# Prediction of clinically significant prostate cancer based on magnetic resonance images and tissue microarrays applying machine learning approaches

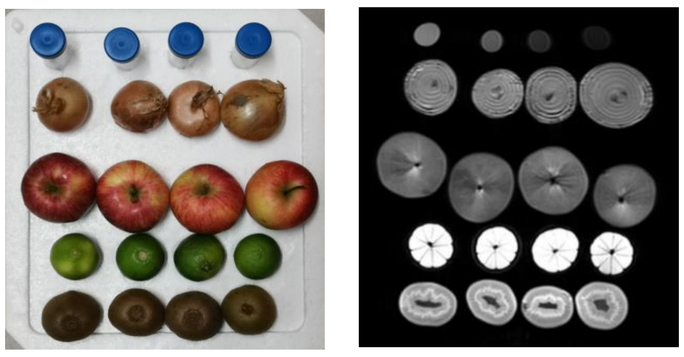
**Yauheniya Zhdanovich**,[1] Jörg Ackermann,[1] Ina Koch,[1] Simon Bernatz, [2] Peter Mandel, [2] Felix Chun, [2] Jens Köllermann, [2] Nicolas Rosbach, [2] Thomas Vogl, [2] and Peter Wild[2]

[1] Molecular Bioinformatics Group, Institute of Computer Science, Goethe-University Frankfurt, Robert-Mayer-Str. 11-15, 60325 Frankfurt am Main, Germany
[2] University Hospital, Goethe-University Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany

**Prostate cancer is the second most common cancer found in men. Definition of clinically significant prostate cancer is a challenging process. Most prostate cancers are slow-growing and do not cause any symptoms, whereas a subset of prostate cancers has an aggressive clinical course. Magnetic resonance imaging (MRI) and bioptic tissue specimen are broadly performed diagnostic procedures. Radiomics is a field of study which proposes methods for extraction and analysis of quantitative imaging features. Machine learning algorithms are widely applied in radiomics. The aim of the work was to predict clinically significant prostate cancer and malignancy of prostate samples applying methods of statistical analysis and machine learning approaches.**

## Multi-fruit phantom

### Construction of a multi-fruit phantom



We reconstructed the multi-fruit phantom of Baeßler *et al* [1]. The phantom contained 16 fruits and four tubes with agarose gel.

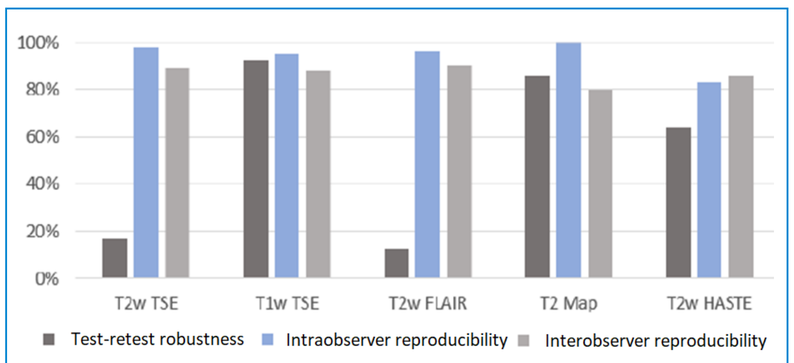
### Analysis of radiomic features

**Aim of the study**

- Investigate the applicability of the radiomic features to the clinical studies
- To assess test-retest robustness, intraobserver reproducibility, and interobserver reproducibility of the features
- Define the most reliable MRI sequences

**Data acquisition and processing**

For the acquisition of the phantom on a 3T scanner, we applied five MRI sequences: T2w TSE, T1w TSE, T2w FLAIR, T2 Map, and T2w HASTE. We segmented images and extracted 107 radiomic features for each volume of interest (VOI) for each image.




All MRI sequences provided more than 80% features with excellent intra- and interobserver reproducibility. For the sequences T1w TSE and T2 Map, more than 90% of the features were robust. With 17% and 12% of robust features, the sequences T2w TSE and T2w FLAIR showed the poorest test-retest robustness.

### Software

- **3D Slicer** (v4.9.0, [2]), extension PyRadiomics for MRI images
- **QuPath** (v0.2.0, [3]) for bioptic tissue specimen
- **Python** (v3.7.6), Jupyter Notebook for data analysis

## Prediction of clinically significant prostate cancer

### Study design

**Aim of the study**

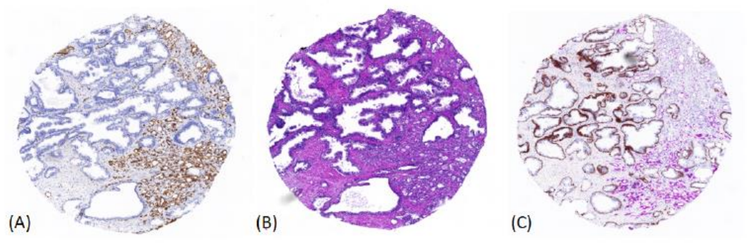
- Select significant features from MRI images and bioptic tissue specimen
- Differentiate between clinically significant and insignificant prostate cancer
- Differentiate between malignant and benign prostate tissues

**Patient cohort**

- 1125 patients, explored from the year 2014 till the year 2019, have been screened for study-inclusion.
- Final dataset included 73 patients (mean age, 66.3 ± 7.6 years)

**Data acquisition and examination of patient data**

We performed multiparametric MRI for 73 patients, radical prostatectomy for 33 patients, and targeted biopsy for 40 patients. From the bioptic tissue specimen, we established histology slides and stained them with Hematoxylin and Eosin (HE), ERG, and PIN-4 staining techniques. We obtained 103 cores (36 malignant, 67 benign). A pathologist annotated the index lesions in MRI ADC and the equivalent histology slides according to the highest Gleason Grade Group (GrG) and Prostate Imaging Reporting and Data System (PI-RADS) [5]. For each ROI of the ADC MRI images, we extracted 105 radiomic features with 3D Slicer. For each core from a histology slide, we extracted up to 172 shape and intensity features with QuPath.



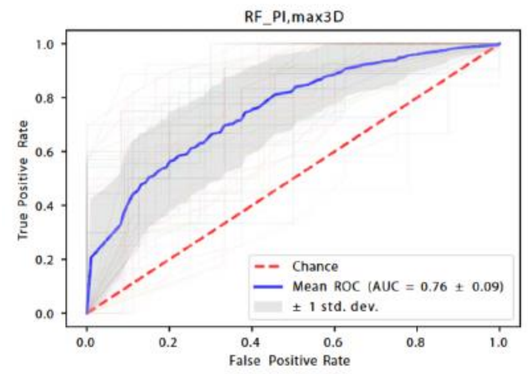

Tissue cores from histology slides stained with (A) the immunohistochemical marker containing anti-ERG antibodies, (B) hematoxylin and eosin, and (C) PIN-4 technique.

### Feature selection and predictive models based on features from MRI images [4]

Out of the 105 features, we selected the top four features with the lowest p-values of the Wilcoxon test: surface to volume ratio (SVR), joint entropy (JE), least axis (LA), and maximum 3D diameter (max3D). We built three classifiers on fifteen feature subsets: support vector machines classifier (SVC), random forest (RF), and neural networks (NN).

| | SVC | RF | NN |
|---|---|---|---|
| PI | $0.71 \pm 0.10$ | $0.71 \pm 0.08$ | $0.72 \pm 0.09$ |
| PI, SVR | $0.75 \pm 0.11$ | $0.58 \pm 0.10$ | $0.64 \pm 0.12$ |
| PI, LA | $0.75 \pm 0.10$ | $0.66 \pm 0.10$ | $0.69 \pm 0.11$ |
| PI, max3D | $0.73 \pm 0.11$ | $0.76 \pm 0.09$ | $0.67 \pm 0.12$ |

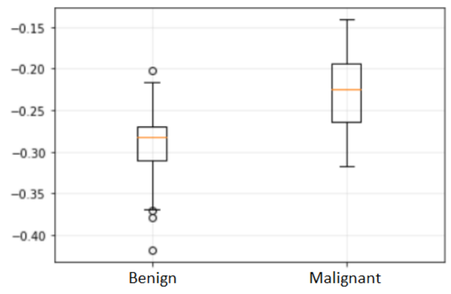
Mean values of area under the curve (AUC) metric for the SVC, RF, and NN models with four feature sets with features PI-RADS (PI), volume ratio (SVR), least axis (LA), and maximum 3D diameter (max3D). For each model, we performed 100-fold cross-validation).




Mean receiver operating curve (ROC) (blue) for the best random forest model trained on the features PI and max3D. The plot is based on the 100-fold cross-validation. The adjacent gray area depicts ± one standard deviation.

### Differentiation between malignant and benign prostate tissues with bioptic tissue specimen

We selected 12 and 13 significant features (Wilcoxon test, $p < 0.001$) of the color channel Brightness for the HE and ERG stainings, respectively. For the PIN-4 staining, we selected 16 features with $p < 0.001$ of the color channel Saturation. We trained SVC, RF, and NN models on the selected features for each staining.




Boxplot for the feature “Brightness: Haralick information measure of correlation” ($p < 0.001$, Gini coefficient=0.71) for HE staining to differentiate between benign and malignant tissues.

| | SVC | RF | NN |
|---|---|---|---|
| HE | $0.91 \pm 0.05$ | $0.83 \pm 0.07$ | $0.88 \pm 0.07$ |
| ERG | $0.86 \pm 0.06$ | $0.85 \pm 0.06$ | $0.86 \pm 0.06$ |
| PIN-4 | $0.94 \pm 0.05$ | $0.92 \pm 0.05$ | $0.94 \pm 0.04$ |

Mean values of ROC AUC metric for the SVC, RF, and NN models trained on datasets with 12, 13, and 16 significant features for HE, ERG, and PIN-4 stainings, respectively. For each model, we performed 100-fold cross-validation).

### Prediction of clinically significant prostate cancer

With the features from bioptic tissue specimen, the differentiation between clinically significant and insignificant prostate cancer was not possible. For the ERG and PIN-4 stainings, no features were significant with $p < 0.05$. For the HE staining, we could identify only 5 features with $p < 0.05$ of the color channel Brightness. For the RF model, the mean AUC value reached 0.69 ± 0.14.

## Conclusion

1. All five MRI sequences performed well in terms of intra- and interobserver reproducibility.
2. The sequences T1w TSE and T2 Map had the highest amount of robust features. The sequences T2w TSE and T2w FLAIR showed the poorest test-retest robustness.
4. We could differentiate between clinically significant and insignificant prostate cancer (AUC = 0.76, RF) with the radiomic features extracted from MRI images of prostate.
5. We could differentiate between malignant and benign bioptic tissue specimen (AUC = 0.94, PIN-4 staining, NN and SVC).
6. Differentiation between clinically significant and insignificant prostate cancer based on bioptic tissue specimen is a challenging process.

**References**

[1] Baeßler, B., Weiss, K., dos Santos, D. P. (2019). Robustness and reproducibility of radiomics in magnetic resonance imaging: a phantom study. Investigative radiology, 54(4), 221-228.

[2] Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J. C., Pujol, S., ... & Buatti, J. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magnetic Resonance Imaging, 30(9), 1323-1341.

[3] Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., ... & James, J. A. (2017). QuPath: Open source software for digital pathology image analysis. Scientific Reports, 7(1), 1-7.

[4] Bernatz, S., Ackermann, J., Mandel P., ..., Zhdanovich, Y., ..., Koch I., Chun F. K.-H., Köllermann J., Wild, P.J., Vogl, T.J. (2020). Comparison of machine learning algorithms to predict clinically significant prostate cancer of the peripheral zone with multiparametric MRI using clinical assessment categories and radiomic features. European Radiology, 1-13.

[5] Weinreb, J. C., Barentsz, J. O., Choyke, P. L., Cornud, F., Haider, M. A., Macura, K. J., ... & Thoeny, H. C. (2016). PI-RADS prostate imaging–reporting and data system: 2015, version 2. European Urology, 69(1), 16-40.