Original Article

Radiomics based Machine Learning Models for Classification of Prostate Cancer Grade Groups from Multi Parametric MRI Images

Fatemeh Zandie[1], Mohammad Salehi[2], Asghar Maziar[1], Mohammad Reza Bayatiani[3], Reza Paydar[1]

[1]Department of Radiation Sciences, School of Allied Medicine, Iran University of Medical Sciences, [2]Department of Medical Physics, School of Medicine, Iran University of Medical Sciences, Tehran, [3]Department of Radiotherapy and Medical Physics, Faculty of Para Medicine, Arak University of Medical Sciences and Khansari Hospital, Arak, Iran

Abstract

Purpose: This study aimed to investigate the performance of multiparametric magnetic resonance imaging (mpMRI) radiomic feature-based machine learning (ML) models in classifying the Gleason grade group (GG) of prostate cancer. **Methods:** In this retrospective study, a total of 203 patients with histopathologically confirmed prostate cancer who underwent mpMRI before prostate biopsy were included. After manual segmentation, radiomic features (RFs) were extracted from T2-weighted, apparent diffusion coefficient, and high b-value diffusion-weighted magnetic resonance imaging (DWMRI). Patients were split into training sets and testing sets according to a ratio of 8:2. A pipeline considering combinations of two feature selection (FS) methods and six ML classifiers was developed and evaluated. The performance of models was assessed using the accuracy, sensitivity, precision, F1-measure, and the area under curve (AUC). **Results:** On high b-value DWMRI-derived features, a combination of FS method recursive feature elimination (RFE) and classifier random forest achieved the highest performance for classification of prostate cancer into five GGs, with 97.0% accuracy, 98.0% sensitivity, 98.0% precision, and 97.0% F1-measure. The method also achieved an average AUC for GG of 98%. **Conclusion:** Preoperative mpMRI radiomic analysis based on ML, as a noninvasive approach, showed good performance for classification of prostate cancer into five GGs. **Advances in Knowledge:** Herein, radiomic models based on preoperative mpMRI and ML were developed to classify prostate cancer into 5 GGs. Our study provides evidence that analysis of quantitative RFs extracted from high b-value DWMRI images based on a combination of FS method RFE and classifier random forest can be applied for multiclass grading of prostate cancer with an accuracy of 97.0%.

Keywords: *Gleason grading, machine learning, multiparametric magnetic resonance imaging, prostate cancer, radiomics*

Submitted: 03-Oct-2023 **Revised:** 24-Aug-2024 **Accepted:** 13-Sep-2024 **Published:** 03-Dec-2024

Introduction

Prostate cancer is the second most prevalent malignancy in the male population worldwide.[1] According to the statistics of the American Cancer Society, 268,490 new prostate cancer cases and 34,500 prostate cancer-related deaths would have been estimated in the United States in 2022.[2] The ability to accurately and timely identify the aggressiveness risk of prostate cancer can improve the selection of the most suitable treatment for these patients; therefore, it can be resulted in decreasing prostate cancer morbidity and mortality.[3] Therefore, it is crucial to preoperatively predict the grade of prostate cancer for treatment decision-making.

Currently, the standard approach for grading of prostate cancer is transrectal ultrasound (TRUS)-guided prostate biopsy.[4] The Gleason score (GS) – identified through the histopathologic analysis of biopsy samples – is the best predictor of the aggressiveness of prostate cancer.[4] In 2014, a new grading system approach was proposed by the International Society of Urological Pathology Consensus Conference on Gleason Grading of Prostatic Carcinoma.[5] A five-grade group (GG) system was adopted based on the GS: GG 1 (GS $\leq$ 6), GG 2 (GS 3 + 4 = 7), GG 3 (GS 4 + 3 = 7), GG 4 (GS 8), and GG 5 (GS 9–10).[5] However, it has been reported that the standard TRUS biopsy for identifying prostate cancer has a poor sensitivity.[6] Moreover, the Gleason grading of histopathology images is an expensive

How to cite this article: Zandie F, Salehi M, Maziar A, Bayatiani MR, Paydar R. Radiomics based machine learning models for classification of prostate cancer grade groups from multi parametric MRI images. J Med Signals Sens 2024;14:33.

Address for correspondence:
Dr. Reza Paydar, Department of Radiation Sciences, School of Allied Medicine, Iran University of Medical Sciences, Tehran, Iran.
E-mail: paydar.r@iums.ac.ir

Access this article online

Website: www.jmssjournal.net

DOI: 10.4103/jmss.jmss_47_23

Quick Response Code:

and time-intensive task and suffers from high inter- and intra-observer variability.[7] Besides, TRUS prostate biopsy is associated with possible side effects including, prostatitis, urinary tract infections, bleeding, pain, and sepsis.[8] Therefore, it is desirable to develop a noninvasive and accurate approach for the preoperative prediction of grade of prostate cancer.

Over the past decade, multiparametric magnetic resonance imaging (mpMRI) has been increasingly utilized in the evaluation, localization, and staging of prostate cancer.[9,10] Nowadays, mpMRI has been established as the imaging gold standard for prostate cancer.[11] Additionally, the Prostate Imaging Reporting and Data System (PI-RADS) is integrated into the clinical evaluation of prostate cancer to standardize the examination process and unify the interpretation of prostate MRI results.[12,13] Interpretation of mpMRI according to the PI-RADS has shown encouraging results in predicting the grade of prostate cancer, with high sensitivity.[14,15] Nevertheless, mpMRI interpretation is a challenging task and prone to inter- and intra-rater variability and is therefore subjective.[16]

To tackle the aforementioned challenges regarding mpMRI interpretation, quantitative imaging parameters extracted from T2-weighted (T2W) and diffusion-weighted images (DWIs) were identified. Researchers have shown that T2W and DW signal intensities can be associated with prostate cancer aggressiveness.[17,18] Furthermore, a correlation between T2W and DWI (including apparent diffusion coefficient [ADC] map) signal intensities with GSs has been found.[19] However, there is considerable overlap in the ranges of ADC values for various GSs, and therefore, such an issue limits their use in distinguishing between these GSs.[18]

In the last decade, radiomics has emerged as a novel tool that can extract a huge number of quantitative features (e.g., texture, size, and shape) from clinical images.[20] In other words, radiomics applies the advanced mathematical algorithms to extract quantitative and mineable imaging features from radiological images in a cost-effective and high-throughput approach, resulting in accurate tumor detection and personalized medical decision-making.[20] Recently, studies applied MRI-based radiomics features (RFs) to differentiate prostate cancer from benign prostate tissue and to evaluate prostate cancer aggressiveness.[21,22] Machine learning (ML), a subset of artificial intelligence, is capable of extracting meaningful patterns from data without being explicitly programmed.[23-26] In recent years, there has been increasing interest in applying ML methods to radiomic studies to improve the assessment of prostate cancer.[27] ML-based mpMRI radiomics as an objective tool has been shown to be helpful in evaluating the grade of prostate cancer.[28,29]

To the best of our knowledge, a limited number of studies have attempted to perform binary classification of prostate cancer aggressiveness using mpMRI-based radiomics using a small number of classification methods, mostly support vector machine (SVM).[27,30] Moreover, ML model performance evaluation was performed using the area under the receiver operating characteristic (ROC) area under curve (AUC) score only. Therefore, the purpose of the current study was to develop an ML-based framework to identify the best-performing classifier for preoperatively identifying the grade of prostate cancer based on mpMRI-derived RFs. Herein, six supervised ML models were used for the grading of prostate cancer into 5 GGs by mpMRI RFs. Furthermore, we comprehensively assessed the performance of candidate classifiers using accuracy, sensitivity, precision, and F1-score in addition to the AUC score.

Materials and Methods

Patient cohorts

Between June 2017 and January 2020, a total of 203 prostate cancer patients were enrolled in this single-institution, retrospective study. We included patients with histopathologic confirmed prostate cancer, mpMRI of the prostate, and TRUS-MRI fusion-guided prostate biopsy within 6 weeks of mpMRI. Patients with a previous history of prostatic surgery or hormonal therapy before the MRI scans, poor quality of the MR images (e.g., owing to severe artifacts from hip prostheses), confirmed diagnosis of a tumor other than prostate cancer, and incomplete clinical data were excluded from this study. For patients with multiple tumor sites identified on mpMRI, the site with the dominant lesion was used for analysis. The included patients were randomly split into training sets (n = 162) and test sets (n = 41) using a ratio of 8:2. Figure 1 illustrates the flowchart of the present study.

Image acquisition

All MRI examinations were acquired using two different models of Siemens 3T MRI scanner, MAGNETOM Trio and MAGNETOM Skyra (Siemens Healthineers, Erlangen, Germany), using a body transmit coil. Herein, mpMRI sequences, including axial T2W, ADC, and high b-value diffusion-weighted MRI (DWMRI), were chosen to extract RFs. ADC maps were automatically reconstructed on a designated workstation.

Segmentation

Segmentation is the crucial first step in the radiomic pipeline.[31] Highly distinctive RFs were obtained from the delineated region of interest in two-dimensional or of the volume of interest (VOI) in three-dimensional (3D) approaches. Herein, the mpMRI examinations were retrieved from the Picture Archiving and Communication System. Then, all images were imported into 3D slicer, a free open-source software platform for biomedical imaging research (https://slicer.org). A 3D VOI of clinical lesions was manually delineated by a board-certified radiologist (with 15 years of experience in prostate imaging)

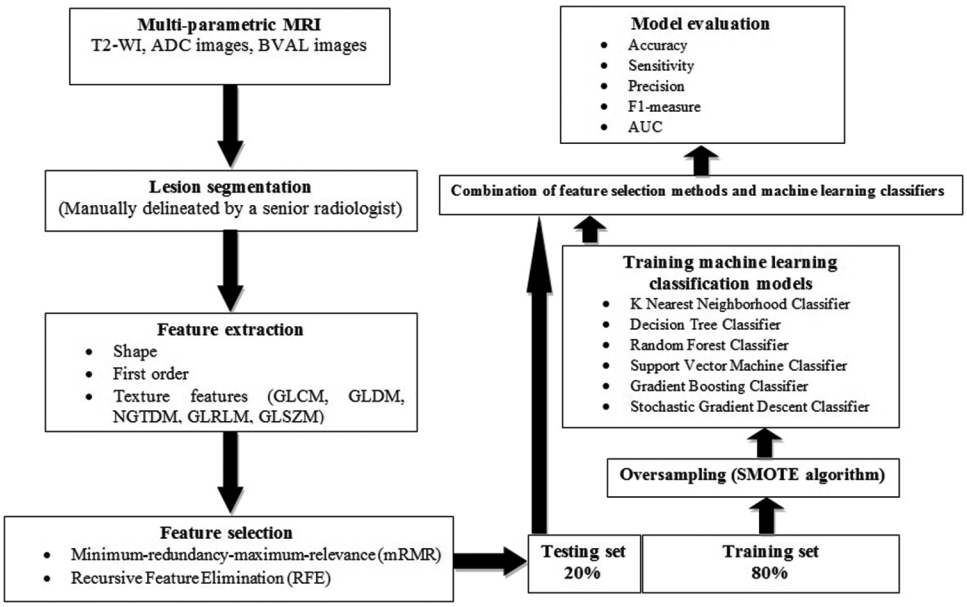

Figure 1: Flow diagram of our radiomics-based machine learning framework. T2W1 – T2-weighted; ADC – Apparent diffusion coefficient; AUC – Area under curve; GLCM – Gray-level co-occurrence matrix; GLDM – Gray-level dependence matrix; NGTDM – Neighboring gray-tone difference matrix; GLRLM – Gray-level run-length matrix; GLSZM – Gray-level size-zone matrix; SMOTE – Synthetic Minority Over-Sampling Technique; mRMR – Minimum redundancy maximum relevance; RFE – Recursive feature elimination

Table 1: Features extracted

Class	Feature
3D shape-based features	Mesh volume, major axis length, minor axis length, flatness, maximum 2D diameter, voxel volume, maximum 3D diameter, surface ratio volume
First-order features	Energy, total energy, minimum, maximum, mean, median, IQR, root mean squared, skewness, kurtosis, entropy, robust mean absolute deviation
GLCM	Difference entropy, difference variance, sum entropy, cluster tendency, correlation, maximum probability
GLSZM	LALGLE, SAHGLE, zone variance, SZN, LAHGLE
GLRLM	RLN, SRHGLE, LRHGLE, RLNN, LAHGLE, run entropy, LRLGLE
NGTDM	Busyness, contrast, complexity, DNN, HGLE
GLDM	GLN, GLNN, dependence entropy, LDE, LGLE

GLCM – Gray-level co-occurrence matrix; 2D – Two-dimensional; 3D – Three-dimensional; IQR – interquartile range; GLSZM – Gray-level size-zone matrix; LALGLE – Large area low gray level emphasis; SAHGLE – Small area high gray level emphasis; SZN – Size-zone nonuniformity; LAHGLE – Large area high gray level emphasis; GLRLM – Gray-level run-length matrix; RLN – Run-length nonuniformity; SRHGLE – Short run high gray level emphasis; LRHGLE – Long run high gray level emphasis; RLNN – RLN normalized; LRLGLE – Long run low gray level emphasis; NGTDM – Neighboring gray-tone difference matrix; DNN – Dependence nonuniformity normalized; HGLE – High gray level emphasis; GLDM – Gray-level dependence matrix; GLN – Gray level nonuniformity; GLNN – GLN normalized; LDE – Large dependence emphasis; LGLE – Low gray level emphasis

on the axial T2W, ADC, and high b-value DWMRI images on each slice, as depicted in Figure 2.

Feature extraction and selection

All RFs were extracted from the VOIs on both the T2W and ADC images of each patient using Pyradiomics v. 3.0., an open-source Python package.[32] The extracted RFs were divided into three classes:[32] (1) 8 shape-based features; (2) 12 first-order features; and (3) 28 texture features, including the gray-level co-occurrence matrix (GLCM), gray-level dependence matrix, neighboring gray-tone difference matrix, gray-level run-length matrix, and gray-level size-zone matrix, as shown in Table 1. In total, 113 RFs were calculated for every patient on both the T2W and ADC images. In this study, two different feature selection (FS) methods were applied to decrease the number of features used for training. The minimum redundancy maximum relevance (mRMR) was conducted to select the top-ranked RFs and eliminate the redundant and irrelevant RFs.[33] Furthermore, recursive feature elimination (RFE) was used to eliminate RFs that are redundant.[34] The RF dimension after utilizing both the methods was decreased to ten features per each MR sequence (i.e. T2, ADC, and high b-value DWMRI).

To tackle the class imbalance of the training samples, we applied the Synthetic Minority Over-Sampling Technique (SMOTE) algorithm, an oversampling method that artificially adds the minor class instances to match the number of training samples of the major class instances.[35]

Classification prediction

After FS, ML classifiers were applied to construct a predictive model that most accurately classifies different grades of prostate cancer (grade 1/2/3/4/5) based on mpMRI-derived

RFs. Initially, we normalized the input data before building ML models.[36] Herein, 6 well-supervised classification models were used as classifiers, as outlined in Table 2. Our ML classifiers combined with two FS methods (i.e., mRMR and RFE) were trained and validated on training dataset and tested on the test dataset. Our ML classifiers were implemented in the Python Scikit-learn ML package (version 0.20.4).

Performance evaluation

The performance of our RF-based ML classifiers was assessed using the test cohort. A variety of metrics, including accuracy, weighted average-based sensitivity, precision, and F1-measure, were calculated to evaluate the performance of our ML models for multiclass grading of prostate cancer. Furthermore, we outlined confusion matrix and ROC curve for our best-performing model.

Results

Table 3 summarizes the detailed clinical characteristics of study participants. A total of 203 patients with prostate cancer disease, including 59 Grade 1, 71 Grade 2, 43 Grade 3, 16 Grade 4, and 14 Grade 5, were included according to the inclusion criteria. The mean ± standard deviation (range) age of patients in the study cohort was 65.6 ± 5.6 (52–80) years. As shown in Table 3, 41.87%, 38.43%, and 40.70% of lesions were localized in the peripheral zone, transitional zone, and anterior stroma, respectively.

Table 2: Machine learning classification algorithms

Algorithm	Abbreviation
K-nearest neighborhood classifier	KNNC
Decision tree classifier	DTC
Random forest classifier	RFC
Support vector machine classifier	SVMC
Gradient boosting classifier	GBC
Stochastic gradient descent classifier	SGDC

We further used six ML classification methods to each of the two FS methods to classify prostate cancer to different GGs. Therefore, for each group of MRI images, 12 combinations of 2 FS methods and 6 ML classifiers were systematically assessed. Hence, the performance of 36 proposed ML models was assessed on the test set.

Table 4 outlines the performance of the different FS methods and ML classifiers for classification of prostate cancer GGs from each group of MRI images on the test set using the accuracy, sensitivity, specificity, and F1-score. As observable in Table 4, among all of the selected groups, the group with high b-value DWMRI features achieved the highest performance with a combination of FS method RFE and classifier RFC with an accuracy of 97.0%, a sensitivity of 98.0%, a precision of 98.0%, and a F1-score of 97.0%. Figure 3 shows the performance of best-performing pipeline for five-class classification in the form of confusion matrix. ROC curve for best-performing model on test set is depicted in Figure 4. On the T2W features, the combination of

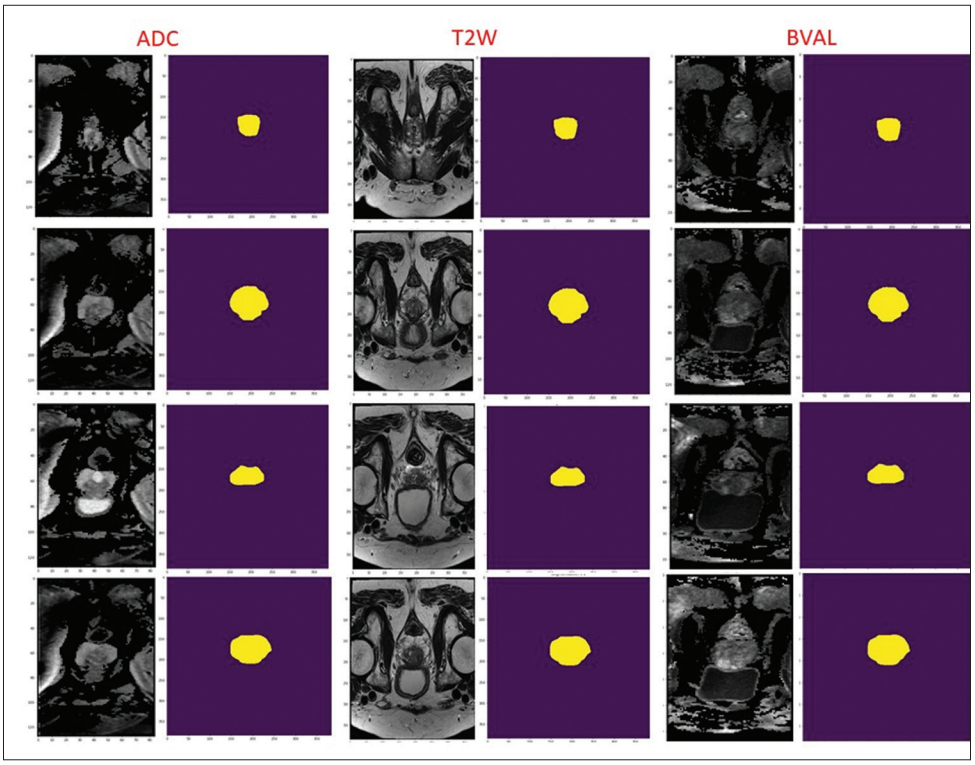

Figure 2: An example segmentation of prostate lesion in a representative patient. T2W – T2-weighted; ADC – Apparent diffusion coefficient

Table 3: The clinical characteristics of study participants

	n=203 patients				
	Grade 1	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**
Number of patients	59	71	43	16	14
Age (years), mean±SD	65.9±6.0	66.9±5.9	65.1±4.6	61.3±4.2	64.35±3.5
PSA (ng/mL), median (range)	7.0 (2.8–17.5)	8.6 (2–26)	10.9 (2.3–26)	20 (2.5–35)	30.5 (15–50.6)
Location					
PZ	23	32	16	9	5
AS	31	21	12	6	8
TZ	5	18	15	1	1

PZ – Peripheral zone; AS – Anterior stroma; TZ – Transitional zone; PSA – Prostate specific antigen; SD – Standard deviation

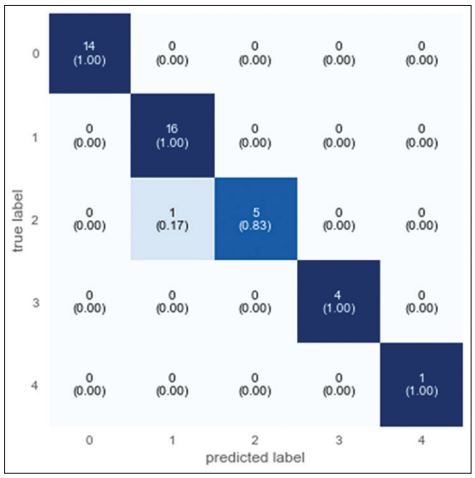

Figure 3: Confusion matrix for best-performing model

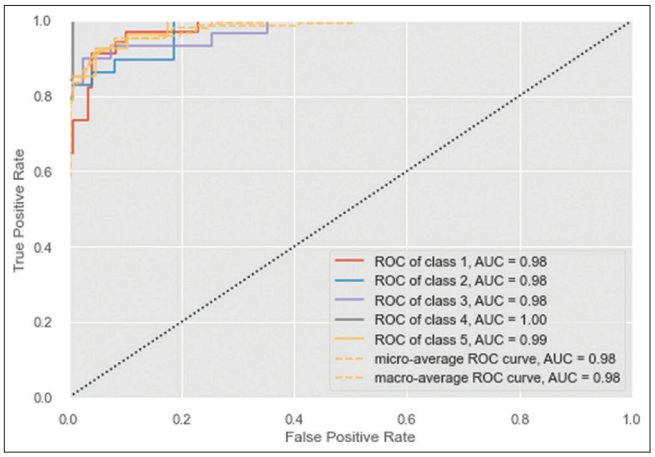

Figure 4: Receiver operating characteristic curve for best-performing model. ROC – Receiver operating characteristic; AUC – Area under curve

FS method RFE and classifier RFC achieved the best performance with accuracy = 87.0%, sensitivity = 88.0%, precision = 88.0%, and F1-score = 87.0%. On the ADC features, the best performance was achieved when the RFE as FS method combined with GBC as ML classifier, showing an accuracy of 87.0%, a sensitivity of 90.0%, a precision of 90.0%, and an F1-score of 90.0%, as shown in Table 4.

Discussion

A preoperatively accurate identifying the grade of prostate cancer is very important for clinical decision-making and selection of treatment modalities. Although TRUS-guided prostate biopsy is a popular tool in predicting prostate cancer grade invasiveness, it suffers from different side effects that restrict its clinical use.[8] Hence, mpMRI has been introduced as a noninvasive, useful tool for identifying prostate cancer grade; however, there are variations in inter-observer interpretation of mpMRI images, depending on personal knowledge and reader experience.[14,37] ML based on the radiomic approach is capable of extracting a huge number of quantitative imaging features beyond those attained through visual analysis by clinicians, providing an objective prediction model.[10]

The main purpose of the current study was to develop an ML-based framework for multiclass classification of prostate cancer into 5 GGs based on preoperative mpMRI RFs. Our findings showed that radiomic model incorporating high b-value DWMRI images had the highest performance in classification of prostate cancer GGs, suggesting that the clinical application of radiomics is a promising noninvasive approach in terms of the preoperative prediction of the grade of prostate cancer. Moreover, we applied various combinations of two FS methods and six ML classifiers on T2W, ADC, and B-Value (BVAL) images separately to build radiomic signatures for multiclass grading of prostate cancer. From our data, it can be seen that the algorithm combination of FS method RFE and ML classifier RFC achieved the highest performance for multiclass classification of prostate cancer between 5 GGs with satisfying accuracy, sensitivity, precision, and F1-score of 97.0%, 98.0%, 98.0%, and 97.0%, respectively, based on high b-value DWMRI features. As a consequence, BVAL images can reflect the heterogeneity of tumors better and that RFs can be applied to discriminate small signal discrepancies in prostate tumors.

Our study demonstrated the efficacy of using mpMRI radiomic feature-based ML models to classify Gleason GGs of prostate cancer, achieving an impressive accuracy of 97.0%, sensitivity of 98.0%, precision of 98.0%, F1-measure of 97.0%, and an AUC of 98%. These results underscore the robustness of our approach in providing a

Table 4: The performance of the different feature selection methods and machine learning classifiers on the test dataset for multiclass grading of prostate cancer

Image	Algorithm	Precision	Sensitivity	F1-score	Accuracy
T2W	mRMR				
	KNNC	0.81	0.76	0.77	0.75
	DTC	0.81	0.73	0.74	0.73
	RFC	0.83	0.76	0.77	0.74
	SVMC	0.60	0.56	0.55	0.56
	GBC	0.75	0.71	0.70	0.70
	SGDC	0.31	0.29	0.29	0.30
	RFE				
	KNNC	0.77	0.76	0.76	0.75
	DTC	0.74	0.71	0.66	0.70
	RFC	0.88	0.88	0.87	0.87
	SVMC	0.87	0.85	0.85	0.85
	GBC	0.86	0.83	0.82	0.83
	SGDC	0.65	0.61	0.60	0.61
ADC	mRMR				
	KNNC	0.58	0.51	0.54	0.51
	DTC	0.72	0.63	0.65	0.63
	RFC	0.68	0.59	0.61	0.58
	SVMC	0.52	0.27	0.33	0.27
	GBC	0.79	0.68	0.73	0.68
	SGDC	0.42	0.17	0.23	0.17
	RFE				
	KNNC	0.48	0.47	0.44	0.46
	DTC	0.86	0.86	0.86	0.85
	RFC	0.87	0.87	0.87	0.86
	SVMC	0.47	0.46	0.44	0.46
	GBC	0.90	0.90	0.90	0.87
	SGDC	0.31	0.38	0.34	0.38
BVAL	mRMR				
	KNNC	0.72	0.71	0.70	0.70
	DTC	0.92	0.91	0.91	0.91
	RFC	0.96	0.95	0.95	0.95
	SVMC	0.91	0.90	0.90	0.90
	GBC	0.96	0.95	0.95	0.95
	SGDC	0.63	0.54	0.57	0.53
	RFE				
	KNNC	0.82	0.80	0.81	0.80
	DTC	0.75	0.59	0.51	0.58
	RFC	0.98	0.98	0.97	0.97
	SVMC	0.80	0.80	0.80	0.80
	GBC	0.95	0.93	0.94	0.92
	SGDC	0.83	0.73	0.76	0.73

mRMR – Minimum redundancy maximum relevance; T2W – T2-weighted; KNNC – K-nearest neighborhood classifier; DTC – Decision tree classifier; RFC – Random forest classifier; SVMC – Support vector machine classifier; GBC – Gradient boosting classifier; SGDC – Stochastic gradient descent classifier; RPE – Recursive feature elimination; ADC – Apparent diffusion coefficient; RFE – Recursive feature elimination, BVAL – B-Value

noninvasive and detailed assessment of prostate cancer aggressiveness. In contrast, the study employing a deep learning-based AI workflow for automated EPE grading from MRI reported moderate performance metrics, with a balanced accuracy of 0.39, ROC AUCs ranging from 0.55 to 0.70 for different EPE grades, and an overall accuracy of 0.72.[38] The AI model's sensitivity (0.67) and specificity (0.73) also indicated potential utility but highlighted the need for further refinement. The superior performance of our radiomics-based model in classifying Gleason GGs can significantly aid in personalized treatment planning by providing critical insights into the cancer's aggressiveness. In comparison, accurate preoperative assessment of EPE, as targeted by the deep learning model, is crucial for informing surgical strategies and improving postoperative outcomes, though the current performance metrics suggest room for improvement to achieve clinical utility comparable to our model. Methodologically, our study emphasized the importance of manual segmentation and FS, utilizing RFE and random forest classifiers to achieve high accuracy. Conversely, the deep learning study leveraged automated segmentation to extract relevant features for EPE grading, reflecting a different yet complementary approach to prostate cancer assessment.

Integrating our detailed radiomic feature analysis with the deep learning capabilities from the EPE grading study could potentially enhance both Gleason grading and EPE assessment, leading to more accurate and clinically useful diagnostic tools. This integrative approach could harness the strengths of both methodologies, paving the way for advancements in prostate cancer diagnosis and treatment planning.

Most existing literature on methods of prostate cancer grading focus on binary classification of low-grade (GG = 1) and high-grade (GG ≥ 2) prostate cancer.[22,39-44] Furthermore, few studies applied ML classifiers in predicting the grade of prostate cancer from MRI,[39,40] whereas most of the studies used statistical analysis to show the correlation between features and prostate cancer aggressiveness.[22,41,42] In a study, Nezzo *et al*. applied statistical analysis using ANOVA test to compare mean diffusivity (MD) values obtained from diffusion tensor imaging (DTI) between different prostate cancer GGs.[41] The results showed that MD values derived from DTI acquired at high B values could be applied for discriminating between prostate cancer of GG 1 and 2 and GG ≥ 3.[41] Citak-Er *et al*. investigated linear discriminant analysis (LDA) and SVM classifiers to predict final GS preoperatively using mpMRI and clinical factors.[39] They observed that SVM classifier achieved a slightly higher sensitivity but a lower specificity in comparison with LDA method on their private dataset.[39] Using a pioneering method Fehr *et al*. classified low-grade and high-grade prostate cancer.[40] Their method applied GLCM texture features and RFE-SVM classifier and SMOTE-based oversampling of instances in minority classes and achieved an accuracy of 93.0% for binary classification (GG1 vs. GG ≥ 2).[40] Jensen *et al*. utilized image histogram

and textural features in combination with K-nearest neighborhood classifier to perform a binary classification of GG1, GG2, GG (1 + 2), GG 3, and GG (4 + 5) and achieved AUC of 0.85, 0.89, 0.83, 0.94, and 0.86, respectively.[43] In a study, Abraham and Nair applied high-level features extracted using sparse auto-encoders and the random forest (RF) classifier for discriminating clinically significant from clinically insignificant prostate cancer, achieving an accuracy of 93.65%.[44] They also used the SMOTE and Weka resample algorithm and adaptive synthetic sampling approach to address the class-imbalance problem.[44] In another study by the same investigators, a stacked sparse auto-encoder (SSAE) framework was used to classify prostate cancer into 5 GGs.[45] Using a SSAE, as a deep learning architecture, high-level features were extracted and classified using a Softmax classifier. Their framework achieved a positive predictive value of 80.26% and an accuracy of 47.3% in predicting prostate cancer with GG >1.[45] Herein, we applied radiomics-based ML models to perform multiclass classification of prostate cancer GGs, which achieved good accuracy on test cohort. Of note, multiclass classification is more difficult and challenging than a simple binary classification. It is worthwhile to mention that there was an unequal distribution of classes in our dataset. Therefore, we applied SMOTE algorithms, an oversampling technique, to address class imbalance problem. In a previous study, it has also been shown that the SMOTE algorithm was useful to solve the class imbalance problem for discriminating GS = 6 from GS ≥ 7.[40]

Our study has some limitations. In this study, our dataset had a small number of subjects. Herein, any data augmentation method was used. A relatively small patient data size may trigger overfitting of the classifiers. We will apply oversampling techniques along with data augmentation techniques to solve the issue of limited data and class imbalance, as well as overfitting for future studies. Our data were collected from two kinds of scanner from a single MRI vendor, achieved using a similar imaging protocol. It is worthwhile to mention that RFs need to be reliable across various scanners and imaging parameters. Therefore, further multicenter, multi-MRI vendor studies are warranted on the purposes of our study.

Conclusion

Our data suggested that mpMRI radiomic models based on optimal combinations of FS methods and ML classifiers have great potentials in predicting the grade of prostate cancer. Our results showed that the combination of FS method RFE and classifier RFC is particularly promising for multiclass classification of prostate cancer into 5 GGs on the high b-value DWMRI-extracted RFs. The proposed ML-based radiomic model, as a noninvasive approach, can assist clinicians in reporting prostate cancer grade and facilitate clinical decision-making. Our framework may be of fundamental importance to decrease the number of biopsies. This study demonstrates the potential of multiparametric MRI radiomics and ML for accurately classifying prostate cancer Gleason GGs. Our model achieved a high classification accuracy of 97% using high b-value DWMRI features. By leveraging preoperative mpMRI, we offer a noninvasive approach to prostate cancer grading, potentially improving patient care. While the study shows promise, limitations include retrospective design, moderate sample size, and the complexity of the methodology. Future research should address these to enhance generalizability and clinical impact.

Acknowledgment

This study was granted by the research chancellor of Iran University of Medical Sciences, Tehran, Iran.

Financial support and sponsorship

This study has received funding from the Iran University of Medical Sciences, Tehran, Iran, and the grant number is 99-3-15-19520.

Conflicts of interest

There are no conflicts of interest.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209-49.
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin 2022;72:7-33.
3. Chang AJ, Autio KA, Roach M 3rd, Scher HI. High-risk prostate cancer-classification and therapy. Nat Rev Clin Oncol 2014;11:308-23.
4. Hwang I, Kim SY, Cho JY, Lee MS, Kim SH. The diagnostic ability of an additional midline peripheral zone biopsy in transrectal ultrasonography-guided 12-core prostate biopsy to detect midline prostate cancer. Ultrasonography 2016;35:61-8.
5. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA, *et al.* The 2014 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma: Definition of grading patterns and proposal for a new grading system. Am J Surg Pathol 2016;40:244-52.
6. Epstein JI, Feng Z, Trock BJ, Pierorazio PM. Upgrading and downgrading of prostate cancer from biopsy to radical prostatectomy: Incidence and predictive factors using the modified Gleason grading system and factoring in tertiary grades. Eur Urol 2012;61:1019-24.
7. Karimi D, Nir G, Fazli L, Black PC, Goldenberg L, Salcudean SE. Deep learning-based Gleason grading of prostate cancer from histopathology images-role of multiscale decision aggregation and data augmentation. IEEE J Biomed Health Inform 2020;24:1413-26.
8. Loeb S, Vellekoop A, Ahmed HU, Catto J, Emberton M, Nam R, *et al.* Systematic review of complications of prostate biopsy. Eur Urol 2013;64:876-92.
9. Wang Q, Li H, Yan X, Wu CJ, Liu XS, Shi HB, *et al.* Histogram analysis of diffusion kurtosis magnetic resonance imaging in differentiation of pathologic Gleason grade of prostate cancer. Urol Oncol 2015;33:337.e15-24.

10. Fütterer JJ, Briganti A, De Visschere P, Emberton M, Giannarini G, Kirkham A, *et al.* Can clinically significant prostate cancer be detected with multiparametric magnetic resonance imaging? A systematic review of the literature. Eur Urol 2015;68:1045-53.
11. Regmi SK, Sathianathen N, Stout TE, Konety BR. MRI/PET imaging in elevated PSA and localized prostate cancer: A narrative review. Transl Androl Urol 2021;10:3117-29.
12. Barentsz JO, Richenberg J, Clements R, Choyke P, Verma S, Villeirs G, *et al.* ESUR prostate MR guidelines 2012. Eur Radiol 2012;22:746-57.
13. Michaely HJ, Aringhieri G, Cioni D, Neri E. Current value of biparametric prostate MRI with machine-learning or deep-learning in the detection, grading, and characterization of prostate cancer: A systematic review. Diagnostics (Basel) 2022;12:799.
14. Chen F, Cen S, Palmer S. Application of prostate imaging reporting and data system version 2 (PI-RADS v2): Interobserver agreement and positive predictive value for localization of intermediate- and high-grade prostate cancers on multiparametric magnetic resonance imaging. Acad Radiol 2017;24:1101-6.
15. Greer MD, Brown AM, Shih JH, Summers RM, Marko J, Law YM, *et al.* Accuracy and agreement of PIRADSv2 for prostate cancer mpMRI: A multireader study. J Magn Reson Imaging 2017;45:579-85.
16. Ahmed HU, El-Shater Bosaily A, Brown LC, Gabe R, Kaplan R, Parmar MK, *et al.* Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): A paired validating confirmatory study. Lancet 2017;389:815-22.
17. Donati OF, Mazaheri Y, Afaq A, Vargas HA, Zheng J, Moskowitz CS, *et al.* Prostate cancer aggressiveness: Assessment with whole-lesion histogram analysis of the apparent diffusion coefficient. Radiology 2014;271:143-52.
18. Jung SI, Donati OF, Vargas HA, Goldman D, Hricak H, Akin O. Transition zone prostate cancer: Incremental value of diffusion-weighted endorectal MR imaging in tumor detection and assessment of aggressiveness. Radiology 2013;269:493-503.
19. Donati OF, Afaq A, Vargas HA, Mazaheri Y, Zheng J, Moskowitz CS, *et al.* Prostate MRI: Evaluating tumor volume and apparent diffusion coefficient as surrogate biomarkers for predicting tumor Gleason score. Clin Cancer Res 2014;20:3705-11.
20. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. Radiology 2016;278:563-77.
21. Xu M, Fang M, Zou J, Yang S, Yu D, Zhong L, *et al.* Using biparametric MRI radiomics signature to differentiate between benign and malignant prostate lesions. Eur J Radiol 2019;114:38-44.
22. Wibmer A, Hricak H, Gondo T, Matsumoto K, Veeraraghavan H, Fehr D, *et al.* Haralick texture analysis of prostate MRI: Utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. Eur Radiol 2015;25:2840-50.
23. Salehi M, Mohammadi R, Ghaffari H, Sadighi N, Reiazi R. Automated detection of pneumonia cases using deep transfer learning with paediatric chest X-ray images. Br J Radiol 2021;94:20201263. doi: 10.1259/bjr.20201263.
24. Ghaffari H, Tavakoli H, Pirzad Jahromi G. Deep transfer learning-based fully automated detection and classification of Alzheimer's disease on brain MRI. Br J Radiol 2022;95:20211253. doi: 10.1259/bjr.20211253.
25. Showkatian E, Salehi M, Ghaffari H, Reiazi R, Sadighi N. Deep learning-based automatic detection of tuberculosis disease in chest X-ray images. Pol J Radiol 2022;87:e118-24.
26. Mousavie Anijdan SH, Reiazi R, Fallah Tafti H, Moslemi D, Moghadamnia AA, Paydar R. Application of Radiomics in Radiotherapy: Challenges and Future Prospects. Journal of Babol University of Medical Sciences. 2022;24:127-40.
27. Smith CP, Czarniecki M, Mehralivand S, Stoyanova R, Choyke PL, Harmon S, *et al.* Radiomics and radiogenomics of prostate cancer. Abdom Radiol (NY) 2019;44:2021-9.
28. Castillo TJ, Starmans MP, Arif M, Niessen WJ, Klein S, Bangma CH, *et al*. A multi-center, multi-vendor study to evaluate the generalizability of a radiomics model for classifying prostate cancer: High grade versus. Low grade. Diagnostics (Basel) 2021;11:369.
29. Transin S, Souchon R, Gonindard-Melodelima C, de Rozario R, Walker P, Funes de la Vega M, *et al.* Computer-aided diagnosis system for characterizing ISUP grade ≥2 prostate cancers at multiparametric MRI: A cross-vendor evaluation. Diagn Interv Imaging 2019;100:801-11.
30. Stoyanova R, Takhar M, Tschudi Y, Ford JC, Solórzano G, Erho N, *et al.* Prostate cancer radiomics and the promise of radiogenomics. Transl Cancer Res 2016;5:432-47.
31. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging-"how-to" guide and critical reflection. Insights Imaging 2020;11:91.
32. van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, *et al.* Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017;77:e104-7.
33. Peng H, Long F, Ding C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 2005;27:1226-38.
34. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn 2002;46:389-422.
35. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321-57.
36. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJ. Machine learning methods for quantitative radiomic biomarkers. Sci Rep 2015;5:13087.
37. Thompson JE, van Leeuwen PJ, Moses D, Shnier R, Brenner P, Delprado W, *et al.* The diagnostic performance of multiparametric magnetic resonance imaging to detect significant prostate cancer. J Urol 2016;195:1428-35.
38. Simon BD, Merriman KM, Harmon SA, Tetreault J, Yilmaz EC, Blake Z, *et al*. Automated Detection and Grading of Extraprostatic Extension of Prostate Cancer at MRI via Cascaded Deep Learning and Random Forest Classification. Acad Radiol 2024:S1076-6332(24)00220-4. doi: 10.1016/j.acra.2024.04.011.
39. Citak-Er F, Vural M, Acar O, Esen T, Onay A, Ozturk-Isik E. Final Gleason score prediction using discriminant analysis and support vector machine based on preoperative multiparametric MR imaging of prostate cancer at 3T. Biomed Res Int 2014;2014:690787.
40. Fehr D, Veeraraghavan H, Wibmer A, Gondo T, Matsumoto K, Vargas HA, *et al.* Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. Proc Natl Acad Sci U S A 2015;112:E6265-73.
41. Nezzo M, Di Trani MG, Caporale A, Miano R, Mauriello A, Bove P, *et al.* Mean diffusivity discriminates between prostate cancer with grade group 1 and 2 and grade groups equal to or greater than 3. Eur J Radiol 2016;85:1794-801.
42. Nketiah G, Elschot M, Kim E, Teruel JR, Scheenen TW, Bathen TF, *et al.* T2-weighted MRI-derived textural features

reflect prostate cancer aggressiveness: Preliminary results. Eur Radiol 2017;27:3050-9.
43. Jensen C, Carl J, Boesen L, Langkilde NC, Østergaard LR. Assessment of prostate cancer prognostic Gleason grade group using zonal-specific features extracted from biparametric MRI using a KNN classifier. J Appl Clin Med Phys 2019;20:146-53.
44. Abraham B, Nair MS. Computer-aided diagnosis of clinically significant prostate cancer from MRI images using sparse autoencoder and random forest classifier. Biocybern Biomed Eng 2018;38:733-44.
45. Abraham B, Nair MS. Computer-aided classification of prostate cancer grade groups from MRI images using texture features and stacked sparse autoencoder. Comput Med Imaging Graph 2018;69:60-8.