



UPPSALA  
UNIVERSITET

Project in Computational  
Science  
Fall term 2020

### Project members

Emil Åberg  
emil.berg96@gmail.com

Maja Arvola  
majaarvola@gmail.com

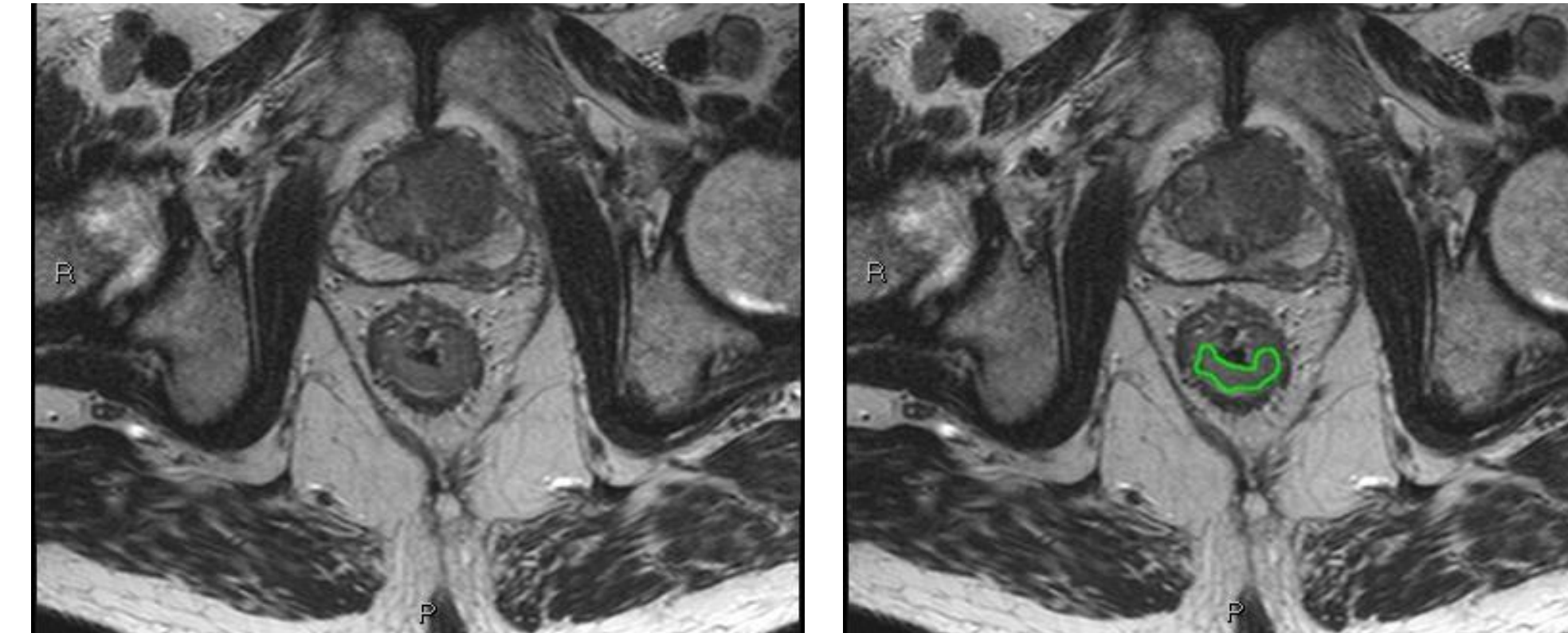
### Supervisor

Filip Malmberg

Department of  
Surgical Sciences,  
Section of Radiology

# Will treatment cure my cancer?

**Can the outcome of radiation therapy be predicted for individuals with colorectal cancer?** The purpose of this project is to investigate if a particular method, involving Radiomic feature extraction, can give a positive answer to this question. The workflow from data set to prediction model is described in this poster.



An example image layer from MRI of patient with colorectal cancer, and corresponding segmentation of the tumor region.

## Image data

MRI prior to therapy was available for 39 patients, together with known treatment outcome ranging from no response to complete response.

Additional images was supplied containing green pixels, marking the tumor region. The segmentation was made by an expert radiologist.

## Extract features

Features, such as

- Volume
- Sphericity
- Entropy

was extracted from the images using Radiomic analysis. By extracting features from several filtered copies of the images, a total of 1578 features was extracted for each patient.

Additional features related to patient information, such as age, gender, weight and type of treatment was added as well.

## Select feature subset

A subset of the features was selected to reduce problem dimensionality to avoid overfitting. Following three methods for feature selection was compared:

- Minimum Redundancy Maximum Relevance (MRMR)
- Least Absolute Shrinkage and Selection Operator (LASSO)
- Logistic Regression with L1-penalty.

## Train prediction model

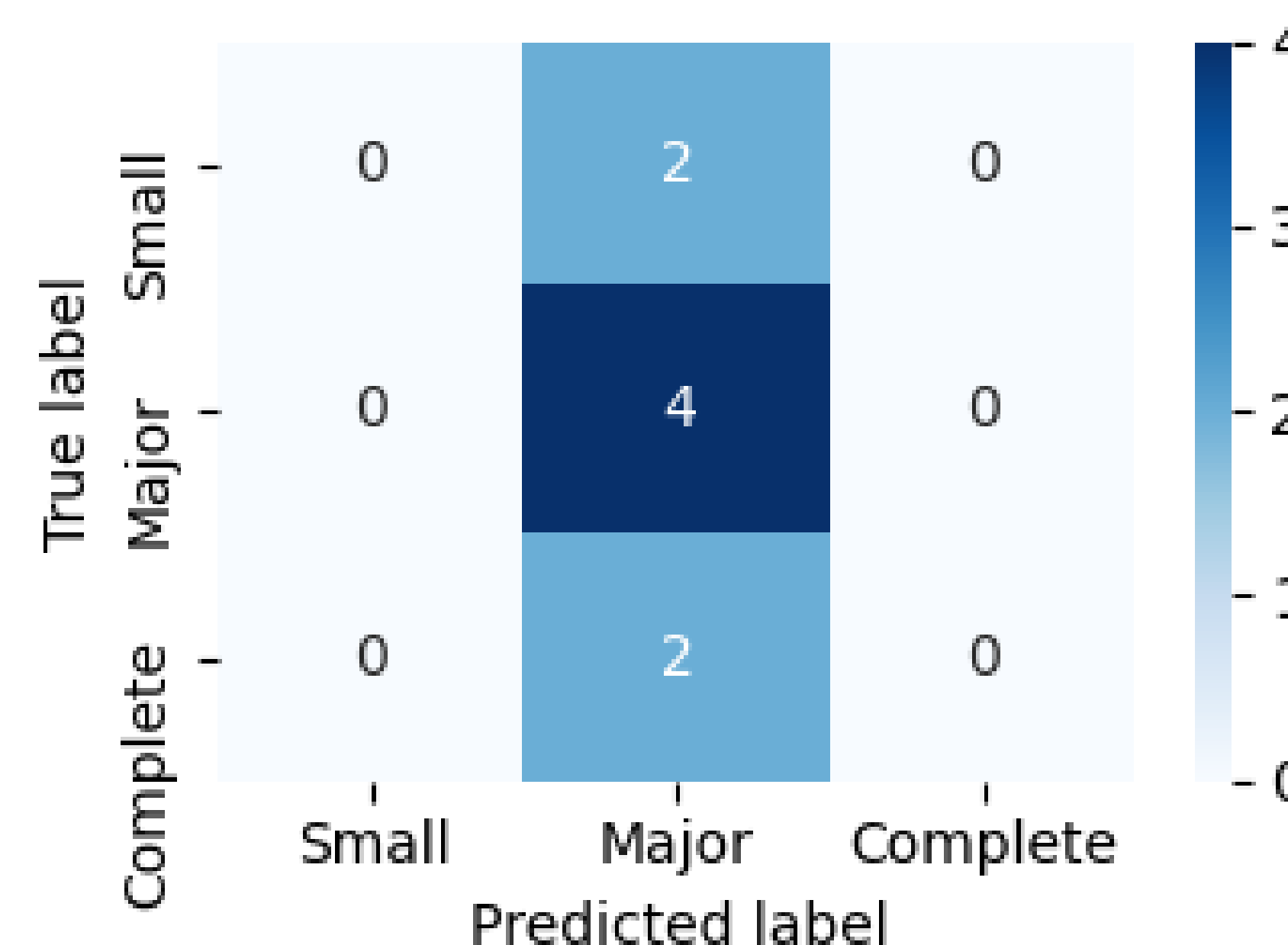
Three different prediction models was trained and validated using 31 patients:

- Random Forest (regression)
- Random Forest (classification)
- Logistic Regression

## Select the best model

Selecting one method for feature selection and one for prediction results in 9 different combinations. All was evaluated using cross-validation on the training data.

LASSO for feature selection and Random Forest Regression for prediction yielded the best performance on the training data with an accuracy score of 61.3%.



Confusion matrix for the selected prediction model when evaluated on test data. The categories describes the grade of tumor reduction.

## Performance on test data

The best performing prediction model was tested on 8 patients and resulted in an accuracy score of 50%. The model predicted the same outcome for all patients which also was the majority class in the data set. This indicates a poor performance of the prediction model, possibly caused by a relatively small data set and class imbalanced.

## Answer:

The model is unable to make a trustworthy prediction of the treatment outcome.