Classification of retinoblastoma-1 gene mutation with machine learning-based models in bladder cancer

Okan İnce a,*, Hülya Yıldız a, Tanju Kismet a, Şükrü Mehmet Ertürk b, Hakan Önder a

a Health Sciences University Prof. Dr. Cemil Tascioglu City Hospital, Department of Radiology, Turkey
b Istanbul University Istanbul Medical Faculty, Department of Radiology, Turkey

ARTICLE INFO

Keywords:
Bladder
Computer applications-detection
Diagnosis
CT

ABSTRACT

Purpose: This study aims to evaluate the potential of machine learning algorithms built with radiomics features from computed tomography urography (CTU) images that classify RB1 gene mutation status in bladder cancer.

Method: The study enrolled CTU images of 18 patients with and 54 without RB1 mutation from a public database. Image and data preprocessing were performed after data augmentation. Feature selection steps were consisted of filter and wrapper methods. Pearson’s correlation analysis was the filter, and a wrapper-based sequential feature selection algorithm was the wrapper. Models with XGBoost, Random Forest (RF), and k-Nearest Neighbors (kNN) algorithms were developed. Performance metrics of the models were calculated. Models’ performances were compared by using Friedman’s test.

Results: 8 features were selected from 851 total extracted features. Accuracy, sensitivity, specificity, precision, recall, F1 measure and AUC were 84%, 80%, 88%, 86%, 80%, 0.83 and 0.84, for XGBoost; 72%, 80%, 65%, 67%, 80%, 0.73 and 0.72 for RF; 66%, 53%, 76%, 67%, 53%, 0.60 and 0.65 for kNN, respectively. XGBoost model had outperformed kNN model in Friedman’s test (p = 0.006).

Conclusions: Machine learning algorithms with radiomics features from CTU images show promising results in classifying bladder cancer by RB1 mutation status non-invasively.

1. Introduction

Bladder cancer is the most common cancer in the urinary system. It is in ninth place among all cancers [1]. Clinical staging and pathological grading systems have been widely used to evaluate patients and decide the best treatment protocol. Muscular invasion is one of the most critical clinical criteria for staging [2]. Tumors with muscular invasion are considered advanced cancers prone to recurrence and metastasize. Advanced bladder cancers are recently treated with neoadjuvant chemotherapy protocols before surgery [3]. Therefore, it is crucial to determine advanced bladder cancers preoperatively.

Multiple gene mutations from different signal pathways were reported in bladder cancers [4]. However, retinoblastoma-1 (RB1) gene mutations cause alterations in the activity of the members of the retinoblastoma (RB) protein family (e.g., Rb, p107, and p130), leading to advanced-stage bladder cancers with high recurrence and low survival rates [5]. That oncogenic alteration leads to the failure of dephosphorylation of RB protein, leading to failure of arrest in the mitotic cycle.

Consequently, uncontrolled cell proliferation proceeds rapidly in an aggressive process [6].

Radiomics is a rapidly emerging field in radiology that makes it possible to analyze minimal differences between pixels and their relation to each other, which are unseeable to the human eye [7]. With radiomics, quantitative texture analysis of the tumors is being done by radiologists [8]. Moreover, mining the radiomic data with machine learning algorithms allows building various models to classify high-low-grade tumors or accurately predict response to treatment.

This study evaluates the potential of machine learning-based models that classify RB1 gene mutation status, with radiomics from computed tomography urography (CTU) images in bladder cancer.

2. Material and methods

2.1. Ethics and data source

No ethical approval was needed for this study because the patients’ data were obtained from a freely available public dataset.

* Corresponding author.
E-mail address: okan_ince@yahoo.com (O. Ince).
From the Cancer Genome Atlas – Bladder Carcinoma (TGCA-BLCA) database [9], cases with and without RB1 mutations were selected, and their imaging data were downloaded from The Cancer Imaging Archive (TCIA) for scientific purposes [10]. For better visualization of the tumor in segmentation, CTU images of the dataset were enrolled. Cases without CTU images, non-diagnostic images due to various artifacts, and patients with multiple tumors were excluded. Radiomics workflow is summarized in Figure 1. The methodological quality of the workflow based on the Radiomics Quality Score [8] is assessed in Table 1.

2.2. CT acquisition parameters

Since the TGCA-BLCA project consisted of multicenter data, scanning parameters were mean slice thickness: 3.31 mm (IQR 2.37–5), mean mAs: 155 mAs (IQR 84–216), and mean kV: 120 kV (IQR 120-120). The images consisted of 11 and 14 different scanners from four vendors as RB1-mutated and non-mutated patients, respectively. To minimize the differences, the images have undergone multiple image preprocessing steps. The range between contrast delay time was 10 min–35 min.

2.3. Segmentations

Two radiologists, one who has more than twenty years of experience in abdominal radiology and the other one is a fourth-year radiology resident, have manually segmented the largest cross-sectional area of the tumor in CTU images in the 2D plane by using freely accessible 3D Slicer software (v.4.10.2) with consensus. To avoid high-density contrast material in the region of interest (ROI), 2 mm shrinkage was applied in every segmentation label (Figure 2). For data augmentation to remove the imbalance between groups, one slice above and below of the tumor’s largest cross-sectional area was also segmented in the RB1 mutation group. Segmented labels were shared in the corresponding author’s github repository (https://github.com/okanince/HELIYON-OkanInce).

2.4. Image preprocessing and feature extraction

Images were normalized with the ±3 sigma technique for preprocessing steps [11]. Then, pixels were rescaled to $1 \times 1 \text{mm}^2$ with cubic B-spline interpolation method, and gray-levels were discretized with a fixed bin width of $3$ [12]. An optimal bin-width value was selected to keep total bins between 10-100.

Six separate feature subgroups were extracted from the original and wavelet filtered images using PyRadiomics [13], a built-in extension pack in 3D Slicer software. A detailed list of the extracted features was included in Table 2.

2.5. Data preprocessing and feature selection

Data preprocessing steps were observed for the stability and reliability of machine learning algorithms. All the data were standardized and discretized to 18 bins with a uniform bin width. The dataset was split...
to train and test sets with a ratio of 70/30. To avoid the injection of the train set to the test set, the data splitting process was performed before data augmentation. Consequently, a patient-based dataset splitting was ensured.

Feature selection is an essential step for building machine learning algorithms. Hence, exceeding the number of features in the model can cause overfitting bias [14]. To avoid that and reduce the multidimensional model input, we have followed two feature selection steps, the filter method, and the wrapper method, respectively.

In the filter method, features having high collinearity in Pearson’s correlation analysis were excluded. The r threshold was selected as 0.7 [15]. The remaining features were the input of the second step. A wrapper-based sequential feature selection algorithm was built with backward propagated 5-fold cross-validation. The learning classifier was chosen as XGBoost with default hyperparameters [16, 17, 18, 19]. The wrapper method evaluates multiple models by including or excluding features in the wrapper-based sequential feature selection to reach the best feature combination. In the backward propagation, which the current study uses, the initial model is built with all features and validated by 5-fold cross-validation. Subsequent models were evaluated by excluding one of each feature respectively. The most relevant features were determined after multiple evaluations. Using cross-validation technique important features were selected utilizing only training folds. Thus, the “double-dipping” phenomenon was avoided [20].

Moreover, since the data were split to train and test sets before, as mentioned above, the test set was not used in any part of the feature selection.

2.6. Machine-learning algorithms based classification

The selected final features were used in the machine learning models. Three models were built by coding in python language (3.7.11). The first model’s classifier was selected as XGBoost with hyperparameters of maximum estimators, learning rate, gamma, subsample ratio of columns by tree, and maximum depth as 200, 0.03, 0.3, 1, and 7, respectively. Second, the Random Forest (RF) classifier was selected with the hyperparameters of a number of estimators, criterion, maximum depth, minimum samples in leaf, minimum samples to split, and maximum features as 200, “entropy,” 6, 2, 3 and “none”, respectively. Third, a k-Nearest Neighbors (kNN) classifier was selected with the hyperparameters of the number of neighbors, weights, algorithm, power parameter, and distance metric as 6, “uniform,” “auto,” 3 and “Minkowski.” The grid search algorithm tuned the hyperparameters of the models with 10-fold cross-validation by using the train set. The models were trained with the train set, and the test set evaluated their performances. Accuracy, sensitivity, specificity, precision, recall, F1 measure (a harmonic calculation of precision and recall), and area under the receiver operating characteristics curve (AUC) were calculated. The models’ performances

Figure 2. 2D segmentation of tumor from anterior bladder wall in the preprocessed image is presented.
Table 2. Total features extracted from original and wavelet filtered images. (GLDM: Gray-level dependence matrix, GLCM: Gray-level co-occurrence matrix, GLRLM: Gray-level run length matrix, GLSZM: Gray-level size-zone matrix, NGTDM: Neighboring gray-tone difference matrix).

| Shape                  | First Order | GLDM | GLCM | GLRLM | GLSZM | NGTDM |
|------------------------|-------------|------|------|-------|-------|-------|
| Voxel Volume           | Interquartile Range | Gray Level Variance | Joint Average | Short Run Low Gray Level Emphasis | Gray Level Variance | Coarseness |
| Maximum 3D Diameter    | Skewness     | High Gray Level Emphasis | Sum Average | Gray Level Variance | Zone Variance | Complexity |
| Mesh Volume            | Uniformity   | Dependence Entropy | Joint Entropy | Low Gray Level Run Emphasis | Gray Level Non Uniformity Normalized | Strength |
| Major Axis Length      | Median       | Dependence Non Uniformity | Cluster Shade | Gray Level Non Uniformity Normalized | Size Zone Non Uniformity Normalized | Contrast |
| Sphericity             | Energy       | Gray Level Non Uniformity | Maximum Probability | Run Variance | Size Zone Non Uniformity | Busyness |
| Least Axis Length      | Robust Mean Absolute Deviation | Small Dependence Emphasis | Idmn | Gray Level Non Uniformity | Gray Level Non Uniformity | |
| Elongation             | Mean Absolute Deviation | Small Dependence High Gray Level Emphasis | Joint Energy | Long Run Emphasis | Large Area Emphasis | |
| Surface Volume Ratio   | Total Energy | Dependence Non Uniformity Normalized | Contrast | Short Run High Gray Level Emphasis | Small Area High Gray Level Emphasis | |
| Maximum 2D Diameter Slice | Maximum | Large Dependence Emphasis | Difference Entropy | Run Length Non Uniformity | Zone Percentage | |
| Flatness               | Root Mean Squared | Large Dependence Low Gray Level Emphasis | Inverse Variance | Short Run Emphasis | Large Area Low Gray Level Emphasis | |
| Surface Area           | 90 Percentile | Dependence Variance | Difference Variance | Long Run High Gray Level Emphasis | Large Area High Gray Level Emphasis | |
| Minor Axis Length      | Minimum      | Large Dependence High Gray Level Emphasis | Idn | Run Percentage | High Gray Level Zone Emphasis | |
| Maximum 2D Diameter Column | Entropy | Small Dependence Low Gray Level Emphasis | Idm | Long Run Low Gray Level Emphasis | Small Area Emphasis | |
| Maximum 2D Diameter Row | Range Variance | 10 Percentile Kurtosis Mean | Low Gray Level Emphasis | Correlation Autocorrelation Sum Entropy MCC Sum Squares Cluster Prominence Im2 Imc1 Difference Average Id Cluster Tendency | Run Entropy High Gray Level Run Emphasis Run Length Non Uniformity Normalized | Low Gray Level Zone Emphasis Zone Entropy Small Area Low Gray Level Emphasis |

were compared with Friedman’s test in SPSS v.23 (IBM Corp, Armonk, NY, USA) [21]. Post-hoc pairwise analysis was performed if a significant difference was found. The threshold of the statistical significance was set as 0.05.

3. Results

3.1. Data source

There were 78 patients with RB1 mutation and 334 without RB1 mutation of 412 in the TCGA-BLCA dataset. Images from 28 patients with RB1 mutation and 82 patients without RB1 mutation were available in TCGA. After exclusions, CTU images from 18 and 54 patients with and without RB1 mutation, were enrolled in the study. After data augmentation, 54 labeled data from both groups were in the study. Patients’ demographics are described in Table 3.

3.2. Feature extraction and selection

In total, 851 features were extracted (14 shape-based, 18 first order, 14 gray-level-dependence matrix (GLDM), 24 gray-level co-occurrence matrix (GLCM), 16 gray-level run-length matrix (GLRLM), 16 gray-level size-zone-matrix (GLSZM), five neighboring gray-tone difference matrix (NGTDM) and 744 wavelet derived texture features).

In Pearson's correlation analysis, 95 non-redundant features were selected. After the wrapper-based sequential feature selection step, the selected features were reduced to 12. Details of the selected features are included in Table 4 and Figure 3.

3.3. Machine-learning algorithms based classification

XGBoost, RF, and kNN models classified RB mutation status with accuracy rates of 84%, 72%, and 66%, respectively. Sensitivity, specificity, and AUC were 80%, 88% and 0.84 for XGBoost; 80%, 65% and 0.72 for RF; 53%, 76% and 0.65 for kNN. Detailed performance metrics and confusion matrices are shown in Table 5. In Friedman’s test, the XGBoost model showed better performance score than kNN model (p = 0.006). There was no statistically significant difference between performances of RF - kNN and RF – XGBoost models in posthoc pairwise analysis (p = 0.54, 0.25, respectively). The calibration plot of the models is shown in Figure 4.

4. Discussion

This study has evaluated machine learning-based models in classifying RB1 mutation presence from CTU images of bladder cancer. Three different models were built, and each of them achieved discriminative AUC scores. Achieving those scores from three different models indicates the feasibility of the machine learning models in RB1 mutation classification. According to these results, machine learning-based models would be utilizable for detecting RB1 mutation preoperatively in bladder cancer.
In the literature, radiomics with texture analysis has been studied in bladder cancer on differentiating low-grade and high-grade tumors [22, 23, 24] and tumors with the status of perivesical infiltration [25], prediction of disease-free survival [26], recurrences [27, 28], and response to the treatment [29]. Any gene mutation dedicated study could not be seen in the literature. To our knowledge, this study is the first that evaluates machine learning models performances in the classification of bladder cancer by RB1 gene mutation status.

Histopathological examination of cystoscopic punch biopsy is the gold standard in the staging of bladder cancers. However, this procedure is invasive, requires hospitalization, and has challenges such as biopsy specimen inadequacy and sampling error [30, 31]. Our study shows that machine learning-based models may contribute to the diagnosis preoperatively and non-invasively. Furthermore, it would be possible for patients to preoperatively benefit from rapidly developing precise medicine applications [32], which target the cyclin-dependent kinase (CDK) 4/6 pathway. There are currently promising new immunotherapy and vaccine trials targeting different molecular oncogenesis pathways in bladder cancer in the literature [33, 34, 35]. Considering these novel applications, determining the RB1 gene mutation directly at low cost and non-invasively would be pivotal in clinical practice.

Our study has several limitations. Firstly, this is a retrospective study, and all the data were obtained from previous recordings that could lead to a selection bias. Secondly, the patient population was small, and there was an imbalance between groups. Class imbalance is an issue for most machine learning algorithms that run around the assumption that all the classes are distributed equally. In such a case, models tend to predict in favor of the majority class. Data augmentation is a proven technique that can be used for preventing the imbalance between the classes or reducing the risk of overfitting where the number of samples is small [36]. Using various synthetic oversampling methods like Synthetic Minority Over-sampling (SMOTE) and Adaptive Synthetic (ADASYN) oversampling algorithm would be more timesaving [37, 38]. However, that would have caused the majority of the data to be synthetic.

For this reason, the authors performed multiple segmentations from the actual images of the minority class. Thirdly, two radiologists segmented the tumors in consensus to increase the segmentation accuracy. Therefore, interobserver analysis in segmentation could not be done. Fourthly, segmentations were done in the largest slice of the tumor in a 2D plane. Volumetric segmentation could represent the tumor better but requires an exceeding amount of time to perform [39]. Also, most of the studies in the literature are based on 2D segmentation. Several automated segmentation software is being used in the studies in the literature [40]. The authors acknowledge that they are currently in the development process of automated segmentation software for the bladder. Using such software, volumetric segmentations could have been performed more effortlessly and faster in the future. Fifthly, the TCGA-BLCA dataset consisted of various scanners from different centers with different protocols. Although it is essential since that represents clinical practice, this multiplicity may be challenging for machine learning algorithms. However, similar results from different models show the importance and necessity of image and data preprocessing steps. Lastly, we split the dataset to train and test sets that caused the models to

Table 3. Demographic characteristics of included patients by their RB1 mutation status.

| Feature | with RB1 mutation | without RB1 mutation |
|---------|-------------------|----------------------|
| Age (mean ± SD) | 69.1 ± 7.5 | 69.1 ± 11 |
| Sex (female/male) | 4/14 | 14/40 |
| Scanner (Vendor/Model) | 4/11 | 4/14 |

Table 4. Selected features after both filter and wrapper methods.

| Feature Label | Image Type | Feature Class | Feature Name |
|---------------|------------|---------------|--------------|
| TexF1         | wavelet-HLL | GLDM          | Dependence   |
| TexF2         | wavelet-LHL | FIRST ORDER   | Maximum      |
| TexF3         | wavelet-LHH | GLRLM         | Low Gray Level Run Emphasis |
| TexF4         | wavelet-LHH | GLSZM         | Gray Level Variance |
| TexF5         | wavelet-LHH | GLSZM         | High Gray Level Zone Emphasis |
| TexF6         | wavelet-LHH | FIRST ORDER   | Energy       |
| TexF7         | wavelet-LHH | GLSZM         | Gray Level Variance |
| TexF8         | wavelet-LHH | NGTDM         | Strength     |
| TexF9         | wavelet-HHH | FIRST ORDER   | Skewness     |
| TexF10        | wavelet-HHH | GLSZM         | Gray Level Non Uniformity |
| TexF11        | wavelet-LLL | GLCM          | Correlation  |
| TexF12        | original    | GLCM          | Correlation  |

L: Low, H: High, GLDM: Gray-level dependence matrix, GLRLM: Gray-level run-length matrix, GLSZM: Gray-level size-zone-matrix, NGTDM: Neighboring Gray-tone difference matrix, GLCM: Gray-level co-occurrence matrix.

Figure 3. Correlation matrix of selected features are shown in the heatmap. None of the features are correlated to each other (r < 0.7).
test with smaller data sets. It is needed to test the models with larger datasets from external centers.

In conclusion, machine learning-based models with radiomics from CTU images show promising results in classifying bladder cancer by their RB1 gene mutation status non-invasively. Nevertheless, further studies with larger datasets are needed to test the models from external centers before their clinical use. Radiomics will have great potential when combined with artificial intelligence techniques like machine learning in the future.

Declarations

Author contribution statement

Okan Ince: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Hülya Yıldız: Conceived and designed the experiments; Performed the experiments.

Tanju Kisbet: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Şükrü Mehmet Ertürk: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Hakan Onder: Analyzed and interpreted the data.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data included in article/supplementary material/referenced in article.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.
