A classification for urinary bladder epithelial cancer cell basing on random forest algorithm

Yeyu Huang,† Yeen Huang‡

1Department of Information Science and Technology, Jinan University, Guangzhou, China; School of Chemistry, Sun Yat-sen University, Guangzhou, China
2Department of Epidemiology and BioStatistics, Sun Yat-sen University, Guangzhou, China

Abstract

Objectives: Classification diagnosis of urinary bladder epithelial cancer cell was strongly influenced by subjective judgments, which is made by pathologists, resulting in the presence of high false positive rate and false negative rate of diagnosis. Therefore, Random Forest was performed to diagnosis and classification urinary bladder epithelial cancer cell for exploring the feasibility and application value of its method.

Methods: A total number of 258 urinary bladder epithelial cancer samples were collected and diagnosed. Morphological and colorimetric features of samples were evaluated by the application of imageJ. Random Forest algorithm, integrated with Weka 3.6.6 was performed to training samples and modeling. Test accuracy was calculated by 10-fold Cross-validation.

Results: The overall classification accuracy performed by random forest was 98.13% between normal group and lesions group, 98.95% between urothelium dysplastic exfoliated cells and bladder urothelial cancer exfoliated cells. For the classification diagnosis of urinary bladder epithelial cancer cell, the classification diagnostic effect performed by Random Forest was the best while distinguishing lesions cells from normal cells, and bladder urothelial cancer exfoliated cells from urothelium dysplastic exfoliated cells, respectively.

Conclusion: It was indicated that Random Forest can be considered as an effective classification method to classified urinary bladder epithelial cancer cells.

Keywords: Colorimetric parameters, image analysis, morphological parameters, random forest algorithm, urinary bladder epithelial cancer

Address for correspondence: Yeen Huang, MD. Department of Epidemiology and BioStatistics, Sun Yat-sen University, Guangzhou, China
Phone: +86 0755-83243394 E-mail: 743159984@qq.com
Submitted Date: January 20, 2020 Accepted Date: February 19, 2020 Available Online Date: March 12, 2020

©Copyright 2020 by Eurasian Journal of Medicine and Oncology - Available online at www.ejmo.org
OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.
measurement and analysis of medical data through machine learning, also accumulation and analysis of knowledge could be continually executed through self-learning machine. Therefore, AI is commonly used to assist doctors through accomplishing more precise and reliable diagnostic decisions.[3] Represented as a machine-learning algorithm of AI, Random Forest had been widely used in classification researches of biomedical and bioinformatics, such as figure parts identification, metabolomics data classification and medical imaging data analysis, owing to its highly prediction accuracy of categories, greater tolerance for abnormal values and noise, as well as less prone to overfitting.[4–7] Researches of classification diagnosis based on urinary bladder epithelial cancer were focused on mathematical statistics methods such as linear discriminate function (LDF), K-Nearest Neighbor (KNN), or other AI algorithm such as Artificial Neural Networks (ANN), Support Vector Machine (SVM) and etc.[8–10] Researches that were related to random forest and the applications of classification based on urinary bladder epithelial cancer were limited. The reports of the related research of Random Forest application in urinary bladder epithelial cancer were few and only focused on certain types other than the common classification criterion of urinary bladder epithelial cancer.

To solve above problems, Random Forest was used for training samples and modeling with twenty-three morphological and colorimetric characteristics extracted from urinary bladder epithelial cancer samples. The accuracy of the classification was tested with 10-fold cross-validation so as to explore the feasibility and the application value of this category method.

**Methods**

**Urinary Bladder Epithelial Cancer Samples**

This study was approved by the Jinan University Institutional Review Board, and each study patient provided oral informed consent by telephone calls. Samples included in this study were selected randomly from the overseas Chinese Hospital from August 2018 to August 2019. A total of 20 cases of normal and 238 cases of patients who diagnosed with urinary bladder epithelial cancer were collected. All slices were diagnosed by two pathologists with at least five-years working experiences in clinical diagnosis.

**Cells Image Acquisition**

The slices were divided into three types, including uroepithelium normal exfoliated cells (UNC), urothelium dysplastic exfoliated cells (UDC), and bladder urothelial cancer exfoliated cells (UCC). Images saved in the format of TIFF with the resolution sizes of 1360×1024 pixels. The categorization process of three types urinary bladder epithelial cancer cells were shown in figure 1.

**Parameters Measurement**

The morphological and colorimetric parameters were measured by application of ImageJ 1.45.[11] The selected morphological parameters contained Area, Major Axis, Minor Axis, Perimeter, Form Factor PE (PE), Form Factor AR (AR), Regular Form Factor (RFF), Form Irregular Index (FII), Nucleus/Cytoplasm Ratio (NCR). The selected colorimetric parameters were red (R), green (G), blue (B) and red color coefficient (r), green color coefficient (g) and blue color coefficient (b).[12]

**Diagnosis and Classification Using Random Forest**

With the consideration of the diagnostic habits and processes (Fig. 2) of urinary bladder epithelial cancer for pathologists, three types of urinary bladder epithelial cancer cells were divided into Normal group and Lesions group.[14] With the application of Weka 3.6.6 (Waikato Environment for Knowledge Analysis, New Zealand), an open source machine learning and data mining software, Random Forest was used to training samples and modeling based on the morphological and colorimetric features of the above two groups and then test accuracy was calculated by 10-fold cross-validation.

![Figure 1](http://example.com/figure1.png)  
*Figure 1. Categorization process of three type’s urinary bladder epithelial cancer cells.*

Normal: Normal group; Lesions: Lesions group.

![Figure 2](http://example.com/figure2.png)  
*Figure 2. The flowchart of diagnosing urinary bladder epithelial cancer.*

First, determined whether the urinary bladder epithelial cancer samples were lesions, if not, diagnosed as UNC, if yes, then judged if its were cancer cell, if not, diagnosed as UDC, if yes, then diagnosed as UCC.
Results

Statistical Description of Samples
The age and cell counts of three types of urinary bladder epithelial cancer cells are shown in Table 1. Totally 1041 UNC cells, 1022 UDC cells, 1075 UCC cells were collected. Among all types of cell, the average age of the oldest was UDC (59.08±6.42 years old), the youngest was UNC (45.10±5.69 years old).

Classification Results Between Normal Group and Lesions Group
By using Random Forest, the classification results between Normal group (UNC) and Lesions group (UDC, UCC) are shown in Table 2. The overall classified accuracy was 98.13% and Kappa=0.936. Classification accuracy and F-Score was 93.6% and 0.948, 99.1% and 0.989 for Normal group and Lesions group, respectively. True positive of two groups were both exceeded 96.0%.

Classification Results Between Urothelium Dysplastic Exfoliated Cells and Bladder Urothelial Cancer Exfoliated Cells
Table 3 shows the classification results between urothelium dysplastic exfoliated cells and bladder urothelial cancer exfoliated cells. The overall classified accuracy was 98.95% and Kappa=0.979. Classification accuracy and F-Score was 98.4% and 0.989, 99.5% and 0.990 for urothelium dysplastic exfoliated cells and bladder urothelial cancer exfoliated cells respectively. True positive of two cells both exceeded 98.0%.

Discussion
Urinary bladder epithelial cancer cell was currently extensively used in the administration of early bladder cancer screening programmers. However, drawbacks such as missed diagnosis and misdiagnosis caused by subjective judgment which is made by pathologists were inevitably occurred occasionally.[15] To confront the challenge mentioned above, this study explored the application prospects of Random Forest in classification diagnosis of Urinary bladder epithelial cancer cell, by applying Random Forest, which was widely accepted in the field of artificial intelligence, to independent learning and categorizing.

According to the classification results of Random Forest, three conclusions were made as following: I. Random Forest could well distinguish uroepithelium normal exfoliated cells (overall accuracy=98.13%, Table 2). We speculated it was caused by the obvious differences of shape or color of normal cells and lesions cells. II. For urothelium dysplastic exfoliated cells and bladder urothelial cancer exfoliated cells, Random Forest had a high diagnostic accuracy to discriminate them (classification accuracy=99.4% for the former, 98.5% for the later, Table 3), and we believe this categorization tool could perfectly distinguish mild atypical lesions from high-grade bladder intraepithelial neoplasia and severe atypical lesions.

According to the results of this study, three conclusions were made as followed I. Random Forest can best distinguish lesions cells from uroepithelium normal exfoliated cells (overall accuracy=98.13%, Table 2). We speculated it was caused by the obvious differences of shape or color of normal cells and lesions cells. II. For urothelium dysplastic exfoliated cells and bladder urothelial cancer exfoliated cells, Random Forest had a high diagnostic accuracy to discriminate them (classification accuracy=99.4% for the former, 98.5% for the later, Table 3), and we believe this categorization tool could perfectly distinguish mild atypical lesions from high-grade bladder intraepithelial neoplasia and severe atypical lesions.

Some researchers[16-18] reported that Classification and Regression Tree algorithm was performed to classify and predict the prevalence risk of bladder cancer, the defect of the study was that it was only focusing on classifying slight bladder cancer from sever bladder cancer, neglecting the atypical lesions. Similarly, artificial neural networks (ANN)
was considered as the priority interest during the further distinguishing atypical lesions and cancer lesions, which moreover, the classification accuracy was imprecise while categorized and analyzed by Random Forest, regardless the features of Urinary bladder epithelial cancer cell. The categories included normal group and lesions group, urothelium dysplastic exfoliated cells and bladder urothelial cancer exfoliated cells, respectively. Random Forest algorithm was performed to classify each category described above sequentially; this phenomenon met the pathologists' daily diagnosing workflow. The deficiency of this study was that only the morphological and colorimetric features of Urinary bladder epithelial cancer cell were categorized and analyzed by Random Forest, regardless the impacts of classification effect caused by other features, moreover, the classification accuracy was imprecise while distinguishing atypical lesions and cancer lesions, which was considered as the priority interest during the further researches.

Conclusion

For the diagnosis of Urinary bladder epithelial cancer cell, Random Forest performs satisfying classification capacity in distinguish uroepithelium normal exfoliated cells and lesions cells, as well as in discriminating urothelium dysplastic exfoliated cells from bladder urothelial cancer exfoliated cells. It showed us that Random Forest was a feasible approach to distinguish patients from normal people.

Disclosures

Ethics Committee Approval: The study was approved by the Jinan University Institutional Review Board (Code: EC20180503).

Peer-review: Externally peer-reviewed.

Conflict of Interest: None declared.


References


trocin, a sesquiterpene lactone isolated from Antrodia cinnamomea, and its molecular mechanisms. Cancer Letters 2015;373:174–84. [CrossRef]