008	0.141	<b>0.226</b>	0.332	0.245	0.047	0.000
	4	0.000	0.000	0.004	<b>0.295</b>	0.607	0.092	0.001
	5	0.001	0.010	0.063	0.380	<b>0.474</b>	0.068	0.004
	6	0.000	0.004	0.064	0.282	0.505	<b>0.138</b>	0.006
	7	0.000	0.002	0.005	0.016	0.182	0.435	<b>0.360</b>

$NN_{31}^p$		Predicted Class					
		1	2	3	4	5	6
True class	1	<b>0.045</b>	0.191	0.317	0.351	0.088	0.008
	2	0.056	<b>0.506</b>	0.355	0.082	0.001	0.000
	3	0.006	0.282	<b>0.524</b>	0.180	0.008	0.000
	4	0.000	0.005	0.325	<b>0.621</b>	0.049	0.000
	5	0.000	0.004	0.160	0.588	<b>0.230</b>	0.018
	6	0.003	0.015	0.126	0.373	0.419	<b>0.064</b>

$NN_{32}^p$		Predicted Class				
		1	2	3	4	5
True class	1	<b>0.076</b>	0.318	0.356	0.231	0.019
	2	0.040	<b>0.825</b>	0.134	0.001	0.000
	3	0.008	0.266	<b>0.689</b>	0.036	0.001
	4	0.000	0.005	0.341	<b>0.644</b>	0.009
	5	0.001	0.015	0.182	0.692	<b>0.109</b>

$NN_{33}^p$		Predicted Class			
		1	2	3	4
True class	1	<b>0.390</b>	0.419	0.159	0.032
	2	0.022	<b>0.951</b>	0.027	0.000
	3	0.003	0.377	<b>0.569</b>	0.051
	4	0.003	0.017	0.357	<b>0.624</b>

$NN_{34}^p$		Predicted Class		
		1	2	3
True class	1	<b>0.505</b>	0.432	0.063
	2	0.017	<b>0.981</b>	0.003
	3	0.004	0.576	<b>0.420</b>

$NN_{35}^p$		Predicted Class	
		1	2
True class	1	<b>0.873</b>	0.127
	2	0.807	<b>0.193</b>

**Figure 3.49:** ADC map - NNs cascade with proportional sampling. Normalised confusion matrices regarding training sets from  $NN_{25}^p$  to  $NN_{35}^p$ . Rows enclosed in the bold boxes highlight the class which is not included in the training set of the successive neural network. Dashed boxes indicate the NN whose test set undergoes dendrogram clustering (1: malignant label, benign labels from 2 to 7)

$NN_{36}^p$		<i>Predicted Class</i>						
		1	2	3	4	5	6	7
<i>True class</i>	1	<b>0.048</b>	0.228	0.261	0.242	0.153	0.068	0.014
	2	0.055	<b>0.713</b>	0.226	0.006	0.000	0.000	0.000
	3	0.004	0.176	<b>0.732</b>	0.089	0.000	0.000	0.000
	4	0.000	0.001	0.236	<b>0.679</b>	0.082	0.001	0.000
	5	0.000	0.010	0.210	0.422	<b>0.278</b>	0.080	0.008
	6	0.000	0.001	0.020	0.191	0.463	<b>0.324</b>	0.064
	7	0.000	0.000	0.004	0.015	0.202	0.780	<b>0.471</b>

$NN_{37}^p$		<i>Predicted Class</i>					
		1	2	3	4	5	6
<i>True class</i>	1	<b>0.079</b>	0.250	0.309	0.249	0.095	0.018
	2	0.097	<b>0.659</b>	0.236	0.008	0.000	0.000
	3	0.010	0.227	<b>0.648</b>	0.110	0.004	0.000
	4	0.001	0.024	0.372	<b>0.515</b>	0.083	0.004
	5	0.001	0.021	0.181	0.489	<b>0.285</b>	0.023
	6	0.000	0.003	0.033	0.246	0.506	<b>0.212</b>

$NN_{38}^p$		<i>Predicted Class</i>				
		1	2	3	4	5
<i>True class</i>	1	<b>0.106</b>	0.279	0.418	0.192	0.004
	2	0.059	<b>0.701</b>	0.237	0.003	0.000
	3	0.000	0.154	<b>0.809</b>	0.037	0.000
	4	0.001	0.014	0.629	<b>0.354</b>	0.001
	5	0.000	0.009	0.325	0.634	<b>0.032</b>

$NN_{39}^p$		<i>Predicted Class</i>			
		1	2	3	4
<i>True class</i>	1	<b>0.333</b>	0.431	0.199	0.037
	2	0.056	<b>0.853</b>	0.091	0.000
	3	0.001	0.279	<b>0.678</b>	0.042
	4	0.000	0.027	0.513	<b>0.460</b>

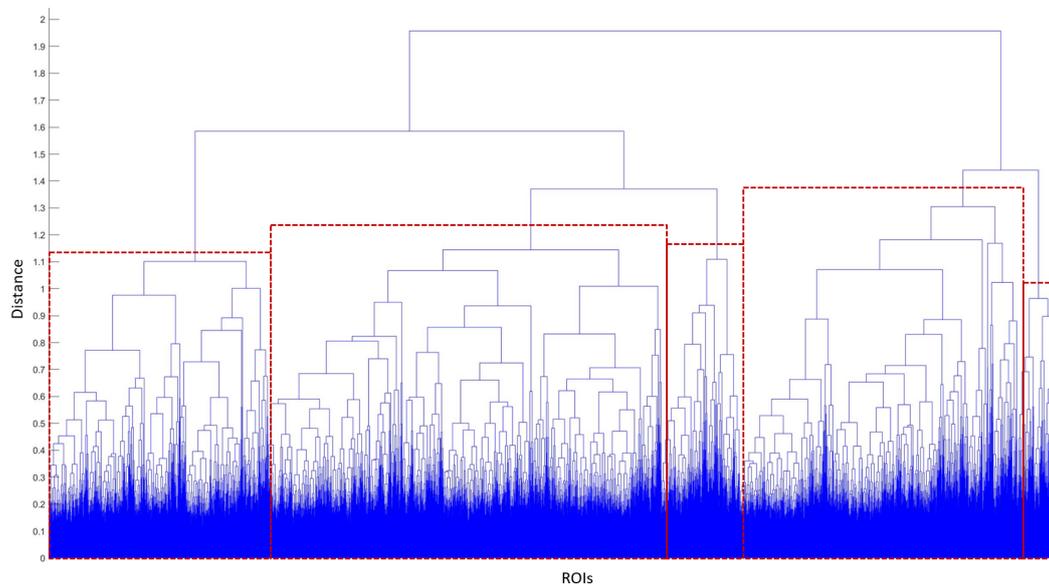
$NN_{40}^p$		<i>Predicted Class</i>		
		1	2	3
<i>True class</i>	1	<b>0.577</b>	0.390	0.033
	2	0.042	<b>0.941</b>	0.017
	3	0.000	0.508	<b>0.492</b>

$NN_{41}^p$		<i>Predicted Class</i>	
		1	2
<i>True class</i>	1	<b>0.762</b>	0.238
	2	0.793	<b>0.207</b>

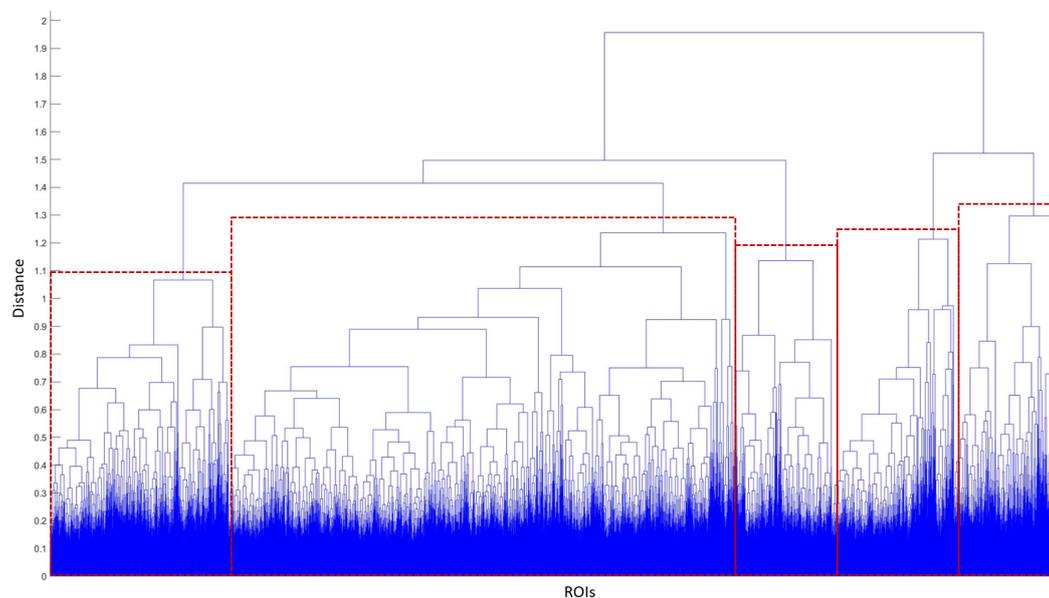
**Figure 3.50:** ADC map - NNs cascade with proportional sampling. Normalised confusion matrices regarding training sets from  $NN_{36}^p$  to  $NN_{41}^p$ . Rows enclosed in the bold boxes highlight the class which is not included in the training set of the successive neural network (1: malignant label, benign labels from 2 to 7)

## Dendrograms of NNs Cascade

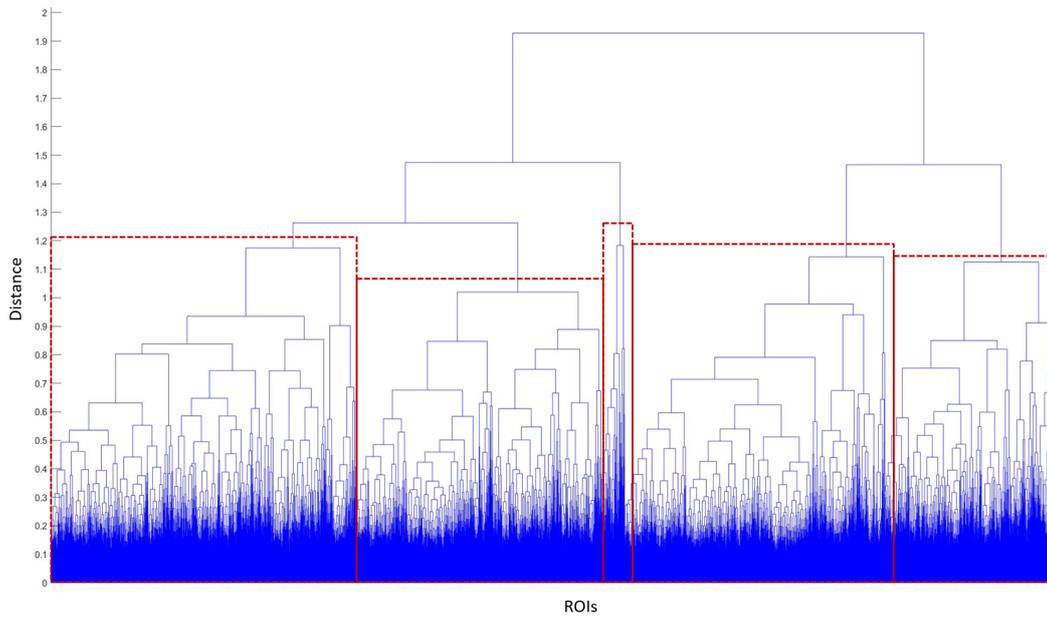
Dendrograms and their clusters belonging to test sets of NNs cascade with proportional sampling are reported in this section.



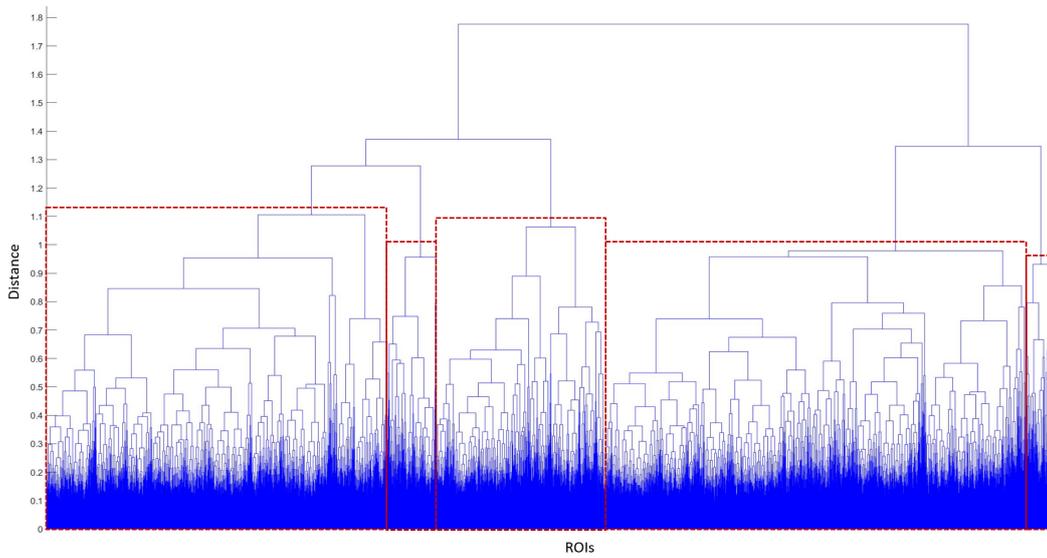
**Figure 3.51:** ADC map - NNs cascade with proportional sampling. Dendrogram and clusters belonging to test set of  $NN_8^p$ . Dashed red boxes enclose clusters



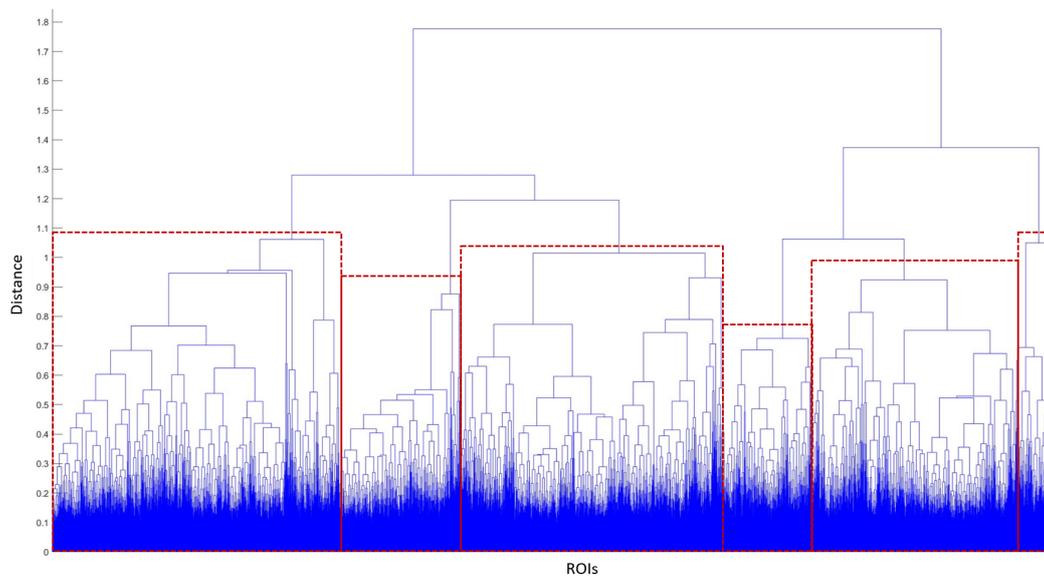
**Figure 3.52:** ADC map - NNs cascade with proportional sampling. Dendrogram and clusters belonging to test set of  $NN_{13}^p$ . Dashed red boxes enclose clusters



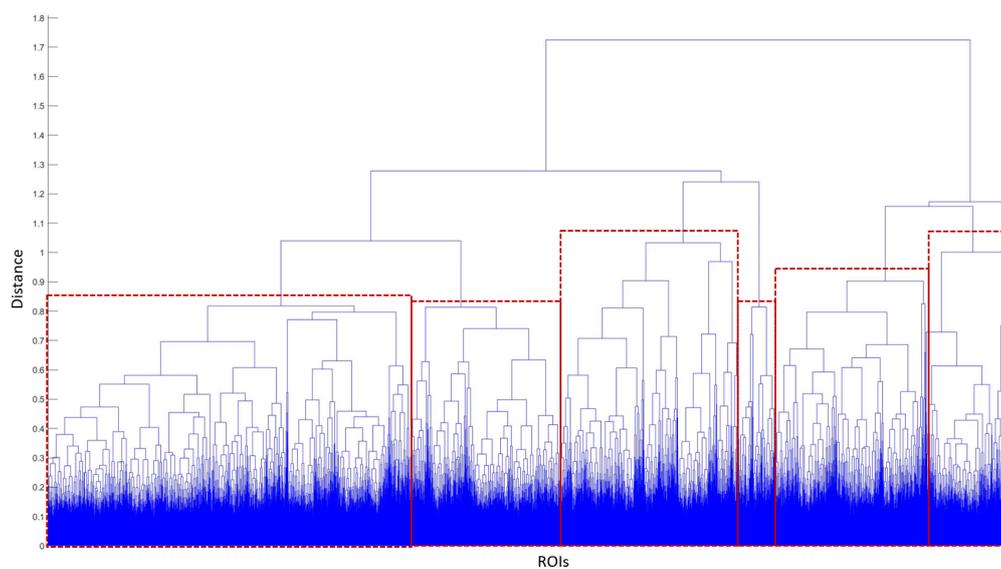
**Figure 3.53:** ADC map - NNs cascade with proportional sampling. Dendrogram and clusters belonging to test set of  $NN_{19}^p$ . Dashed red boxes enclose clusters



**Figure 3.54:** ADC map - NNs cascade with proportional sampling. Dendrogram and clusters belonging to test set of  $NN_{23}^p$ . Dashed red boxes enclose clusters



**Figure 3.55:** ADC map - NNs cascade with proportional sampling. Dendrogram and clusters belonging to test set of  $NN_{28}^p$ . Dashed red boxes enclose clusters



**Figure 3.56:** ADC map - NNs cascade with proportional sampling. Dendrogram and clusters belonging to test set of  $NN_{33}^p$ . Dashed red boxes enclose clusters

# List of Abbreviations

<b>Abbreviation</b>	<b>Definition</b>
ADC	Apparent Diffusion Coefficient
AM	Acquisition Matrix
AS	Anterior fibromuscular Stroma
BB	Bounding Box
bp-MR	bi-parametric Magnetic Resonance
CAD	Computer Aided Diagnosis
CZ	Central Zone
Dend	Dendrogram based training set
DRE	Digital Rectal Examination
DT	Decision Tree
DW	Diffusion Weighted
ERC	Endorectal coil
FE	Feature Extraction
FN	False Negative
FOV	Field of view
FP	False Positive
FS	Feature Selection
GA	Genetic Algorithm
GLCM	Gray Level Co-occurrence Matrix
GLNU	Gray Level Non-uniformity
GLRLM	Gray Level Run Length Matrix
GS	Gleason Score
HGRE	High Gray-Level Run Emphasis
ITK	Insight Segmentation and Registration Toolkit
k-NN	k-Nearest Neighbours
LGRE	Low Gray-Level Run Emphasis
LOE	Largest Object Extraction
MaxMedian	Maximum median sensitivity criterion
MinDiff	Minimum difference criterion
MLP	Multilayer Perceptron
mp-MR	multi-parametric Magnetic Resonance
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MRSI	Magnetic Resonance Spectroscopy Imaging
NEX	Number of excitations
NN	Neural Network

<b>Abbreviation</b>	<b>Definition</b>
P	Label for tissue belonging to Prostate
PAC	Pelvic phased Array Coil
PCa	Prostate Cancer
PI-RADS	Prostate Imaging Reporting And Data System
Prop	Proportional sampling training set
PSA	Prostate Specific Antigen
PZ	Peripheral Zone
RLNU	Run Length Non-uniformity
RM	Reconstruction Matrix
ROI	Region of Interest
SD	Standard Deviation
Sens	Sensitivity
Spec	Specificity
ST	Slice Thickness
SV	Seminal Vesicles
SVM	Support Vector Machine
T	Label for Tissue other than prostate
T2-w	T2-weighted
TRUS	Transrectal Ultrasound
TN	True Negative
TP	True Positive
TZ	Transition Zones
US	Urethral Sphincter

# List of Figures

1.1	Representation of the prostate along with surrounding organs [7] . . . . .	4
1.2	Detailed illustration of prostate areas (PZ: peripheral zones, TZ: transition zones, CZ: central zone, AS: anterior fibromuscular stroma, SV: seminal vesicles, US: urethral sphincter, a: anterior, mp: medial posterior, lp: lateral posterior). The prostate is divided laterally by a vertical dashed line and into anterior/posterior by a horizontal dashed line passing through the centre of the gland [10] . . . . .	5
1.3	Most widespread cancers in men considering 1984, 2007 and a prediction for 2030, showed using pie charts with areas proportional to cases number [3]	6
1.4	Detection of prostate cancer by means of a computer aided diagnosis system [38] . . . . .	9
2.1	Schematic representation of the study population for prostate segmentation	10
2.2	Representative examples to visualise the problem of misalignment between image modalities . . . . .	12
2.3	Phases of segmentation algorithm . . . . .	13
2.4	From left to right: ADC map given as input, image resulting from preprocessing, image representing k-means pixel clustering . . . . .	14
2.5	On the left: result of thresholding applied on output of k-means algorithm. On the right: image obtained after largest object extraction . . . . .	15
2.6	Flowchart for retrieving largest object in each slice based on centroid coordinates ( $[C_x, C_y]$ : centroid coordinates respectively along $x$ and $y$ , $L_1$ : 25% of horizontal image dimension, $L_2$ : 75% of horizontal image dimension, $n$ : objects counter) . . . . .	16
2.7	Phases of bounding box computation . . . . .	16
2.8	Logical operations performed beginning from k-means cluster comprising the edges of present structures (on the upper left) and <i>LOE mask</i> (on the upper right) . . . . .	17
2.9	Flowchart for retrieving median bounding box coordinates ( $L_1$ : 70% of vertical image dimension, $C$ can be either $X$ or $Y$ , subscript $m$ indicates median coordinates considering slices $[n - 1, n + 1]$ , while subscript $u$ indicates median coordinates considering slices $[n + 2, n + 4]$ ) . . . . .	18
2.10	From left to right: final mask on ADC map and T2-w image . . . . .	19
2.11	Median value of precision considering base, midgland and apex for each patient . . . . .	22
2.12	Median value of Dice coefficient considering base, midgland and apex for each patient . . . . .	23
2.13	Minimum, median and maximum value of precision considering base, midgland and apex for each patient . . . . .	24

2.14	Minimum, median and maximum value of Dice coefficient considering base, midgland and apex for each patient . . . . .	25
2.15	Outcome of the segmentation algorithm. On the left, prostate mask applied on ADC map. On the right, prostate mask applied on T2-w . . . . .	26
3.1	Schematic representation highlighting every step implemented for tumour detection . . . . .	27
3.2	Schematic representation of the study population for tumour detection . . .	28
3.3	Simplified example about GLCM calculation. 1) An image of 4×4 grey tones is taken into account and transformed into a numeric grid. 2) Considering a pixel with intensity 1 and its immediate neighbour to the right (i.e. 3), there are two occurrences in the image, thus the entry (1, 3) of the GLCM matrix is filled with 2. Following the same criterion, the entire GLCM is filled. 3) The gray-level co-occurrence matrix is thus normalised so that each entry holds the probability of every pixel pair. Adapted from [62] . . . . .	30
3.4	Representation of the four directions along with texture features are computed. The red cross designates the center of the 5 × 5 ROI . . . . .	31
3.5	Phases of decision tree based approach for training set construction . . . . .	33
3.6	Emblematic examples of decision trees for patients with different number of malignant samples and a comparable number of benign samples (superior to 2500). Red boxes highlight leaf nodes assigned malignant label . . . . .	34
3.7	Occurrence of selected attributes across all decision trees . . . . .	36
3.8	Schematic representation of an evolutionary algorithm . . . . .	38
3.9	Soft-margin support vector machine in a 2-D example [65] . . . . .	39
3.10	Schematic representation of GA solution encoding with SVM parameters. Each square represents a single bit . . . . .	40
3.11	Implementation of the genetic algorithm to set classifier parameters and selected features . . . . .	41
3.12	<i>SVM</i> . Sensitivity and specificity across patients . . . . .	44
3.13	<i>SVM</i> . Examples of tumour detection. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue . . . . .	45
3.14	Dendrograms with overlapping red boxes representing clusters . . . . .	47
3.15	<i>SVM vs Multiclass SVM</i> . Classification performances across patients on ADC map . . . . .	50
3.16	<i>SVM vs Multiclass SVM</i> . Classification performances across patients on T2-weighted modality . . . . .	51
3.17	<i>SVM vs Multiclass SVM</i> . Examples of tumour detection for a patient with several false positives. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue . . . . .	52
3.18	Schematic representation of GA solution encoding with k-NN parameters. Each square represents a single bit . . . . .	53
3.19	<i>Multiclass SVM vs Multiclass k-NN</i> . Classification performances across patients on ADC map . . . . .	55

3.20	<i>Multiclass SVM vs Multiclass k-NN</i> . Classification performances across patients on T2-weighted modality . . . . .	56
3.21	<i>SVM vs Multiclass SVM</i> . Examples of tumour detection referred to T2-w. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous areas . . . . .	57
3.22	<i>SVM vs Multiclass SVM</i> . Examples of tumour detection referred to ADC map. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous areas . . . . .	57
3.23	Representation of an artificial neuron [65] . . . . .	58
3.24	Example of a two-layer MLP [65] . . . . .	59
3.25	Input neurons log-sigmoid activation function . . . . .	61
3.26	<i>DT based training set - MaxMedian criterion</i> . Performance proper of each NN structure in terms of specificity and sensitivity evaluated on training and test sets. Dashed boxes indicate the structure elected as best . . . . .	62
3.27	<i>DT based training set - MinDiff criterion</i> . Performance proper of each NN structure in terms of specificity and sensitivity evaluated on training and test sets. Dashed boxes indicate the structure elected as best . . . . .	63
3.28	<i>DT based training set</i> . Sensitivity and specificity across patients considering the best elected structures . . . . .	64
3.29	<i>DT based training set - T2-w</i> . Examples of tumour detection according to best elected NN structures. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue . . . . .	66
3.30	<i>DT based training set - ADC map</i> . Examples of tumour detection according to best elected NN structures. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue . . . . .	66
3.31	<i>Dendrogram based training set - MaxMedian criterion</i> . Performance proper of each NN structure in terms of specificity and sensitivity evaluated on training and test sets. Dashed boxes indicate the structure elected as best . . . . .	67
3.32	<i>Dendrogram based training set - MinDiff criterion</i> . Performance proper of each NN structure in terms of specificity and sensitivity evaluated on training and test sets. Dashed boxes indicate the structure elected as best . . . . .	68
3.33	<i>Dendrogram based training set</i> . Sensitivity and specificity across patients considering the best elected structures . . . . .	69
3.34	<i>Dendrogram based training set - T2-w</i> . Examples of tumour detection according to best elected NN structures. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue . . . . .	71
3.35	<i>Dendrogram based training set - ADC map</i> . Examples of tumour detection according to best elected NN structures. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue . . . . .	71
3.36	<i>ADC map</i> . Dendrogram clustering on benign samples of $NN_4$ test set. Dashed red boxes enclose clusters . . . . .	74
3.37	<i>ADC map</i> . Dendrogram on class 2 from remaining test set of $NN_4$ . Dashed red boxes enclose clusters . . . . .	76

3.38	<i>ADC map</i> . Dendrogram on class 3 from remaining test set of $NN_4$ . Dashed red boxes enclose clusters . . . . .	77
3.39	<i>ADC map - NNs cascade with proportional sampling</i> . Schematic representation of the procedure adopted to implement cascade of NNs . . . . .	80
3.40	<i>ADC map - NNs cascade with proportional sampling</i> . Each row comprises a NNs cascade. The bars represent the normalised number of malignant and benign samples according to NN prediction on test set. Dashed boxes highlight the test set whose samples will be used in the creation of the training set for the following cascade (1: malignant label, benign labels from 2 to 7) . . . . .	81
3.41	<i>ADC map - NNs Cascade with Proportional Sampling</i> . Patient-wise sensitivity ( <i>Sens</i> ) and specificity ( <i>Spec</i> ) . . . . .	82
3.42	<i>ADC map - NNs cascade with proportional sampling</i> . Examples of tumour detection. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous areas . . . . .	83
3.43	<i>ADC map</i> . Comparison in terms of number of false positives, sensitivity and specificity considering all classifiers. (DT: Decision tree based training set, Dend: Dendrogram based training set, Prop: Proportional sampling training set) . . . . .	84
3.44	<i>T2-weighted</i> . Comparison in terms of number of false positives, sensitivity and specificity considering all classifiers. (DT: Decision tree based training set, Dend: Dendrogram based training set) . . . . .	85
3.45	<i>ADC map</i> . Examples of tumour detection performed by all classifiers, referring to large, medium and small cancerous areas. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue . . . . .	86
3.46	<i>T2-weighted</i> . Examples of tumour detection performed by all classifiers, referring to large, medium and small cancerous areas. Red dots indicate the centre of each ROI classified as malignant, while blank squares correspond to cancerous tissue . . . . .	87
3.47	<i>ADC map - NNs cascade with proportional sampling</i> . Normalised confusion matrices regarding training sets from $NN_5^p$ to $NN_{14}^p$ . Rows enclosed in the bold boxes highlight the class which is not included in the training set of the successive neural network. Dashed boxes indicate the NN whose test set undergoes dendrogram clustering (1: malignant label, benign labels from 2 to 6) . . . . .	122
3.48	<i>ADC map - NNs cascade with proportional sampling</i> . Normalised confusion matrices regarding training sets from $NN_{15}^p$ to $NN_{24}^p$ . Rows enclosed in the bold boxes highlight the class which is not included in the training set of the successive neural network. Dashed boxes indicate the NN whose test set undergoes dendrogram clustering (1: malignant label, benign labels from 2 to 6) . . . . .	123

3.49	<i>ADC map - NNs cascade with proportional sampling.</i> Normalised confusion matrices regarding training sets from $NN_{25}^p$ to $NN_{35}^p$ . Rows enclosed in the bold boxes highlight the class which is not included in the training set of the successive neural network. Dashed boxes indicate the NN whose test set undergoes dendrogram clustering (1: malignant label, benign labels from 2 to 7) . . . . .	124
3.50	<i>ADC map - NNs cascade with proportional sampling.</i> Normalised confusion matrices regarding training sets from $NN_{36}^p$ to $NN_{41}^p$ . Rows enclosed in the bold boxes highlight the class which is not included in the training set of the successive neural network (1: malignant label, benign labels from 2 to 7) . .	125
3.51	<i>ADC map - NNs cascade with proportional sampling.</i> Dendrogram and clusters belonging to test set of $NN_8^p$ . Dashed red boxes enclose clusters . . . .	126
3.52	<i>ADC map - NNs cascade with proportional sampling.</i> Dendrogram and clusters belonging to test set of $NN_{13}^p$ . Dashed red boxes enclose clusters . . . .	126
3.53	<i>ADC map - NNs cascade with proportional sampling.</i> Dendrogram and clusters belonging to test set of $NN_{19}^p$ . Dashed red boxes enclose clusters . . . .	127
3.54	<i>ADC map - NNs cascade with proportional sampling.</i> Dendrogram and clusters belonging to test set of $NN_{23}^p$ . Dashed red boxes enclose clusters . . . .	127
3.55	<i>ADC map - NNs cascade with proportional sampling.</i> Dendrogram and clusters belonging to test set of $NN_{28}^p$ . Dashed red boxes enclose clusters . . . .	128
3.56	<i>ADC map - NNs cascade with proportional sampling.</i> Dendrogram and clusters belonging to test set of $NN_{33}^p$ . Dashed red boxes enclose clusters . . . .	128

# List of Tables

2.1	Characteristics of the study population. Measurements are given as <i>median (range)</i>	10
2.2	MRI protocol for T2-weighted and diffusion weighted modalities (ST: slice thickness, FOV: field of view, NEX: number of excitations, AM: acquisition matrix, RM: reconstruction matrix, SE: spin echo, EPI: echo planar imaging)	11
2.3	Confusion matrix used to compute metrics (P: label for prostate, T: label for tissue other than prostate, TP: true positive, FP: false positive, TN: true negative, FN: false negative)	20
2.4	Metrics to evaluate segmentation performances considering the entire set of patients. Measurements are given as <i>median (range)</i>	21
3.1	Characteristics of patients and lesions. *Measurements are given as <i>median (range)</i>	28
3.2	List of extracted features (SD: standard deviation, GLCM: Gray Level Co-occurrence Matrix, GLRLM: Gray Level Run Length Matrix, HGRE: High Gray-Level Run Emphasis, LGRE: Low Gray-Level Run Emphasis, GLNU: Gray-Level Non-uniformity, RLNU: Run Length Non-uniformity)	29
3.3	Total number of ROIs separated per class and image type	32
3.4	Number of ROIs separated per class and per image type constituting decision tree based training set	35
3.5	<i>SVM</i> . Parameters selected by the genetic algorithm and best fitness values	41
3.6	<i>SVM</i> . Binary representation of FS performed by the genetic algorithms. White cells correspond to selected variables, black ones indicate those features that are not considered	42
3.7	<i>SVM</i> . Classification performances considering training and test sets per image modality. The reference confusion matrix is reported on the right (1: malignant label, 0: benign label)	42
3.8	Number of ROIs reported per class and per image type constituting dendrogram based training set	46
3.9	<i>Multiclass SVM</i> . Parameters selected by the genetic algorithm and best fitness values	48
3.10	<i>Multiclass SVM</i> . Binary representation of FS performed by the genetic algorithms. White cells correspond to selected variables, black ones indicate those features that are not considered	49
3.11	<i>Multiclass SVM</i> . Classification performances considering training and test sets per image modality. The reference confusion matrix is reported on the right (1: malignant label, 0: benign label)	49
3.12	<i>Multiclass k-NN</i> . Parameters selected by the genetic algorithm and best fitness values	53
3.13	<i>Multiclass k-NN</i> . Binary representation of FS performed by the genetic algorithm. White cells correspond to selected variables, black ones indicate those features that are not considered	54

3.14	<i>Multiclass k-NN</i> . Classification performances considering training and test sets per image modality. The reference confusion matrix is reported on the right (1: malignant label, 0: benign label) . . . . .	54
3.15	List of tested hidden layers (number of neurons in each layer is provided) . . . . .	60
3.16	<i>DT based training set</i> . Classification performances obtained considering training and test sets per image modality and best elected NN structures. The reference confusion matrix is reported on the right (1: malignant label, 0: benign label) . . . . .	65
3.17	<i>Dendrogram based training set</i> . Classification performances considering training and test sets per image modality and best elected NN structures. The reference confusion matrix is reported on the right (1: malignant label, 0: benign label) . . . . .	70
3.18	NNs performance comparison between the two employed training sets, given per image modality. *The reported number is the upper or lower bound of existing values (regarding sensitivity and specificity) . . . . .	72
3.19	<i>ADC map - 1<sup>st</sup> NNs Cascade</i> . Normalised confusion matrices belonging to each neural network. Rows enclosed in the bold boxes highlight the class which is not included in the training set of the successive neural network (1: malignant label, benign labels from 2 to 5) . . . . .	73
3.20	<i>ADC map - 1<sup>st</sup> NNs Cascade</i> . Normalised confusion matrix belonging to test set (1: malignant class, 0: benign class) . . . . .	74
3.21	<i>ADC map - 2<sup>nd</sup> NNs Cascade</i> . Normalised confusion matrices belonging to each neural network training set. Rows enclosed in the bold boxes highlight the class which is not included in training set of the successive neural network (1: malignant label, benign labels from 2 to 6) . . . . .	75
3.22	<i>ADC map - 1<sup>st</sup> and 2<sup>nd</sup> NNs Cascades</i> . Normalised confusion matrix considering complete test set (1: malignant class, 0: benign class) . . . . .	76
3.23	<i>ADC map - 3<sup>rd</sup> NNs Cascade</i> . Normalised confusion matrices belonging to each neural network. Rows enclosed in the bold boxes highlight the class which is removed in the training set of the successive neural network (1: malignant label, benign labels from 2 to 6) . . . . .	78
3.24	<i>ADC map - 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> NNs Cascade</i> . Normalised confusion matrix considering complete test set (1: malignant class, 0: benign class) . . . . .	79
3.25	Template providing normalised number of malignant and benign elements per NN prediction. Class 1 denotes malignant samples, while benign classes are denoted from 2 to $n$ . . . . .	81
3.26	<i>ADC map - NNs cascade with proportional sampling</i> . Normalised confusion matrix considering complete test set (1: malignant class, 0: benign class) . . . . .	82
3.27	Parameters notation adopted in the computation of GLCM texture features . . . . .	91
3.28	Abbreviations for GLCM texture features . . . . .	91
3.29	Parameters notation adopted in the computation of GLRLM texture features . . . . .	95
3.30	Abbreviations for GLRLM texture features . . . . .	95

# Bibliography

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA Cancer J Clin*, 2016.
- [2] J. Ferlay, E. Steliarova-Foucher, J. Lortet-Tieulent, S. Rosso, J. W. W. Coebergh, H. Comber, D. Forman, and F. Bray, "Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012," *Eur J Cancer*, 2013.
- [3] M. Mistry, D. M. Parkin, A. S. Ahmad, and P. Sasieni, "Cancer incidence in the United Kingdom: projections to the year 2030," *Br J Cancer*, 2011.
- [4] N. Arumainayagam, H. U. Ahmed, C. M. Moore, A. Freeman, C. Allen, S. A. Sohaib, A. Kirkham, J. van der Meulen, and M. Emberton, "Multiparametric MR imaging for detection of clinically significant prostate cancer: a validation cohort study with transperineal template prostate mapping as the reference standard," *Radiology*, 2013.
- [5] F. Russo, D. Regge, E. Armando, V. Giannini, A. Vignati, S. Mazzetti, M. Manfredi, E. Bollito, L. Correale, and F. Porpiglia, "Detection of prostate cancer index lesions with multiparametric magnetic resonance imaging (mp-MRI) using whole-mount histological sections as the reference standard," *BJU Int*, 2016.
- [6] "Comparison of multi-parametric and bi-parametric magnetic resonance imaging in men with suspicion of prostate cancer." <https://clinicaltrials.gov/ct2/show/NCT03693703>. Accessed: 4 March 2019.
- [7] "What is prostate cancer?" <https://www.cancer.org/cancer/prostate-cancer/about/what-is-prostate-cancer.html>. Accessed: 12 January 2019.
- [8] J. E. McNeal, *Prostate*. Alan R. Liss, Inc., 1981.
- [9] L. Aaron, O. Franco, and S. W. Hayward, "Review of Prostate Anatomy and Embryology and the Etiology of BPH," *Urologic Clinics of North America*, 2016.
- [10] "Pi-rads prostate imaging - reporting data system 2015 version 2." <https://www.acr.org/-/media/ACR/Files/RADS/Pi-RADS/PIRADS-V2.pdf?la=en>. Accessed: 12 January 2019.
- [11] S. Thomas and A. Oto, "Multiparametric MR imaging of the Prostate Pitfalls in Interpretation," *Radiol Clin N Am*, 2017.
- [12] "Key statistics for prostate cancer." <https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html>. Accessed: 12 January 2019.

- [13] “Tumore della prostata.” <https://www.airc.it/cancro/informazioni-tumori/guida-ai-tumori/tumore-della-prostata>. Accessed: 12 January 2019.
- [14] N. R. M. Hoffman, F. D. Gillil, J. W. Eley, L. C. Harlan, R. A. Stephenson, J. L. Stanford, P. C. Albertson, A. S. Hamilton, W. C. Hunt, and A. Potosky, “Racial and ethnic differences in advanced-stage prostate cancer: the Prostate Cancer Outcomes Study,” *Journal of the National Cancer Institute, Volume 93*, 2001.
- [15] G. Oster, L. Lamerato, A. G. Glass, K. E. Richert-Boe, A. Lopez, K. Chung, A. Richhariya, T. Dodge, G. G. Wolff, A. Balakumaran, and J. Edelsberg, “What every doctor who treats male patients should know,” *PCRI Insights 8*, 2005.
- [16] G. L. Lu-Yao, P. C. Albertsen, D. F. Moore, W. Shih, Y. Lin, R. S. DiPaola, M. J. Barry, A. Zietman, M. O’Leary, E. Walker-Corkery, and S. L. Yao, “Outcomes of localized prostate cancer following conservative management,” *Journal of the National Cancer Institute, Volume 93*, 2001.
- [17] G. Oster, L. Lamerato, A. G. Glass, K. E. Richert-Boe, A. Lopez, K. Chung, A. Richhariya, T. Dodge, G. G. Wolff, A. Balakumaran, and J. Edelsberg, “Natural history of skeletal-related events in patients with breast, lung, or prostate cancer and metastases to bone: a 15-year study in two large US health systems,” *Support Care Cancer 21*, 2013.
- [18] “Gleason score and isup grade group.” <https://www.pcf.org/about-prostate-cancer/diagnosis-staging-prostate-cancer/gleason-score-isup-grade/>. Accessed: 4 March 2019.
- [19] J. E. McNeal, E. A. Redwine, F. S. Freiha, and T. A. Stamey, “Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread,” *Am.J.Surg.Pathol.12*, 1988.
- [20] R. J. Cohen, B. A. Shannon, M. Phillips, R. E. Moorin, T. M. Wheeler, and K. L. Garrett, “Central Zone Carcinoma of the Prostate Gland: A Distinct Tumor Type With Poor Prognostic Features,” *J.Urol.179(5)*, 2008.
- [21] T. A. Stamey, A. N. Donaldson, C. E. Yemoto, J. E. McNeal, S. Sozen, and H. Gill, “Histological and clinical findings in 896 consecutive prostates treated only with radical retropubic prostatectomy: epidemiologic significance of annual changes,” *J.Urol.160(6Pt2)*, 1998.
- [22] R. Fusco, M. Sansone, V. Granata, S. V. Setola, and A. Petrillo, “A systematic review on multiparametric MR imaging in prostate cancer detection,” *Infectious Agents and Cancer*, 2017.
- [23] P. J. L. D. Visschere, A. Briganti, J. J. Fütterer, P. Ghadjar, H. Isbarn, C. Massard, P. Ost, P. Sooriakumaran, C. I. Surcel, M. Valerio, R. C. N. van den Bergh, G. Plousard, G. Giannarini, and G. M. Villeirs, “Role of multiparametric magnetic resonance imaging in early detection of prostate cancer,” *Insights into Imaging*, 2016.

- [24] M. R. Engelbrecht, G. J. Jager, R. Laheij, A. L. Verbeek, H. J. van Lier, and J. O. Barentsz, "Local staging of prostate cancer using magnetic resonance imaging: a meta-analysis," *European Radiology*, 2002.
- [25] M. R. Engelbrecht, H. J. Huisman, R. J. Laheij, G. J. Jager, G. J. van Leenders, C. A. H.-V. D. Kaa, J. J. de la Rosette, J. G. Blickman, and J. O. Barentsz, "Discrimination of prostate cancer from normal peripheral zone and central gland tissue by using dynamic contrast-enhanced MR imaging," *Radiology*, 2003.
- [26] L. W. Turnbull, D. L. Buckley, L. S. Turnbull, G. P. Liney, and A. J. Knowles, "Differentiation of prostatic carcinoma and benign prostatic hyperplasia: correlation between dynamic Gd-DTPA-enhanced MRI and histopathology," *J Magn Reson Imaging*, 1999.
- [27] T. Sugahara, Y. Korogi, M. Kochi, I. Ikushima, Y. Shigematu, T. Hirai, Okuda, L. Liang, Y. Ge, Y. Komohara, Y. Ushio, and M. Takahashi, "Usefulness of diffusion-weighted MRI with echo-planar technique in the evaluation of cellularity in gliomas," *J Magn Reson Imaging*, 1999.
- [28] M. Castillo, J. K. Smith, L. Kwock, and K. Wilber, "Apparent diffusion coefficients in the evaluation of highgrade cerebral gliomas," *AJNR Am J Neuroradiol*, 2001.
- [29] B. N. Bloch, E. M. Genega, D. N. Costa, I. Pedrosa, M. P. Smith, H. Y. Kressel, L. Ngo, M. G. Sanda, W. C. DeWolf, and N. M. Rofsky, "Prediction of prostate cancer extracapsular extension with high spatial resolution dynamic contrast-enhanced 3-T MRI," *European Radiology*, 2012.
- [30] L. Wang, J. Zhang, L. H. Schwartz, H. Eisenberg, N. M. Ishill, C. S. Moskowitz, P. Scardino, and H. Hricak, "Incremental value of multiplanar cross-referencing for prostate cancer staging with endorectal MRI," *American Journal of Roentgenology*, 2007.
- [31] F. V. Coakley, J. Kurhanewicz, Y. Lu, K. D. Jones, M. G. Swanson, S. D. Chang, P. R. Carroll, and H. Hricak, "Prostate cancer tumor volume: measurement with endorectal MR and MR spectroscopic imaging," *radiology*, 2002.
- [32] R. Fusco, M. Sansone, M. Petrillo, S. V. Setola, V. Granata, G. Botti, S. Perdonà, V. Borzillo, P. Muto, and A. Petrillo, "Multiparametric MRI for prostate cancer detection: Preliminary results on quantitative analysis of dynamic contrast enhanced imaging, diffusion-weighted imaging and spectroscopy imaging," *Magnetic resonance Imaging*, 2016.
- [33] T. W. J. Scheenen, A. B. Rosenkrantz, M. A. Haider, and J. J. Fütterer, "Multiparametric Magnetic Resonance Imaging in Prostate Cancer Management Current Status and Future Perspectives," *Investigative Radiology*, 2015.
- [34] E. K. Vos, T. Kobus, G. J. Litjens, T. Hambroek, C. A. H. van de Kaa, J. O. Barentsz, M. C. Maas, and T. W. Scheenen, "Multiparametric Magnetic Resonance Imaging for Discriminating Low-Grade From High-Grade Prostate Cancer," *Investigative Radiology*, 2015.

- [35] C. V. Dinh, P. Steenbergen, G. Ghobadi, S. W. T. J. P. Heijmink, F. J. Pos, K. Haustermans, and U. A. van der Heide, "Magnetic resonance imaging for prostate cancer radiotherapy," *Physica Medica*, 2016.
- [36] E. H. Hamoen, M. de Rooij, J. A. Witjes, J. O. Barentsz, and M. M. Rovers, "Use of the Prostate Imaging Reporting and Data System (PI-RADS) for Prostate Cancer Detection with Multiparametric Magnetic Resonance Imaging: A Diagnostic Meta-analysis," *European Urology*, 2015.
- [37] A. B. Rosenkratz and S. S. Taneja, "Radiologist, Be Aware: Ten Pitfalls That Confound the Interpretation of Multiparametric Prostate MRI," *AJR*, 2014.
- [38] G. Lemaître, R. Martí, J. Freixenet, J. C. Vilanova, P. M. Walker, and F. Meriaudeau, "Computer-Aided Detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: A review ," *Computers in Biology and Medicine*, 2015.
- [39] M. Fascelli, S. Rais-Bahrami, S. Sankineni, A. M. Brown, A. K. George, R. Ho, T. Frye, A. Kilchevsky, R. Chelluri, S. Abboud, M. M. Siddiqui, M. J. Merino, B. J. Wood, P. L. Choyke, P. A. Pinto, and B. Turkbey, "Combined Biparametric Prostate Magnetic Resonance Imaging and Prostate-specific Antigen in the Detection of Prostate Cancer: A Validation Study in a Biopsy-naïve Patient Population," *Urology* 88:125-34, 2015.
- [40] S. Rais-Bahrami, M. M. Siddiqui, S. Vourganti, B. Turkbey, A. R. Rastinehad, L. Stamatakis, H. Truong, A. Walton-Diaz, A. N. Hoang, J. W. Nix, M. J. Merino, B. J. Wood, R. M. Simon, P. L. Choyke, and P. A. Pinto, "Diagnostic value of biparametric magnetic resonance imaging (MRI) as an adjunct to prostate-specific antigen (PSA)-based detection of prostate cancer in men without prior biopsies," *BJUI* 115(3):381-8, 2015.
- [41] K. C. D. Thestrup, V. Logager, I. Baslev, J. M. Møller, R. H. Hansen, and H. S. Thomsen, "Biparametric versus multiparametric MRI in the diagnosis of prostate cancer," *Acta Radiologica* 2016 Aug; 5(8): 2058460116663046, 2016.
- [42] E. D. Campli, A. D. Pizzi, B. Seccia, R. Cianci, M. d'Annibale, A. Colasante, S. Cinalli, P. Castellani, R. Navarra, R. Iantorno, D. Gabrielli, A. Buffone, M. Caulo, and R. Basilico, "Diagnostic accuracy of biparametric vs multiparametric MRI in clinically significant prostate cancer: Comparison between readers with different experience," *European Journal of Radiology* 101, pp. 17-23, 2018.
- [43] R. Cuocolo, A. Stanzione, G. Rusconi, M. Petretta, A. Ponsiglione, F. Fusco, N. Longo, F. Persico, S. Coccozza, A. Brunetti, and M. Imbriaco, "PSA-density does not improve bi-parametric prostate MR detection of prostate cancer in a biopsy naïve patient population," *European Journal of Radiology* 104, pp. 64-70, 2018.
- [44] X. Niu, Z. Chen, L. Chen, J. Li, T. Peng, and X. Li, "Clinical Application of Biparametric MRI Texture Analysis for Detection and Evaluation of High-Grade Prostate Cancer in Zone-Specific Regions," *AJR*, 2018.

- [45] J. O. Barentsz, J. C. Weinreb, S. Verma, H. C. Thoeny, C. M. Tempany, F. Shtern, A. R. Padhani, D. Margolis, K. J. Macura, M. A. Haider, F. Cornud, and P. L. Choyke, "Synopsis of the PIRADS v2 guidelines for multiparametric prostate magnetic resonance imaging and recommendations for use," *European Urology*, Volume 69, Issue 1, Pages 41-49, 2016.
- [46] T. L. Chenevert, C. J. Galbán, M. K. Ivancevic, S. E. Rohrer, F. J. Londy, T. C. Kwee, C. R. Meyer, T. D. Johnson, A. Rehemtulla, and B. D. Ross, "Diffusion Coefficient Measurement Using a Temperature-Controlled Fluid for Quality Control in Multicenter Studies," *J Magn Reson Imaging* 34(4):983-987, 2011.
- [47] Y. Peng, T. Antic, M. L. Giger, S. E. Eggener, and A. Oto, "Validation of quantitative analysis of multiparametric prostate MR images for prostate cancer detection and aggressiveness assessment: a cross-imager study," *Radiology*, 2014.
- [48] H. J. Johnson and M. M. McCormick and L. Ibáñez and the Insight Software Consortium, *The ITK Software Guide Book 1: Introduction and Development Guidelines Fourth Edition Updated for ITK version 4.13.0*. 2017.
- [49] M. Kirschner, F. Jung, and S. Wesarg, "Automatic Prostate Segmentation in MR Images with a Probabilistic Active Shape Model," *Semantic Scholar*, 2012.
- [50] A. Fenster and B. Chiu, "Evaluation of Segmentation algorithms for Medical Imaging," *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005.
- [51] M. Shahedi, D. W. Cool, C. Romagnoli, G. S. Bauman, M. Bastian-Jordan, G. Rodrigues, B. Ahmad, M. Lock, A. Fenster, and A. D. Ward, "Postediting prostate magnetic resonance imaging segmentation consistency and operator time using manual and computer-assisted segmentation: multiobserver study," *J. Med. Imag.* 3(4), 046002, 2016.
- [52] S. Klein, U. A. van der Heide, I. M. Lips, M. van Vulpen, M. Staring, and J. P. Pluim, "Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information," *Medical Physics*, 2008.
- [53] Z. Tian, L. Liu, Z. Zhang, and B. Fei, "PSNet: prostate segmentation on MRI based on a convolutional neural network," *J Med Imaging (Bellingham)*, 2018.
- [54] Ş. Öztürk and B. Akdemir, "Application of Feature Extraction and Classification Methods for Histopathological Image using GLCM, LBP, LBGLCM, GLRLM and SFTA," *Procedia Computer Science*, 2018.
- [55] V. Giannini, A. Vignati, S. Mirasole, S. Mazzetti, F. Russo, M. Stasi, and D. Regge, "MR-T2-weighted signal intensity: a new imaging biomarker of prostate cancer aggressiveness," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 4:3-4, 130-134, 2016.

- [56] P. C. Vos, J. O. Barentsz, N. Karssemeijer, and H. J. Huisman, "Automatic computer-aided detection of prostate cancer based on multiparametric magnetic resonance image analysis," *Phys. Med. Biol.* 57 1527, 2012.
- [57] T. Tamada, T. Sone, Y. Jo, S. Toshimitsu, T. Yamashita, A. Yamamoto, D. Tanimoto, and K. Ito, "Apparent diffusion coefficient values in peripheral and transition zones of the prostate: comparison between normal and malignant prostatic tissues and correlation with histologic grade," *J. Magn. Reson. Imaging* 28 720-6, 2008.
- [58] L. Wang, Y. Mazaheri, J. Zhang, N. M. Ishill, K. Kuroiwa, and H. Hricak, "Assessment of biologic aggressiveness of prostate cancer: correlation of MR signal intensity with gleason grade after radical prostatectomy," *Radiology* 246 168-76, 2008.
- [59] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics ( Volume: SMC-3 , Issue: 6)*, 1973.
- [60] A. Vignati, S. Mazzetti, V. Giannini, F. Russo, E. Bollito, F. Porpiglia, M. Stasi, and D. Regge, "Texture features on T2-weighted magnetic resonance imaging: new potential biomarkers for prostate cancer aggressiveness," *Phys. Med. Biol.* 60 (2015) 2685-2701, 2015.
- [61] W. Chen, M. L. Giger, H. Li, U. Bick, and G. M. Newstead, "Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images," *Magn. Reson. Med.*, 2007.
- [62] A. Wibmer, H. Hricak, T. Gondo, K. Matsumoto, H. Veeraraghavan, D. Fehr, J. Zheng, D. Goldman, C. Moskowitz, S. Fine, V. E. Reuter, J. Eastham, E. Sala, and H. A. Vargas, "Haralick Texture Analysis of prostate MRI: Utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason Scores," *Eur Radiol.* 25(10): 2840-2850, 2015.
- [63] R. M. haralick, "Statistical and Structural Approaches to Texture," *Proceedings of the IEEE*, 67, pp. 786-804, 1979.
- [64] M. M. Galloway, "Texture analysis using gray level run lengths," *Computer Graphics and Image Processing*, 4(2), 172-179, 1975.
- [65] R. Begg and D. T. H. lai and M. Palaniswami, *Computational Intelligence in Biomedical Engineering*. CRC Press taylor & Francis Group, 2008.
- [66] M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker, and G. D. Tourassi, "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," *Neural Networks* 21, 427-436, 2008.
- [67] L. Roach and O. Maimon, *Data mining with decision trees: theory and applications*. World Scientific Pub Co Inc., 2008.
- [68] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution ," *Proc. 20th Int. Conf. on machine learning (ICML-2003)*, 2003.

- 
- [69] P. S. Kumar and D. Lopez, "A review on feature selection methods for high dimensional data," *International Journal of Engineering and Technology (IJET)*, 2016.
- [70] S. Lessmann, R. Stahlbock, and S. F. Crone, "Genetic Algorithms for Support Vector Machine Model Selection," *The 2006 IEEE International Joint Conference on Neural Network Proceedings, 2006*, pp. 3063-3069, 2006.
- [71] V. Giannini, S. Rosati, C. Castagneri, L. Martincich, D. Regge, and G. Balestra, "Radiomics for pretreatment prediction of pathological response to neoadjuvant therapy using magnetic resonance imaging: Influence of feature selection," *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* pp. 285-288, 2018.
- [72] S. Rosati, M. C. Gianfreda, G. Balestra, V. Giannini, S. Mazzetti, and D. Regge, "Radiomics to Predict Response to Neoadjuvant Chemotherapy in Rectal Cancer: Influence of Simultaneous Feature Selection and Classifier Optimization," *2018 IEEE Life Sciences Conference (LSC)* pp. 65-68, 2018.
- [73] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, 20, 273-297, 1995.
- [74] E. Allwein, R. Schapire, and Y. Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *Journal of Machine Learning Research*, Vol. 1, pp. 113-141, 2000.
- [75] D. Coomans and D. L. Massart, "Alternative k-nearest neighbour rules in supervised pattern recognition : Part 1. k-Nearest neighbour classification by using alternative voting rules," *Analytica Chimica Acta Volume 136*, pp. 15-27, 1982.
- [76] V. Giannini, S. Rosati, D. Regge, and G. Balestra, "Specificity improvement of a CAD system for multiparametric MR prostate cancer using texture features and artificial neural networks," *Health Technol. Vol. 7*, pp. 71-80, 2017.
- [77] D. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *SIAM Journal on Applied Mathematics*, Vol. 11, No. 2, pp. 431-441, 1963.
- [78] X. Yang, S. Tridandapani, J. J. Beitler, D. S. Yu, E. J. Yoshida, W. J. Curran, and T. Liu, "Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: An in vivo study of late toxicity," *Medical Physics*, 2012.
- [79] B. V. Dasarathy and E. B. Holder, "Image characterizations based on joint gray level-run length distributions," *Pattern Recognition Letters 12*, pp. 497-502, 1991.