Medical Image Analysis 55 (2019) 27–40

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

# Unsupervised tumor detection in Dynamic PET/CT imaging of the prostate

Eldad Rubinstein[a,1,*], Moshe Salhov[a,1], Meital Nidam-Leshem[c], Valerie White[d], Shay Golan[b], Jack Baniel[b], Hanna Bernstine[c], David Groshar[c], Amir Averbuch[a]

[a] School of Computer Science, Tel Aviv University, Tel Aviv, Israel
[b] Department of Urology, Rabin Medical Center, Petach Tikva, Israel
[c] Department of Nuclear Medicine, Rabin Medical Center, Petach Tikva, Israel
[d] The Institute of Pathology, Ichilov Medical Center, Tel Aviv, Israel

ARTICLE INFO

*Article history:*
Received 2 June 2018
Revised 30 March 2019
Accepted 5 April 2019
Available online 6 April 2019

*Keywords:*
PET
Prostate
Kinetic modeling
Autoencoder
Density estimation

ABSTRACT

Early detection and localization of prostate tumors pose a challenge to the medical community. Several imaging techniques, including PET, have shown some success. But no robust and accurate solution has yet been reached. This work aims to detect prostate cancer foci in Dynamic PET images using an unsupervised learning approach. The proposed method extracts three feature classes from 4D imaging data that include statistical, kinetic biological and deep features that are learned by a deep stacked convolutional autoencoder. Anomalies, which are classified as tumors, are detected in feature space using density estimation. The proposed algorithm generates promising results for sufficiently large cancer foci in real PET scans imaging where the foci is not viewed by the tomographic devices used for detection.

© 2019 Elsevier B.V. All rights reserved.

## 1. Introduction

Prostate cancer is a common form of cancer (Schwarzenböck et al., 2012). Its standard diagnosis procedure starts by measuring the presence of the protein Prostate Specific Antigen (PSA) in the blood. It is considered alarming if the measured PSA is above a certain threshold. The next exploratory step for patients with high PSA values is to apply an ultrasound-guided biopsy, during which prostate tissue is randomly sampled. However, this process is considered inaccurate and indecisive. It can produce both misdetection and false alarm especially in early stages of a tumor growth. Furthermore, this procedure can also inflame the patient.

Several imaging techniques such as Positron Emission Tomography (PET) (Conti, 2014; Umbehr et al., 2013) and Magnetic Resonance Imaging (MRI) (Hricak et al., 2007) have been proposed to improve the prostate cancer detection rate. PET, which is an imaging modality that produces a functional image of the body or part of it, is based on injecting a radioactive tracer into the patient's body. $^{11}$C-choline is a commonly used tracer in prostate cancer imaging. The ability of this tracer to highlight tumors in a PET image is related to its participation in the cells membrane synthesis process (de Jong et al., 2002).

PET or PET/CT imaging-based detection and localization of prostate tumors have been studied in Beheshti et al. (2010), Farsad et al. (2005), Giovacchini et al. (2008), de Jong et al. (2002), Piert et al. (2009), Testa et al. (2007), Tuncel et al. (2008) and Umbehr et al. (2013). Farsad et al. (2005) discuss a sextant based analysis, which is based on division of the prostate into six parts, while achieving 66% sensitivity and 81% specificity. Testa et al. (2007) perform a similar analysis and report 55% sensitivity and 86% specificity.

Ultrasound signals of the prostate were analyzed using various machine learning techniques in Feleppa et al. (2009). The ultrasound images were acquired from each patient immediately before prostatic biopsies. By using stacked restricted Boltzmann machines to detect the cancer foci, an Area Under the Curve (AUC) of $0.91 \pm 0.04$ was achieved in Feleppa et al. (2009).

Deep learning is part of a broad family of methods for Representation Learning (Goodfellow et al., 2016; Srivastava et al., 2014). It has enabled to achieve state-of-the-art results in many diverse domains (Goodfellow et al., 2016), that include medical image analysis (Litjens et al., 2017). Deep learning can be used for unsupervised feature extraction. For example, a stacked convolutional autoencoder is utilized to extract feature from data in Masci et al. (2011). Furthermore, Autoencoders have been used

* Corresponding author.
*E-mail address:* eldada333@gmail.com (E. Rubinstein).
[1] Equal contribution.

https://doi.org/10.1016/j.media.2019.04.001
1361-8415/© 2019 Elsevier B.V. All rights reserved.

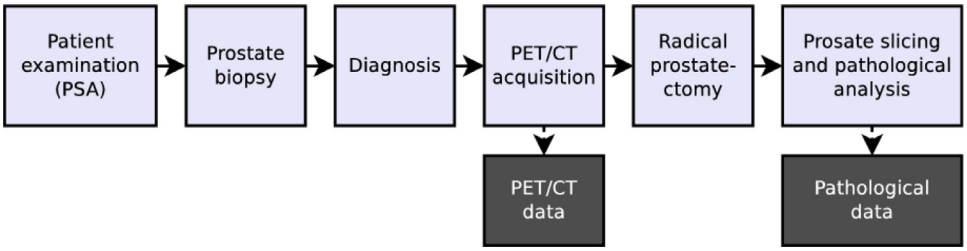



**Fig. 1.** Medical flow of the patient screening, imaging and prostate prostatectomy.

in Shin et al. (2011) for time series analysis with a large dataset of unlabeled MRI images. Their goal was automation of tissue characterization. Stacked sparse autoencoders are used for prostate MR images segmentation in Guo et al. (2016). Comparison of hand-crafted features with autoencoder based features for the task of oil wells failures prediction is given in Liu et al. (2015). This comparison concludes that autoencoder-based features provide better results than hand-crafted features.

This work aims to improve early detection and localization of prostate tumors by combining hand-crafted perfusion-based features with deep autoencoder-based features generated from a Dynamic $^{11}$C-choline-PET/CT imaging. While the proposed methodology is demonstrated for $^{11}$C-choline tracer, it can be applied to any radioactive tracer such as Gallium-68 PSMA and others.

The remainder of the paper is organized as follows: Section 2 describes the overall research flow, including the medical flow and materials. Section 3 describes the proposed unsupervised detection algorithm including the preprocessing stage, feature extraction and detection methods. It also explains the evaluation process. Section 4 presents the dataset and the experimental results. Finally, Section 5 includes conclusions and proposals for further research.

## 2. Materials and methods

This paper describes a research work that has been conducted in cooperation with Rabin Medical Center, Petach Tikva, Israel. Three departments participated: urology/oncology, nuclear medicine and pathology. The patient screening, imaging and prostatectomy flow is described in Fig. 1. Initially, patients with PSA above a threshold go through a prostate biopsy. If the pathological diagnosis of the prostate biopsy recommends to perform a radical prostatectomy and the patient agrees for it, then the patient is injected with $^{11}$C-choline tracer and scanned using a Dynamic PET/CT imaging prior to the prostatectomy operation. Then, a prostate removal operation takes place. The harvested prostate is sliced using a whole mount procedure, diagnosed and the cancerous cells are marked by a pathologist. In terms of data, the output of this procedure is twofold and includes both the imaging data and the marked pathological data.

### 2.1. Patients

A prospective case-series design was used. The study group used in this paper consists of consecutive patients that have localized biopsy-proven adenocarcinoma of the prostate who underwent robot-assisted radical prostatectomy performed by a single surgeon at a tertiary medical center in 2012–2014. The local institutional review board approved the study. All patients provided written informed consent to participate in the study. Data were collected on demographics, PSA values, and preoperative biopsy results. Standard Dynamic $^{11}$C-choline PET/CT was performed 2 months or less before surgery and at least 3 months after prostate biopsy.

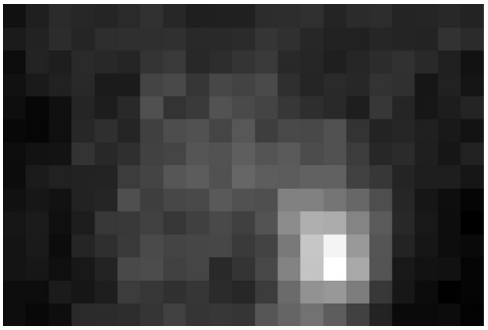


**Fig. 2.** An example of PET image. The image is trimmed around the prostate and averaged over the axes $z$ and $t$. A large tumor can be easily spotted in the bright pixels in the bottom-right corner.

### 2.2. PET/CT protocol

The acquired Dynamic PET data is a 4D volume, which is originally stored in DICOM format. An example of PET image is shown in Fig. 2. Images were obtained using an integrated 8-slice PET-CT scanner (Discovery ST, GE Medical Systems, Milwaukee, WI). The protocol included a Dynamic PET acquisition limited to a single bed position and centered on the prostate, starting with the injection of 10–20 mCi (370–740 MBq) of $^{11}$C-choline with a low-dose CT scan. After a scout view of the pelvis was obtained, the study centered on the prostate with PET coverage of 15.3 cm. A non-diagnostic low-dose (30 mA) CT scan was acquired. $^{11}$C-choline was injected as a rapid bolus flushed with 50 cc saline 0.9% at a rate of 5.0 mL/s using an automatic power injector (Dual-shot, Nemoto, Japan). The dynamic study was acquired as a 3D scan (matrix size $64 \times 64$ and $256 \times 256$, 3.27-slice thickness) consisting of 18 sequential frames of 10 s each followed by 7 frames of 60 s each. After re-sampling the data, each time series contains 61 members of equally sampled time intervals. The times series of the Dynamic PET are called Time Activity Curves (TACs). Section 3.1 provides more details. PET emission data after attenuation correction was reconstructed with a 3D OSEM algorithm (Hudson and Larkin, 1994) (2 iterations, 20 subsets).

The acquisitions of PET data and the corresponding CT data, although not occurring simultaneously, are done in the same study on the same PET/CT machine without any movement of the patient between CT/PET acquisitions. This acquisition procedure resulted in registered CT and PET imaging data.

#### 2.2.1. CT data and prostate boundary

The acquired CT scan is a 3D volume with a resolution of $512 \times 512 \times 60$ voxels that represents the patient anatomy. The boundary of the prostate was marked over the PET-registered (Fig. 3) CT scan by two expert radiologists as shown in Fig. 4.

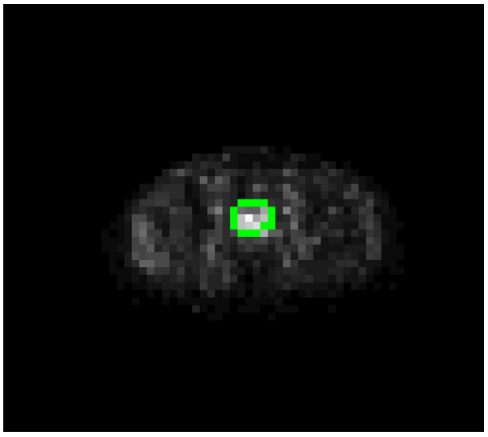


**Fig. 3.** A prostate boundary on top of a PET imaging slice. The prostate boundary (green), which was drawn over a CT image, was aligned to the PET image using the PET-CT registration from Section 2.2 . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

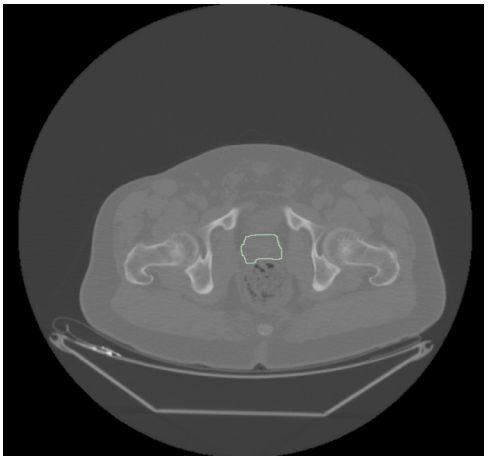


**Fig. 4.** A hand marked (by a radiologist) of the prostate boundary on top of a CT imaging slice. The prostate boundary is the light green contour line in the middle of the image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 2.3. Standard of reference

The surgical specimens served as the standard of reference. The harvested prostates were submitted for whole-mount histology. Axial sections of 3–5 mm were obtained in a plane perpendicular to a theoretical line between the base and the apex. Orientation was preserved by inking right and left margins with different colors (black-right, red-left) and by anatomical landmarks (i.e., seminal vesicle and urethra). The whole-mount sections of 4 um each were stained with hematoxylin and eosin according to a standard procedure. A Board-certified uropathologist reviewed the slides and identified cancer foci by laterality and prostatic areas (base, mid-gland, and apex) and with respect to anatomic landmarks, as previously described. The identified cancer foci were marked on a 2D map that corresponded and registered to each sliced prostate and served as 2D label. The 2D label includes the boundary of the prostate, the boundary of the tumors and the anatomical landmarks. Fig. 5 presents an example of a 2D label generated from a prostate slice.

### 2.4. Matching PET/CT and histopathology

A radiologist expert, specialized in nuclear medicine, assessed the correspondence between the PET/CT images and the histopathology findings. The most evident locations of a prostate tumor and a normal prostate tissue segment were visually identified on the PET/CT images. The location of the prostate tumor was then confirmed by the pathological report with respect to anatomic landmarks.

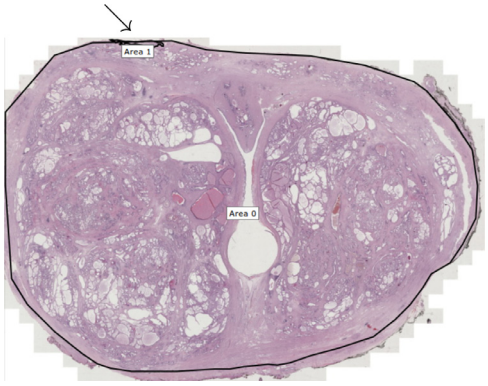



**Fig. 5.** A 2D label generated from a prostate whole-mount slice that contains two marked areas. Area 0 is the contour of the whole prostate. Area 1 in the upper left corner (see arrow) is a small tumor.

In order to reduce the registration-based errors effect when comparing the anomaly detection results to the Board-certified uropathologist hand made label, all the labels, which correspond to the same prostate octant, were aggregated into a single label similar to Beheshti et al. (2010) and Farsad et al. (2005).

The 2D pathology labels are used to compute the label of each octant as follows: the pathological volume, which consists of the pathological slices, was split by half in the middle of each dimension such that there are total of $2^3 = 8$ octants. Each octant was marked by one of three labels: big tumor, small tumor and normal tissue (no cancerous cells)[2]. The anatomical landmarks are used to verify that the pathological slices are rotated to align with the PET image.

### 2.5. Detection and evaluation process

An anomaly detection procedure was applied to features extracted from the Dynamic PET data. The extracted features were engineered by using three methodologies: statistical features based on TACs and their histograms, kinetic modeling features and deep learning-based features.

A flow chart of tumors detection and its evaluation process is outlined in Fig. 6. Boxes (a)–(c) in the figure describe the input data. The detection process starts in box (d) with preprocessing of the images. In box (f), features are extracted from each voxel and a classifier algorithm decides whether each voxel is anomalous or not. The results from all voxels in the same octant (1/8 part) are then aggregated to produce a single label for each octant in the PET imaging. These labels are compared with the ground truth labels and a total performance score is computed – see boxes (e) and (g). Fig. 7 presents a graphical illustration of this flow.

## 3. Unsupervised detection method

Each acquired Dynamic PET imaging is treated as a single dataset. It consists of a 4D-volume whose dimensions are ($x$, $y$, $z$, $t$). Each voxel (3D pixel) in coordinates ($x$, $y$, $z$) is a data point in the dataset. The fourth dimension is time, i.e., each voxel is

[2] According to the pathological marks, tumors larger than 36 $mm^2$ are considered to be big.

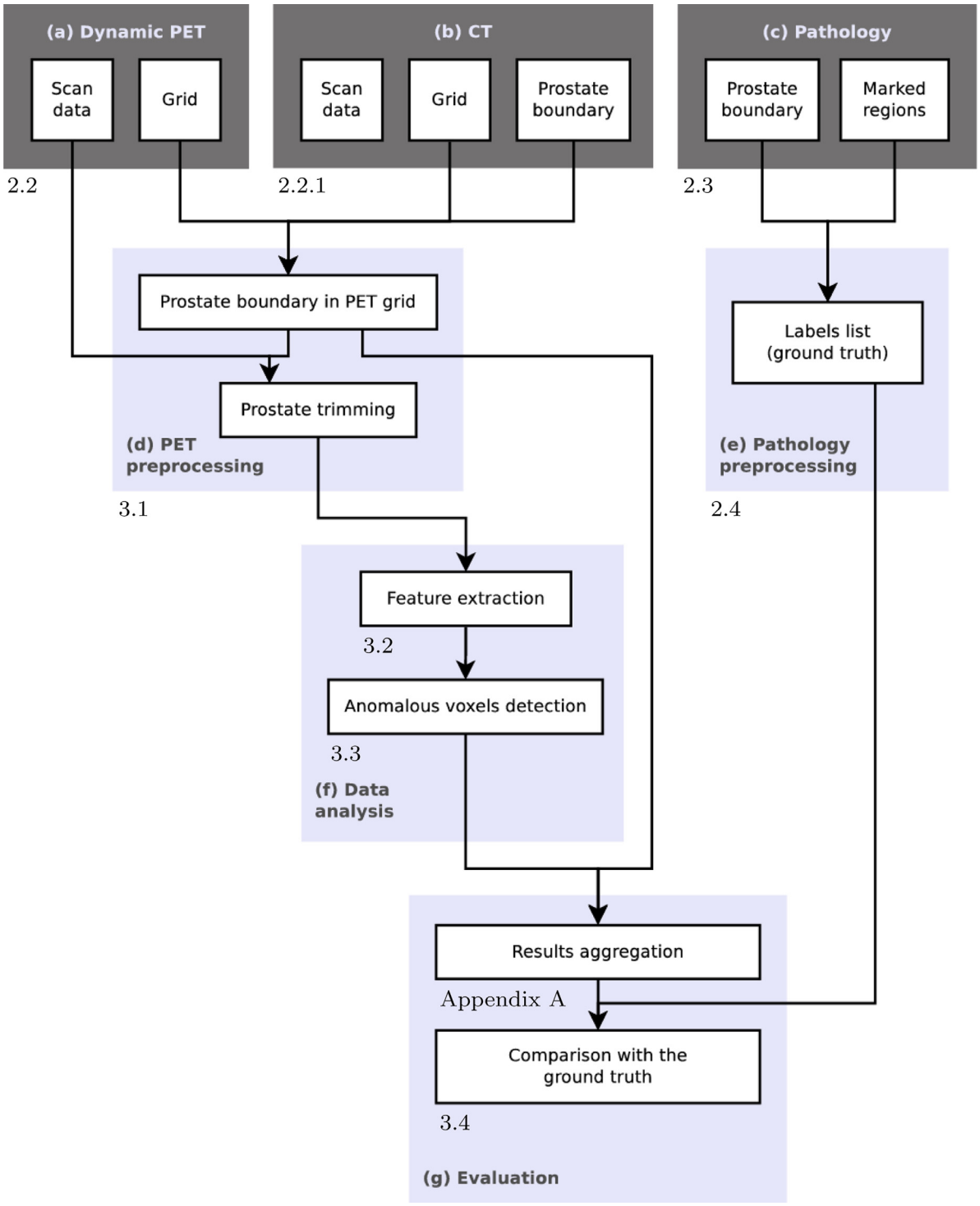



**Fig. 6.** Flow of the anomalies detection and evaluation process. The Dynamic PET/CT data is used for detecting anomalies. The marked pathological data is used as the ground truth. The numbers next to each box refer to the relevant section in the paper. (See Sections 2.2, 2.2.1, 2.3, 2.4, 3.1–3.4.)

measured over a time interval. The function $f_{x,y,z}(t)$ describes the time measurements of the $(x, y, z)$ voxel during the PET scan.

### 3.1. Dynamic PET preprocessing

The PET measurements are scaled by using the radioactive decay of the PET tracer. The Standard Uptake Values (SUVs) are computed from all the prostate's voxels. More specifically, the SUVs are computed as described in Dimitrakopoulou-Strauss et al. (2012) and Häggström (2014) to be

$$f_{x,y,z}(t) = SUV_{x,y,z}(t) = \frac{c_{x,y,z}(t)}{ID/BW}, \quad (1)$$

where $c_{x,y,z}(t)$ are the values of the tracer concentration in different times, $ID$ is the injected dose of the tracer and $BW$ is the patient's body weight. Finally, the PET time series is re-sampled to have a uniform time difference between sequential frames.

Each normalized SUV PET image is trimmed to a box that contains the prostate with one voxel margin around it. The prostate

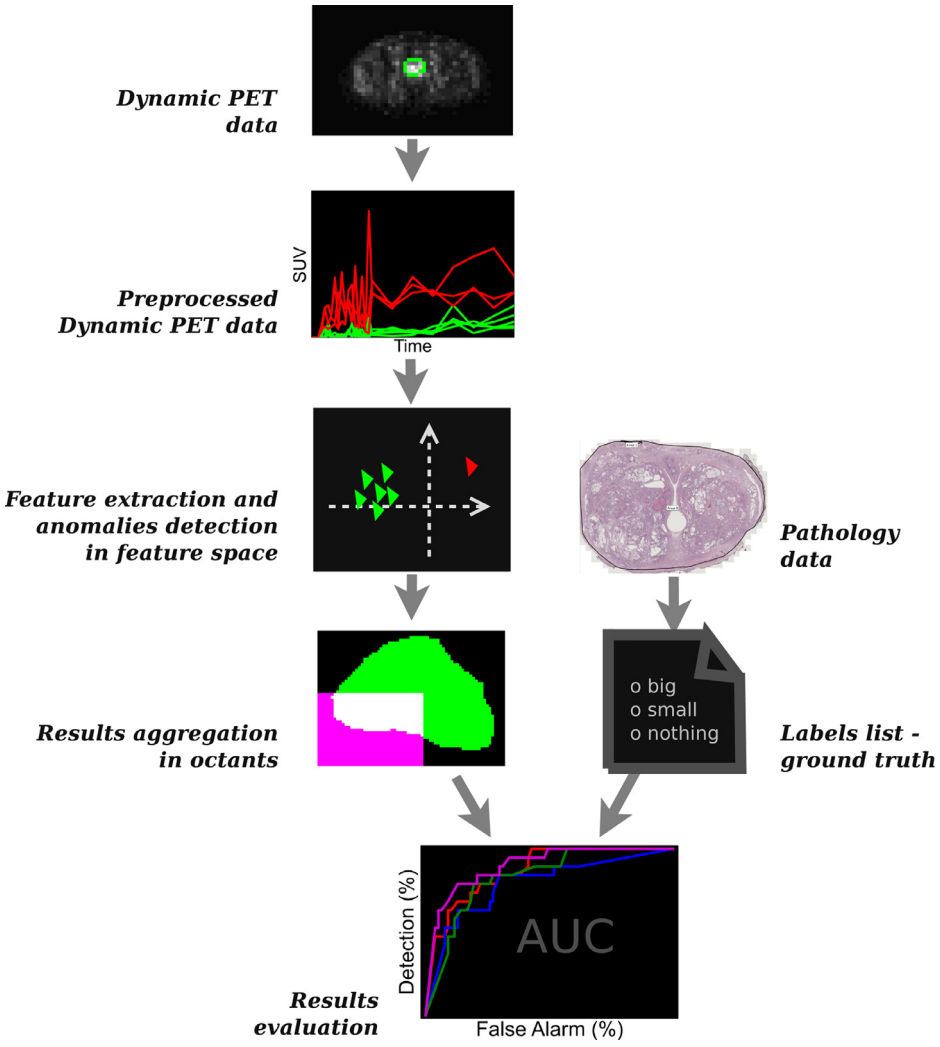



**Fig. 7.** Graphical illustration of the anomalies detection process and its evaluation.

**Table 1**
Extracted features from each voxel.

| Class | Feature Name | Description |
|---|---|---|
| Statistical | SUV TACs | The values $f(t_1)$, $f(t_2)$, ..., which constitute the original TAC of each voxel. |
| | Mean SUV | $\frac{1}{p}\sum_{i=1}^{p} f(t_i)$, where $p$ is the length of the TAC. |
| | Partial SUV mean | $\frac{1}{p'}\sum_{i=1}^{p'} f(t_i)$ for some $p' < p$. |
| | SUV histogram vector (SHV) | $\ell$ counters (bins) $c_1, \ldots, c_\ell$, such that each is used as a feature. Each counter contains the number of $f(t)$ values in a certain range. The bins are not necessarily evenly spaced and not necessarily evenly weighted. |
| | SHV mean | $\frac{1}{\ell}\sum_{i=1}^{\ell} c_i$ |
| | SHV skewness | The skewness (normalized third moment) of the vector $(c_1, \ldots, c_\ell)$. |
| | SHV 75% | The 75% of the vector $(c_1, \ldots, c_\ell)$. |
| Kinetic Modeling | $k_1$ coefficient from the 2-compartments model (Morris et al., 2004) | It represents the flow of the tracer from the blood into the first compartment, after fitting the data to two compartments Kinetic Model. |
| Deep autoencoder | Reconstruction errors | A stacked convolutional autoencoder (Masci et al., 2011) is trained using PET data. The reconstruction errors in different training epochs are used as features. |

bounding box is derived from a boundary contour that was generated and agreed upon by two expert radiologists using the registered CT imaging.

### 3.2. Feature extraction

The raw (measured) and normalized features of each PET image are the time series $f_{x,y,z}(t)$ from Eq. (1). The features, which are extracted from $f_{x,y,z}(t)$, are listed in Table 1. The extracted features were engineered using three methodologies: I. Statistical features that are based on TACs and their histograms, II. Kinetic modeling features that measure the flow of the tracer under a compartment model assumption and III. Deep learning-based features that are learned in an unsupervised manner using a stacked convolutional autoencoder architecture.

#### 3.2.1. Statistical features

The class of statistical features include the SUV-normalized TACs and their statistics. The Mean SUV feature is a measure of

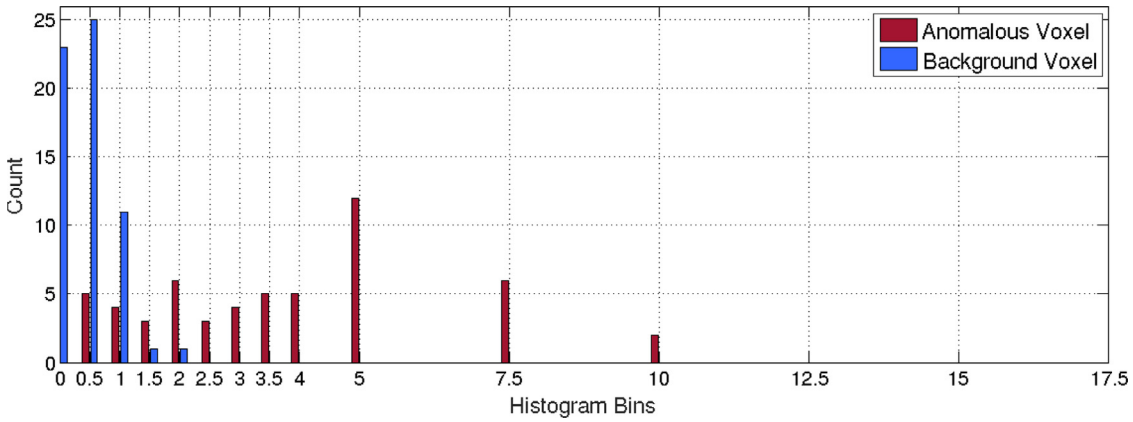



**Fig. 8.** Histograms of a background voxel (blue) and an anomalous voxel (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the total amount of tracer that passed through a voxel during the whole PET scan.

The SUV histogram features (Fig. 8) measure the distribution of SUV TAC values over its time span. The first and the last histogram bins are chosen to cover at least 99.5% of the elements of all time series in the dataset. Therefore, the value of each histogram bin can be viewed as the unnormalized probability of a time sample to be inside the range that the bin represents.

In addition to the usage of histogram bins (SHV), several statistical features (SHV mean, skewness and 70% percentile) are extracted from the histograms.

The assembly of the different features is described in Section 3.3.1.

*3.2.2. Kinetic modeling-based features*

Kinetic Models, or Compartment Models, are a common way to analyze Dynamic PET images (Coxson et al., 1992; Häggström, 2014; Morris et al., 2004). They are based on fitting the Dynamic PET data to biological temporal models. In this work, the Dynamic PET data was fitted to a two compartments model (Godfrey, 1983). The Arterial Input Function (AIF) for the model is extracted from the PET scan. This function represents the concentration of the tracer in the blood. The extraction was done before the image was trimmed along the prostate boundary. The output of this fitting procedure is the model coefficients for each voxel also known as *parametric maps* of the image. The most useful coefficient in the compartment model is $k_1$, which represents the flow of the tracer from the blood into the first compartment. The fitting was achieved by using the Pmod software (Pmod Technologies Ltd., 2016).

*3.2.3. Deep (Autoencoder Based) features*

Unsupervised feature extraction, that is based on a stacked convolutional autoencoder, is described in Masci et al. (2011). The main idea is that informative features can be extracted from the data by learning a transformation into a low dimensional space (encoder) while reducing the reconstruction error between the original data and the reconstructed data from the low dimensional space (decoder). We use the reconstruction errors as features, under the assumption that they will be higher for anomalous voxels (Chandola et al., 2009; Japkowicz et al., 1995; Manevitz and Yousef, 2007).

In order to compute the features, an autoencoder is trained using a unified dataset that combines all patient data.

**Creation of a unified dataset:** In order to generate the dataset for training the autoencoder, all patients data were combined together into a single dataset. All TACs in the unified dataset were deconvoluted from their corresponding AIF by using a circular deconvolution. This was done in order to normalize each TAC according to the AIF of the corresponding patient (Morris et al., 2004).

**Autoencoder setup and training:** The stacked convolutional autoencoder architecture is detailed in Fig. 10. Its upper half of the layers (encoder) consists of three sets of a convolution layer, a max pool layer and a dropout layer (Goodfellow et al., 2016; Srivastava et al., 2014), followed by a small fully connected (dense) layer. The bottom half of the layers decodes the data back to its original form (decoder). The weights of the network nodes in the bottom half are kept tied to the nodes of the matching layers in the upper half. The numbers next to each layer name in Fig. 10 represent the layer size. For instance, the first convolutional layer is of size $60 \times 12$, which means that its input is a vector of size 60 and it consists of 12 convolutional filters.

The last layer is an output layer whose values are almost identical to the values of the input layer, except an additive Gaussian noise that is added to the input and re-generated in every training epoch of the network.

The $L_2$ reconstruction error $\zeta_{i,k}$, at the $k$th training epoch, is computed for the $i$th voxel of the unified dataset as

$$\zeta_{i,k} = \sqrt{\sum_{j=1}^{p} \left(v_{i,j} - \tilde{v}_{i,j,k}\right)^2} \quad i = 1, \ldots, \ell, \quad k = k', \ldots, n \tag{2}$$

where $\ell$ is the total number of voxels, $p$ is the length of each TAC, $n$ is the total number of training epochs, $v_{i,j}$ is the value of the $j$th dimension of the $i$th deconvoluted TAC and $\tilde{v}_{i,j,k}$ is its autoencoder reconstruction in the $k$th epoch. The $\zeta_{i,k}$ are stored to be used later as features.

The difference between the reconstruction of a background voxel and the reconstruction of an anomalous voxel is illustrated in Fig. 9. The convolutional autoencoder is trained using Stochastic Gradient Descent with Nesterov momentum (Sutskever et al., 2013).

### 3.3. Anomalous voxels detection

Given the engineered features, the anomaly detection process is initialized by fusing the features into a single dataset matrix. The fused features are analyzed to compute an anomaly score for each voxel.

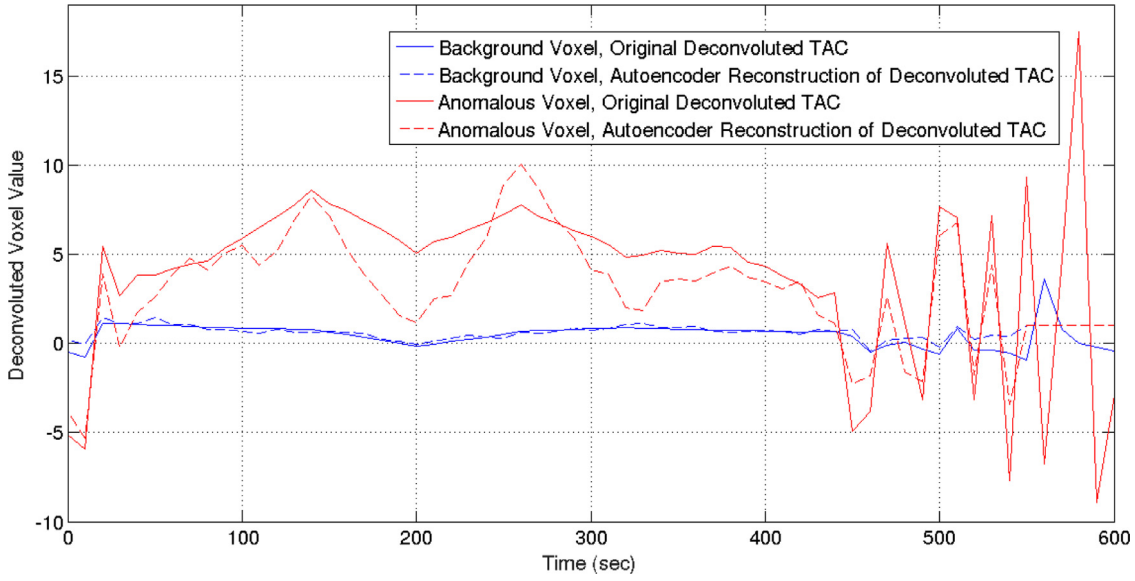



**Fig. 9.** Autoencoder reconstruction example of background and anomalous voxels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

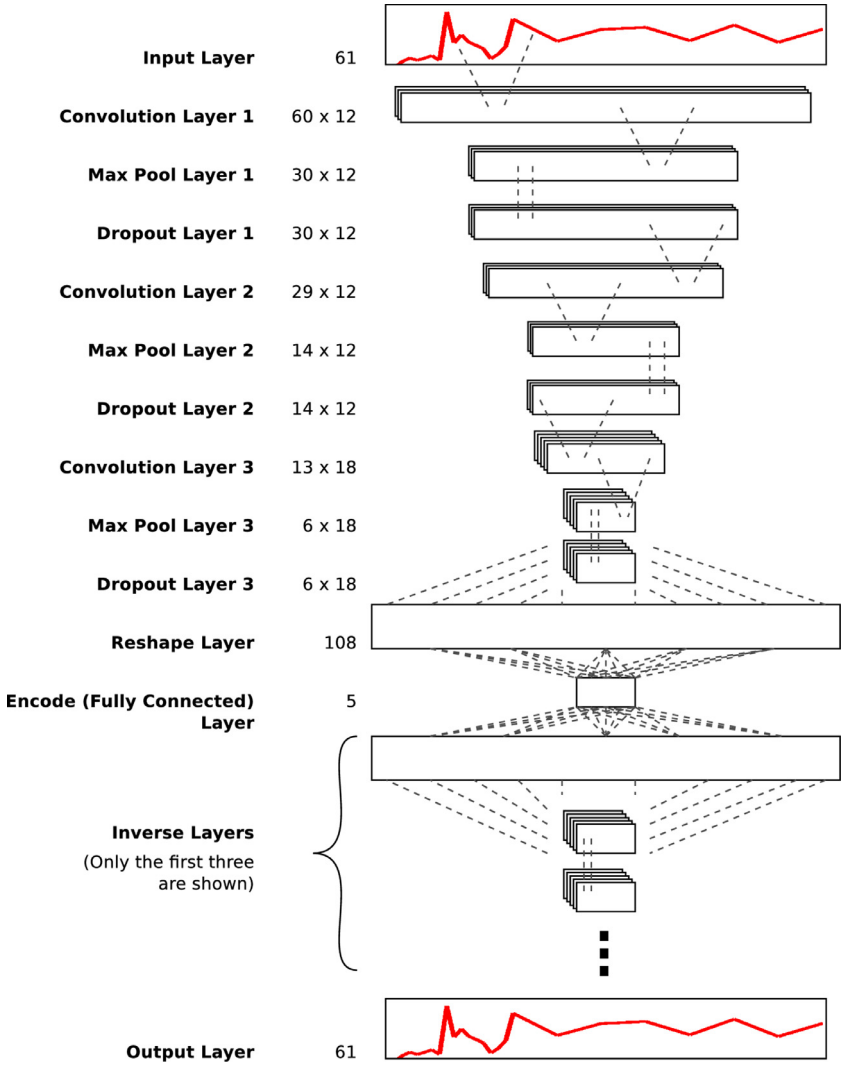



**Fig. 10.** Autoencoder network structure.

**Algorithm 1** Spectral Features Fusion.

**Input:** matrices $A_1, A_2, \ldots, A_\ell$ of sizes $n \times m_1$, $n \times m_2$, $\ldots$, $n \times m_\ell$, respectively, where $n$ is the number of data points (voxels), $m_i$ is the number of features in the $i$th matrix, $\ell$ is the number of features families, $\eta$ is the ratio between the largest eigenvalue and the smallest relevant eigenvalue and $P$% is the percentile scale factor.
**Output:** matrix $F$ of size $n \times m_T$, where $m_T$ is the total number of features.

$F \leftarrow$ empty matrix
**for** $i \leftarrow 1$ to $\ell$ **do**
  **if** $m_i > 1$ **then**
    Apply SVD to $A_i$ s.t $A_i = USV^T$ and $\lambda_k = [S]_{k,k}$, $k = 1, \cdots, m_i$
    Find $k$ s.t. $\lambda_{k+1} \leq \eta \cdot \lambda_1 \leq \lambda_k$
    $U_i' \leftarrow \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_k \\ | & | & & | \end{pmatrix}$
  **else**
    $U_i' \leftarrow A_i$
  **end if**
  Compute $\rho_i \leftarrow P$% of the first column of $U_i'$
  Scale $A_i' \leftarrow \frac{1}{\rho} U_i'$
  Concatenate $F \leftarrow \left[F\ A_i'\right]$
**end for**
**return** $F$

#### 3.3.1. Assembly of different features

The procedure for fusing all features is detailed in Algorithm 1. Each computed feature family is stored as a matrix $A_i$, $i = 1, \ldots, \ell$, where $i$ is the index of the feature family. The feature families include SUV TACs, SUV histogram vector, autoencoder reconstruction errors, kinetic modeling coefficients and other features that are described in Section 3.2. The rows of these matrices represent different voxels while the columns represent different features.

For each input matrix $A_i$, the spectral decomposition of $A_i$ is used to find the matrix $U_i'$ by extracting the significant columns from the left eigenvectors of $A_i$. The matrices $U_i'$, $i = 1, \ldots, \ell$, are scaled and concatenated to the output matrix $F$. Scaling is performed by dividing $U_i'$ with a scale factor $\rho$ which is $P$% of the first left eigenvector of $A_i$. The result matrix $F$ is used in Algorithm 2.

**Algorithm 2** Anomalies detection.

**Input:** matrix $F$ of size $n \times m_T$, where $n$ is the number of voxels and $m_T$ is the number of features, threshold $\beta$, function $f(F, \beta)$.
**Output:** $I_a$ - list of anomalous voxels indices.

Compute scores and thresholds - $\alpha_i, \xi \leftarrow f(F, \beta)$
Apply threshold - $I_a \leftarrow$ indices of $\alpha_i$ s.t. $\alpha_i > \xi$
**return** $I_a$

#### 3.3.2. Computing Voxels Anomaly Scores

Algorithm 2 detects the anomalies on $F$, under the assumption that isolated data points in the features space, where the density is relatively low, represent anomalies.

The first step in Algorithm 2 computes the voxels anomaly scores $\alpha_i$, $i = 1, \ldots, n$, where $n$ is the number of voxels, and the decision threshold $\xi$. It uses a given function $f(F, \beta)$ where $\beta$ is an input threshold. There are several functions that can be used as $f(F, \beta)$:

**Algorithm 3** Anomalous octants detection.

**Input:** $I_a$ is the list of anomalous voxels indices, $P_r$, $r = 1, \ldots, 8$ is the list of voxels indices in each octant, threshold $\nu$.
**Output:** $I_c$ - list of anomalous octants.

**for** $r \leftarrow 1$ to $8$ **do**
  Compute score per octant: $s_r \leftarrow \#(P_r \cap I_a)$
**end for**
Compute maximal score: $\tau \leftarrow \max_{1 \leq r \leq 8} s_r$
Find $I_c$ as the set of indices $r = 1, \ldots, 8$, s.t. $s_r > \tau \cdot \nu$
**return** $I_c$

1. **Threshold function:** The decision is made by setting a threshold $\beta$ to a certain % of the feature data. Given $\beta$ all the voxels whose feature values are larger than the threshold are classified as anomalous. It is possible to use this function only if $F$ has a single column, which means that the total number of features is $m_T = 1$.
2. ***k*-Nearest Neighbors** (Bishop, 2006; Chandola et al., 2009): The $i$th voxel is considered anomalous if its distance to its $k$th neighbor $\alpha_i$ is above a threshold $\xi$. The parameter $k$ governs the degree of smoothing of the density estimation (Bishop, 2006). The decision threshold is set to $\xi = \mu + \beta \cdot \sigma$, where $\mu$ is the mean of the distances from all data voxels to their $k$th neighbor and $\sigma$ is the standard deviation of these distances. The $k$-Nearest Neighbors method uses the Euclidean distance as its metric.
3. **Parzen Windows** (Bishop, 2006; Chandola et al., 2009): In this case, the threshold $\beta$ is a certain % of the density estimation value. The Parzen Windows method uses the Euclidean distance and the Gaussian kernel. Scott's rule (Härdle et al., 2004; Scott and Sain, 2004) is used to determine the window size in each dimension. The implementation of Parzen Windows is taken from Ihler and Mandel (2007).

Section 4 explains how each method is used as a function $f(F, \beta)$. It also details the used parameters.

### 3.4. Prostate cancer detection evaluation

In order to compare the voxels' anomaly detection results with the ground truth pathological slices, we aggregate the results in the 8 octants $P_r$, $r = 1, \ldots, 8$. Each octant contains many PET voxels. Details about the division of the prostate into octants are given in Appendix A.

An anomaly score is calculated by Algorithm 3 for each octant. This algorithm counts the anomalous voxels in each octant and returns a list of the anomalous octants.

For performance evaluation, we compare the agreement between the detected anomalous octants and the pathology-based ground truth octant labels. We label these octants as either *anomalous* or *normal tissue (no cancerous cells)*. The pathological ground truth labels in each octant can be either *big tumor*, *small tumor* or *normal tissue (no cancerous cells)*.

The performance evaluation consists of two detection tasks: firstly, the task of separating *big tumor* labeled[3] octants from *small tumor* and *no cancerous cells* labeled octants, and secondly, the task of separating *small tumor* labeled octants from *no cancerous cells* labeled octants while ignoring the *big tumor* labeled octants.

[3] Here the term *labeled* refers to the ground truth labels. A tumor area threshold is used to decide which label is used in each octant.

**Table 2**
All possible scenarios for describing the outcome in prostate octants. Each two-letters-combination *XY* stands for the count of octants in all tested cases whose detection result is $X \in \{A, N\}$ and ground truth is $Y \in \{N, S, B\}$.

| Two-letters-combination | Detection result | Ground truth |
|---|---|---|
| *AN* | Anomaly | No cancerous cells |
| *NN* | Normal | No cancerous cells |
| *AS* | Anomaly | Small tumor |
| *NS* | Normal | Small tumor |
| *AB* | Anomaly | Big tumor |
| *NB* | Normal | Big tumor |

The chosen method, which measures the detection results, is the area under the receiver operating characteristic (ROC) curve also known as the Area Under Curve (AUC).

In order to explain the measurement process, six two-letters-combinations are defined. Together, they cover all possible options for each prostate octant. The first letter describes the anomalies detection algorithm result: *A* for anomaly and *N* for normal. The second letter describes the ground truth: *N* stands for no cancerous cells, *S* for small and *B* for big. For example, *NB* stands for the number of octants in all tested cases that seem to contain no tumors according to Algorithm 3, but actually contain a big one. The other five combinations can be defined accordingly. A list of all possible combinations is given in Table 2.

The false alarm rate *F*, which lies in the *x*-axis of the ROC graph, is given by

$$F = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}}.$$

In our case, it is given by

$$F_B = \frac{AN + AS}{AN + AS + NN + NS}$$

and

$$F_S = \frac{AN}{AN + NN},$$

for big ($F_B$) and small ($F_S$) tumors, respectively. The false alarm rate is closely related to the *specificity* value that is given by

$$\text{Specificity} = 1 - F.$$

The sensitivity *D*, which lies in the *y*-axis of the ROC graph, is given by

$$D = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}.$$

In our case, it becomes

$$D_B = \frac{AB}{AB + NB}$$

and

$$D_S = \frac{AS}{AS + NS},$$

for big and small tumors, respectively.

In order to produce the ROC curve, the sensitivity and the false alarm rates are generated and recorded for each value of the aggregation stage threshold, which is denoted by $\nu$ in Algorithm 3. This threshold sets the border between anomalous and background prostate octants. The threshold for the individual voxels decision, which is denoted by $\beta$ in Algorithm 2, stays constant through this process.

**Table 3**
Clinical and pathologic characteristics.

| Characteristic | Value |
|---|---|
| Age, years, mean (SD) | 64.1 (5.4) |
| PSA, ng/ml (SD) | 7.5 (5.3) |
| Prostate volume, $cm^3$, average (SD) | 52.7 (28.3) |
| Final pathologic grade, n (%) | |
| Gleason $(3+3)=6$ | 6 (25) |
| Gleason $(3+4)=7$ | 13 (55) |
| Gleason $(4+3)=7$ | 3 (12) |
| Gleason $(4+5)=9$ | 1 (4) |
| Gleason $(5+4)=9$ | 1 (4) |
| Final pathologic stage, n (%) | |
| pT2a | 5 (21) |
| pT2b | 1 (4) |
| pT2c | 10 (42) |
| pT3a | 6 (25) |
| pT3b | 2 (8) |

## 4. Results

### 4.1. Clinical and pathological characteristics

The study cohort includes 24 patients of mean age $64 \pm 5.4$ years that went through the flow in Fig. 1. The average time interval from PET/CT scanning to surgery was 14 days (range, 3–61 days). Mean PSA level was $7.5 \pm 5.3$ ng/mL. On final surgical pathology, 16 patients (66%) had tumor confined to the prostate (AJCC stage II). Gleason score was 6 in 6 patients (25%), 7 in 16 patients (66%), and 9 in 2 patients (9%). A summary of the clinical and pathologic characteristics is presented in Table 3.

From the 24 cases, only 20 were usable for the proposed method evaluation. In 13 cases, the maximal available PET resolution in the *xy* plane, before trimming the prostate, was $256 \times 256$. For the other 7 cases, the maximal resolution was only $64 \times 64$. Furthermore, in the 7 low resolution cases, only the first 540 s of the scan are available, while for the other cases, 600 s are available.

A match between pathological marks and the PET and CT scans was problematic for one case out of the 13 cases that have a high resolution - see Section 2.4. Therefore, the ground truth correctness is unreliable for this case. This leaves a set of 12 cases that have a PET resolution of $256 \times 256$, a scan length of 600 s and an appropriate registration between the PET and the pathology.

The total number of voxels in the 12 high resolution cases, which are usable for detection and evaluation, is 57169.

### 4.2. Tumor detection performance

This section describes the results of applying the method described in Sections 2 and 3. The results are based on different sets of extracted features and on different parts of the data. Algorithm 1 is used with $\eta = 0.2$, $P = 60$ in all experiments described in this section.

**Experiment I – Mean thresholding:** In this experimental setup, we used SUV mean as the only feature, which is similar to using static PET information. In this setup big tumors are detected by using 98% percentile of the SUV mean as a threshold in each case. Detection performance of AUC = 0.812 is achieved for the benchmark set of 12 cases. However, this feature with the same threshold scheme achieves low performance of AUC = 0.535 in detecting small tumors.

**Experiment II – Comparing between different features:** The next experimental setups use additional features such as deep autoencoder based features and hand-crafted features from Section 3.2. ROC curves for the combined set of

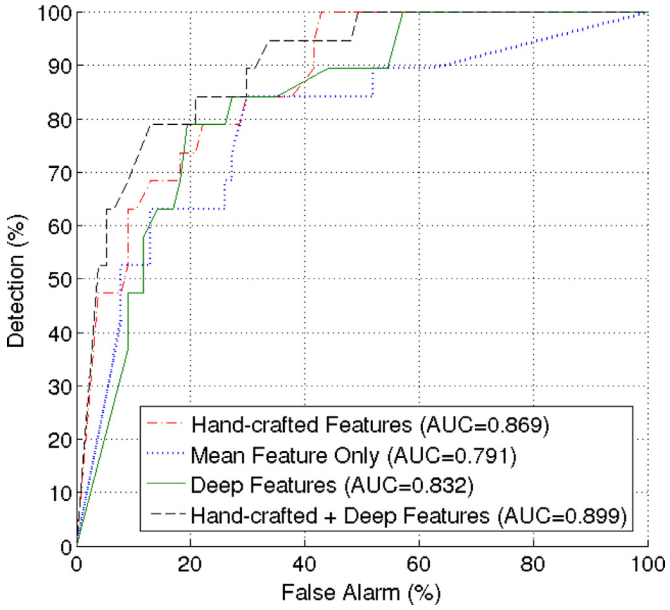



**Fig. 11.** ROC graphs for big anomalies detection with different features combinations using the 12 cases dataset with a resolution of $256 \times 256$.

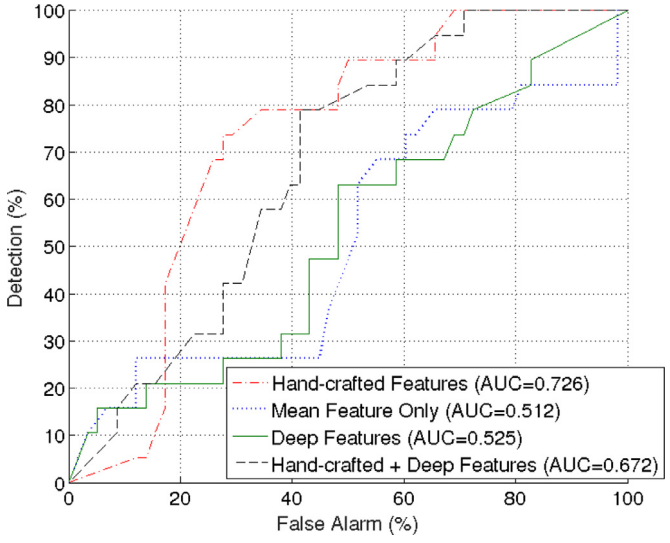



**Fig. 12.** ROC graphs of small anomalies detection with different features combinations using a 12 cases dataset with a $256 \times 256$ resolution.

features, which consists of both hand-crafted and deep features, are presented in Figs. 11 and 12. For big anomalies, this combination yields an overall score of AUC = 0.899. However, for small anomalies, the hand-crafted feature set prevails with an overall score of AUC = 0.726. Results with mean feature only are also presented.

The AUC = 0.899 score in big anomalies detection can be translated to 78.95% sensitivity and 12.99% false alarms, and the AUC = 0.726 score in small anomalies detection can be translated to 73.68% and 27.59%, respectively.

The hand-crafted features set and the detection technique, which produced the best result, are different between big and small anomalies. In particular, the small anomalies search is based on a 2 Compartments Kinetic Model $k_1$ feature (see Section 3.2.2) in addition to the statistical features and on the Parzen Windows detection method. However, the big anomalies search is based only on statistical features as hand-crafted features and on the $k$-NN as a search method. Full details about these two configurations are given in Table 5.

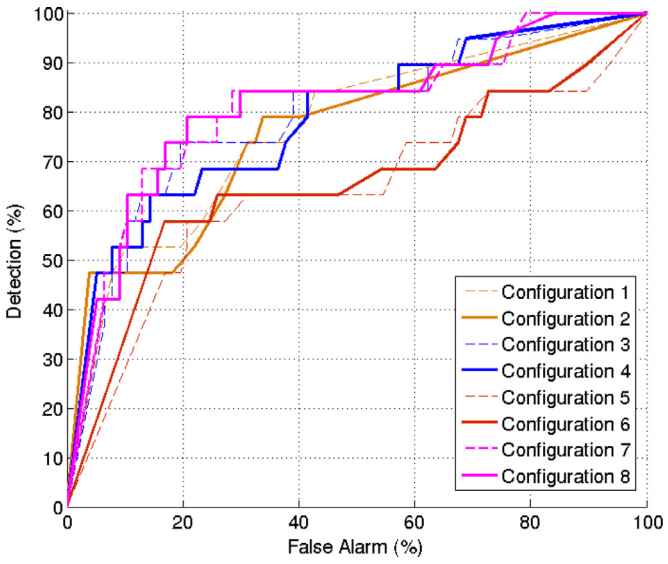



**Fig. 13.** ROC graphs of big anomalies detection with different resolutions using the 12 cases dataset with the mean feature. See Table 4 for details about each configuration. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 4**
Results of big anomalies detection with different resolutions using the 12 cases dataset and the mean feature.

| Configuration | Resolution | Overlap | Threshold | AUC |
|---|---|---|---|---|
| 1 | $64 \times 64$ | 0 | 97.5% | 0.765 |
| 2 | $64 \times 64$ | 0 | 98% | 0.758 |
| 3 | $256 \times 256$ | 0 | 97.5% | 0.789 |
| 4 | $256 \times 256$ | 0 | 98% | 0.785 |
| 5 | $64 \times 64$ | 2 | 97.5% | 0.641 |
| 6 | $64 \times 64$ | 2 | 98% | 0.658 |
| 7 | $256 \times 256$ | 2 | 97.5% | 0.806 |
| 8 | $256 \times 256$ | 2 | 98% | 0.812 |

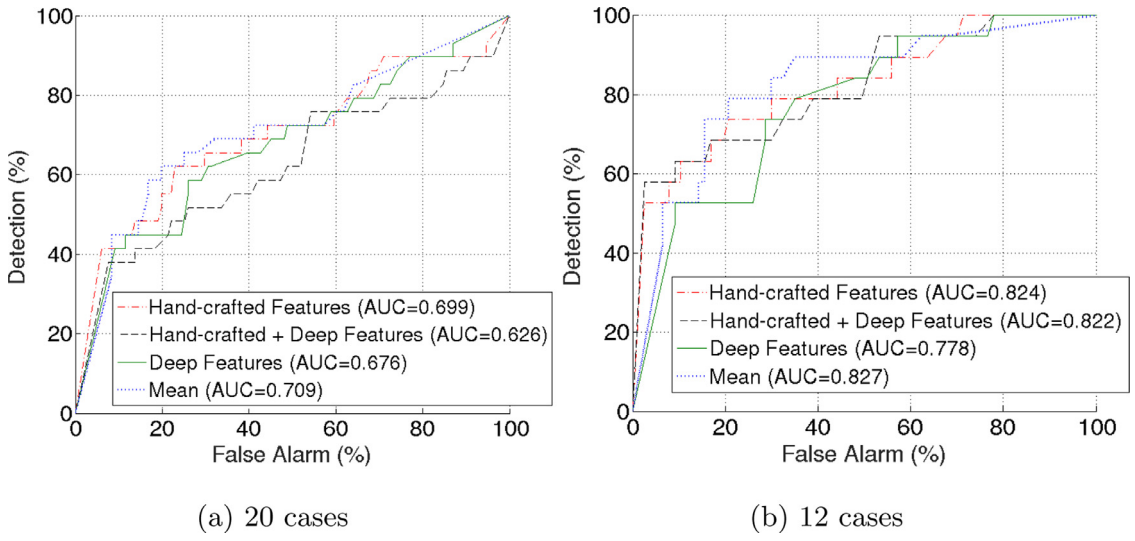
**Experiment III – Comparing between 64 and 256 PET resolutions:** Fig. 13 displays the results of big anomalies detection in 12 cases for which both resolutions in the $xy$ plane ($64 \times 64$ and $256 \times 256$) are given. The detection is based on the mean of the time series of each voxel as the only feature[4].

When we apply the best known detection configuration used on the $256 \times 256$ resolution to the $64 \times 64$ resolution (Configuration 6), the output is AUC = 0.658, compared with AUC = 0.812 for the $256 \times 256$ resolution. One of the reasons for this low score is the 2 pixel overlapping between the octants (see Appendix A), which is probably too high for a small resolution where each voxel is large. Disabling this overlapping and applying a small change to the threshold results in AUC = 0.765 (Configuration 1), which is better than Configuration 6 but still lower than the results from the $256 \times 256$ resolution.

[4] Recall that $64 \times 64$ and $256 \times 256$ represent the $xy$ resolution of the whole PET image before the prostate was trimmed.

**Table 5**
Results summary. The autoencoder numbers refer to Table 6. The features names refer to Table 1.

| Dataset | Task | Features | Hist bins | Hist weights | Auto-encoder # | Search method | Search threshold | Result AUC |
|---|---|---|---|---|---|---|---|---|
| 12 cases | Big anomalies | SUV TACs, SUV histogram vector (SHV), SHV mean, SHV skewness, SHV 75th percentile, deep | (0, 0.5, 1, ..., 4, 5, 7.5, ..., 18) | All ones | 1 | $k$-NN ($k = 8$) | $2.5 \times$ std | 0.899 |
| 12 cases | Big anomalies | SUV TACs, SHV, SHV mean, SHV skewness, SHV 75th percentile | (0, 0.5, 1, ..., 4, 5, 7.5, ..., 18) | All ones | – | $k$-NN ($k = 8$) | $2.5 \times$ std | 0.869 |
| 12 cases | Big anomalies | Deep | – | – | 2 | $k$-NN ($k = 8$) | $2.5 \times$ std | 0.848 |
| 12 cases | Big anomalies | Deep | – | – | 1 | $k$-NN ($k = 8$) | $2.5 \times$ std | 0.832 |
| 12 cases | Big anomalies | Mean SUV | – | – | – | Simple threshold | 98% | 0.812 |
| 20 cases | Big anomalies | Mean SUV | – | – | – | $k$-NN ($k = 8$) | $2.5 \times$ std | 0.709 |
| 20 cases | Big anomalies | SUV TACs, SHV, SHV mean, SHV skewness, SHV 75th percentile, deep | (0, 0.5, 1, ..., 4, 5, 7.5, ..., 18) | All ones | 3 | $k$-NN ($k = 8$) | $2.5 \times$ std | 0.626 |
| 12 cases | Small anomalies | SHV, SHV skewness, SHV 75th percentile, 2-comp $K_1$ | (0, 0.5, 1, ..., 4, 5, 7.5, ..., 20) | (2, 2, 1, 1, 1, ..., 1) | – | Parzen Windows | 98.3% | 0.726 |
| 12 cases | Small anomalies | Mean SUV | – | – | – | Simple threshold | 98% | 0.535 |




(a) 20 cases (b) 12 cases

**Fig. 14.** ROC graphs of big anomalies detection, in different cases.

**Experiment IV – Using the extended dataset:** The results for the extended dataset of 20 cases are shown in Fig. 14. Some of the 8 additional cases have a smaller resolution in comparison with the other 12 cases, some are measured over a shorter time period, or are suspected to have partly wrong labels, as it was described in Section 4.1.

In order to apply an autoencoder (Masci et al., 2011) to the extended dataset, cases with a resolution of $64 \times 64$ were artificially re-sampled to be four times larger[5], and the cases with a 600 s sampling period were trimmed to the last common time (540 s). An autoencoder, which is similar to the one used in Experiment II without the exact same parametrization, was trained with this new extended dataset.

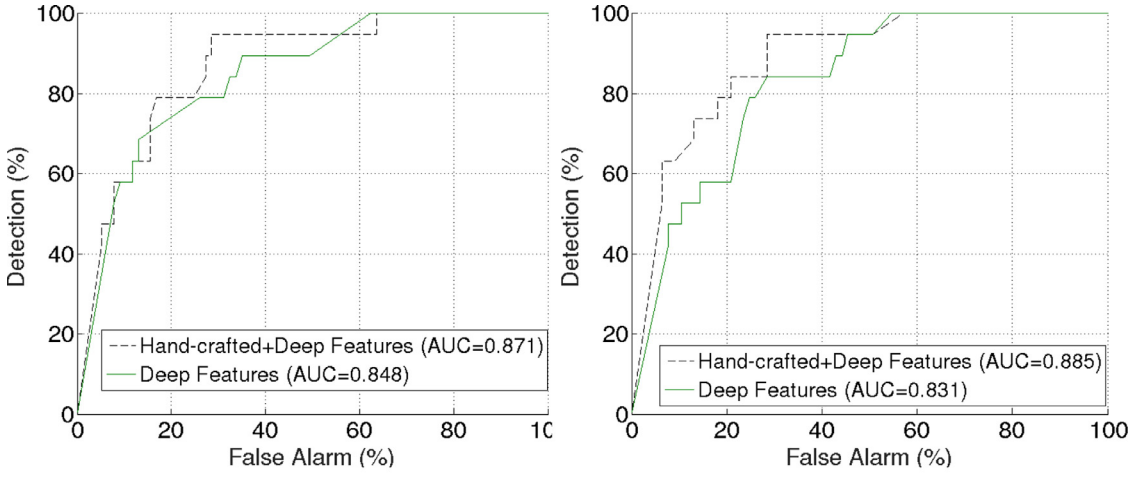
Fig. 14b shows the results for the original 12 cases as a basis for comparison under these new conditions described above. There is a significant AUC drop when the extended dataset is used. See the comparison between Fig. 14a and b. One of the reasons for this drop is a resolution difference, however, due to the findings in Experiment III this effect is expected to be relatively moderate at least for the mean feature. The rest of the drop is supposedly caused by mistakes in the PET-pathology matching or by inaccuracies introduced by the artificial resolution enlargements.

**Experiment V – Autoencoder modification:** A modified autoencoder was trained on the 12 cases set. In this autoencoder, the convolutional filter size was increased from 2 to 3. In addition, the dropout ratios and the additional noise were lowered – see Table 6 for details. The AUC of deep features with this modified autoencoder (AUC = 0.848) is better than what was introduced in Experiment II. However, when these results are combined with hand crafted features or with the features of the original autoencoder, there is no improvement in the overall AUC – see Fig. 15.

[5] This enlargement is performed because the whole dataset reduction to a $64 \times 64$ resolution makes it too small to be processed by a neural network. Leaving the original resolutions as is makes the count of anomalous voxels biased.

**Table 6**
Autoencoders parameters.

| Auto-encoder # | # of Conv filters | Conv filters size | Encode size | Noise variance | # Epochs | Dropout ratio | Start learning rate | Conv layers activation func | Dense layer activation func |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 12,12,18 | 2 | 5 | 0.08 | 100 | 0.05 | 0.03 | Linear | Rectifier |
| 2 | 12,12,18 | 3 | 5 | 0.05 | 100 | 0.02 | 0.03 | Linear | Rectifier |
| 3 | 12,12,18 | 2 | 5 | 0.02 | 100 | 0 | 0.02 | Linear | Rectifier |




(a) Modified autoencoder (conv filter size = 3)

(b) Modified + original autoencoders

**Fig. 15.** ROC graphs of big anomalies detection with different autoencoder networks whose structure is described in Fig 10.

## 5. Conclusions

### 5.1. Findings discussion

In this paper, we present a method for primary detection and localization of prostate tumors using PET images. The presented method was evaluated using pathological specimen as a reference.

The results of this research show the advantage of a learning approach to medical imaging analysis, which includes the use of several classes of features including “deep” features while combining them into a single detector. This approach yields AUC = 0.899 in the task of finding big anomalies while using only the mean SUV feature resulted in AUC = 0.812 in the same task, as it is explained in Section 4.

Specifically, the presented algorithm can contribute to early diagnosis of prostate cancer when MRI and ultrasound do not identify the cancerous cells. It can also help in directing biopsy to regions of the prostate that are suspected as cancerous according to Algorithm 2.

Although the features that are extracted by the autoencoder contribute to the best result in big tumors detection, they did not outperform the hand crafted features, in contrast to numerous other results, especially in the field of computer vision. This is due to several reasons such as the relatively small size of the dataset and the unsupervised approach used in the detection. It is “easier” to perform semi- or full supervised detection. Usually, hand crafted is complicated. Here it is feasible because there are not many features but in high dimensional data it is more complicated.

### 5.2. Future work

This work can be extended in several directions. A bigger dataset, which requires additional participants in the medical research, will be used to further analyze the performance of the proposed algorithm and to investigate new algorithms. In addition, it is expected that the performance of the neural network based features will improve when the dataset size grows.

The identification of smaller tumors was inferior to that of larger ones. This might suggest that a different set of features or a different algorithm are needed to achieve a better performance.

Another possible change is to replace the octants approach with a voxel-by-voxel registration of the pathological findings to the PET images, which can be done manually or by utilizing a 3D shape registration algorithm. If the registration accuracy is sufficient, it will improve the precision of the tumors localization evaluation. In addition, and perhaps more importantly, it will make it possible to address the tumor detection problem with supervised approaches such as SVM and supervised neural network based algorithms.

Adding a preliminary stage to the algorithm, which detects images with no anomalies at all, may improve the achieved results for such images.

Using autoencoders can be studied further beyond the scope of this work, by trying different activation functions and network structures.

More sophisticated biological models can be used as well, by introducing spatial relations into differential equations of the model in addition to the temporal ones.

In order to enable full automation of the process, the prostate must be identified in the CT imaging and isolated from its surroundings. This can be achieved by utilizing automatic segmentation, which has been proposed by Li et al. (2011) and Ma et al. (2017).

## Acknowledgments

This research was partially supported by the Israel Ministry of Science & Technology (Grant no 3-8701), Israel Science Foundation

(ISF, Grant no 1556/17) and Blavatnik Computer Science Research Fund.

## Conflict of interest

None.

## Appendix A. Detection results aggregation

As part of the detection results evaluation, we aggregate these detection results in the octants.

The first step in the aggregation procedure decides to which octant each voxel belongs. This decision should take into account the observation that the prostate is not necessarily located symmetrically inside a 3D box that contains it. Therefore, splitting this box by half in the middle of each dimension is insufficiently accurate.

Let $s_1, \ldots, s_n$ be the 2D slices of the prostate boundary, where $n$ is the number of slices. Each slice $s_k$ is a binary image, where each voxel equals to 1 iff it is inside the prostate. Algorithm 4 describes

**Algorithm 4** Division into octants.

**Input:** 3D volume with elements $v_{i,j,k} \in \{0,1\}, \quad i = 1, \ldots, \ell, \quad j = 1, \ldots, m, \quad k = 1, \ldots, n.$

**Output:** List of voxels members in each octant $P_r, \quad r = 1, \ldots, 8$

Choose $k^*$ s.t. $\sum_{\substack{i,j,k \\ 1 \le k \le k^*}} v_{i,j,k} = \sum_{\substack{i,j,k \\ k^*+1 \le k \le n}} v_{i,j,k}$

**for** $k \leftarrow 1$ to $n$ **do**

Choose $i_k^*$ s.t. $\sum_{\substack{i,j \\ 1 \le i \le i_k^*}} v_{i,j,k} = \sum_{\substack{i,j \\ i_k^*+1 \le i \le \ell}} v_{i,j,k}$

Choose $j_k^*$ s.t. $\sum_{\substack{i,j \\ 1 \le j \le j_k^*}} v_{i,j,k} = \sum_{\substack{i,j \\ j_k^*+1 \le j \le m}} v_{i,j,k}$

**end for**

$P_1 \leftarrow \{v_{i,j,k} \mid i \le i_k^* + 1 \text{ and } j \le j_k^* + 1 \text{ and } k \le k^* + 1\}$

$P_2 \leftarrow \{v_{i,j,k} \mid i \ge i_k^* \text{ and } j \le j_k^* + 1 \text{ and } k \le k^* + 1\}$

/* Skipping the pseudocode of the next 5 octants */

$P_8 \leftarrow \{v_{i,j,k} \mid i \ge i_k^* \text{ and } j \ge j_k^* \text{ and } k \ge k^*\}$

**return** $P_1, P_2, \ldots, P_8$

how this volume is divided into octants. The middle slice, which defines the border between octants in the z-axis, represents the binary map center of mass in this axis. The border coordinates in

**Fig. A.16.** A single CT slice of a prostate boundary binary map. Voxels inside the prostate marked by green. Voxels inside the bottom-left octant marked by pink. Intersected voxels marked by white. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

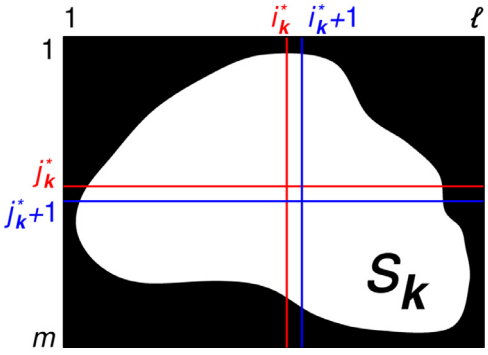



**Fig. A.17.** Division of the prostate slice $s_k$, $1 \le k \le n$. $i_k^*$ is the slice center of mass in the horizontal axis and $j_k^*$ is the center of mass in the vertical axis. Coordinates $i_k^* + 1$ and $j_k^* + 1$ are the overlap between octants.

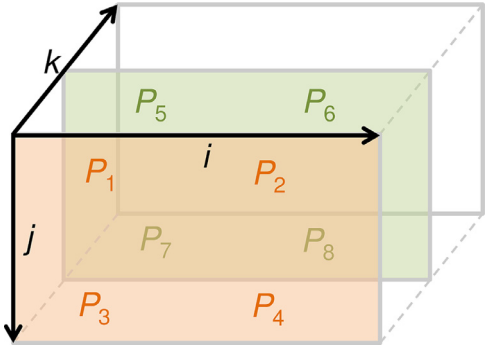



**Fig. A.18.** Illustration of the 3D octants numbering. The front octants are $P_1$, $P_2$, $P_3$ and $P_4$. The other four octants are at the back.

the other directions are chosen in a similar way separately in each slice.

In addition, in order to overcome the inaccuracies in the PET-pathology match, the two middle voxels in each direction are assumed to belong to both bordering octants, thus they create a two voxels overlap between bordering octants.

Voxels, which lie inside the 3D trimmed box but lie outside the prostate boundary, are nevertheless taken into account. This is done to overcome the inaccuracies in the prostate boundary drawn by the physician.

The prostate division process is illustrated in Figs. A.16–A.18.

## References

Beheshti, M., Imamovic, L., Broinger, G., Vali, R., Waldenberger, P., Stoiber, F., Nader, M., Gruy, B., Janetschek, G., Langsteger, W., 2010. 18F choline PET/CT in the preoperative staging of prostate cancer in patients with intermediate or high risk of extracapsular disease: a prospective study of 130 patients. Radiology 254 (3), 925–933. doi:10.1148/radiol.09090413.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, New York, NY.

Chandola, V., Banerjee, A., Kumar, V., Lin, T.Y., 2009. Anomaly detection: a survey. ACM Comput. Surv. 41 (3), 15. doi:10.1145/1541880.1541882.

Conti, M., 2014. New prospects for PET in prostate cancer imaging: a physicist's viewpoint. EJNMMI Phys. 1 (1), 1–11. doi:10.1186/2197-7364-1-11.

Coxson, P.G., Salmeron, E.M., Huesman, R.H., Mazoyer, B.M., 1992. Simulation of compartmental models for kinetic data from a positron emission tomograph. Comput. Methods Programs Biomed. 37 (3), 205–214. doi:10.1016/0169-2607(92)90116-O.

Dimitrakopoulou-Strauss, A., Pan, L., Strauss, L.G., 2012. Quantitative approaches of dynamic FDG-PET and PET/CT studies (dPET/CT) for the evaluation of oncological patients. Cancer Imaging 12, 283–289. doi:10.1102/1470-7330.2012.0033.

Farsad, M., Schiavina, R., Castellucci, P., Nanni, C., Corti, B., Martorana, G., Canini, R., Grigioni, W., Boschi, S., Marengo, M., Pettinato, C., Salizzoni, E., Monetti, N., Franchi, R., Fanti, S., 2005. Detection and localization of prostate cancer: correlation of 11C-choline PET/CT with histopathologic step-section Analysis. J. Nucl. Med. 46 (10), 1642–1649.

Feleppa, E.J., Rondeau, M.J., Lee, P., Porter, C.R., 2009. Prostate-cancer imaging using machine-learning classifiers: potential value for guiding biopsies, targeting therapy, and monitoring treatment. In: Ultrasonics Symposium (IUS), 2009 IEEE International. IEEE, pp. 527–529.

Giovacchini, G., Picchio, M., Coradeschi, E., Scattoni, V., Bettinardi, V., Cozzarini, C., Freschi, M., Fazio, F., Messa, C., 2008. [(11)C]choline uptake with PET/CT for the initial diagnosis of prostate cancer: relation to PSA levels, tumour stage and anti-androgenic therapy. Eur. J. Nucl. Med. Mol. Imaging 35 (6), 1065–1073. doi:10.1007/s00259-008-0716-2.

Godfrey, K., 1983. Compartmental Models and Their Application. Academic Press.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Guo, Y., Gao, Y., Shen, D., 2016. Deformable MRprostate segmentation via deep feature learning and sparse patch matching. IEEE Trans. Med. Imaging 35 (4), 1077–1089.

Häggström, I., 2014. Quantitative Methods for Tumor Imaging with Dynamic PET. Umeå University.

Härdle, W.K., Müller, M., Sperlich, S., Werwatz, A., 2004. Nonparametric and Semiparametric Models. Springer Science & Business Media.

Hricak, H., Choyke, P.L., Eberhardt, S.C., Leibel, S.A., Scardino, P.T., 2007. Imaging prostate cancer: a multidisciplinary perspective 1. Radiology 243 (1), 28–53. doi:10.1148/radiol.2431030580.

Hudson, H.M., Larkin, R.S., 1994. Accelerated image reconstruction using ordered subsets of projection data. IEEE Trans. Med. Imaging 13 (4), 601–609.

Ihler, A., Mandel, M., 2007. Kernel density estimation toolbox for MATLAB (R13).

Japkowicz, N., Myers, C., Gluck, M., 1995. A novelty detection approach to classification. In: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, pp. 518–523.

de Jong, I.J., Pruim, J., Elsinga, P.H., Vaalburg, W., Mensink, H.J.A., 2002. Visualization of prostate cancer with 11C-choline positron emission tomography. Eur. Urol. 42 (1), 18–23.

Li, W., Liao, S., Feng, Q., Chen, W., Shen, D., 2011. Learning image context for segmentation of prostate in CT-guided radiotherapy. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 570–578.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42 (1995), 60–88. doi:10.1016/j.media.2017.07.005.

Liu, J., Jaiswal, A., Yao, K.-T., Raghavendra, C.S., 2015. Autoencoder-derived features as inputs to classification algorithms for predicting well failures. SPE Western Regional Meeting. Society of Petroleum Engineers.

Ma, L., Guo, R., Zhang, G., Tade, F., Schuster, D.M., Nieh, P., Master, V., Fei, B., 2017. Automatic segmentation of the prostate on CT images using deep learning and multi-atlas fusion. In: Medical Imaging 2017: Image Processing, 10133. International Society for Optics and Photonics, p. 101332O.

Manevitz, L., Yousef, M., 2007. One-class document classification via Neural Networks. Neurocomputing 70 (7-9), 1466–1481. doi:10.1016/j.neucom.2006.05.013.

Masci, J., Meier, U., Cirean, D., Schmidhuber, J., 2011. Stacked convolutional auto-encoders for hierarchical feature extraction. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). LNCS 6791 (Part 1), 52–59. doi:10.1007/978-3-642-21735-7_7.

Morris, E.D., Endres, C.J., Schmidt, K.C., Christian, B.T., Jr, R.F.M., Fisher, R.E., Muzic Jr, R.F., 2004. Kinetic modeling in positron emission tomography. Emission Tomography: The Fundamentals of PET and SPECT, Academic, San Diego.

Piert, M., Park, H., Khan, A., Siddiqui, J., Hussain, H., Chenevert, T., Wood, D., Johnson, T., Shah, R.B., Meyer, C., 2009. Detection of aggressive primary prostate cancer with 11C-choline PET/CT using multimodality fusion techniques. J. Nucl. Med. 50 (10), 1585–1593. doi:10.2967/jnumed.109.063396.

Pmod Technologies Ltd., 2016. Pmod.

Schwarzenböck, S., Souvatzoglou, M., Krause, B.J., 2012. Choline PET and PET/CT in primary diagnosis and staging of prostate cancer. Theranostics 2 (3), 318–330. doi:10.7150/thno.4008.

Scott, D.W., Sain, S.R., 2004. Multidimensional density estimation. Handb. Stat. 24, 229–261. doi:10.1016/S0169-7161(04)24009-3.

Shin, H.-C., Orton, M., Collins, D.J., Doran, S., Leach, M.O., 2011. Autoencoder in time-series analysis for unsupervised tissues characterisation in a large unlabelled medical image dataset. In: 2011 10th International Conference on Machine Learning and Applications and Workshops. IEEE, pp. 259–264.

Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958. doi:10.1214/12-AOS1000.

Sutskever, I., Martens, J., Dahl, G., Hinton, G., 2013. On the importance of initialization and momentum in deep learning. In: Proceedings of the 30th International Conference on Machine Learning (ICML-13), pp. 1139–1147.

Testa, C., Schiavina, R., Lodi, R., Salizzoni, E., Corti, B., Farsad, M., Kurhanewicz, J., Manferrari, F., Brunocilla, E., Tonon, C., Monetti, N., Castellucci, P., Fanti, S., Coe, M., Grigioni, W.F., Martorana, G., Canini, R., Barbiroli, B., 2007. Prostate cancer: sextant localization with MR imaging, MR spectroscopy, and 11C-choline PET/CT. Radiology 244 (3), 797–806. doi:10.1148/radiol.2443061063.

Tuncel, M., Souvatzoglou, M., Herrmann, K., Stollfuss, J., Schuster, T., Weirich, G., Wester, H.-J., Schwaiger, M., Krause, B.J., 2008. [11C] Choline positron emission tomography/computed tomography for staging and restaging of patients with advanced prostate cancer. Nucl. Med. Biol. 35 (6), 689–695. doi:10.1016/j.nucmedbio.2008.05.006.

Umbehr, M.H., Müntener, M., Hany, T., Sulser, T., Bachmann, L.M., 2013. The role of 11C-choline and 18F-fluorocholine Positron Emission Tomography (PET) and PET/CT in prostate cancer: a systematic review and meta-analysis. Eur. Urol. 64 (1), 106–117. doi:10.1016/j.eururo.2013.04.019.