

## PROBABILITY ASSESSMENT OF RECTAL TUMOR PRESENCE USING CLASSICAL MACHINE LEARNING ALGORITHMS

Aleksandra Bibić<sup>1</sup>, Ivana Mišković<sup>2</sup>, Mira Sarvan<sup>3</sup>, Mladen Marinković<sup>2</sup>, Edib Dobardžić<sup>3</sup>

<sup>1</sup>University of Belgrade, Faculty of Medicine, dr Subotića starijeg, 11000 Belgrade, Serbia,  
[aleksandra.bibic@med.bg.ac.rs](mailto:aleksandra.bibic@med.bg.ac.rs)

<sup>2</sup>Clinic for Radiation Oncology, Institute of Oncology and Radiology of Serbia, Pasterova 14, 11000  
Belgrade, Serbia, [aainv.m@gmail.com](mailto:aainv.m@gmail.com), [mladen309@gmail.com](mailto:mladen309@gmail.com)

<sup>3</sup> University of Belgrade, Faculty of Physics, Cara Dušana 13, 11000 Belgrade, Serbia,  
[miras@ff.bg.ac.rs](mailto:miras@ff.bg.ac.rs), [edib.dobardzic@ff.bg.ac.rs](mailto:edib.dobardzic@ff.bg.ac.rs)

### ABSTRACT

The aim of this study was to investigate the application of classical machine learning algorithms for the detection of rectal tumors on computed tomography (CT) images of patients following neoadjuvant chemoradiotherapy. The study included 138 patients, of whom 136 had a confirmed tumor, while two subjects with a healthy rectum were included to balance the dataset. We extracted CT slices in which a portion of the rectum was visualized, resulting in a dataset of 3566 images. Due to variations in rectal volume, statistical distribution functions of tissue density along the contour and within the interior of the segmented region were used as input features. The most relevant features were selected using feature selection and dimensionality reduction methods, namely Mutual Information (MI) and Principal Component Analysis (PCA), and used for training and testing multiple machine learning algorithms. We used several machine learning algorithms: Logistic Regression (LR), Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), Classification and Regression Trees (CART), Naive Bayes (NB), k-Nearest Neighbors (KNN), and Random Forest (RF). We trained all models on the same feature sets, and evaluated their performance using sensitivity and specificity, which are crucial in medical diagnostics. We achieved the highest accuracy (~80%) using RF models trained on six and seven features selected by MI. We obtained the highest sensitivity (~85%) with RF and NB models using seven features. We observed that specificity was highest for the LDA model with three features and the RF model with seven features (~80%).

**Keywords:** rectal tumor detection, computed tomography, machine learning, feature selection.