Faculty of
Mathematics
and Informatics

VILNIUS UNIVERSITY
FACULTY OF MATHEMATICS AND INFORMATICS
MODELLING AND DATA ANALYSIS
MASTER'S STUDY PROGRAMME

EVALUATION OF DYNAMIC CONTRAST IN PROSTATE MRI FOR CANCEROUS TISSUE IDENTIFICATION

Master's thesis

Author: Roman Surkant
VU email address: roman.surkant@mif.stud.vu.lt
Supervisor: Sr. Res., Dr. Jolita Bernatavičienė

Vilnius

2022

1 Abstract

Prostate cancer is one of the leading causes of cancer death worldwide and early diagnosis and treatment is critical. Cancer evaluation is done by using different imaging sequences of MRI, each having own acquisition methods and purpose. Dynamic contrast enhancement, one of such imaging types, is useful for detecting tumors due to higher vascular permeability and density. Contrast images indicate which areas of patient's body have higher concentration of contrast agent and how that concentration is changing over time. The pattern of such changes is evaluated by segmenting the prostate using SLIC algorithm applied on a Temporal Variation Matrix (TVM), a novel approach indicating high signal variation zones, and constructing time-signal intensity curves from segmented regions. Such discrete signal curves are then transformed into continuous form by applying B-spline functional data smoothing to create functional curves. Two classification approaches - functional (K-Nearest Neighbors and Nearest Centroid) and machine learning (Support Vector Machine) - are used to model signal curve behavior for malignant tissue identification. K-Nearest Neighbor resulted in highest accuracy of 95.67% and highest recall of 75%, but lowest precision, while one of SVM models achieved precision of 93.75% at the cost of the lowest recall - 9.00%.

Keywords: dynamic contrast, MRI, segmentation, classification, functional data analysis, support vector machine

2 Notation and abbreviations

DCE – dynamic contrast-enhancement

DWI – diffusion-weighted imaging

FDA – functional data analysis

ML – machine learning

MRI – magnetic resonance imaging

PI-RADS – prostate imaging - reporting and data system

PSA – prostate-specific antigen

SVM – support vector machine

T2w – T2 weighted imaging

SLIC – simple linear iterative clustering

AUC – area under curve

Patient $p \in P$, where P - set of all patients

Cycle, or time moment t – a point in time when a DCE image is captured

Slice $s^{p,t}$ – a grayscale cross-sectional image of patient's pelvic region at particular coordinate at cycle t and patient p

Pixel intensity $i^{(x,y)} \in [0, 255]$ – signal value at image coordinates (x,y)

Segment $seg^{s,t}$ - a collection of pixels belonging to a specific area in an image $(x, y) \in seg$

Contents

3 Introduction

Prostate cancer is one of the leading causes of cancer death worldwide. Among males, prostate cancer has second highest incidence rate after lung cancer [3]. Although death rates have been decreasing due to medical progress, it still remains a considerable disease affecting many patients. Due to nature of cancer, early diagnosis and treatment is critical.

Preliminary identification of cancer involves PSA screening measuring concentration of a protein produced by the prostate, and the concentration is elevated in patients with prostate cancer [21]. Due to high level of false-negative and false-positive cases in PSA testing [16] which lead to incorrect biopsies, a less invasive and more reliable procedure is needed.

Typically cancer diagnosis is based on multiparametric magnetic resonance imaging (mpMRI) [8] which incorporates several imaging sequences for a medical professional to evaluate: T2-weighted imaging (T2w), diffusion-weighted imaging (DWI), dynamic contrast-enhanced imaging (DCE), and MR spectroscopic imaging (MRSI). T2w sequence help separating the anatomy of different prostate zones (e.g. peripheral zone, transition zone). Due to higher water content, the peripheral zone has higher signal intensity which helps with detecting tumors which appear as areas of lower intensity. However, such zones also may be benign abnormalities due to external factors such as fibrosis or prostatitis. For the transition zone, T2w sequence provides bigger diagnostic accuracy due to well-know morphological features. Signal decrease is also associated with Gleason score (a grade that represents tumor severity consisting of two scores on a 1-5 scale) of the tumor, helping with evaluating tumor classification. DWI measures Brownian motion of water molecules in the tissue. Net displacement of water molecules is measured through an apparent diffusion coefficient (ADC) that is derived from DWI. ADC is also inversely correlated with the Gleason score, providing useful information for diagnosis. DWI sequence helps with differentiating malignant tumors from benign abnormalities like prostatitis and fibrosis, and is overall considered as the most useful imaging sequence for cancer detection in prostate. DCE sequence requires intravenous injection of a contrast agent to assess tumor angiogenesis. Typically, tumors show early enhancement of the contrast agent and rapid wash-out compared to healthy tissues. DCE is considered as non-dominant sequence because of enhancement pattern overlap between malignant and benign tumors and is used only as an adjunct to T2w and DWI findings. MRSI is a technique that analyzes tissue metabolism, instead of the anatomy for previously mentioned imaging sequences. MRSI is mostly used for evaluating various characteristics of tumors for cancer localization [20].

Due to subjectivity of prostate cancer evaluation using mpMRI, a structured reporting scheme PI-RADS has been developed. With its introduction, the accuracy of cancer identification has considerably improved [1]. This system is based on literature evidence and consensus expert opinion which defines recommendations for global standardization and diminishes variation in acquisition, interpretation, and reporting of mpMRI analysis. The newest, second version of PI-RADS, contains a single protocol for patient assessment recommending to use T2w, DCE, and DWI (excluding MRSI) MRI data, biopsy status, and other historical information (family history, treatment history, etc.) [2].

This research is focused on the application of DCE MRI sequence in isolation from other MRI data. Prostate DCE MRI data is gathered by capturing imaging sequences of the entire region of prostate during an intravenous injection of a contrast agent (typically gadolinium). Over a course of several minutes, a set of cross-sectional images is created at different time moments, usually every few seconds. The role of the contrast agent is to enhance angiogenesis of tumor in DCE imaging. Since blood vessels are essential to cancer growth, tumors typically have higher vascular permeability and density which attracts higher amount of contrast medium. This way DCE MRI can be used to detect, characterize, and monitor tumors.

The aim of this thesis is to develop a dynamic contrast imaging evaluation method for cancer localization. Such model would have potential to be used in medical institutions in a role of technical aid to radiologists for lowering the necessary time to evaluate patient's status and/or improve the accuracy of cancer diagnosis.

The following tasks are defined to achieve that:

1. Conduct literature review to evaluate application options for analyzing DCE MRI-type data both for prostate and other body regions (e.g. brain, breast, lung)
2. Define the scope and pre-process data: initial data is extracted directly from MRI software and requires reformatting/data type restructuring with additional attention to missing or inconsistent information
3. Select methods and approaches most applicable and relevant to this work
4. Identify advantages and disadvantages of selected methods and propose potential improvement actions

Even with introduction PI-RADS, a framework for prostate cancer evaluation and reporting standardization, more advanced methods have the potential to improve the quality of tumor identification even further. Current DCE-MRI research in both prostate and other cancer areas (e.g. breast) is extensive and authors propose a wide variety of approaches to capture the effect of contrast enhancement, either using DCE sequence alone or in conjunction with other MRI sequences, and use that information for both cancer classification and localization.

In general sense, there are three paths for contrast enhancement analysis as per [19]:

- quantitative - a method that is based on the pharmacokinetic features of how tissues react to contrast agent and calculates certain quantitative perfusion parameters
- semiquantitative method focuses on the derivation of parameters that express the form of time-signal curves; this method is also called "curveology"
- qualitative - a visual method of DCE evaluation that assumes a certain trajectory for time-signal curves based on physiological behaviour differences of contrast agent absorption in malignant and benign tissues

A qualitative method of generating automatic wash-in maps in [22] eliminates the need of manual segmentation and approaches the problem in a visual way. Such visual representation may also be used for parameter calculation similar to quantitative or semiquantitative methods. For instance, research following the semiquantitative path (such as [11], [12]) provides good examples of parameters in use. Such parameters are then used for modeling with possible additional feature engineering in between. [14] and [13] explore such possibility for a range of model frameworks (e.g. a Support Vector Machine) where authors construct feature sets that best discriminate malignant and benign tissue regions.

4 Data

Data was acquired in collaboration with National Cancer Institute (NVI). The dataset consists of four data types: DCE MRI images, prostate region masks, cancer region masks, and biopsy results. Examples of each data type are shown in Figure 1 and more detailed characteristics explained below.

DCE-MRI. DCE-MRI data are grayscale images that are captured during an MRI procedure. Such images indicate the distribution of contrast agent in a particular cross-section of patient's pelvic region. The differentiating aspect of dynamic contrast imaging compared to other image sequence types is its bi-dimensional nature. DCE-MRI sequence has a temporal dimension indicating the time moment when the DCE image was recorded ("cycle"), and a spacial dimension linking the DCE image to a location of the cross-section in patient's body ("slice"). This allows to interpret the data in a 3D time series-based manner.

DCE-MRI is the primary target of this research. Current dataset contains 135660 anonymized images of 144 patients. Each patient having on average 41 slices (total range: 25 - 102) and 26 cycles (total range: 5 - 55). Figure 1a shows an example of dynamic contrast image of patient 66, slice 17, cycle 2.

Prostate region masks. Prostate masks are binary image-type data which indicates the region of prostate for each DCE-MRI slice (all cycles of the same slice have the same prostate region indicated by the same prostate mask). Such masks were segmented manually by medical experts at NVI. Since MRI covers areas both above and below the prostate, some prostate masks are blank (e.g. fully dark). An example of prostate mask for slice 17 of patient 66 is shown in Figure 1b.

Cancer region masks. Similarly to prostate region masks, cancer region masks are binary images showing the location of cancer tumour which were also manually segmented by medical experts. Tumors can be of three types: malignant, undetermined, and benign. Cancer masks indicate only the suspected region of cancer according to medical experts implying a possibility of false information (e.g. a missed tumor not shown in a mask); the factual tumor type is diagnosed though a biopsy. An example of cancer mask for slice 17 of patient 66 is shown in Figure 1c.

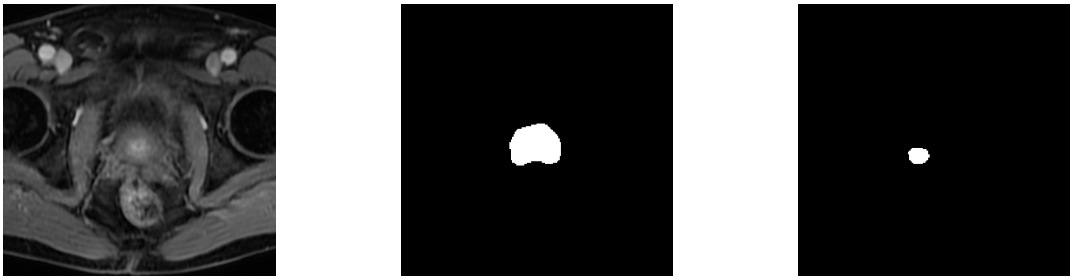
Biopsy results. Each patient underwent on average 15 biopsies (total range: 3 - 25), each biopsy showing tumour type (malignant, undetermined, benign) and tumour severity based on a Gleason score. The biopsy outcome data is split into two datasets of different format:

- Tabular: contains numerical identifier of patient, slice, biopsy, three Gleason scores (first, second, and combined)
- Mask: multi-label mask showing the location of biopsies; each biopsy has a numerical identifier linking to the tabular dataset. Typically a biopsy mask is one or at most several pixels wide and has a circular shape

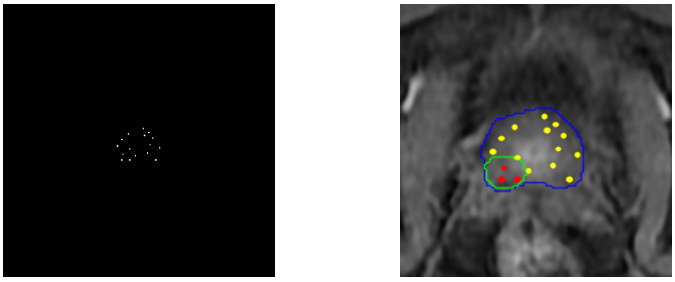
While biopsies provide ground truth label for tumours, there are two flaws:

- Biopsies provide information only in a point-wise manner. The gaps between biopsies are essentially missing data where ground truth information is unavailable.
- To perform a biopsy, a medical expert inserts a needle into prostate and extracts a tissue sample for examination. The needle punctures several slices without knowing which slice exactly was the tissue taken from. Such process introduces a systematic uncertainty, for instance, a slice having no tumors may show positive diagnosis if cancerous tissue was extracted from a different slice above or below

Example of a biopsy mask for slice 17 of patient 66 is shown in Figure 1d, while Figure 1e shows combined view of DCE-MRI image with prostate and cancer region contours and biopsies colored by



(a) DCE-MRI (b) Prostate mask (c) Cancer mask



(d) Biopsy mask (e) Overlay, zoomed

Figure 1: Data type examples. Overlay image combines all previous images: prostate region (blue), cancer region (green), malignant biopsy (red), undetermined or benign biopsy (yellow)

type. For some patients, there is disagreement between biopsy results and cancer masks, i.e. a malignant biopsy located outside of a cancer mask, or vice versa - a cancer mask containing an undetermined or a benign biopsy. This aspect of data inconsistency introduces an additional degree of uncertainty for what should be considered as ground truth.

5 Theoretical part

5.1 Literature review

The authors in [16] explain that tumors have increased vascular permeability due to angiogenesis. This makes DCE-MRI data useful for detection and analysis of tumors. In addition, dynamic contrast imaging might provide information for treatment development.

The prostate produces a prostate-specific antigen (PSA). The levels of PSA are increased in patients with prostate cancer, which may act as indicator for cancer diagnosis. Unfortunately, due to high rates of false-positive and false-negative [16], PSA measurements are inconsistent and other measurement alternatives are necessary. MRI is used in conjunction with PSA for localization and severity classification of tumors.

Currently, the standard practice for MRI evaluation is to to perform multiparametric MRI examination, e.g. to use multiple imaging sequences (T2w, DWI, DCE-MRI, etc.). Each sequence has advantages and disadvantages, so utilizing several sequences in conjunction allows more accurate diagnostics. According to author in [19], DCE-MRI is often used in multiparametric analysis, since tumors in general tend to show earlier and more intense enhancement in comparison to surrounding tissue due to tumor angiogenesis. The main characteristic of angiogenesis is higher number of vessels while newly formed vessels have weaker vessel wall integrity (high permeability).

The author of [19] also defines three methods of analyzing DCE-MRI data: qualitative, semiquantitative, and quantitative. Qualitative analysis is a visual evaluation of DCE imaging based on the assumption of weaker vessel wall integrity which implies faster tumor enhancement with contrast agent due to more rapid blood exchange between cancerous tissues and blood vessels. According to qualitative analysis, malignant tumors show faster enhancement after contrast injection and faster decline in enhancement compared to healthy tissue. However, there is a significant overlap between malignant and benign tissue in different parts of the tumor, which decreases the accuracy of such method.

Semiquantitative method (also called "curveology") is based on the same assumption as above, but instead focuses on constructing curve parameters. Such parameters describe the form and behaviour of time-signal curves, for example:

- Time to peak
- Maximum slope
- Time to contrast uptake
- Peak intensity

Such parameters are then used for tumor identification. The downside of semiquantitative approach is its ability to be generalized. Different MRI machines under varying conditions (temporal resolution, contrast injection rate) produce output that may be inconsistent which makes comparison difficult. In addition, variability between patients also introduces another degree of inconsistency.

Quantitative method is focusing on pharmacokinetic modeling by measuring contrast concentration change. The usual approach is to calculate the rate of contrast exchange between plasma and extracellular space. Such metrics help with discriminating malignant and benign tissues.

There are also disadvantages to DCE-MRI:

- there is misregistration between slices due to possible movement of internal organs or/and the patient, which may distort the results of further analysis
- DCE-MRI is non-specific, i.e. there are other causes that may indicate the same behaviour as a benign or malignant tumor which leads to false diagnosis

The first disadvantage is solved by applying image registration techniques, while the second disadvantage is avoided by using other MRI sequence data for cross-validation [19].

Prostate is not the only area where DCE-MRI data provides critical information in diagnostics. Currently researched dynamic contrast applications are for kidney, heart, breast, and prostate [11]. The author introduces two quantitative approaches for DCE-MRI analysis - non-parametric and parametric. The former approach derives various empirical indices that describes the shape and structure of time-signal curve $S(t)$. The main advantage is the simplicity in calculation of such indices, however some factors - MR parameters, scanner type, varying image acquisition method, and possible correlation with physiological factors (e.g. organ functionality) - make this approach less reliable and generalizable. The parametric approach focuses on the estimation of kinetic parameters by fitting time-signal curves to well-knows pharmacokinetic models. The parameters of such models, such as capillary permeability, are easy to interpret, but highly dependent on the tissue type or tumor characteristics. Such models might be useful only in specific applications.

The authors of [11] mention the following set of perfusion-based parameters of non-parametric approach:

- T_0 - onset time (the time between contrast injection and its appearance in the tissue)
- $S_m = \max_t S(t)$ - maximum signal intensity
- $\Delta S = S_m - S_0$ - peak enhancement (intensity difference between maximum and base enhancement)
- $\frac{\Delta S}{S_0}$ - peak enhancement ratio
- T_p - time to peak
- T_{90} - time to reach 90% of maximum enhancement
- $\frac{\Delta S}{T_p}$ - ratio between peak enhancement and time to peak
- $\frac{\Delta S}{T_p - T_0}$ - wash-in slope
- $\frac{S_{max} - S_{final}}{T_{max} - T_p}$ - wash-out slope
- $\frac{S_E - S_0}{S_L - S_0}$ - early to late enhancement ratio

The author also mentions a set of metrics that are more resistance to signal noise that are based on AUC:

- $(1/T_{max} - T_p) \int_{T_p}^{T_{max}} S(t)dt$ - average plateau (average signal change during wash-out)
- AUC of $S(t)$, or in a specified time interval

Regarding the parametric approach, the choice of model depends on factors such as the type of tissue (prostate, brain, etc.) and MR signal variability conditions. Parametric models are separated into two categories:

- Compartmental models focus on analysis of blood-tissue exchange of contrast with other components ("compartments"). Separation into mode compartments leads to higher model accuracy at the cost of higher model complexity.
- Spatially distributed kinetic models focus on modeling the flow of contrast by separating the tissue into series of infinitesimal compartments that interact only with nearby locations. Such models require higher data quality to maintain model accuracy.

Due to higher complexity and lower degree of explainability of parametric models, this approach is not integrated into this research's workflow.

Currently existing research in prostate identification using dynamic contrast imaging is extensive, covering a wide range of methods and approaches. To expand the overview, the work on other types of cancer (heart, breast, lung, etc.) is also reviewed, since the principles of contrast enhancement for different types of tissue may be similar.

Authors in [12] analyze prostate DCE-MRI by separating transitional and peripheral zone images into distinct datasets. Based on the methodology of PI-RADS sector mapping, tumors were localized, but for each patient only the most prominent tumor was selected for further analysis. Identification of tumor and severity (Gleason score) were confirmed through pathological examination. For each identified lesion, two fixed-radius zones were singled out - one for malignant tissue, and one for benign - using visual inspection of expert uroradiologist. The authors evaluate the characteristics of contrast agent wash-in pattern by calculating two metrics for each voxel: maximum wash-in slope (MWS) and delay phase slope (DPS). MWS is maximum difference between two sequential time points (cycles) while DPS - intensity from terminal point of MWS to the last time moment of image acquisition. From resulting MWS and DPS, the histograms are created and a number of statistical parameters are extracted for each label: mean, standard deviation, coefficient of variation, skewness, curtosis, interquartile range, percentiles (10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}), and modified full width at half-maximum (a measure indicating histogram width - a difference between maximum and minimum value of a sample). The differences of histogram parameters between benign and malignant tissues were assessed by Wilcoxon signed rank test for paired data. For MWS, the model indicates significant discrimination between malignant and benign tissues for coefficient of variation parameter, and 10^{th} percentile for bigger lesions. For DPS, such parameters are mean, 25^{th}, 50^{th}, 75^{th}, 90^{th} percentiles. In addition, coefficient of variation for smaller lesions and skewness, 10^{th} percentile for bigger lesions shows significant difference.

The analysis of breast cancer using two public datasets is explained in [7]. DCE-MRI is manually segment expert radiologist, delineating tumor contours in each imaging slice outlining three-dimensional volume of interest. The authors define 50 textural features for segmented tumors that are divided into two groups: first-order features and second-order features. The former consists of such metrics: mean, median, mode, standard deviation, mean absolute deviation, total range, skewness, kurtosis, interquartile range. Next, time-intensity curves are constructed and subdivided into three parts: first part shows the time for contrast agent to reach the lesion when the signal equals pre-contrast injection level; second part represents enhancement increase due to contrast absorption (wash-in); third part shows contrast agent transfer back into blood (wash-out). Using such curves the following semiquantitative metrics are defined: maximum signal difference (MSD), time to peak between wash-in and wash-out segments, wash-in slope (WIS), wash-out slope (WOS), wash-in intercept, wash-out intercept, wash-in AUC, wash-out AUC, wash-in and wash-out AUC, linear percentage change of MSD and WOS (SIS). The final model contains the following significant textural parameters: entropy, long-run emphasis (LRE), busyness. Significant semiquantitative parameters are: MSD, WOS, WIS, SIS.

In [22], the authors take an approach of generating prostate wash-in maps which reflect the percentage change of signal intensity between the first image and images acquired for one minute after contrast injection. The highest percentage change in this time frame is used for colored wash-in map construction. Two radiologists (one expert, one trainee), independently from each other, manually identified the severity of tumors and rated their diagnosis confidence using a custom scale during two sessions: first session having T2w, DWI, DCE information, and second session having T2w, DWI, wash-in map information. During both session critical clinical information (PSA level, Gleason grade, pathological conclusions, etc.) were also available. For both radiologists, wash-in maps significantly decreased the time to evaluate patients and showed moderate degree of agreement with the Gleason score ground truth for cancer severity.

Machine learning (ML) approaches are a popular choice for DCE-MRI analysis for breast cancer. Authors in [13] introduce a model that separates malignant and benign breast lesions based on SVM

algorithm. For data preprocessing, image registration is done to compensate for patient movement and all region of interests (ROI) are manually delineated. ROI were drawn only around the most enhanced areas to avoid capturing non-lesion and necrotic tissue. For each lesion's ROI, a time series vector is constructed representing signal intensity change over time. Such vectors are then used for feature vector creation for the SVM. The following four feature vectors are generated:

1. relative signal intensity $f_1 = \frac{s(n)}{s(0)}$ - a time series where each element of initial signal vector is divided by signal intensity at first time moment

2. derivative signal intensity $f_2 = \frac{d}{dt} f_1 = \frac{s(n+1)-s(n)}{t_{n+1}-t_n}$ - a derivative vector of the relative signal intensity feature. This feature is based on the assumption that the rate of increase in enhancement reflect the differences in permeability between benign and malignant tumors

3. relative signal intensity vector combined with its derivative $f_3 = \{f_1, f_2\}$. The derivative signal intensity feature ignores useful information that the first feature might provide, so both are merged

4. a vector containing three parameters: maximum signal intensity (%), time of maximum enhancement, maximum wash-out (%)

Application of other ML models is also shown in [14] where the authors investigate cancer aggressiveness modeling for prostate cancer. Tumor location were manually identified by four medical experts: 2 radiologists, a histopathologist, and a clinical urologist. For cancer severity ground truth, Gleason score are confirmed by biopsy. Time-signal curves were constructed which are used for manual segmentation of two types of ROIs: according to the first phase of enhancement, and the strongest enhancement phase. Three datasets are defines: the first for the first enhancement phase ROIs, the second for the strongest enhancement phase ROIs, and the third for both ROI types combined. Four categories of feature are defined for modeling: category 1 features contain first order statistics which quantitatively describe voxel intensity distribution within MRI images; category 2 features describe shape- and size-based characteristics of cancer region; category 3 explain texture-based features; category 4 features contain higher-order statistics that explain texture- and intensity-based characteristics of filter-transformed images. The feature set was reduced to a subset of eigenvalues that have variance greater than the threshold of 0.8. In addition, a chi-square test was used to measure the link to Gleason scores and a LASSO algorithm for regularization. after eigenvalue normalization, five machine learning (ML) algorithms were used: linear kernel SVM, logistic regression (LR), random forest (RF), decision tree (DT), k-nearest neighbor (KNN), having ROC-AUC as performance metric. Different sets of features were defined for each of three datasets. The best models for each of datasets are as follows: for first dataset, all models except for DT performed similarly with AUC $0.83 - 0.88$; similar results for the second dataset with AUC range $0.80 - 0.84$; for the third dataset, LR, KNN, and SVM showed higher AUC with $0.90 - 0.93$.

Based on the reviewed literature and completed research by other authors, two problems are formulated and novel solutions to each are proposed which are explained and implemented in this work:

1. A common task performed in aforementioned papers is manual segmentation/localization of objects of interest. This aspect of evaluation removes the possibility of a fully-automatic cancer identification system. A workaround is proposed in a form of a segmentation algorithm - Simple Linear Iterative Clustering, or SLIC - that utilizes the information of both the proximity of voxels and voxel enhancement levels to divide an image into regions of interest, each of which are then separately evaluated. The input for such algorithm is a novel concept of a Temporal Variation Matrix which represents the variation of voxel intensities over time for a given DCE slice.

2. Time-signal intensity curves have limited temporal resolution due to having a discrete form which potentially lowers the quality of analysis. This issue is solved by applying Functional Data Analy-

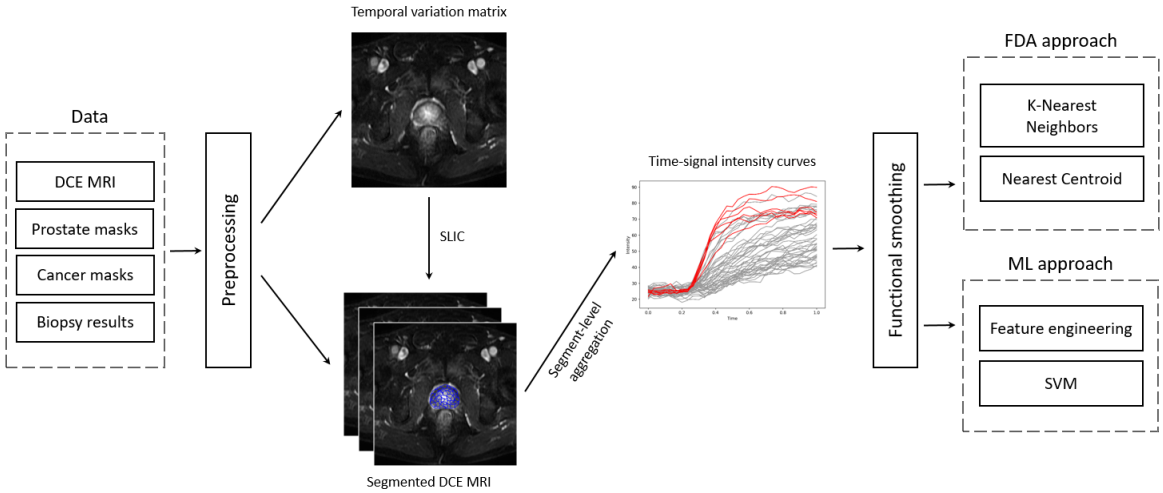


Figure 2: Diagram of the workflow

sis (FDA) techniques for vector smoothing to create a functional (and continuous) representation of time-signal curves.

Further explanation for each of the solutions is provided in the Workflow section.

5.2 Workflow

This research is based on the assumption that contrast agent concentration is a reliable proxy for cancerous tissue detection. The general aim is to construct time-signal intensity curves that represent the concentration of contrast agent in various zones of patient's prostate and to create a classifier which uses curve information for discrimination between healthy and cancerous tissue.

Based on available data and tasks, a workflow is defined which is shown in Figure 2. The workflow consists of four primary blocks - data preparation, data preprocessing, curve construction, and modeling.

The data preparation block is responsible for capturing important metadata from image files and aggregating unstructured data into more convenient form. Data preprocessing block adjusts raw data to a format necessary for the computational methods used later in the workflow. In the curve construction block DCE images are segmented using an automatic segmentation algorithm SLIC, the input for which is a Temporal Variation Matrix, a novel concept of calculating voxel-level signal variance. The voxels inside each segment are aggregated and formed into time-series vectors (curves). The modeling block explains how constructed curves are used from malignant/benign tissue classification.

5.2.1 Data preparation

Data preparation block helps with capturing necessary information of the data (e.g. image resolution and contents) for better understanding of data and quality control, and transformation of data into more meaningful format.

Metadata reader extracts the metadata from each MRI image object. The following metadata values are captured:

- patient ID
- cycle ID
- slice ID
- resolution - height
- resolution - width
- pixel intensity - minimum
- pixel intensity - maximum
- pixel intensity - mean

Patient, cycle, and slice ID are necessary for alignment with other data. Resolution and pixel intensity statistics are used for quality checks: is DCE image's resolution identical to corresponding mask's resolution, and whether DCE image's pixel intensity is completely uniform (blank image).

Biopsy aggregation is collection and combination of biopsy information from multiple CSV-type files (one for each patient) into a single source for more convenient processing and analysis. The resulting dataset contains the following features:

- patient ID
- tumor ID
- tumor type
- Gleason score - first
- Gleason score - second
- Gleason score - combined

Tumor ID links to biopsy masks, while tumor type indicates tumor category - benign, undetermined, or malignant. Tumor severity is quantified by using the Gleason score, a single case containing 3 values - first and seconds Gleason scores on a scale of 1 to 5, and a combined Gleason score equalling the sum of both.

5.2.2 Data preprocessing

Data preprocessing block is responsible for data transformation/reformatting to fit the requirements of used algorithms/code modules, for example, a certain bit depth (i.e. color granularity) for image input.

Mask rescaling - transformation of prostate and cancer mask files into binary format for compatibility with algorithms used. Raw mask files have inconsistent pixel depth, so the following pixel-level remapping done: foreground - from [1,65535] to 1, background remaining the same (zero intensity).

Image rescaling - transformation of MRI images into appropriate format with 8-bit pixel precision (256 color values).

5.2.3 Curve construction

Curve construction block introduces a novel approach - a Temporal Variation Matrix (TVM) - which transforms time series set of images $\{s_k^1, s_k^2, ..., s_k^3\}$ into a single matrix representing slice's pixel-wise intensity variation across all cycles. TVM is then used for prostate segmentation, which is used for creating time-signal intensity curves.

Temporal variation matrix - construction of a matrix with identical resolution as source MRI image which represents statistical variance of signal value of each pixel between all cycles of a selected slice:

$$\text{for patient } p \in P: \qquad i_{TVM}^{(x,y),s_k} = \sigma^2(i^{(x,y),s_k,t_1}, i^{(x,y),s_k,t_2}, ..., i^{(x,y),s_k,t_T}) \tag{1}$$

TVM shows regions with high (bright) and low (dark) change in signal intensity over time. Example of a TVM for patient 66 is shown in Figure 4d together with three DCE-MRI images at different time moments of the same slice (Figures 4a, 4b, 4c) for reference. TVM highlights three objects: two arteries in upper left and upper right corners, and the prostate region (including tumors). All three objects circulate blood that contains the contrast agent, which is the cause for having higher signal intensity variation in DCE-MRI images.

Image segmentation - application of Simple Linear Iterative Clustering (SLIC) algorithm to separate the prostate region into segments. It is an iterative segmentation algorithm based on k-means clustering which groups pixels according to two criteria: color similarity and pixel proximity.

As per [10], the algorithm for segmenting a specific region (mask) of the image follows three steps:

Step 1) Having N segments, a Euclidean distance transform $D(x)$ is used to iteratively place cluster centers spaced at maximum distance from the boundary and other cluster centers. B is

a set of background (non-mask) labels, L is a set containing background B and labelled points P ($L = B \cup P$). During initialization $P = \{\}$.

$$D(x) = \min_{y \in L}(\sum_{i}^{n}(x_i - y_i)^2)^{\frac{1}{2}} \tag{2}$$

In Equation 2, $D(x)$ - distance transform, x - location, n - number of spacial dimensions. The furthest distance p^* is calculated:

$$p^* = \underset{x}{\text{argmax}} D(x) \tag{3}$$

After this, first iteration is completed, P is updated $P := \{p^*\}$, and the next iteration starts with a total of N iterations thus creating N cluster centers.

Step 2) K-means clustering iteratively redistributes placed cluster centers in the masked area so that they become equidistant based on spacial distance

Step 3) K-means clustering is applied to voxels inside the mask based on distance in Equation 4, where r - weight between feature distance d_f and spacial distance d_s.

$$d = \sqrt{(d_f)^2 + (d_s/r)^2} \tag{4}$$

This way the prostate region of an image is separated into segments which are then projected back into MRI images that were used for TVM construction (Figure 4e).

The implementation of SLIC algorithm was done using scikit-image[1] package.

Segment aggregation. Signal intensities inside each segment of MRI images are transformed into a single representative pixel intensity value through an aggregation function.

$$\text{for } p \in P, s \in S: \qquad i^{seg,s} = agg(i^{(x,y),s}) \tag{5}$$

The choice of the aggregation function agg is determined experimentally from a pool of common descriptive statistics: mean, maximum, minimum, 10^{th} percentile, 90^{th} percentile.

Curve construction. A time series curve for each segment is created, having cycle ID and aggregated pixel intensity as dimensions. Each curve labeled into binary classes: 1 - segment has a significant ($\geq 50\%$) overlap with cancer mask and non-zero overlap with a malignant biopsy mask, 0 - otherwise.

An example of such curve is shown in Figure 7. Due to the fact that the initial data was acquired under varying conditions (multiple medical institutions with different MRI scanners, likely using different hardware and configuration), DCE images between different patients have inconsistent characteristics:

- temporal resolution - varying number of cycles; e.g. low temporal resolution may "cut off" important information on wash-out
- spacial resolution - varying pixel resolution of the image; low quality images decreases the accuracy of analysis
- contrast agent injection time moment; early injection "pushes" the curve to the left and hides curve pre-enhancement information, while late injection "pushes" the curve to the right hiding curve's wash-out trajectory

[1] https://scikit-image.org/docs/stable/api/skimage.segmentation.html

Curve smoothing. Generated time series that the curves are based on represent the signal intensity progression in different prostate zones over time. Such data is discrete, having equal data points as the number of cycles for a selected patient. To obtain a more natural, smooth representation of such time series, functional data analysis (FDA) methods are applied.

FDA contains various smoothing techniques that allow the transformation of discrete data into a functional (continuous) form. According to [4], functions are built following two steps:

- defining a set of basis functions ϕ_k that act as functional building blocks
- setting up a collection of coefficients which define a function as a linear combination of such basis functions

A function should be complex enough to properly fit the data, but also general enough to conserve computational power and not use too many parameters. Mathematical notation of a function $x(t)$ is shown in Equation 6 and is called functional basis expansion.

$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t) \tag{6}$$

In Equation 6, $\phi_k, k = 1, ..., K$ - basis functions, $c_1, ..., c_K$ - coefficients of expansion.

Since DCE signal is non-periodic, a spline-based (or piecewise polynomial) smoothing, specifically B-spline, is selected [4]. When and interval of observation is divided into subintervals, a polynomial is constructed over every subinterval. Splines (or piecewise polynomials) are defined through three variables: range of validity, knots, and order. The order indicates the degree (highest power) of the polynomial plus one. The knots are related to breakpoints according to which the subinterval splitting is done.

5.2.4 Modeling

Modeling block uses the constructed time-signal intensity curves to build a set of classification models. As per Figure 3, two approaches are explored: a functional data (FDA) analysis approach and a machine learning (ML) approach.

The FDA approach uses functional data as input. In this work, such data is the first-order derivative of functional time-signal curves. For classification, two models are used: K-Nearest Neighbors and Nearest Centroid.

The ML approach contains an intermediate step of feature engineering where different feature vectors are constructed based on literature review. Such feature vectors are then used as input for a Support Vector Machine model.

Both approaches are classification modeling tasks where the output label y is defined as the following:

$$y_i = \begin{cases} 1 & \text{if } \geq 50\% \text{ segment overlap with a cancer mask \& non-zero overlap with a malignant biopsy} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Functional data analysis approach uses functional data as input and a single-value vector y for label. In this work, functional data are the smoothed functional curves obtained in the previous chapter and the label vector contains binary classes as per Equation 7.

K-Nearest Neighbors[2] - the first of the two FDA models - is based on a general statistical (non-functional) algorithm of K-Nearest Neighbors (KNN). This model is non-parametric and can be used for both classification and regression tasks. KNN calculates Euclidean distances between each of

[2]https://fda.readthedocs.io/en/latest/modules/ml/autosummary/skfda.ml.classification.KNeighborsClassifier.html

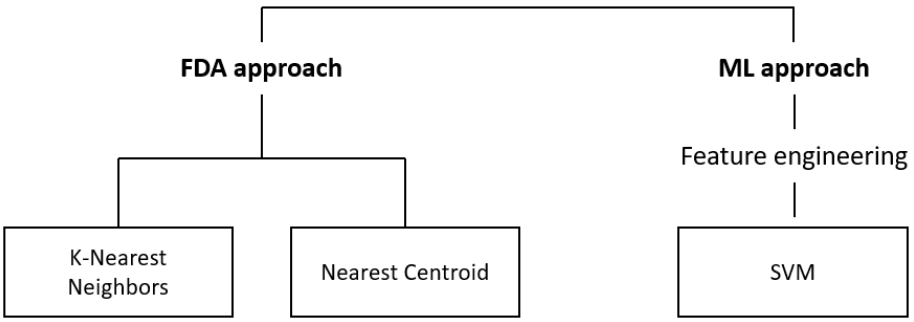


Figure 3: Diagram of the workflow

data points, and the data point is then assigned to a specific. As per [18], having a feature vector $[x_1^i, x_2^i, ..., x_N^i]$ where x_k^i - the values of data point x_i's k-th attribute, the distance between x_i and x_j is $d(x_i, x_j) = \sqrt{\sum_{k=1}^{N}(x_k^i - x_k^j)^2}$. The class label assigned to the selected data point equals to the majority label of k points closest according the the distance function.

K is the primary parameter for KNN and it is highly data-dependent: larger K values diminish the effect of noise in data and makes classification boundary less specific. A certain grid search-based algorithm is used for finding the optimal K for current data. This is explained in greater detail in Experimental results section.

The second functional classifier is Nearest Centroid[3]. This algorithm is based on centroids that are computed according to the mean of observations, each class label having own centroid. Observation prediction takes into account the closest centroid and assumes its label. As per [17], given observations $\{\overrightarrow{x_1}, ..., \overrightarrow{x_n}\}$ and labels $\{y_1, ..., y_n\}$, centroids are calculated $\overrightarrow{\mu_l} = \frac{1}{|C_L|}\sum_{i \in C_l} \overrightarrow{x_i}$, C_l - set of indices of class l. Predicted observations are then assigned to a class as such: $\hat{y} = \underset{l}{\operatorname{argmin}} ||\overrightarrow{\mu_l} - \overrightarrow{x}||$

Machine learning approach uses a feature set for input. A single feature set may contain one or more numerical or categorical vectors representing a particular behavior of a signal curve. As mentioned above in previous chapters, various publications describe a range of different metrics. The choice for which ones should be grouped for classification is done experimentally. In the Feature engineering & classification chapter, several sets of features are defined and an SVM-based model is created for each of the feature sets.

The SVM itself is a supervised machine learning technique which searches for a decision boundary (or hyperplane) that separates one class from another while having some distance margin around a decision boundary. Given training vectors $x_i \in R^n, i = 1, ..., l$, and an indicator vector $y \in R^l, y \in \{1, -1\}$, a C-SVM (C-Support Vector Classification) [5] solves a primal optimization problem:

$$\min_{w,b,\xi} \quad \frac{1}{2}w^T w + C\sum_{i=1}^{l}\xi_i \tag{8}$$

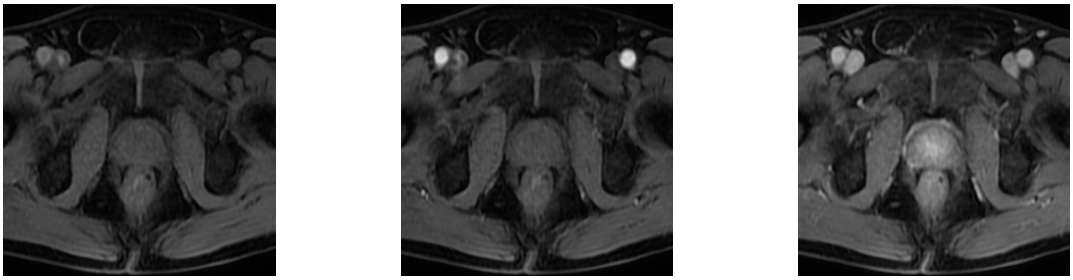
In Equation 8, C - regularization parameter, and $\phi(x_i)$ maps x_i into a higher-dimensional space. This optimization problem also assumes conditions $y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i$ and $\xi \geq 0, i = 1, ..., l$.

Due to possible high dimensionality of vector variable w, in practice an alternative (non-primal) optimization problem is solved:

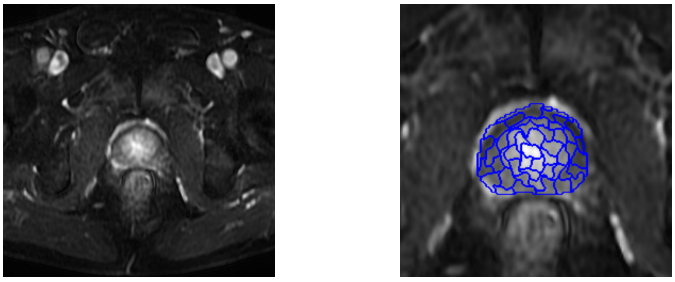
$$\min_{\alpha} \quad \frac{1}{2}\alpha^T Q\alpha - e^T\alpha \tag{9}$$

where e is unit vector $e = [1, ..., 1]^T$, Q - l by l matrix, $Q_{ij} \equiv y_i y_j K(x_i, x_j)$, K - kernel function $K(x_i, x_j) \equiv \phi(x_i)^T\phi(x_j)$. Equation 9 also assumes $y^T\alpha = 0$ and $0 \leq \alpha_i \leq C, i = 1, ..., l$.

[3]https://fda.readthedocs.io/en/latest/modules/ml/autosummary/skfda.ml.classification.NearestCentroid.html



(a) DCE-MRI, t=0 (b) DCE-MRI, t=5 (c) DCE-MRI, t=15



(d) TMV (e) TVM SLIC, zoomed

Figure 4: Temporal variation matrix

After finding the solution to optimization problem 9, the optimal vector variable w and the decision function are as follows:

$$w = \sum_{i=1}^{l} y_i \alpha_i \phi(x_i) \tag{10}$$

$$sgn(w^T \phi(x) + b) = sgn\left(\sum_{i=1}^{l} y_i \alpha_i K(x_i, x) +\right) \tag{11}$$

The main parameter of SVM - regularization C - controls the tolerance for misclassification. Smaller C values (close to zero) expands the margin around the decision boundary accepting a higher rate of misclassification, while higher C values the boundary is narrower.

The implementation of SVM was done using scikit-learn[4] package which is based on LIBSVM software.

[4]https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

6 Experimental results

In the aforementioned workflow, a number of optimization objectives were formulated:

1. image segmentation: parameter selection for SLIC algorithm
2. segment aggregation: selection of segment aggregation method
3. curve construction: parameter selection for curve smoothing

6.1 Image segmentation

The SLIC algorithm has two main parameters:

- Number of segments - the (approximate) number of desired segments that the image or region is separated into. In some circumstances, e.g. in combination with very low compactness, the resulting number of segments is lower than the parameter value.
- Compactness - a value which balances color proximity and space proximity. Higher values give more weight to space proximity, making segment's shape more square/cubic, while lower values prioritize color proximity resulting in more homogeneous color grouping, but distorting segment's form

The algorithm is applied exclusively for the prostate region of each TVM. The aim is to separate the prostate into a number of segments with each segment having similar pixel intensities. The behaviour of SLIC segments - the form and the granularity - for slice 17 of patient 66 can be seen in Figure 5. Left-most images, having low compactness, show more irregular segment shapes compared to high compactness segmentation on the right. High granularity segmentation divides the prostate region into more parts (lower half of example images), but at a higher computational cost and marginal additional useful information for analysis.

SLIC parameters were selected using expert judgement and basing the decision on the following criteria:

- number of segments parameter should be high enough to have separation between cancerous and healthy tissue, but low enough to remain computationally relevant and keep segments visually distinguishable
- compactness parameter should capture similar intensities, but still prioritize color proximity in favor of maintaining circular shape due to the fact that cancerous growth typically does not have too irregular formation

Besides the visual evaluation of SLIC with different parameters in Figure 5, a quantitative evaluation is also done based on the mean and standard deviation of voxel values in each of the segments. Figure 6 shows a scatterplot between signal mean and signal standard deviation for 9 SLIC parameter combinations from Figure 5. The aim is to select such parameter combination with the lowest deviation, i.e. intra-segment intensity is maximally uniform. Segmentation into 10 segments leads to smaller number of observations with SD range of approx. $\sigma \in [10, 30]$. With segmentation into 50 and 100 segments, observations are heavily concentrated in the rage $\sigma \in [0, 20]$ with relatively small number of outliers with $\sigma > 20$. The change in compactness for these observations doesn't significantly alter the scatterplot distribution.

Based on insights above, the concluding parameters for SLIC were selected as 50 and 7 for segment number and compactness respectively.

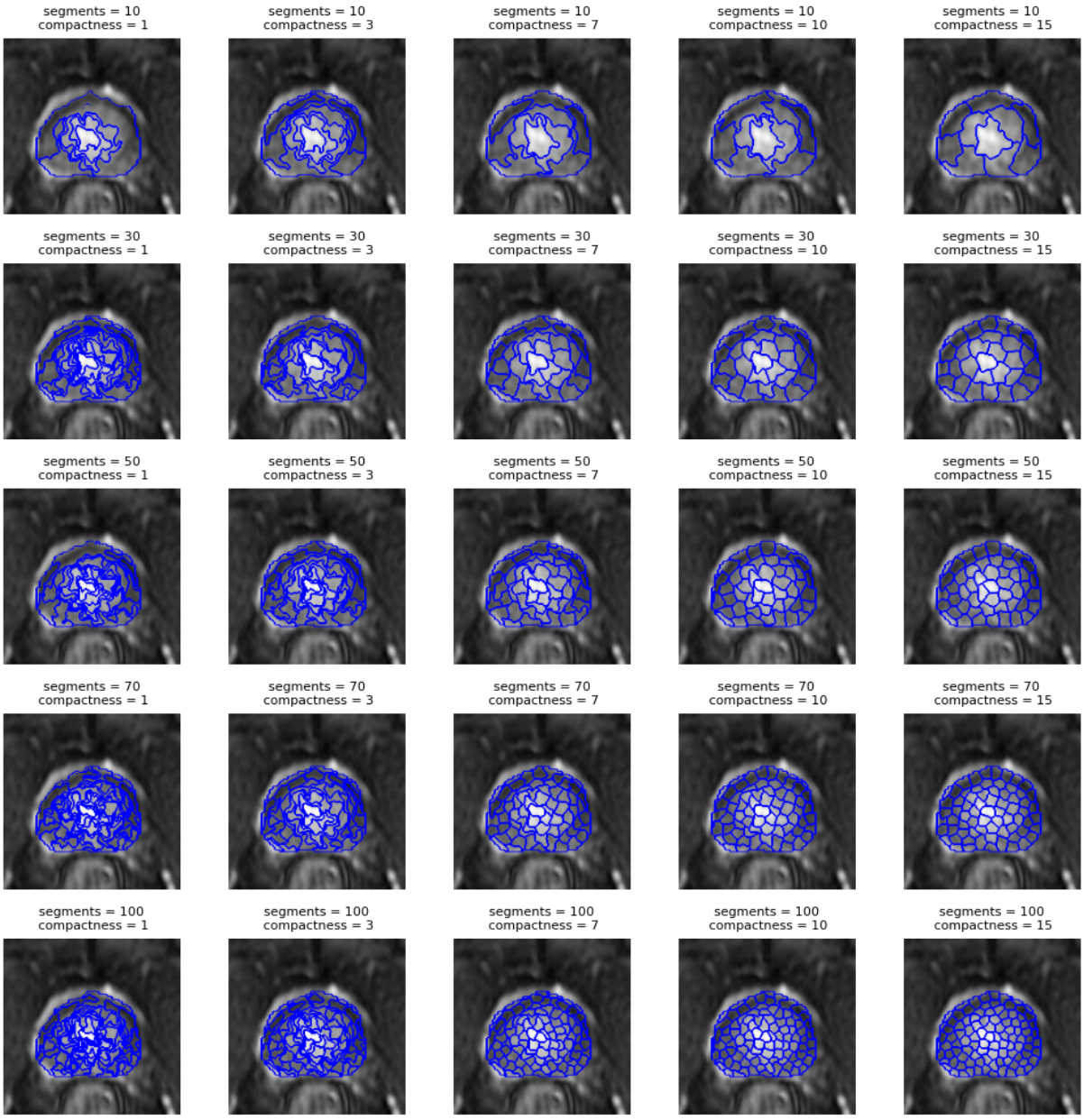


Figure 5: Comparison of different SLIC parameter combinations

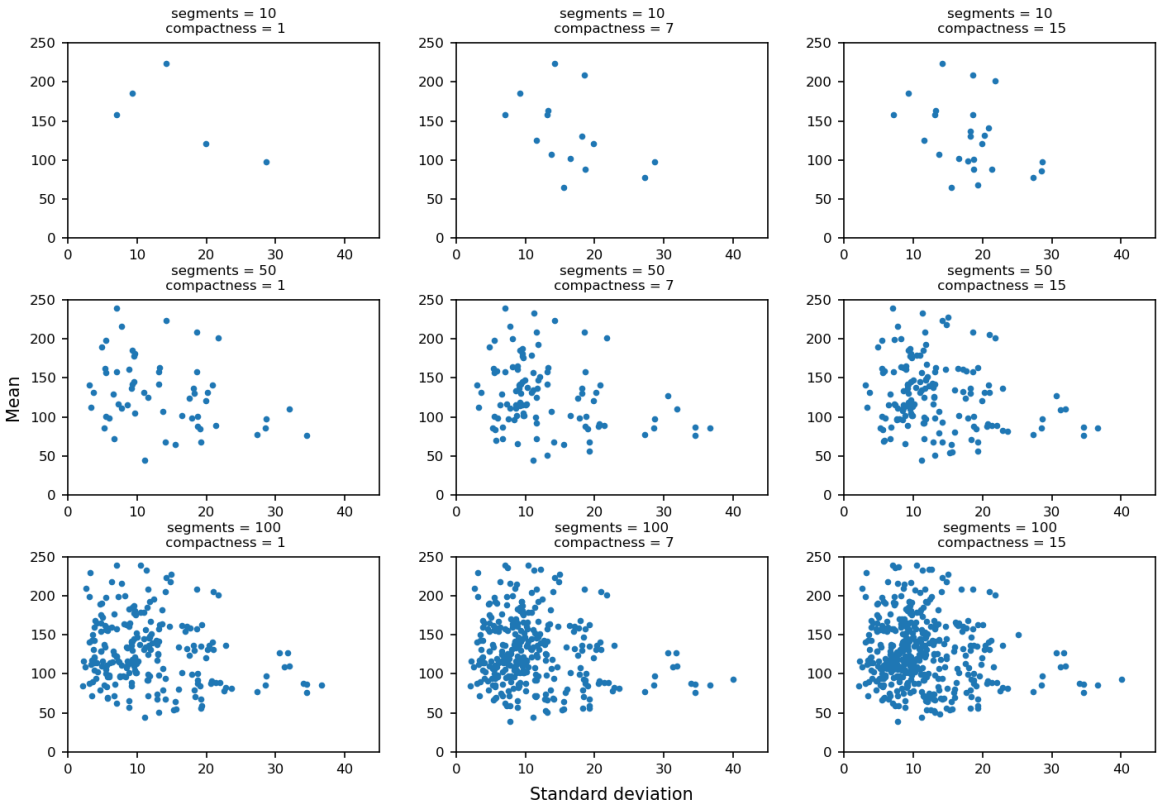


Figure 6: Segment signal mean-standard deviation relationship from different SLIC parameter combinations

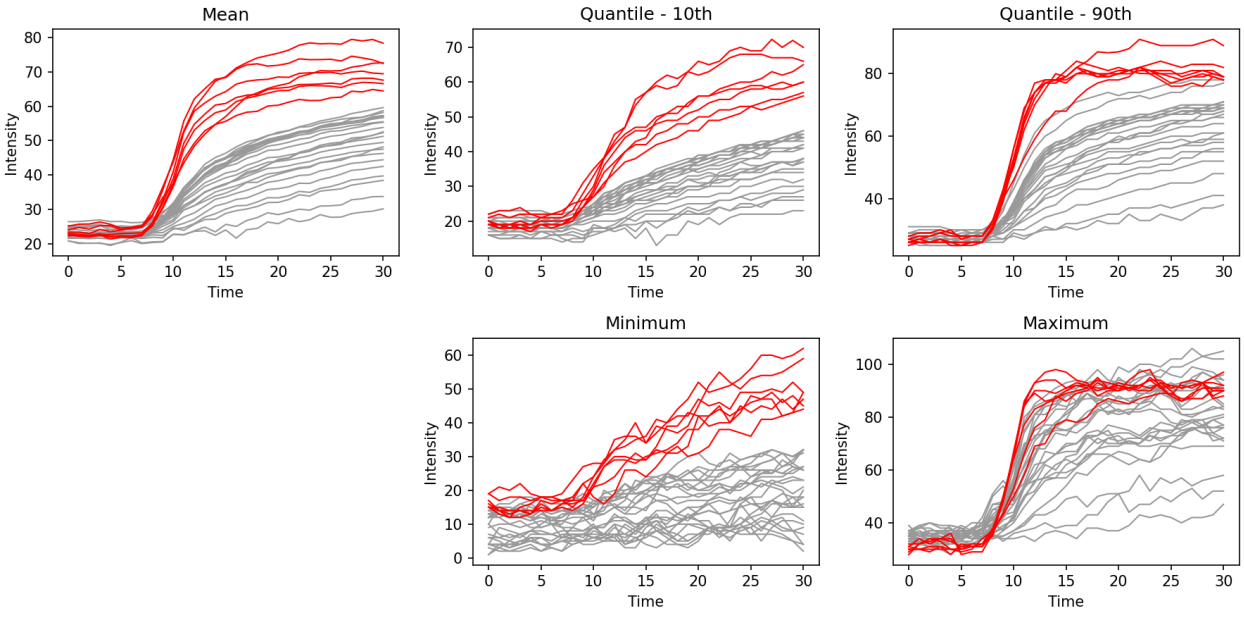
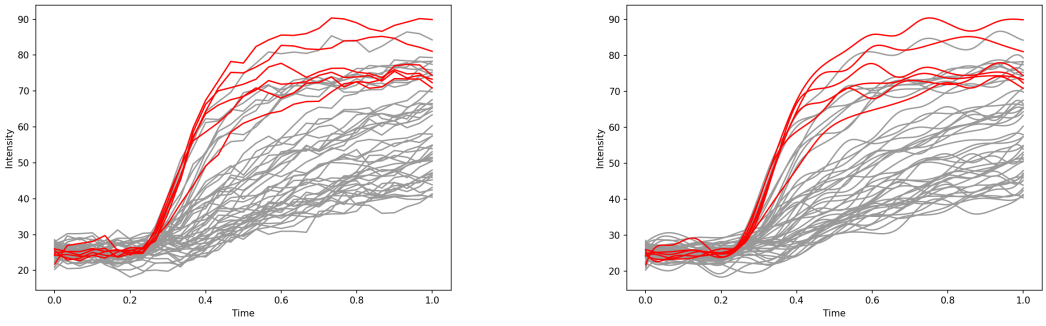


Figure 7: Aggregation function comparison. Cancer region inside prostate aggregated signal (red) and non-cancer region (gray)



(a) Original (b) Smoothed

Figure 8: B-spline smoothing

6.2 Segment aggregation

Optimal aggregation function is determined experimentally from a pool of several common descriptive statistics: mean, minimum, maximum, 10^{th} percentile, 90^{th} percentile.

The selection of segment aggregation method was selected experimentally using expert judgement based on visual evaluation of each approach. An example of aforementioned aggregation methods is shown in Figure 7. The charts combine all slices of patient 66, each curve representing aggregated pixel intensities of cancer (red) and non-cancer (gray) regions of prostate for each slice. The aim is to select a pixel aggregation method so that curves of different class are maximally separated, i.e. there is the clearest distinction between cancerous and healthy tissues.

In Figure 7, maximum aggregation indicates the worst capability of separating the curves, as there is a clear overlap of red and gray curves. Minimum aggregation shows separation that "builds up" over time with clear distinction between red and gray curves only at $t \in [25; 30]$. 90^{th} quantile aggregation shows good curve separation, but the distance between the two classes is smaller compared to mean and 10^{th} quantile methods. The selected method is mean, as the trajectory of the cancerous regions for this method is more favourable in terms of typical histological behaviour.

6.3 Curve construction

Functional data analysis for curve smoothing is applied. Since curves are not periodic, the selected basis is B-spline - a spline (or piecewise polynomial) function that is widely used for data smoothing due to accurate representation and computational efficiency. B-spline has two parameters:

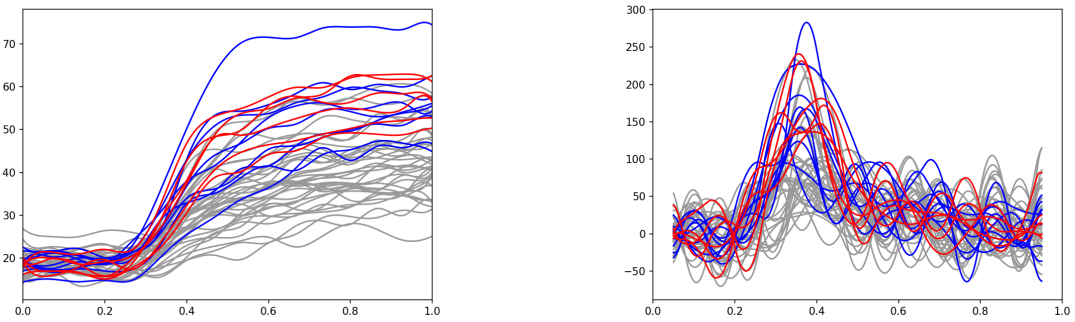
- Number of basis functions
- Order of the splines, i.e. the order of the polynomial

Smoothed curve obtains a non-discrete form which may be used for further transformation, (e.g. curve derivative) and more natural intensity change trajectory.

B-spline parameters were selected based on Generalized Cross Validation score. Using a grid search-type algorithm the following combination of parameter values were evaluated: $\{2, 3, 4, 5\}$ for order and $\{5, 6, ..., 29, 30, 35, 40\}$ for basis function number. Based on GCV score, the optimal parameter combination was determined to be 4 and 18 respectively. Example of smoothing in contrast of original curves are shown in Figure 8, with red and gray curves representing labels 1 and 0 respectively.

6.4 Feature engineering & classification

Due to high degree of variation in the form of curves between different patients, a single patient - patient number 50 - is chosen for a case study. The first criteria for selection was for curves to follow typical



(a) Functional curve

(b) First order derivative curve

Figure 9: Original functional curve and its derivative curve. Segments with confirmed malignant biopsy and $> 50\%$ overlap with cancer mask (red), segments with confirmed malignant biopsy and $\leq 50\%$ overlap with cancer mask, other segments (gray)

histological trajectory for malignant (rapid contrast wash-in and gradual wash-out) and benign (less sudden wash-in and either gradual or no wash-out) regions. The second criteria was for the patient to have no segments belonging non-cancer region that contain a malignant or undetermined biopsy (i.e. label inconsistency due to a cancer mask missing malignant tissue).

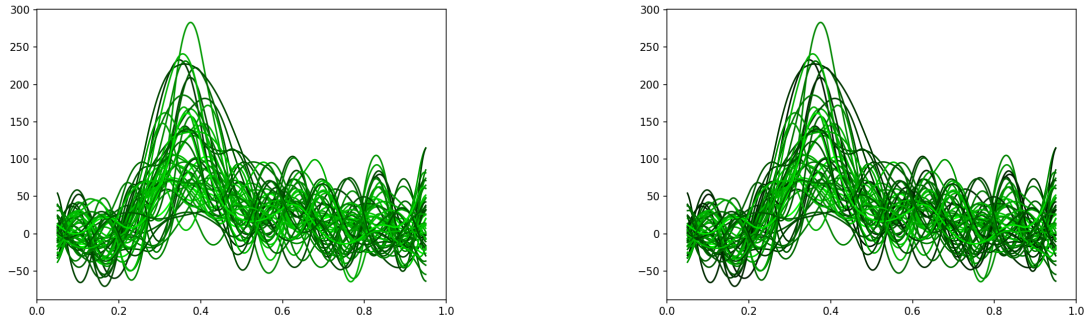
The purpose of narrowing down the data scope to a single patient is to simplify the analysis and explore the application of methods defines in the Workflow section. After reaching an acceptable model accuracy on a case study patient, the data scope will expand to all 144 patients.

For functional data analysis approach, the first order derivative of functional curves are used as input. First derivative curves show the velocity of the curve [4], i.e. the aggressiveness of the change of curve. Since one of primary histological characteristics of a tumor is fast contrast agent wash-in rate [7], the derivative curve with bigger fluctuations should be a reliable indicator. A visual example of curve's first order derivative alongside the original curve is shown in Figure 9. Red curves represent average signal intensity of segments that contain malignant biopsy and significant ($> 50\%$) overlap with the cancer mask, blue curves - malignant biopsy with insignificant cancer mask overlap ($\leq 50\%$), gray curves - other cases. As the red and blue curves follow the expected tumor trajectory in Figure 9a of having stronger wash-in, teh same curves curves in a derivative form in Figure 9b show higher peaks.

For K-Nearest Neighbors a parameter optimization is done: 5-fold cross validation for number of neighbors for values $\{3, 5, ..., 17, 19\}$; the best validation score was shown with $k = 11$ neighbors. For the Nearest Centroid model no parameter selection is done.

For machine learning approach, a total of 10 features are created explaining different aspects of curve behaviour ([11, 13]):

1. integrated depth (ID)
2. modified band depth (MBD)
3. maximum signal intensity (MI)
4. maximum signal intensity of first derivative (MID)
5. difference between maximum and base signal intensities (SD)
6. time-to-maximum intensity (TM)
7. time-to-maximum of first derivative (TMD)
8. slice order (S)
9. discreticized curve (DC)
10. discreticized derivative curve (DCD)



(a) Integrated depth

(b) Modified band depth

Figure 10: Integrated and modified band depth, color-coded: darker shade is higher depth value, lighter color - lower depth value

Model	Features
SVM option 1	ID, MBD
SVM option 2	ID, MBD, MI, MID, SD, TM, TMD, S
SVM option 3	DC, DCD
SVM option 4	ID, MBD, MI, MID, SD, TM, TMD, S, DC, DCD

Table 1: Feature sets

A depth is a concept that helps to determine the "centrality" of a functional curve, i.e. how "deep" it is, similar to a median in descriptive statistics. A depth is a useful measure for quantifying the degree of "outlyingness" of each curve compared to the rest [6, 15]. A visual example of ID and MBD are shown in Figure 10, containing derivative functional curves for the slice 10 of patient 50. The darker shade of the green indicates higher value for either ID or MBD, meaning those curves are closer to being considered as outlying curves compared to the rest of the functional data. The difference between ID and MBD methods is subtle - curves closer to the middle have a slightly lighter shade for ID, implying that ID values are more right-skewed than MBD values.

Discreticization is a transformation of a continuous functional curve into discrete form. The purpose of this is to create a finite-dimensional numerical vector that is used as model input. For DC and DCD features, the discrete form consists of 20 points equally spaced over the time axis.

Features are grouped into 4 feature sets as shown in Table 1 and a separate SVM is trained for each of the feature sets, resulting in 4 models ("SVM options").

SVM algorithm is sensitive to scaling, i.e. features with unequal value ranges have different effect on class discrimination [9]. Prior to modeling, numerical features were preprocessed by my removing the mean and scaling to unit variance and categorical features transformed using ordinal encoding.

The performance of both FDA and machine learning approaches are evaluated using the confusion matrix and three metrics derived from it - accuracy ($\frac{TP+TN}{P+N}$), precision ($\frac{TP}{PP}$), recall ($\frac{TP}{P}$), where TP - true positive, TN - true negative, P - actual positive, N - actual negative, PP - predicted positive.

The resulting confusion matrices are shown in Figure 11 and Figure 12 and performance matrix in Table 2, which are derived from the confusion matrices in Appendices A and B.

Due to the nature of imbalanced data (abundance of non-cancerous curves), the majority of predictions in FDA outcomes (and to a smaller degree, ML outcomes) are true negatives. Among the two functional methods, K-Nearest Neighbors outperformed Nearest Centroid (11b) having higher accuracy and recall, but with significant decrease in precision. For SVM options 2, 3, 4, performance metrics

Model	Accuracy	Precision	Recall
K-Nearest Neighbor	95.67%	18.75%	75.00%
Nearest Centroid	87.62%	87.50%	26.92%
SVM option 1	69.66%	56.25%	9.00%
SVM option 2	86.69%	93.75%	26.32%
SVM option 3	87.62%	81.25%	26.00%
SVM option 4	89.16%	81.25%	28.89%

Table 2: Performance results

showed similar results, while option 1 underperformed compared to the rest.

7 Conclusions and recommendations

A novel approach of Temporal Variation Matrix was developed which helps with defining zones with different signal intensity variation over time. Such concept is then used for fully-automatic segmentation of DCE images using the SLIC algorithm, the parameters for which (compactness and segment number) were optimized for a selected case study patient. Using the resulting segments, discrete time-signal curves were constructed which are transformed into continuous form using Functional Data Analysis B-spline smoothing, including the process of smoothing parameter (polynomial order and basis function number) optimization. Such functional curves were then used for two classification paths: functional and machine learning. Functional classification includes two models: K-Nearest Neighbor and Nearest Centroid. For machine learning approach, feature engineering is done creating 10 feature vectors from which 4 feature groups are formed, each used for training of a separate SVM model. For model evaluation, accuracy, precision, and recall were used. The resulting performance indicates that there is a trade-off between accuracy/precision and recall.

Overall, the concluding models are not performing with sufficient accuracy for real-life application due to high false positive and false negative rates. Potential points of action, both for the modeling and data management parts of the workflow, are proposed for improvement:

1. Imbalanced data - underrepresented classes affect model's ability to segregate different classes, so resampling techniques during or configuring class weights during model training may improve class discrimination; alternatively, data augmentation can balance out class representation by creating artificial data

2. Label definition - current models use binary classes defined in Equation 7; it may be improved by formulating a different label logic which more accurately follows histological and physiological patterns of contrast enhancement

3. Hyperparameter tuning

4. Functional derivative curve registration - registration is a common data processing technique in FDA which warps a set of curves to align them according to particular method's rules; vertically aligning peak enhancement of curves using registration may improve enhancement maximum comparison

5. Image registration - similarly to the point above, registration can also be applied to DCE-MRI images; this step is common in image analysis and is mentioned in [13]

6. Feature engineering - more extensive experimentation (new features, different feature combinations, feature transformations) may improve model learning

7. Image preprocessing - introducing additional transformations in the Data preprocessing block (e.g. de-noising filtering, morphological transformations) may improve DCE imaging quality

8. Segmentation - with current approach, curve construction process is heavily dependent on TVM and SLIC methods; other alternatives have to be explored for more accurate segmentation

9. Other model frameworks - different classifiers may be more suitable for such application, so other types of models (similarly to [14]) have to be explored

Ultimately, the ambition is to create a classification model that will have sufficient predictive capabilities not only for a case study, but the general population, i.e. a model with necessary degree of robustness to evaluate patients of various physiological characteristics under diverse conditions. The next iterations of model development will consist of the implementation of the improvement points above while expanding the scope of the data to all 144 patients.

References

[1] Saeed Alqahtani et al. "Prediction of prostate cancer Gleason score upgrading from biopsy to radical prostatectomy using pre-biopsy multiparametric MRI PIRADS scoring system". In: *Scientific reports* 10.1 (2020), pp. 1–9.

[2] Tristan Barrett, Baris Turkbey, and Peter L Choyke. "PI-RADS version 2: what you need to know". In: *Clinical radiology* 70.11 (2015), pp. 1165–1176.

[3] Freddie Bray et al. "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries". In: *CA: A Cancer Journal for Clinicians* 68.6 (2018), pp. 394–424. DOI: https://doi.org/10.3322/caac.21492. eprint: https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21492. URL: https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21492.

[4] Hervé Cardot. *Functional Data Analysis with R and Matlab by RAMSAY, JO, HOOKER, G., and GRAVES, S.* 2010.

[5] Chih-Chung Chang and Chih-Jen Lin. "LIBSVM: a library for support vector machines". In: *ACM transactions on intelligent systems and technology (TIST)* 2.3 (2011), pp. 1–27.

[6] Ricardo Fraiman and Graciela Muniz. "Trimmed means for functional data". In: *Test* 10.2 (2001), pp. 419–440.

[7] Roberta Fusco et al. "Textural radiomic features and time-intensity curve data analysis by dynamic contrast-enhanced MRI for early prediction of breast cancer therapy response: preliminary data". In: *European radiology experimental* 4.1 (2020), pp. 1–7.

[8] Sangeet Ghai and Masoom A Haider. "Multiparametric-MRI in diagnosis of prostate cancer". In: *Indian journal of urology: IJU: journal of the Urological Society of India* 31.3 (2015), p. 194.

[9] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. *A practical guide to support vector classification*. 2003.

[10] Benjamin Irving. "maskSLIC: regional superpixel generation with application to local pathology characterisation in medical images". In: *arXiv preprint arXiv:1606.09518* (2016).

[11] Fahmi Khalifa et al. "Models and methods for analyzing DCE-MRI: A review". In: *Medical physics* 41.12 (2014), p. 124301.

[12] Chih-Ching Lai et al. "Histogram analysis of prostate cancer on dynamic contrast-enhanced magnetic resonance imaging: A preliminary study emphasizing on zonal difference". In: *PloS one* 14.2 (2019), e0212092.

[13] Jacob Levman et al. "Classification of dynamic contrast-enhanced magnetic resonance breast lesions by support vector machines". In: *IEEE Transactions on Medical Imaging* 27.5 (2008), pp. 688–696.

[14] B Liu et al. "Prediction of prostate cancer aggressiveness with a combination of radiomics and machine learning-based analysis of dynamic contrast-enhanced MRI". In: *Clinical radiology* 74.11 (2019), 896–e1.

[15] Sara López-Pintado and Juan Romo. "On the concept of depth for functional data". In: *Journal of the American statistical Association* 104.486 (2009), pp. 718–734.

[16] Russell N Low, Donald B Fuller, and Naira Muradyan. "Dynamic gadolinium-enhanced perfusion MRI of prostate cancer: assessment of response to hypofractionated robotic stereotactic body radiation therapy". In: *American Journal of Roentgenology* 197.4 (2011), pp. 907–915.

[17] Christopher Manning, Prabhakar Raghavan, and Hinrich Schütze. "Vector space classification". In: *Introduction to Information Retrieval* (2008), pp. 289–317.

[18] Manish Sarkar and Tze-Yun Leong. "Application of K-nearest neighbors algorithm on breast cancer diagnosis problem." In: *Proceedings of the AMIA Symposium*. American Medical Informatics Association. 2000, p. 759.

[19] Sadhna Verma et al. "Overview of dynamic contrast-enhanced MRI in prostate cancer diagnosis and management". In: *American Journal of Roentgenology* 198.6 (2012), pp. 1277–1288.

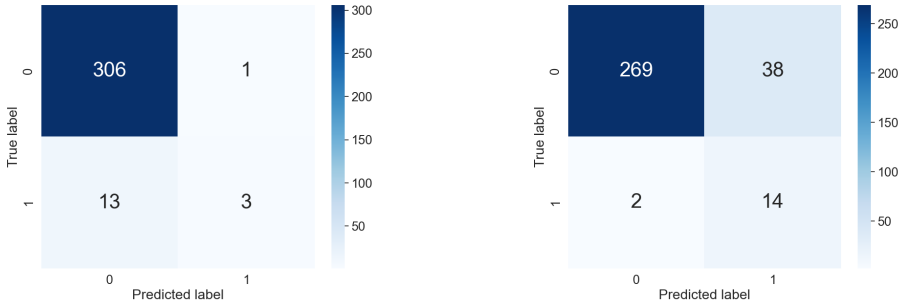
[20] Sadhna Verma et al. "Prostate MRI and 3D MR spectroscopy: how we do it". In: *AJR. American journal of roentgenology* 194.6 (2010), p. 1414.

[21] Thomas E Yankeelov and John C Gore. "Dynamic contrast enhanced magnetic resonance imaging in oncology: theory, data acquisition, analysis, and examples". In: *Current Medical Imaging* 3.2 (2007), pp. 91–107.

[22] Ji Min Yoon et al. "Dynamic contrast-enhanced MRI of the prostate: can auto-generated wash-in color map be useful in detecting focal lesion enhancement?" In: *Investigative Magnetic Resonance Imaging* 23.3 (2019), pp. 220–227.

8 Appendices

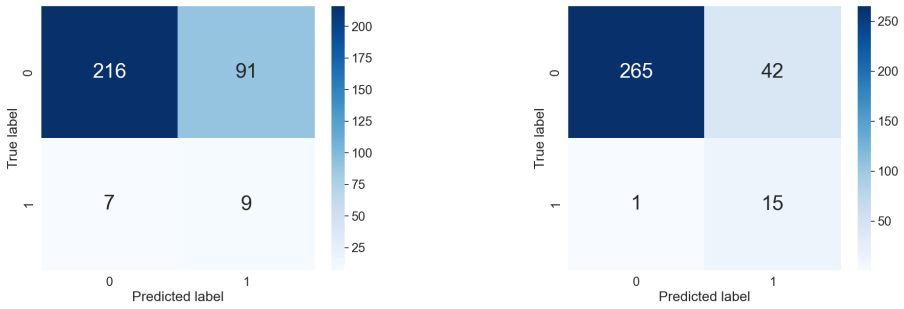
Appendix A



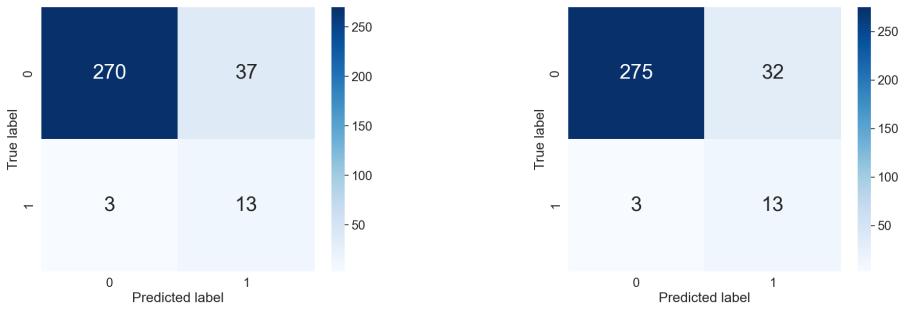
(a) K-Nearest Neighbors (b) Nearest Centroid

Figure 11: Functional data analysis approach performance

Appendix B



(a) SVM option 1 (b) SVM option 2



(c) SVM option 3 (d) SVM option 4

Figure 12: Machine learning approach performance