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 deployed to identify and classify bladder cancer using medical imaging. The following two studies utilized traditional machine learning methods. 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. In the second paper [6] the authors introduced an algorithm based on texture analysis, specifically using the cooccurrence matrix and 2-D Fourier transform to differentiate between cancerous and noncancerous tissue in OCT images. The algorithm used decision tree for classification and was tested on a set of 182 OCT images from 21 patients and achieved a sensitivity of 92% and specificity of 62% in classifying tissue as cancerous or noncancerous.

The following two papers employed deep learning methods for bladder cancer classification. The first 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 second paper referenced [8], the authors investigated the development of a deep learning convolutional neural network (DL-CNN) designed to assist clinicians in staging bladder cancer using computed tomography urography (CTU). The model was trained on a dataset consisting of 84 bladder cancer cases from 76 patients, and it was subsequently tested on an independent dataset of 90 bladder cancer cases from 86 patients. The DL-CNN achieved a classification accuracy of 91% on the independent test set.

In the 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 various methods for feature extraction, including texture analysis, feature descriptors, frequency-based analysis, and pre-trained models to analyse 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 a combination of traditional machine learning and advanced deep learning techniques.

Methodology:  
  
The study employs a systematic workflow, as outlined in **Figure 1**, to analyse CT scans by identifying specific regions of interest (ROIs), extracting meaningful features, and applying machine learning models for classification. Additionally, a genetic algorithm is incorporated to generate synthetic tumor-like ROIs from healthy tissue sections, enabling the identification of characteristic features associated with various cancer stages. The genetic algorithm provides two critical outputs: the number of iterations required for a healthy region to converge to a cancerous region and the absolute distance between the healthy and cancerous regions for a given number of iterations. These outputs were subsequently utilized to perform additional 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. Extracting spatially labeled ROIs using sliding window implementation

From 100 CT scans, the bladder region was isolated using **ImageJ** software as the first step. Within the bladder, ROIs were extracted: 10, 20, 30, 40, and 50 control ROIs alongside one lesion ROI per scan. Lesion ROIs were determined using provided masks, and control ROIs were generated using a sliding window approach. The process flow for this step is depicted in the figure below.

* 1. Feature Extraction using GLCM  
       
     Six feature extraction techniques were evaluated: HOG, GLCM, SIFT, GLOH, Fourier Six feature extraction techniques were evaluated: HOG (Histogram of Oriented Gradients), GLCM (Gray-Level Co-occurrence Matrix), SIFT (Scale-Invariant Feature Transform), GLOH (Gradient Location and Orientation Histogram), Fourier Transform, and Gabor Filters. GLCM was selected due to its superior consistency across regions of interest (ROIs) of varying sizes, making it particularly effective for the tasks at hand. Using GLCM, 20 features were extracted for each ROI based on five key metrics: dissimilarity, correlation, energy, contrast, and homogeneity. For each metric, calculations were performed at four angular orientations (0°, 45°, 90°, and 135°) to capture directional texture patterns. Furthermore, for each orientation, the features were computed across five-pixel displacements (1, 2, 3, 4, and 5), representing varying distances between paired pixels. This systematic approach ensured a comprehensive analysis of texture characteristics across multiple spatial and directional configurations.
  2. Classification of Features  
       
     To classify the extracted features, six machine learning models were employed: Support Vector Machine (SVM), logistic regression, k-nearest neighbors (KNN), decision tree, random forest, and Linear Discriminant Analysis (LDA). Five distinct classification tasks were performed using these features:

1. Classification of cancer invasion (NMIBC vs. MIBC).
2. Classification of lesion vs. control regions.
3. Classification of cancer stages (Ta vs. Tis vs. T1 vs. T2 vs. T3 vs. T4).
4. Classification of early-stage vs. late-stage cancer.
5. Classification of post-treatment changes vs. MIBC.

* 1. Classification using GLCM features.  
       
     Classification was independently conducted for each of the five GLCM metrics: dissimilarity, correlation, energy, contrast, and homogeneity. The extracted GLCM features served as the primary inputs for the machine learning models in this stage.
  2. Classification based on similarity between synthetic healthy section of tissue, and real cancer ROI  
       
     Classification was also performed using the absolute distance outputs from the genetic algorithm, which quantified the similarity between synthetic healthy tissue sections and real cancerous ROIs. This approach was applied to four of the five tasks, excluding the lesion vs. control classification.
  3. Classification based on time taken for algorithm to converge  
       
     The number of iterations (or maximum generations) required by the genetic algorithm for a healthy region to converge to a cancerous region was utilized as an additional feature for classification. Similarly, this approach was applied to four of the five tasks, excluding the lesion vs. control classification.

Results:  
  
The evaluation of the proposed method and baseline approaches was conducted using two key metrics: accuracy and F1-score, ensuring a comprehensive assessment of model performance. For the classification of cancer stages, 3-fold cross-validation was employed due to the relatively limited data available for this task, thereby balancing computational feasibility and robustness of results. In contrast, 5-fold cross-validation was applied for the remaining four tasks, as these datasets had sufficient sample sizes to allow for more detailed validation, enhancing the reliability of performance estimates.

Despite the innovative approach of incorporating a genetic algorithm to generate synthetic tumor-like ROIs, the performance of the proposed method generally fell short compared to baseline approach. Specifically, the proposed method exhibited lower accuracy and F1-score across most tasks, indicating challenges in effectively capturing the nuanced patterns required for robust classification. The following sections present a detailed comparison of results for each classification task, highlighting both the strengths and limitations of the proposed approach.

1. 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:

1.21 Absolute Distance: Among the GLCM metrics evaluated (e.g., dissimilarity, correlation, energy, contrast, and homogeneity), absolute distance using correlation consistently achieved better results compared to other features.

1.22 Maximum Generations: When using the maximum number of iterations required by the genetic algorithm for a synthetic healthy region to converge into a tumor-like region, the energy-based feature yielded the best classification performance.

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

Discussions:

Interpretation of the results

Limitations

One major limitation of this study is the relatively small dataset, consisting of only 100 CT images, distributed across seven cancer stages (as shown in Figure X). The imbalance in class distribution (e.g., only 4 samples for stage T4 compared to 24 for stage T3) have further skewed the results, particularly in tasks requiring stage-wise classification.

To mitigate the effects of class imbalance and ensure fair evaluation, stratified k-fold cross-validation was employed. This method ensured that each fold retained the same proportion of samples from each cancer stage, preventing underrepresented classes from being excluded in any fold. However, while stratified k-fold cross-validation provided robust performance estimates, it could not fully compensate for the limited number of samples in minority classes, which still have affected model training and evaluation.

Another challenge arose from the nature of the GLCM features. Since GLCM captures texture information at specific angular orientations (0°, 45°, 90°, and 135°), standard data augmentation techniques such as rotation and flipping failed to generate diverse features. These transformations resulted in identical GLCM features, as the pixel-pair relationships remained unchanged under such augmentations. Consequently, data augmentation techniques commonly used to enhance small datasets were ineffective in this context.

Future Directions

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).