EINDHOVEN UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering

SLIC segmentation and texture feature extraction for a radiomic analysis of prostate cancer using multiparametric MRI

by

H.Gong

MSc Thesis

Thesis committee

Chair:	Prof.dr.ir.M.Mischi
Member 1:	Dr.S.Turco
Member 2:	Prof.dr.A.Vilanova
Advisory member 1:	Dr.C.Dinis Fernandes

Graduation

Program:	AI & ES
Research group:	Signal Processing Systems
Thesis supervisor:	Dr.S.Turco
Date of defense:	15-08-2024
Student ID:	1897616
Study load (ECTS):	45

CONTENTS

***Abstract*—Prostate cancer (PCa) is a common cancer among male patients worldwide. There has been an urgent demand for efficiently identifying aggressive tumors to provide accurate diagnosis and treatment. Radiomics consists of a quantitative analysis of medical imaging data with the intent to extract relevant information that may have predictive value. Through imaging features, it can further help identify aggressive tumors if distinct tumor phenotypic differences can be captured. The extraction and analysis of features from medical imaging data are the fundamental building blocks of radiomics. Multi-parametric magnetic resonance imaging (mpMRI) can be used to visualize tissue anatomy and physiology, and it is the recommended imaging modality for prostate cancer diagnosis. However, mpMRI has been demonstrated to have low specificity and high inter-observer variability. Quantitative mpMRI radiomics analysis has the potential to uncover aggressive imaging phenotypes.**

The typical radiomics approach is to condense imaging features per region of interest (ROI), which decreases the granularity of the extracted information. On the other side of the spectrum, in a previous study which laid the foundation for the current project, the features were extracted with a moving window approach, resulting in parametric feature maps that were then summarized using histogram statistics for the ROI. Textural imaging features were extracted from multi-parametric MRI sequences (T2W, DWI, and DCE) and combined with DCE-derived parametric pharmacokinetic maps obtained using magnetic resonance dispersion imaging (MRDI). Imaging features were used to train and optimize tumor aggressiveness classifiers.

In this study, Simple Linear Iterative Clustering (SLIC) was investigated as another method of scheduling the area for texture feature extraction. SLIC can be used to identify pixels with similar features in their color space and coordinate space. By clustering these pixels into superpixels, the SLIC algorithm can be more calculation efficient than analyzing the image pixel by pixel. The pixels also have a better spatial consistency in a superpixel than a regular sliding window. In addition, the SLIC algorithm has better noise robustness. Feature extraction and model training on the superpixel-wide have shown to be more efficient and accurate. Three machine learning models including KNN, SVM, and LR were used to train a classifier to detect aggressive tumors on a dataset with 35 PCa patients using 5-fold cross-validation. The best model obtained an accuracy of 71.4%, with sensitivity 80.0% and AUC 60.0%.

I. Introduction

Prostate cancer (PCa) is one of the leading causes of death worldwide and the second-highest incidence of cancer in men [1]. It is necessary to improve the diagnosis of aggressive prostate tumors and provide appropriate treatments. Despite the mainstream diagnosis method for PCa being based on the use of prostate-specific antigen (PSA) and biopsies [2], having MRI as a complement has demonstrated superior performance in testing for PCa, yielding fewer side effects, a higher rate of identifying candidates for treatment, and a reduced incidence of detecting clinically insignificant cancers, thereby minimizing overtreatment [3].

Multi-parametric magnetic resonance imaging (mpMRI) can be used to visualize tissue anatomy and physiology, and it is the recommended imaging modality for PCa diagnosis. mpMRI has been identified to be able to mitigate diagnostic errors in PCa diagnosis [4]. T2-weighted (T2W) imaging, diffusion weighting imaging (DWI), and dynamic contrast-enhanced (DCE) are the three primary sequences [5]. T2W images can accurately depict the details of tissue morphology, and allow easier visibility of fluid intensity [6]. With DWI, it is possible to probe the important information about the movement and functional environment of water in tissue and it reflects the cellular status of normal and pathological tissue, from which an Apparent Diffusion Coefficient (ADC) map can be derived. From the ADC map, tumor-suspected regions can be identified as hypointense regions, reflecting high cellular density content [7]. DCE sequences provide an opportunity to study angiogenesis, and its parameters measure tissue perfusion by assessing microvascular perfusion, permeability, and architecture [8]. Magnetic resonance dispersion imaging (MRDI) has been proposed as a quantitative analysis for DCE-MRI which is based on modeling the transport of the contrast agent from the injection site to the leakage sites as a convective dispersion process. Currently, mpMRI is mainly visually evaluated by radiologists. The visual analysis relies heavily on the interpretation skills and experience of the radiologist, leading to variability in diagnosis and increasing the potential for errors. This variability can result in differences in lesion detection and characterization [9].

Radiomics translates medical images into quantitative data to yield biological information and enable radiologic phenotypic profiling for diagnosis, theragnosis, decision support, and monitoring. Multiple approaches and stages are proposed in radiomics, in which features based on shape, pixel intensities, and texture are extracted from radiographs [10].

This study builds upon previous research investigating the use of mpMRI radiomic features for detecting aggressive prostate cancer (PCa) tumors extracted texture feature maps from T2W, ADC, DCE peak, and DCE wash-in frames. Five categories of features were extracted including global, Gray-level co-occurrence matrix (GLCM), Gray-level run-length matrix (GLRLM), Gray-level size zone matrix (GLSZM), and Neighborhood gray-tone difference matrix (NGTDM) features. Unlike the traditional ROI-based feature extraction method which results in one value of each feature for the full ROI, the texture feature extraction framework employed a moving window approach that defines the region of interest (ROI) as a moving window, enabling the extraction of texture feature values at each imaging pixel, and resulting in a complete parametric map. In this way, it is possible to preserve the spatial characteristics of the features and retain potential texture heterogeneity within the image [11].

The SLIC algorithm was investigated as an approach leading to fewer features than pixel-wide analysis and without losing the granularity. The SLIC algorithm has been widely used for image segmentation and pixel clustering based on features such as texture and color. It can be used to identify regions with similar characteristics in the image which can be clustered in superpixels. Besides, operating on superpixels instead of pixels has been proven to have better speed, memory efficiency, and boundary adherence [12]. In this context, SLIC presents a promising alternative to the moving window strategy employed in the previous study.

With the aim of improving the identification of aggressive tumors with the use of textural features extracted from the prostate superpixels identified on multiparametric MRI, the SLIC algorithm was applied to a dataset of mpMRIs from

PCa patients. The texture features were extracted within the superpixels and three machine-learning models were used to train and validate a classifier for identifying aggressive cancer.

II. Methods

A. Patient data

The dataset consisted of data from 35 PCa patients who were imaged with a mpMRI protocol before being treated with a prostatectomy. For each patient, one slice of MRI was used for further analysis. The tumor region ROIs were identified in one slice for all these patients based on the histopathology samples leading to a total of 35 tumors for the dataset. The sequences including T2W, DWI, and DCE were available for the mpMRIs. Five MRDI maps were also available for each slice and each patient.

B. SLIC segmentation

The SLIC algorithm performs a local clustering of pixels in the color space as well as the pixel coordinates. It creates a desired number of approximately equally sized superpixels as K within an image with N pixels. The approximate size of each superpixel is N/K. The distance between the center of the superpixels is $\sqrt{N/K}$. For the color images, a 5D vector is created for each pixel to store its information, including 3 color space elements and 2 coordinate elements. The color space is LAB or RGB for CIELAB color space. In this case, all the images are grayscales, which makes the color space with only one element - image brightness.

SLIC can be affected by the surrounding image variation. Thus, To address this issue, a square-mask SLIC approach was introduced, which involved applying the SLIC algorithm within a square mask defined by the size of the ROIs.

$$D_s = D_G + \frac{m}{s} \times \sqrt{(x_p - x_c)^2 + (y_p - y_c)^2} \tag{1}$$

, where D_s is the target value for calculating the distance with the center of the superpixels. D_G is the distance in the grayscale space as the brightness differneces. (x_p, y_p) and (x_c, y_c) are the coordinates of pixels and center of superpixels respectively. m is a constant which normally has a value of 10 [12].

Since the standard SLIC method can work more effectively for analysis within defined regions such as a tumor or organ, adding a square mask around the prostate and applying the SLIC algorithm within the mask was considered to have a more promising result [13].

For the SLIC algorithm, there are three main parameters for tuning. The compactness parameter of the superpixel controls the regularity of the superpixel shape. Higher compactness leads to superpixels with more regular shapes, while lower compactness makes the shape of the superpixel closer to the color change. The sigma parameter reduces the noise effects. A high value of sigma helps to reduce noises in the images. Nevertheless, a high sigma value may also result in the loss of critical details. The parameters of the SLIC algorithm were fine-tuned in the study. The number of superpixels within the target area was considered the primary parameter for the segmentation in this case. Since the number of pixels within the mask varies among the patients, a parameter α was created to identify the relation between the total number of pixels within the prostate area and the number of superpixels for segmentation:

$$\alpha = -\ln\left(\frac{n_{\text{segment}}}{\text{pixel\_number}}\right) \tag{2}$$

, where $n\_segment$ is the number of superpixels for segmentation. $pixel\_number$ is the number of pixels within the area and varies from patient to patient depending on the image resolution. Since the $n\_segment$ parameter in SLIC is an integer, the value was rounded to its nearest integer. To quantitatively evaluate the SLIC results and identify the best segmentation parameters, those superpixels that overlapped with the identified prostate area with a large enough percentage over a set threshold were considered as the segmented prostate area. The threshold as the percentage was set to be 0.6 in this case. The value of the Dice Score, Jaccard Index, Hausdorff distance, and Average Surface Distance were calculated to create an evaluation parameter $closeness$. These evaluations were first normalized. Then the Euclidean distance to the ideal solution and the worst solution were calculated. This is a commonly used method in The Technique for Order of Preference by Similarity to Ideal Solution(TOPSIS) when dealing with Multi-Objective Decision Making(MODM). The best solution is expected to have the minimum distance to the ideal solution and the maximum distance to the worst solution. The $closeness$ was defined as:

$$\text{closeness} = \frac{\text{distance to worst}}{\text{distance to worst} + \text{distance to best}} \tag{3}$$

, where $distancetoworst$ is the distance to the worst solution and $distancetobest$ is the distance to the ideal solution. The best SLIC settings corresponded to the largest $closeness$.

C. Feature extraction

Once the optimal SLIC segmentation parameters were identified for each patient, the superpixel segmentation was used to perform feature extraction. Texture feature maps were extracted from T2W, ADC maps, DCE peak, and DCE wash-in sequences. The superpixels selected for texture feature extractions were those that overlapped with the recognized tumor region to the area of the superpixel. To apply texture feature analysis, the radiomics MATLAB toolbox implemented by Vallières et al. was used [14]. The features were extracted within each superpixel. A summary of the extracted texture features is shown in Table I. In addition, the original grayscale images and the 5 MRDI parametric maps were also used as feature maps for each superpixel. To calculate a feature value for the whole tumor, a weighted average was used by giving each of the superpixels a weight according to the proportion of the overlap area with the tumor region.

D. Feature selection

For the reduction of dimensionality of the feature space, correlation analysis was applied to the dataset. For features

Texture type	Features	Number of features
Global	Variance, Skewness, Kurtosis	3
Gray-level co-occurrence matrix (GLCM)	Energy, Contrast, Correlation, Homogeneity, Variance, Sum Average, Entropy, Dissimilarity, Auto Correlation	9
Gray-level run-length matrix (GLRLM)	Short Run Emphasis (SRE), Long Run Emphasis (LRE), Gray-Level Non-uniformity (GLN), RunLengthNon-uniformity (RLN), Run Percentage (RP), Low Gray-Level Run Emphasis (LGRE), High Gray-Level Run Emphasis (HGRE), Short Run Low Gray-Level Emphasis (SRLGE), Short Run High Gray-Level Emphasis (SRHGE), Long Run Low Gray-Level Emphasis (LRLGE), Long Run High Gray-LevelEmphasis (LRHGE), Gray-Level Variance, (GLV) Run-Length Variance (RLV)	13
Gray-level size zone matrix (GLSZM)	Small Zone Emphasis (SZE), Large Zone Emphasis (LZE), Gray-Level Non-uniformity (GLN), Zone-SizeNon-uniformity (ZSN), Zone Percentage (ZP), Low Gray-Level Zone Emphasis (LGZE), High Gray-Level Zone Emphasis (HGZE), Small Zone Low Gray-Level Emphasis (SZLGE), Small Zone High Gray-Level Emphasis (SZHGE), Large Zone Low Gray-Level Emphasis (LZLGE), Large Zone High Gray-LevelEmphasis (LZHGE), Gray-Level Variance (GLV), Zone-Size Variance (ZSV)	13
Neighborhood gray-tone difference matrix (NGTDM)	Coarseness, Contrast, Busyness, Complexity, Strength	5

TABLE I: Table of texture features extracted

with a correlation higher than 0.9, the feature with the lowest correlation to the label was dropped. Then the principal component analysis(PCA) was used for further dimension decrease. The number of components necessary to describe 95% of the variance on the dataset was identified. After that, the features were ranked using the maximum relevance minimum redundancy algorithm (MRMR), and the top-ranking features corresponding to the identified number of components were selected. The data was scaled using a standard scaler.

E. Hyperparameter optimization

The machine learning models including logistic(LR), k-nearest neighbor(kNN), and support vector machine(SVM) were investigated and trained with this subset of features. A nested 5-fold cross-validation was used with a randomized search with 20 iterations for hyperparameter optimization. The performance of the models for different number of features was evaluated using stratified 5-fold over 10 repetitions. The number of features with the highest performance was identified as the best for each fold and every model. The corresponding model settings were reported over the best number of features. The area under the curve (AUC) was chosen for the evaluation due to its ability to comprehensively measure the performance of a classifier on imbalanced datasets without relying solely on the performance at a specific threshold.

F. Machine learning

The 5-fold cross-validation was applied to the dataset which divided it into a 4-fold sub-dataset for training. The rest were used as the test dataset. The process was implemented five times so that all data samples were used both in the training dataset and the validation dataset. However, each data sample was exclusively included in either the training dataset or the validation dataset in each iteration. In every training dataset, a 5-fold cross-validation was applied to find the best model settings and number of features with the highest AUC among all the alternative number of features. The best number of features and model settings were then used to train on the full training folds leading to a total of 5 evaluation metrics including accuracy(ACC), balanced accuracy(bACC), AUC, sensitivity, and specificity. The same settings were subsequently validated on the validation dataset to compare the final performance among the three models.

G. Permutation feature importance

The relevance of features was measured by the Permutation Feature Importance (PFI). PFI shows a performance decrease when a single feature value is randomly shuffled. The PFI was used on each model's optimal number of ranked features with stratified 5-fold cross-validation over 10 repetitions. The feature importance was averaged over 10 repetitions in one fold. The median AUC, median ACC, median sensitivity, median specificity, and median bACC for each model were compared as the condensed results.

H. Evaluation

The three classifiers were trained and tested on the dataset with a stratified 5-fold cross-validation strategy. The ranked feature subset was used to define the best hyperparameters and the optimal number of features for each classifier. All the classifiers were trained on the whole training dataset and validated for their performance of classification prediction on the validation dataset. The evaluation indicators included ACC, bACC, AUC, sensitivity, and specificity.

III. Results

A. Data preparation

The T2W images were used for aligning the results as well as applying the SLIC algorithm. The original images were rescaled and cropped to a more specific region around the prostate contour. Since the images for ADC, DWI, and DCE share the same size as T2W, the location of ROI and superpixels will be the same for the others. An example of a patient for the location of the prostate in the slice in Fig1 and the tumor location was also marked in the prostate region in Fig2.

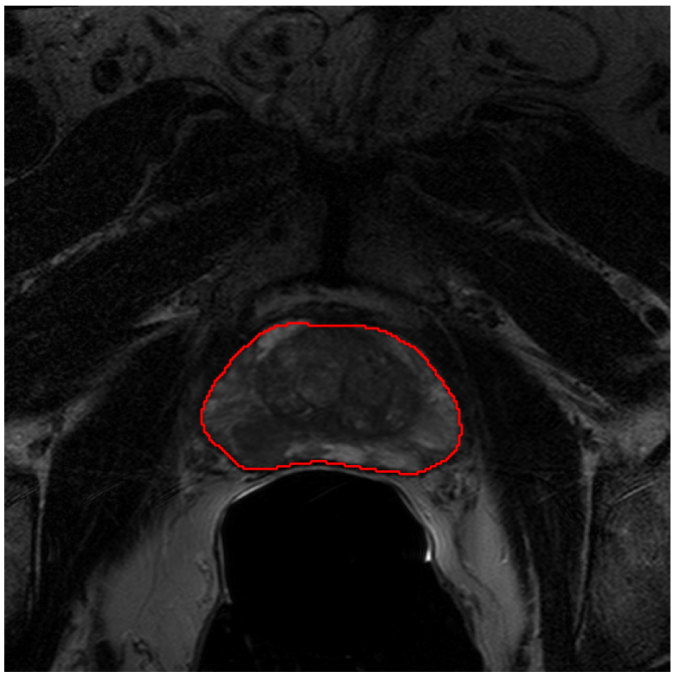


Fig. 1: An example of the prostate contour marked in the original NIFTI image

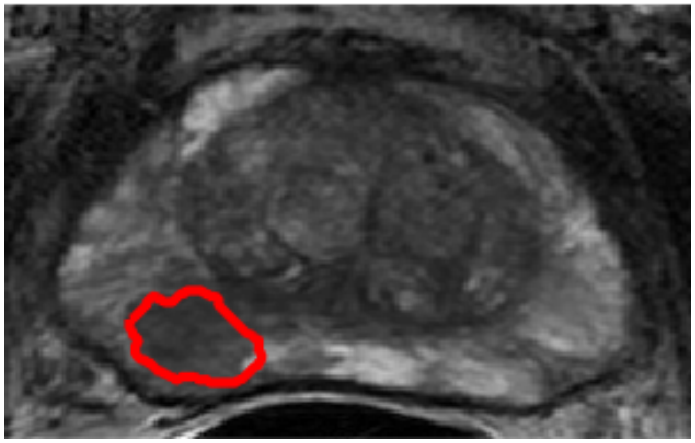


Fig. 2: The tumor shown in the cropped region around the prostate

B. SLIC algorithm

The SLIC algorithm was applied to the mpMRIs on the specific slices. The parameters of the SLIC algorithm were tuned to get a better performance on the segmentation of the superpixel boundaries. The settings of the SLIC algorithm with the highest $closeness$ was considered to be the best model settings for the patient. The best $n_segment$ was found among the iterations of α for each patient. The compactness and sigma parameters exhibited stability and balance in the model at values of 0.05 and 1, respectively.. The main steps and results of SLIC are shown with an example in Figure3.

After the SLIC segmentation for each patient slice, the labels of the superpixels overlapping with the tumor were saved for texture feature extraction.

C. Feature map extraction

A total of 43 texture features were extracted for each superpixel from each T2w, ADC, DCE peak, and DCE wash-in sequence. With the calculated weights of each superpixel, each of the features were condensed to one value for each patient slice. In total, 181 parametric maps were obtained, including 4 grayscale raw images, 43 texture maps for each of the 4 raw images, and 5 MRDI maps. These features were used for further model training.

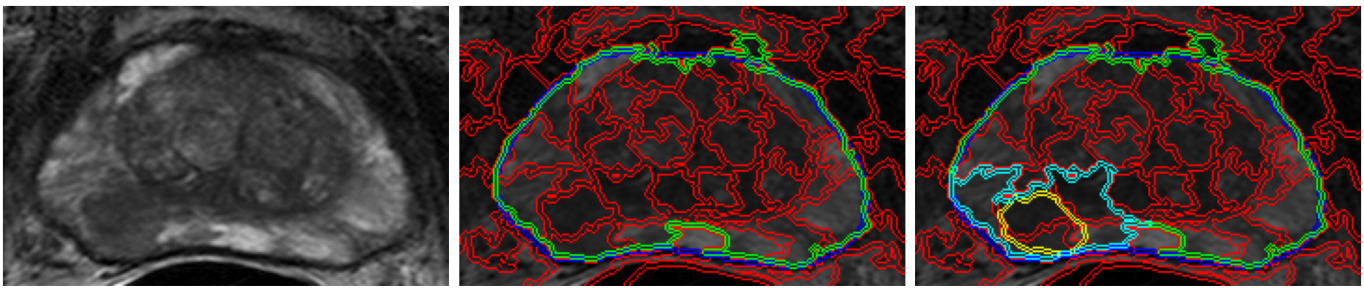
D. Feature selection and hyperparameter optimization

After removing the highly correlated features, a total of 77 features were left for PCA analysis. The number of features for model training was decreased to 18. The classifiers trained with all three models were evaluated on their performance over the number of features. AUC was selected as the evaluation indicator. the mean and standard deviation(STD) of AUC through all folds for each number of features can also be compared to see some characteristics of the models as shown in Fig4.

The overview of the optimal hyperparameters for all the classifiers is shown in Table II. The times of all the features selected for each model are shown in Fig5. The optimal number of features for KNN oscillated between 6 and 14. The feature DCEwashin.GLSZM.SZLGE was selected for all folds. Feature T2.Global.Variance, DCEpeak.GLRLM.LGRE, DECwashin.GLCM.Variance and MRDI.kep were also commonly selected over the folds. The optimized weights, metric, number of neighbors, and optimal number of features are shown in the table. As to the SVM model, the optimal number of features was 5, which was selected 3 times out of 5 folds. DCEwashin.GLSZM.SZLGE was also selected for all folds. T2.Global.Variance, DCEpeak.Global.Kurtosis, DCEpeak.Global.LZLGE and DCEpeak.GLRLM.LGRE were the more common features among the folds. The optimal kernel was the sigmoid. However, the other parameters vary a lot over the folds.The optimal number of features for model LR had the same times of selection on 5 and 13. The feature DCEwashin.GLSZM.SZLGE was still selected for all folds. T2.Global.Variance and DCEwashin.GLCM.Variance were also selected more among all the features. For the LR model, all penalties and solvers had the same selections. However, the regularization parameter showed a large instability. It was observed that DCEwashin.GLSZM.SZLGE was consistently selected across all folds and models. It is also interesting to notice that DCEwashin.GLCM.Variance, DCEpeak.GLRLM.LGRE and T2.Global.Variance were used more than the other features over all the models. While the feature ADC.GLSZM.ZP was never been selected in any folds and the models.

E. Permutation feature importance

The permutation feature importance(PFI) was used to calculate the importance of each feature for all three classifiers with the target scoring of AUC. Every feature was calculated among all five folds and 10 repeats. The PFI,



(a) Region of Interest (b) Predicted Prostate Boundaries (c) Selected Superpixels Overlapping with Tumor

Fig. 3: The process of applying SLIC to a sample image. And the visualized results. (a)The cropped original image of the mpMRI slice around the prostate region; (b)Applying SLIC with the best settings for this patient's mpMRI slice. The red lines represent the boundaries of all the superpixels. The blue line is the contour of the prostate area. The green line shows the segmentation results for the prostate area; (c)Mark the location of the tumor in the ROI(yellow boundaries) and the selected superpixels overlapping with the tumor area(light blue boundaries).

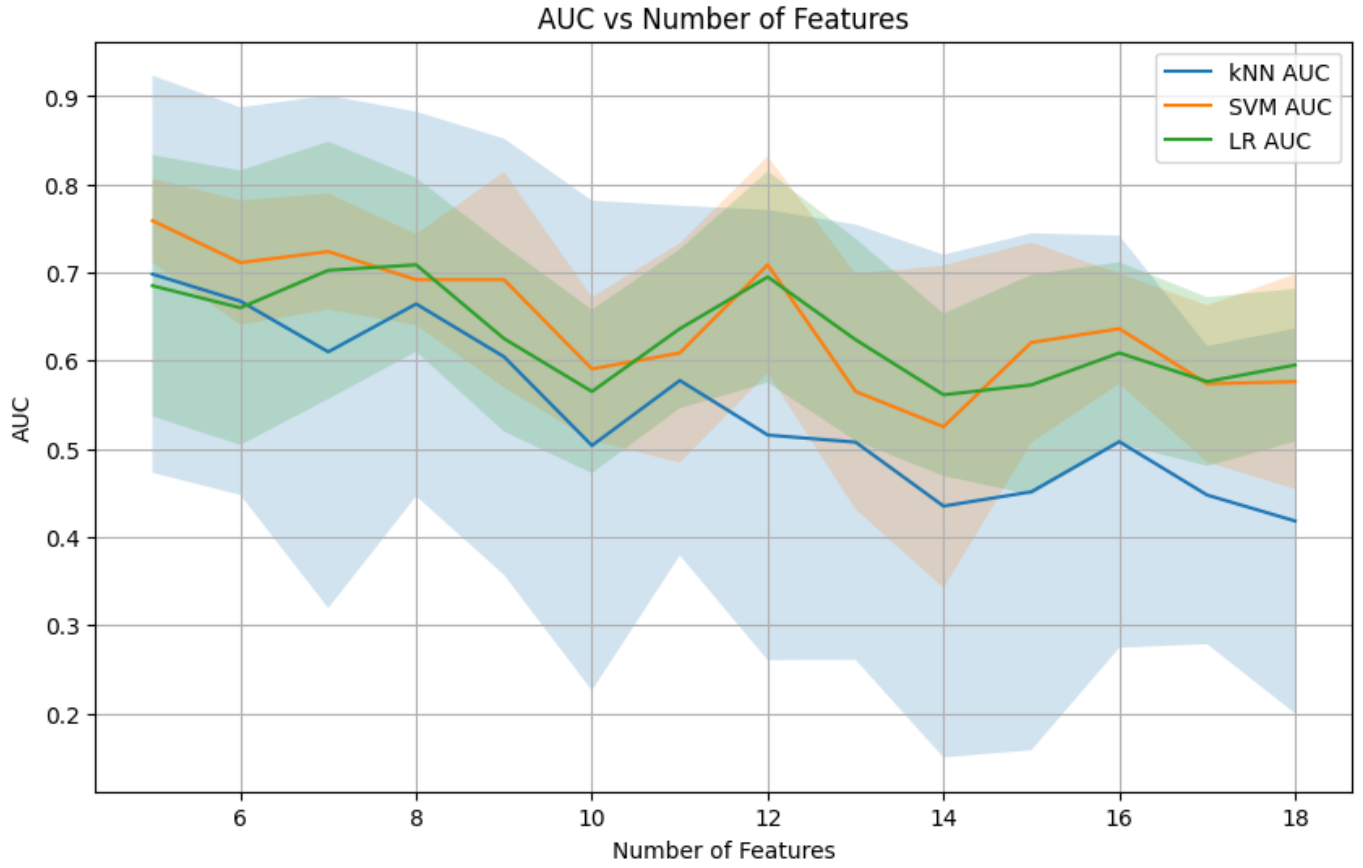


Fig. 4: The mean AUC over the number of features used for training for each model.

normalized to 0 to 1, was then used as the assessment of the importance of all features. All the features and their normalized PFI are shown in Fig6, Fig7, and Fig8 for model kNN, SVM, and LR respectively. After comparing the three models, the feature DCEwashin.GLSZM.SZLGE is found important ranking in the top five for all the models. The features DCEpeak.GLSZM.LZLGE, DCEwashin.Global.Skewness, and DCEwashin.GLSZM.GLN are also more important than the other features among the models. These four features tend to play a key role in making predictions for the models. They contain more information that relates to the aggressive tumor in mpMRIs.

F. Performance evaluation

During the training for all three models with their best model settings, the performance of each evaluation can be seen in Table III which shows the median(IQR) over the folds respectively. The same performance evaluation for the validation dataset is shown in Table IV.

Classifier	Hyperparameter	Values
LR	Solver	liblinear
	Penalty	l2
	Regularization parameter(C)	0.17
	Optimal number of features	5/13
kNN	Weights	distance
	Metric	minkowski
	Number of neighbors(K)	16
	Optimal number of features	oscillated between 6 and 14
SVM	Kernel	Sigmoid
	Kernel coefficient	Scale
	Regularization parameter(C)	0.05
	Optimal number of features	5

TABLE II: The summary of optimal hyperparameters for all classifiers.

Classifier	AUC	Accuracy	Balanced accuracy	Sensitivity	Specificity
LR	0.53(0.27)	0.50(0.33)	0.05(0.03)	0.74(0.95)	0.00(1.00)
kNN	0.61(0.08)	0.68(0.38)	0.50(0.00)	1.00(1.00)	0.00(1.00)
SVM	0.70(0.08)	0.68(0.07)	0.50(0.15)	0.63(0.47)	0.78(1.00)

TABLE III: The performance of three classifiers evaluated on the training dataset for AUC, ACC, bACC, sensitivity, and specificity. The values shown in the table are the median(IQR) over the five folds

Classifier	AUC	Accuracy	Balanced accuracy	Sensitivity	Specificity
LR	0.38(0.20)	0.43(0.17)	0.50(0.13)	0.60(0.75)	0.00(1.00)
kNN	0.50(0.10)	0.63(0.25)	0.50(0.00)	1.00(1.00)	0.00(1.00)
SVM	0.60(0.10)	0.71(0.43)	0.50(0.00)	0.80(0.60)	0.00(0.50)

TABLE IV: The performance of three classifiers evaluated on the validation dataset for AUC, ACC, bACC, sensitivity, and specificity. The values shown in the table are the median(IQR) over the five folds

All the evaluation scores for each model are relatively stable regarding median(IQR). It seems that the models were trained stably. The dataset maintained the same distribution as the training and validation dataset. kNN and LR have high sensitivity and low specificity. This means that they can successfully detect aggressive cancer cases but are poor at finding patients without aggressive tumors. Conversely, SVM demonstrated proficiency in distinguishing between aggressive and non-aggressive tumors. It, however, is similarly poor with the other two models on the test dataset. This might be because of the overfitting.

IV. Discussion

The SLIC algorithm has shown its high efficiency. A square-mask SLIC was utilized in this study aiming at achieving better segmentation results in a smaller area. However, the maskSLIC technique can also have great potential by applying SLIC within the prostate area. Indeed, this method necessitates additional ground truth values beyond the prostate contour.

Since AUC was overall a good metric for an imbalanced dataset, the best model settings were selected based on the highest AUC. As shown in Fig4, the mean AUC of SVM and LR over the number of features are generally higher than that of kNN. At the same time, kNN seems to have larger STD all the time, which means kNN may have worse stability and robustness. SVM and LR exhibit comparable levels of stability and robustness in these aspects.

From the PFI analysis and comparison among all the features, the DCE sequence seems to have more important features than the other two sequences. The models rely on these high PFI features to make decisions to some extent. The models are very sensitive to changes in these features. This may potentially influence the generalization ability of the models, since the models may be overly dependent on these features and ignore other features that may also have predictive power.

In summary of the model performance, kNN has relatively stable scorings compared with SVM and LR, which means kNN has a better generalization ability in this dataset but with the worst overall performance. kNN generally has low IQRs in the training dataset and high ones in the validation dataset. That means it is stable on the training dataset but with oscillation on the validation dataset. SVM shows some generalization ability with better overall performance. It shows a similar stability with kNN. For LR, the performance on the training dataset is relatively obviously better than the test dataset for many evaluation scores. In general, all the models seem to be good at detecting the cases with aggressive tumors, but not good enough on those with non-aggressive tumors. This, on the one hand, can be insufficient data for training and leads to overfitting. On the other hand, the imbalanced dataset with much more aggressive tumor cases than the negative ones made the classifiers focus more on the aggressive tumor features and information.

V. Conclusion

In this study, an approach based on SLIC was investigated to see if it is better to cluster the pixels with similar space information so that the texture features extracted from these clusters can better help train a classifier for aggressive tumor detection. The best-performing classifier achieved an accuracy of 71.4%, a sensitivity of 80.0%, and an AUC of 60.0%. But it has a rather low specificity. Therefore, the best-performing model may demonstrate enhanced accuracy in detecting aggressive tumors, as anticipated in clinical diagnostic scenarios. The model still needs further training on a larger dataset with larger and more balanced data on both classes.

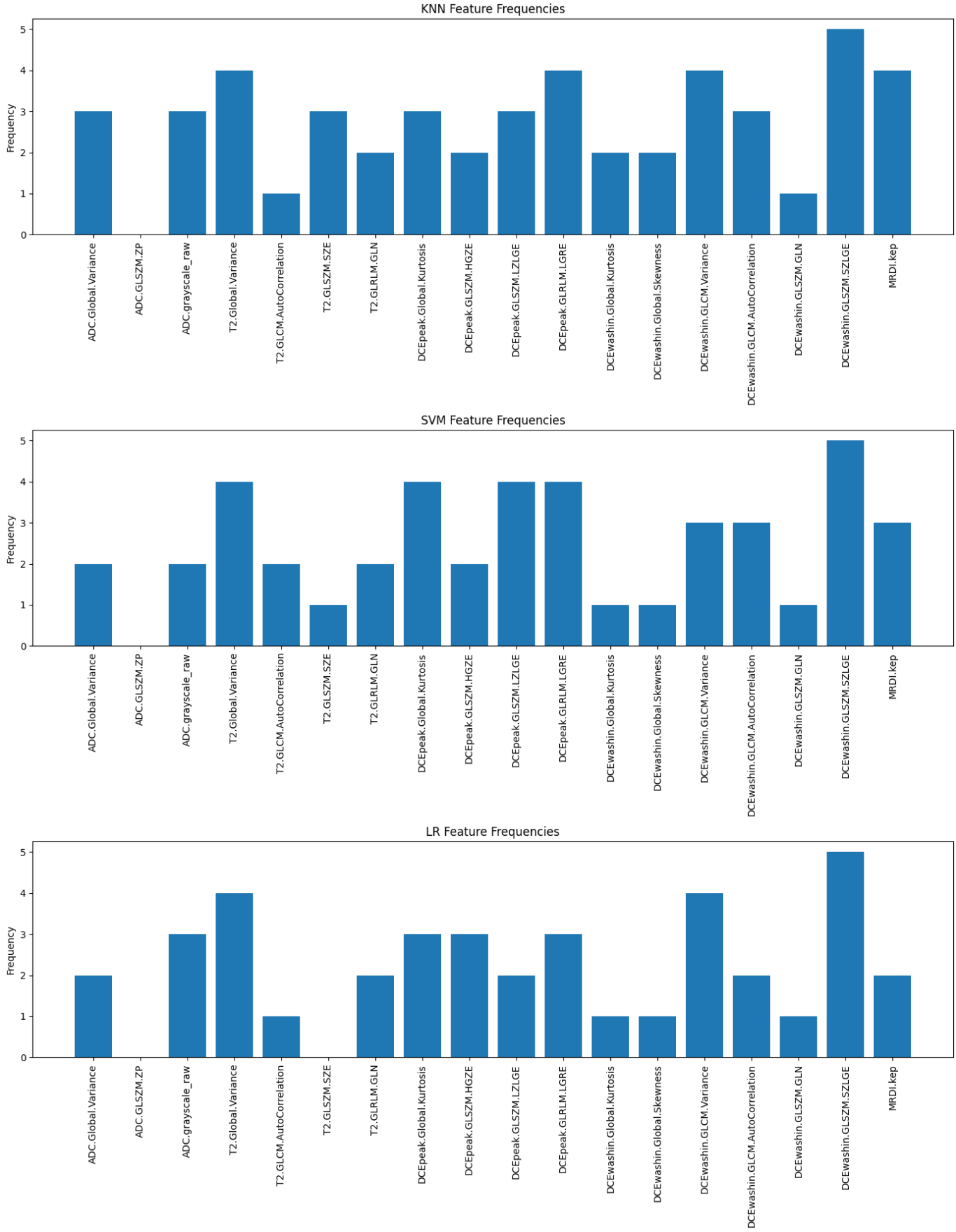


Fig. 5: Most commonly selected over five folds for each model

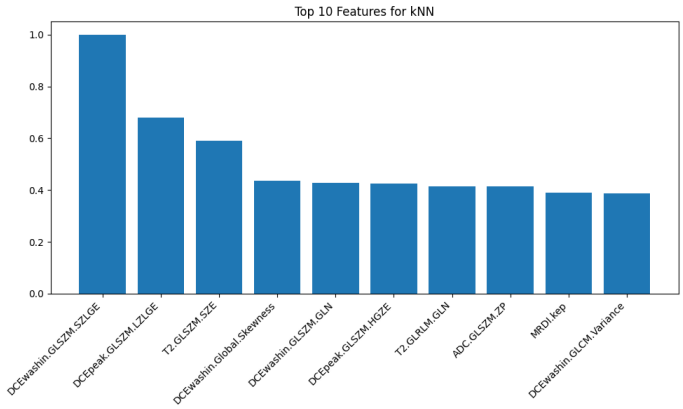


Fig. 6: Features ranked by normalized PFI for kNN

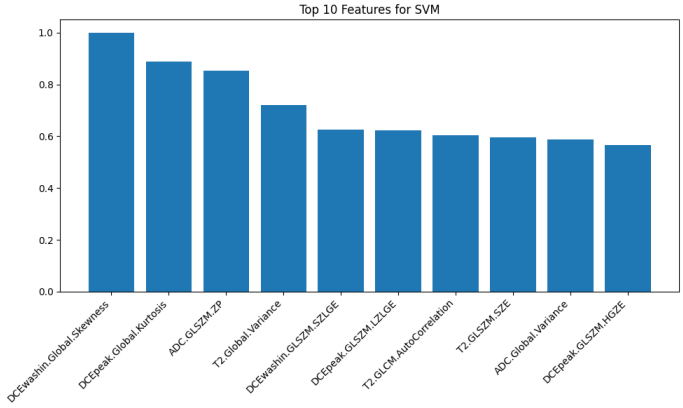


Fig. 7: Features ranked by normalized PFI for SVM

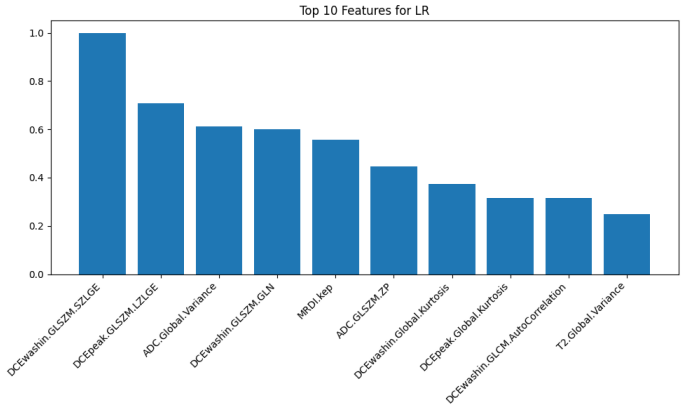


Fig. 8: Features ranked by normalized PFI for LR

REFERENCES

[1] W. Zhang, G. Cao, F. Wu, Y. Wang, Z. Liu, H. Hu, and K. Xu, "Global burden of prostate cancer and association with socioeconomic status, 1990–2019: A systematic analysis from the global burden of disease study," *Journal of Epidemiology and Global Health*, pp. 1–15, 2023.

[2] S. Sharma, J. Zapatero-Rodriguez, and R. O'Kennedy, "Prostate cancer diagnostics: Clinical challenges and the ongoing need for disruptive and effective diagnostic tools," *Biotechnology Advances*, vol. 35, no. 2, pp. 135–149, 2017.

[3] V. Kasivisvanathan, A. S. Rannikko, M. Borghi, V. Panebianco, L. A. Mynderse, M. H. Vaarala, A. Briganti, L. Budäus, G. Hellawell, R. G. Hindley *et al.*, "Mri-targeted or standard biopsy for prostate-cancer diagnosis," *New England Journal of Medicine*, vol. 378, no. 19, pp. 1767–1777, 2018.

[4] A. Stabile, F. Giganti, A. B. Rosenkrantz, S. S. Taneja, G. Villeirs, I. S. Gill, C. Allen, M. Emberton, C. M. Moore, and V. Kasivisvanathan, "Multiparametric mri for prostate cancer diagnosis: current status and future directions," *Nature reviews urology*, vol. 17, no. 1, pp. 41–61, 2020.

[5] A. Chatterjee, C. Harmath, and A. Oto, "New prostate mri techniques and sequences," *Abdominal Radiology*, vol. 45, pp. 4052–4062, 2020.

[6] C. Westra, V. Dialani, T. S. Mehta, and R. L. Eisenberg, "Using t2-weighted sequences to more accurately characterize breast masses seen on mri," *American Journal of Roentgenology*, vol. 202, no. 3, pp. W183–W190, 2014.

[7] M. A. Jacobs, R. Ouwerkerk, B. Kyle Petrowski, and K. J. Macura, "Diffusion weighted imaging with adc mapping and spectroscopy in prostate cancer," *Medicine, Baltimore*, vol. 1500, p. 21205.

[8] A. Lindgren, M. Anttila, O. Arponen, K. Hämäläinen, M. Könönen, R. Vanninen, and H. Sallinen, "Dynamic contrast-enhanced mri to characterize angiogenesis in primary epithelial ovarian cancer: An exploratory study," *European Journal of Radiology*, vol. 165, p. 110925, 2023.

[9] A. B. Rosenkrantz, L. A. Ginocchio, D. Cornfeld, A. T. Froemming, R. T. Gupta, B. Turkbey, A. C. Westphalen, J. S. Babb, and D. J. Margolis, "Interobserver reproducibility of the pi-rads version 2 lexicon: a multicenter study of six experienced prostate radiologists," *Radiology*, vol. 280, no. 3, pp. 793–804, 2016.

[10] W. Rogers, S. Thulasi Seetha, T. A. Refaee, R. I. Lieverse, R. W. Granzier, A. Ibrahim, S. A. Keek, S. Sanduleanu, S. P. Primakov, M. P. Beuque *et al.*, "Radiomics: from qualitative to quantitative imaging," *The British journal of radiology*, vol. 93, no. 1108, p. 20190948, 2020.

[11] C. Dinis Fernandes, A. Schaap, J. Kant, P. van Houdt, H. Wijkstra, E. Bekers, S. Linder, A. M. Bergman, U. van der Heide, M. Mischi *et al.*, "Radiogenomics analysis linking multiparametric mri and transcriptomics in prostate cancer," *Cancers*, vol. 15, no. 12, p. 3074, 2023.

[12] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels," Tech. Rep., 2010.

[13] B. Irving, "maskslic: regional superpixel generation with application to local pathology characterisation in medical images," *arXiv preprint arXiv:1606.09518*, 2016.

[14] M. Vallières, C. R. Freeman, S. R. Skamene, and I. El Naqa, "A radiomics model from joint fdg-pet and mri texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities," *Physics in Medicine & Biology*, vol. 60, no. 14, p. 5471, 2015.