Faculty of physical-mechanics engineering, Systems Engineering and Informatics
Industrial University of Santander

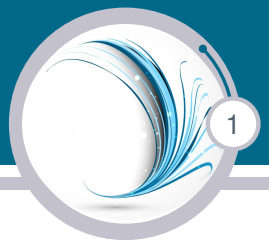
PROSTATEx - Dataset

v. 1.0.0

Yesid Gutiérrez
yesid.gutierrez@correo.uis.edu.co

October 9, 2018

Content



Introduction I
Introduction

The American Association of Physicists in Medicine (AAPM), along with the SPIE (the international society for optics and photonics) and the National Cancer Institute (NCI) conducted a “Grand Challenge” on quantitative image analysis methods for the diagnostic classification of clinically significant prostate lesions.

Introduction I
Data type

The PROSTATEx- Data-set is a set of MR studies, all studies include the following types of images:

- T2-weighted (T2W),were acquired using a turbo spin echo sequence and had a resolution of around 0.5 mm in plane and a slice thickness of 3.6 mm.
- Proton density-weighted (PD-W), was acquired prior to the DCE time series using the same sequence with different echo and repetition times and a different flip angle.
- Dynamic contrast enhanced (DCE), was acquired using a 3-D turbo flash gradient echo sequence with a resolution of around 1.5 mm in-plane, a slice thickness of 4 mm and a temporal resolution of 3.5 s.
- Diffusion-weighted (DW), were acquired with a single-shot echo planar imaging sequence with a resolution of 2 mm in-plane and 3.6 mm slice thickness and with diffusion-encoding gradients in three directions.

All the images were acquired without an endorectal coil.

Introduction I
Numbers

PROSTATEx Data-set in numbers:

Number of Patients	346
Number of Studies	349
Number of Series	18,321
Number of Images	309,251

The data-set contains two sub-sets the Training cohort (204 subjects) and the testing cohort (140 subjects)

Data-set description
Encoding

The images come in two encodings:

- `DICOM (MR images)`
- `MHD (Ktrans images)`

Ktrans is a a key pharmacokinetic parameter computed from the available Dynamic contrast enhanced T1-weighted series. Each patient has 1 Ktrans image and a study with several series of DICOM files, each K trans image is encoded in two files [mhd/zraw].

Data-set description I
Documentation

The DICOM files and the Ktrans images are documented in the "ProstateX-Images.csv" and "ProstateX-Images-KTrans.csv" respectively. The columns in these files encode the following:

- ProxID – ProstateX patient identifier.
- Name – Series Description
- Studydate – Study Date
- fid – Finding ID
- Pos – Scanner Coordinate position of the finding
- WorldMatrix – Matrix describing image orientation and scaling
- ijk – image col,row,slice coordinate of finding
- ImageUID – Image Identifier
- TopLevel
  - 0 - Series forms one image
  - 1 – A set of Series forms a 4D image (e.g. Dynamic MR)
  - NA – Series form one image, but is part of a Level 1 4D image

Data-set description II
Documentation

- SpacingBetweenSlices – Scalar Spacing between slices
- VoxelSpacing – Vector with x,y,z spacing scalars
- Dim – Vector with 4D dimensions of the image
- DCMSerDescr – The original DICOM Series Description
- DCMSerUID – The DICOM Series UID
- DCMSerNum – The DICOM Series Number
- InstanceUIDList – DICOM Instances that make up this series
- ImageUIDList – TopLevel-NA Images the make up this Toplevel 1 image

Data-set description
Findings

The findings are documented in the ProstateX-Findings.csv table. Documentation for the columns in that table is as follows:

- ProxID – ProstateX patient identifier
- fid - Finding ID
- pos - Scanner Coordinate position of the finding
- ClinSig – Identifier available in training set that identifies whether this is a clinically significant finding. Either the biopsy GleasonScore was 7 or higher. Findings with a PIRADS score 2 were not biopsied and are not considered clinically significant. In our center the occurrence of clinically significant cancer in PIRADS 2 lesions is less than 5 %

Data-set description
Acknowledgements

The prostate MR imaging was performed at the Radboud University Medical Centre (Radboudumc) in the Prostate MR Reference Center under supervision of prof. Dr. Barentsz. The Radboudumc is located in Nijmegen, The Netherlands. The dataset was collected and curated for research in computer aided diagnosis of prostate MR under supervision of Dr. Huisman, Radboudumc.

References
References

- https://prostatex.grand-challenge.org/
- https://wiki.cancerimagingarchive.net/display/Public/SPIE-AAPM-NCI+PROSTATEx+Challenges
- https://www.aapm.org/GrandChallenge/PROSTATEx-2/default.asp
- Python Library

Thanks for the attention!