Learning cancer grade in an automated manner

Abstract:

Owing to the molecular heterogeneity of bladder cancers in general, and urothelial carcinoma in particular, it often manifests differently in different patients based on the underlying genetic mutations. And this manifestation is often variable across stages and grades of the disease. The treatment strategies for early-stage tumors are often very different from late-stage ones; likewise, those of metastasized growths are different from localized ones. Muscle invasive tumors often require chemotherapy, whereas for non-muscle invasive ones the preferred course of treatment is surgical intervention, specifically resection. In this project, we employ a genetic algorithm-based approach to synthetically mimic tumor regions of interest from healthy tissue segments, in the process learning characteristic features pertinent with specific stages as they manifest on the CT scans. The proposed method detects cancer stage from images based on cancer and healthy segments. The results are <describe results>.

Introduction:

Bladder cancer is a highly prevalent and heterogeneous disease characterized by a wide range of tumor characteristics, resulting in varying clinical outcomes [1]. Non-muscle invasive bladder cancer (NMIBC), which includes stages Ta, Tis, and T1, typically have a lower risk of recurrence. This type of cancer is usually managed with intravesical therapies and transurethral resection of bladder tumor (TURBT) [2]. In contrast, muscle invasive bladder cancer (MIBC), consisting of stages T2, T3 and T4, has a higher likelihood of metastasis and requires radical cystectomy along with chemotherapy [3]. Tumors classified as low-grade grow at a gradual pace and need prompt endoscopic intervention and supervision. They infrequently present a major threat to patients. On the other hand, high-grade tumors possess a significant likelihood of malignancy and are linked to substantial advancement and higher cancer mortality rates. Therefore, the most appropriate treatment options depend on the stage and grade of the cancer [4].

Various conventional machine learning and deep learning techniques have been employed to identify and classify bladder cancer using medical imaging. This study examines three research papers that explore different approaches to this task. The first paper utilizes texture-based features and traditional machine learning methods. The second paper focuses on deep learning techniques for feature extraction and classification. The third paper leverages pre-trained models for feature extraction, followed by traditional machine learning for classification. In the first paper [5], as part of feature engineering, intensity properties and texture properties from Local Binary Patterns (LBP) and Grey Level Co-occurrence Matrix (GLCM) were extracted. Classification was conducted using a Support Vector Machine (SVM) and involved preoperative T2-weighted MRI scans from 65 consecutive patients undergoing radical cystectomy. The algorithm achieved a sensitivity of 74.2%, a specificity of 82.4%, an accuracy of 78.5%, and an AUC of 80.6% at the patient level. The second paper [7] presented a deep learning convolutional neural network (DL-CNN) designed to differentiate between muscle-invasive bladder cancer (MIBC) and non-muscle invasive bladder cancer (NMIBC) using contrast-enhanced CT scans. The authors trained a small DL-CNN from scratch and evaluated eight additional DL-CNNs that had been pre-trained on the ImageNet dataset. The study utilized a total of 1,200 cross-sectional CT images from 369 bladder cancer patients undergoing radical cystectomy. Among the eight DL-CNNs assessed, the VGG16 model achieved the highest area under the receiver operating characteristic curve (AUROC) at 0.997 in the testing dataset. In the third paper referenced as [9], the authors utilized a pre-trained ResNet-18 model with 71 layers for feature extraction. They implemented five different machine learning classifiers: k-nearest neighbor (KNN), support vector machine (SVM), linear discriminant analysis (LDA), decision tree (DT), and naive bayes (NB). These classifiers were applied to three distinct classification tasks: (1) distinguishing between bladder cancer tissue and normal tissue, (2) differentiating muscle-invasive bladder cancer (MIBC) from non-muscle-invasive bladder cancer (NMIBC), and (3) detecting post-treatment changes (PTC) in comparison to MIBC. The model achieved a high F1-score across all these tasks.

Identifying the cancer stage accurately from medical imaging is crucial for determining the appropriate course of treatment. However, current imaging-based diagnostic techniques face challenges in distinguishing between subtle tissue variations that correspond to different cancer stages and in further identifying the grade. To address this gap, our project employs glcm based texture analysis to extract features from cancerous and non-cancerous regions of interest (ROIs) from 100 bladder cancer CT scans. Subsequently, we apply a genetic algorithm to features derived from non-cancerous tissue, creating synthetic ROIs that mimic cancerous patterns. By analysing and comparing these synthetic cancerous ROIs with actual cancerous regions, we extract meaningful differences that reflect the underlying pathology. Finally, we investigate the relationship between these differences and cancer staging using traditional machine learning techniques.

Methodology:  
  
This study follows a structured workflow, as illustrated in Figure 1, to analyze CT scans by identifying regions of interest (ROIs), extracting relevant features, and applying machine learning models for classification. Instead of examining the entire CT image, the focus is on the urinary bladder region, from which ROIs are extracted and analyzed. As a baseline, Gray-Level Co-Occurrence Matrix (GLCM) features are utilized for feature extraction. Furthermore, a genetic algorithm is used to generate synthetic cancer-like ROIs from healthy tissue sections, facilitating the identification of characteristic features associated with various cancer stages and grades. The genetic algorithm produces three key outputs: (1) the best distance between healthy and cancerous regions for a given number of iterations, (2) the mean distance between healthy and cancerous regions across iterations, and (3) the number of iterations required for a healthy region to converge to a cancerous one. These outputs are subsequently utilized to perform classification, with the aim of demonstrating that classification based on the genetic algorithm's outputs yields improved results compared to traditional feature-based classification.

1.1 Extracting spatially labeled ROIs using a sliding window implementation

In the initial phase of the study, the urinary bladder region was isolated from 100 CT scans utilizing ImageJ software. Within this delineated area, regions of interest (ROIs) were extracted, comprising 10, 20, 30, 40, and 50 healthy ROIs per scan alongside a single lesion ROI. The lesion ROIs were identified using the provided masks, while the control ROIs were generated through a sliding window approach. Figure X showcases sample images with annotated ROIs that correspond to various cancer stages.

* 1. Feature Extraction using GLCM  
       
     Six feature extraction techniques were evaluated: Histogram of Oriented Gradients (HOG), Gray-Level Co-occurrence Matrix (GLCM), Scale-Invariant Feature Transform (SIFT), Gradient Location and Orientation Histogram (GLOH), Fourier Transform, and Gabor Filters. Among these, GLCM was selected for two key reasons:
  2. Effective Texture Analysis – GLCM effectively captures the spatial relationships between pixel intensities, allowing for a detailed characterization of textural variations and directional patterns. This capability is particularly advantageous in medical imaging, where subtle texture differences can provide critical diagnostic insights.
  3. Consistent Feature Representation – Unlike some other texture analysis methods, GLCM produces a fixed number of features irrespective of the region of interest (ROI) size. This ensures a uniform feature representation across different ROIs, facilitating systematic comparisons and analyses of texture characteristics.

Using GLCM, 20 texture features were extracted from each ROI based on five key metrics: dissimilarity, correlation, energy, contrast, and homogeneity. To capture directional texture patterns, feature calculations were carried out at four angular orientations (0°, 45°, 90°, and 135°). Furthermore, for each orientation, features were computed across five pixel displacements (1, 2, 3, 4, and 5), representing varying distances between paired pixels. This structured approach provided a comprehensive analysis of texture characteristics across multiple spatial and directional configurations.

1.3Classification of Features  
  
To classify the extracted features, six supervised machine learning models were utilized: Support Vector Machine (SVM), Logistic Regression, k-nearest Neighbors (KNN), Decision Tree, Random Forest, and Linear Discriminant Analysis (LDA). Four distinct classification tasks were performed to evaluate the predictive capability of the extracted features:

* Cancer Invasion Classification: Differentiating between non-muscle-invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC).
* Cancer Stage Classification: Categorizing cancer into stages Ta, Tis, T1, T2, T3, and T4.
* Early-Stage vs. Late-Stage Classification: Grouping cases of early-stage bladder cancer (Ta, Tis, T1) versus late-stage cases (T2, T3, T4).
* Post-Treatment Changes vs. MIBC Classification: Distinguishing between treatment-induced changes and muscle-invasive bladder cancer.

* 1. Classification using raw GLCM features.  
       
     The classification was independently performed for each of the five GLCM metrics: dissimilarity, correlation, energy, contrast, and homogeneity. In this phase, the extracted GLCM features were used as primary inputs for the machine learning models, establishing a baseline for comparison with classifications derived from genetic algorithm features. Unlike the genetic algorithm-based classification, this method focused solely on tumor ROIs, omitting the inclusion of healthy tissue regions.
  2. Classification based on the similarity between synthetic healthy tissue sections and real cancer Regions of Interest (ROIs)  
       
     This classification method employed two features derived from the genetic algorithm: best distance and mean distance, which were calculated over a fixed number of iterations. These features were assessed by analyzing the similarity between synthetic healthy tissue sections and tumor ROIs. For each GLCM metric—dissimilarity, correlation, energy, contrast, and homogeneity—the genetic algorithm was applied separately using varying sets of healthy ROIs (10, 20, 30, 40, and 50). For each set, the best distance and mean distance were computed independently, providing a comprehensive evaluation of how different texture descriptors contribute to classification performance.

* 1. Classification Based on Algorithm Convergence Time  
       
     A further classification feature was derived from the number of iterations—or maximum generations—required for the genetic algorithm to adjust a healthy region of interest (ROI) to resemble a tumor ROI. The total number of generations needed for this convergence was recorded as a classification feature. This metric was computed for different sets of healthy ROIs (10, 20, 30, 40, and 50) across all GLCM metrics.

Results:  
  
The performance of the proposed method and baseline approaches was evaluated using accuracy and F1-score, ensuring a rigorous assessment across all classification tasks. Given the limited dataset and potential class imbalances, F1-score was prioritized as the primary evaluation metric, as it provides a more reliable measure by balancing precision and recall.

For the classification of cancer stages, a 3-fold cross-validation strategy was employed to mitigate the effects of data scarcity while ensuring computational feasibility and result stability. For the remaining three classification tasks, 5-fold cross-validation was utilized, taking advantage of the larger sample sizes to enhance the reliability and robustness of performance estimates.

To assess the effectiveness of incorporating a genetic algorithm-based approach, its performance was compared against the baseline feature-based classification. While the genetic algorithm introduced novel features derived from synthetic cancer-like ROIs, its impact on classification varied across different tasks. The following sections provide a detailed breakdown of classification performance, analyzing the strengths and limitations of both approaches across different classification tasks.

Classification of cancer invasion (NMIBC vs. MIBC).

1.1 Baseline Performance (GLCM Features)

The baseline classification model relied on GLCM features extracted from the ROIs, including dissimilarity, correlation, energy, contrast, and homogeneity. Among these, correlation emerged as the most predictive feature, achieving the highest F1-score compared to other features.

1.2 Performance Using Genetic Algorithm Features

In the genetic algorithm-based approach, two features derived from the algorithm were used for classification:

* 1. Best Distance: Among the GLCM metrics evaluated (e.g., dissimilarity, correlation, energy, contrast, and homogeneity), best distance using correlation consistently achieved better results compared to other features.
  2. Mean Distance: Among the GLCM metrics evaluated (e.g., dissimilarity, correlation, energy, contrast, and homogeneity), mean distance using correlation consistently achieved better results compared to other features.
  3. Maximum Generations: When using the maximum number of iterations required by the genetic algorithm for a synthetic healthy region to converge into a cancer region, the energy-based feature yielded the best classification performance in terms of F1 score.

2. Classification of cancer stages (Ta vs. Tis vs. T1 vs. T2 vs. T3 vs. T4).

Discussions:

The findings of this study are influenced by several limitations, particularly the dataset size and class distribution. The dataset consisted of 100 CT images spanning seven cancer stages, with an imbalanced class distribution that affected stage-wise classification performance. A stratified k-fold cross-validation approach was implemented to ensure proportional representation across training sets. However, this method did not fully address the challenge posed by underrepresented classes (Ta, Tis, and T4), which likely impacted classification performance.

The use of Gray-Level Co-occurrence Matrix (GLCM) features introduced constraints due to their sensitivity to pixel relationships at predefined angular orientations (0°, 45°, 90°, and 135°). Conventional data augmentation methods, such as rotation and flipping, did not alter pixel-pair relationships, limiting their effectiveness in enhancing model generalization. Additionally, this study focused on five selected GLCM metrics, and the exclusion of other texture descriptors may have reduced the model’s ability to capture more complex texture patterns.

To address these limitations, expanding the dataset is essential to improve class balance and support deep learning-based methods. Handcrafted feature extraction methods, while useful, may not fully capture intricate texture variations. Deep learning approaches, particularly Convolutional Neural Networks (CNNs), offer a potential alternative for feature extraction. CNNs can learn hierarchical texture and structural representations, which could enhance classification performance. However, traditional CNN architectures often require fixed input sizes, presenting challenges when dealing with varying region of interest (ROI) dimensions across different cancer stages.

Future work could focus on integrating adaptive CNN architectures that accommodate varying ROI sizes while leveraging both handcrafted and deep learning-based features. Additionally, alternative feature descriptors, such as wavelet-based texture analysis or frequency-domain representations, could be explored to enhance model robustness. Incorporating larger and more diverse datasets will be crucial for improving classification performance and ensuring the generalizability of the proposed approach.

This study was limited by the small dataset size, consisting of only 100 CT images distributed across seven cancer stages. The class distribution was imbalanced, affecting stage-wise classification. A stratified k-fold cross-validation approach was used to ensure proportional representation across training sets. However, this strategy could not fully compensate for the lack of sufficient training examples, particularly in underrepresented classes (Ta, Tis, and T4), which likely impacted overall classification performance.

GLCM features were also constrained by their sensitivity to pixel relationships at specific angular orientations (0°, 45°, 90°, and 135°). Conventional data augmentation techniques (e.g., rotation, flipping) were ineffective because they did not alter pixel-pair relationships, limiting the ability to improve model generalization through augmentation. Additionally, this study focused on five key GLCM metrics. The exclusion of other texture descriptors may have restricted the model’s ability to capture more complex texture variations.

To improve the proposed approach, expanding the dataset is necessary to enhance model generalization by providing better class balance and enabling deep learning-based methods. Given the limitations of handcrafted features, deep learning techniques, particularly Convolutional Neural Networks (CNNs), could be explored for feature extraction. CNNs learn hierarchical texture and structural features, which may lead to better classification performance. Traditional CNN architectures often require fixed input sizes, making them less adaptable to varying ROI dimensions across different cancer stages.

Conclusions:

References

1. Wong, V.K., et al., *Imaging and Management of Bladder Cancer.* Cancers, 2021. **13**(6): p. 1396.

2. Chang, S.S., et al., *Diagnosis and Treatment of Non-Muscle Invasive Bladder Cancer: AUA/SUO Guideline.* J Urol, 2016. **196**(4): p. 1021-9.

3. Chang, S.S., et al., *Treatment of Non-Metastatic Muscle-Invasive Bladder Cancer: AUA/ASCO/ASTRO/SUO Guideline.* J Urol, 2017. **198**(3): p. 552-559.

4. Kirkali, Z., et al., *Bladder cancer: epidemiology, staging and grading, and diagnosis.* Urology, 2005. **66**(6 Suppl 1): p. 4-34.

5. Tong, Y., et al., *Radiomics-guided therapy for bladder cancer: Using an optimal biomarker approach to determine extent of bladder cancer invasion from t2-weighted magnetic resonance images.* Adv Radiat Oncol, 2018. **3**(3): p. 331-338.

6. Lingley-Papadopoulos, C.A., et al., *Computer recognition of cancer in the urinary bladder using optical coherence tomography and texture analysis.* J Biomed Opt, 2008. **13**(2): p. 024003.

7. Yuhan Yang, X.Z., Yixi Wang,Xuelei Ma, *Application of deep learning as a noninvasive tool to differentiate muscle-invasive bladder cancer and non–muscle-invasive bladder cancer with CT* 2021.

8. Chapman-Sung, D.H., *Convolutional Neural Network-Based Decision Support System for Bladder Cancer Staging in CT Urography: Decision Threshold Estimation and Validation.* 2020.

9. Sarkar, S., et al., *Performing Automatic Identification and Staging of Urothelial Carcinoma in Bladder Cancer Patients Using a Hybrid Deep-Machine Learning Approach.* Cancers (Basel), 2023. **15**(6).

**5. Discussion**

**5.1 Interpretation of Results**

This study focused on the urinary bladder region in CT scans, selecting **10–50 control ROIs** to model tumor-like regions via a genetic algorithm. **GLCM-based classification** effectively captured texture variations, serving as a baseline. The genetic algorithm introduced complexity by simulating the transition from healthy to tumor tissue, with performance influenced by the similarity between selected ROIs.

**5.2 Limitations**

**5.2.1 Limited Dataset and Class Imbalance**

The dataset comprised **100 CT images across seven cancer stages**, with severe class imbalance (e.g., **Ta: 6, Tis: 9, T4: 4 samples**). Despite **stratified k-fold cross-validation**, minority classes remained challenging to classify.

**5.2.2 Feature Extraction and Augmentation Challenges**

GLCM features, computed at **0°, 45°, 90°, and 135°**, were invariant to common augmentation techniques like rotation and flipping, limiting data expansion strategies.

**5.2.3 Constraints in GLCM Feature Selection**

Only five GLCM metrics (**dissimilarity, correlation, energy, contrast, homogeneity**) were used, excluding others like **entropy and variance**, which could improve texture representation. Similarly, **five pixel displacements** may have restricted spatial analysis.

**5.3 Future Directions**

**5.3.1 Dataset Expansion**

A larger dataset would enhance class balance and enable **deep learning-based methods**, improving generalization.

**5.3.2 Deep Learning for Feature Extraction**

CNNs could replace handcrafted features, learning hierarchical representations. However, **ROI size variability** poses a challenge, which can be addressed through:

* **Fully Convolutional Networks (FCNs):** Replace fully connected layers with convolutions, allowing flexible input sizes.
* **Global Average Pooling (GAP):** Summarizes feature maps spatially, making CNNs adaptable to varying ROI dimensions.