

Computer aided detection of ureteral stones in thin slice computed tomography volumes using Convolutional Neural Networks

Martin Långkvist^{a,*}, Johan Jendeborg^b, Per Thunberg^c, Amy Loutfi^a, Mats Lidén^b

^a Center for Applied Autonomous Sensor Systems, Örebro University, Örebro, Sweden

^b Department of Radiology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

^c Department of Medical Physics, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

ARTICLE INFO

Keywords:

Computer aided detection
Ureteral stone
Convolutional neural networks
Computed tomography
Training set selection
False positive reduction

ABSTRACT

Computed tomography (CT) is the method of choice for diagnosing ureteral stones - kidney stones that obstruct the ureter. The purpose of this study is to develop a computer aided detection (CAD) algorithm for identifying a ureteral stone in thin slice CT volumes. The challenge in CAD for urinary stones lies in the similarity in shape and intensity of stones with non-stone structures and how to efficiently deal with large high-resolution CT volumes. We address these challenges by using a Convolutional Neural Network (CNN) that works directly on the high resolution CT volumes. The method is evaluated on a large data base of 465 clinically acquired high-resolution CT volumes of the urinary tract with labeling of ureteral stones performed by a radiologist. The best model using 2.5D input data and anatomical information achieved a sensitivity of 100% and an average of 2.68 false-positives per patient on a test set of 88 scans.

1. Introduction

During the last decades, computed tomography (CT) has emerged as the method of choice for diagnosing ureteral stones [1]. The formation of crystals in the urinary system - urinary stone disease - is a common condition with a lifetime risk of approximately 10% [2]. A ureteral stone is a kidney stone that has been displaced and passed into and thereby obstructing the ureter causing severe pain. An obstructing ureteral stone can further be complicated by infection and renal failure.

Small ureteral stones are likely to pass spontaneously while larger stones may need interventional therapy such as extracorporeal shock wave lithotripsy or endoscopic lithotripsy. Diagnosis, localization and estimation of prognosis for spontaneous passage of ureteral stones are performed by a radiologist by reviewing the cross-sectional slices of the CT scan, that produces a grayscale 3D volume of the examined body [3].

Computer assisted detection (CAD) tools for assisting radiologists are gradually being introduced in the everyday image review [4]. A few previous works related to CAD in urinary stone disease have focused on detection of kidney stones that have not passed into the ureters [5,6], and image features for discrimination among already detected calcifications between ureteral stones and mimickers [7]. To the best of our knowledge there are no previous reports of CAD systems for the more complex task of detecting ureteral stones.

While previous works mainly have analyzed 2D images by extracting features from segmented regions, this work instead operates on the 3D volumes directly on the raw pixel data using a Convolutional Neural Network (CNN) [8,9]. Pixel/voxel-based machine learning (PML) has recently emerged in medical image processing as an alternative to traditional feature-based ML [10]. The advantage of PML methods is that the algorithms learn directly from the raw data instead of using extracted features from segmented objects and thus no information is lost and, in addition, removes the need for researching, extracting, and selecting features and segmentation methods.

A challenge with PML methods for high-resolution CT data is how to efficiently train the model and how to deal with the large amount of data. In this work, we propose a method that operates directly on the raw data of selected volumes-of-interest (VOI) and present a strategy for reducing the number of false positives using anatomical information, which is a challenge for many CAD systems for medical applications where a false negative is more critical than a false positive.

This paper has the following contributions:

- Performs ureteral stone detection from raw high-resolution CT scans using Convolutional Neural Networks
- Proposes a probabilistic approach to include anatomical information in the decision process

* Corresponding author.

E-mail address: martin.langkvist@oru.se (M. Långkvist).

- Present a training set selection strategy using similarity ranking
- Evaluates various design choices and model parameters for the proposed method

The rest of the paper is organized as follows. The CT scans and the proposed method is described in Section 2. The experimental results is performed in Section 3. Finally, comparison with related work and conclusions are given in Section 4.

2. Materials and method

2.1. Input data

The full data set consists of 465 clinically acquired unenhanced abdominal CT scans with $0.8 \times 0.8 \times 1$ mm voxel size. The study group included both men and women ≥ 18 years old with one scan per patient. The CT scans were obtained with three different CT systems (Brilliance, Philips Medical Systems Best, The Netherlands; Somatom Definition Flash, Siemens, Erlangen, Germany; Somatom Definition AS, Siemens, Erlangen, Germany). All scans were obtained with an intermediate dose CT protocol for the urinary tract (120 kVp, CTDI 5mGy, no dose modulation). The weight of the included patients is unknown and the noise depends on the body size since no dose modulation was used. A solitary ureteral stone is present in all of the 465 CT scans and the position has been provided by a radiologist. Each CT scan, $I \in \mathbb{N}^3$, is $512 \times 512 \times \sim 400$ pixels where the intensity c in pixel position (i, j, k) is defined by $I^{ijk} = c$ where the pixel intensities in the data sets are in the range $c = [-1024 \ 3071]$. The intensity of each pixel represents the attenuation value in Hounsfield Units (HU) scale and is related to the material composition at that location inside the body with air having an attenuation value of -1000 HU and water 0 HU.

The Regional Research Ethics Board approved the study protocol and waived the informed consent requirement.

2.2. Convolutional Neural Networks

Convolutional Neural Networks (CNNs or ConvNets) [11] have been successful in computer vision tasks with structured multiple arrays such as object recognition in images [12], speech recognition from time-series data [13], and videos [14].

A conventional Artificial Neural Network (ANN) consists of an input layer, hidden layer, and output layer. ANNs are fully connected, which means that each hidden unit has a connection to all input units and each output unit has a connection to each hidden unit. For an input unit I_j , bias unit b_i , and a weight matrix W_{ij} , the hidden unit h_i is calculated as $h_i = \sum_{j=1}^J W_{ij} I_j + b_i$. For high-dimensional inputs, such as images, where J is large, the number of parameters to learn in the weight matrix becomes too high and the model suffers from the curse of dimensionality [15].

A CNN, instead, uses local connections with tied weights and

therefore require fewer parameters to learn and takes advantage of the spatial structure in the input data. The CNN is also capable of learning slightly translational and rotational invariant features (also called filters), which is a benefit for structured data.

An overview of how a CNN is used to classify a CT volume is seen in Fig. 1. The steps of a single-layer CNN consists of a convolving layer, followed by a non-linear activation function, and finally subsampling with pooling. A fully-connected layer (FC-layer) is connected to the output from the pooling layer, followed by a classifier attached to the FC-layer for the final classification. The model parameters to be trained are $\theta = \{W_1, \dots, W_k, b_1, \dots, b_k, W_s, b_s\}$ and the hyperparameters to set are the input size m , filter size n , pooling dimension p , and number of filters k . The selection of these parameters is further discussed in Section 3.3.

The convolutional layer f_{ij}^k for filter k for an input image I^c with c number of channels is computed as:

$$f_{ij}^k = \sigma \left(b_k + \sum_c \sum_{a=0}^{n-1} \sum_{b=0}^{n-1} W_k^{abc} I_{i+a, j+b}^c \right) \quad (1)$$

where W_i and b_i are the filter and bias for the i th filter, n is the filter size, and $\sigma(\cdot)$ is the activation function. The sigmoid activation function, $\sigma(x) = 1/(1 + e^{-x})$, is used in this work.

The pooling layer is calculated by taking the max over non-overlapping subregions of the convolutional layer with pooling size p . The pooling layer g_{ij}^k for the k th filter is calculated as:

$$g_{ij}^k = \max \left(f_{1+p(i-1):pi, 1+p(j-1):pj}^k \right) \quad (2)$$

The concatenated pooling layer is then used as input to a FC-layer. The dimension of the pooling layer depends on the input size m , filter size n , number of filters k , and pooling dimension p according to $k \left(\frac{m-n+1}{p} \right)^3$. The FC-layer has h number of hidden units, which we set to $h = 100$ in this work.

Finally, a classifier is attached to the FC-layer for the final classification. We use a softmax classifier which outputs a vector with K elements that represents the normalized classification probabilities for each class K . In this work we set $K = 2$ for the class of stone and non-stone. The probabilities are calculated as:

$$P(y = j|x) = \frac{\exp(x^T W_j)}{\sum_{k=1}^K \exp(x^T W_k)} \quad (3)$$

2.2.1. Calculating probability map

Locating a calcification within or outside the ureter is the key principle that radiologists use when differentiating between stones and phleboliths. When surrounding anatomical landmarks are removed, the differentiation is more difficult.

One approach to incorporating more information from the

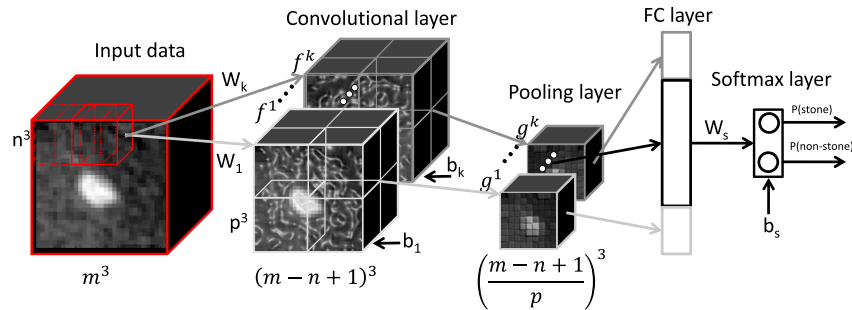


Fig. 1. Process of classifying a volume-of-interest. The convolutional layer is formed by convolving the filters over the input volume and the pooling layer is a subsampling of the convolutional layer. The input vector that is fed to the classifier is formed by concatenating all pooling layers for each filter.

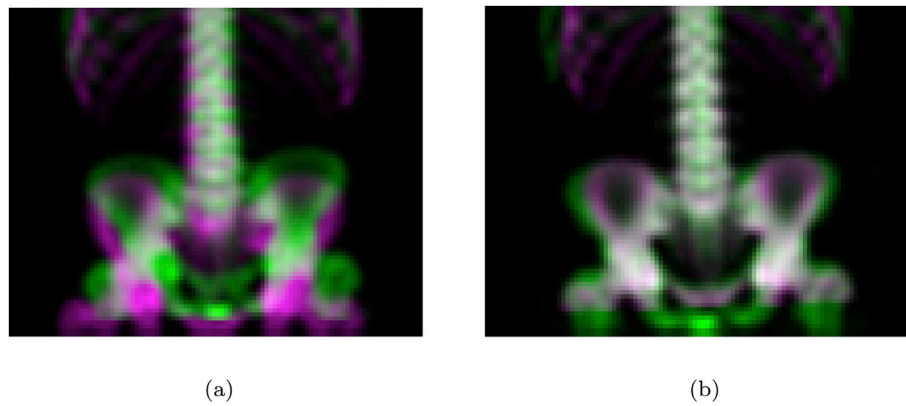


Fig. 2. (a) Before scan alignment (b) After scan alignment. White regions show where the reference scan (magenta) and registered scan (green) have the same intensities. Only translational and rotational transformation is used. The image shows the result projected in the frontal plane for clarity. Best seen in color. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

surrounding area is to use a multi-scale CNN that uses downsampled images [16]. Since we have a relatively large database of scans and annotated stone positions, we will instead calculate a probability map of the location of real stones and use this information in a post-processing step to filter out the number of false-positives.

Before calculating the probability map, the skeletal structures of all scans were first automatically aligned with a translational and rotational transformation in all three dimensions to a randomly selected reference scan. The ureters are located in the retroperitoneal space, stretching from the kidneys cranially and down to the bladder. Skeletal registration was chosen, since the relation between the ureters in the retroperitoneal space and the spine, is only to a minor degree dependent on body size. A non-deforming registration could therefore be used.

To achieve the skeletal registration, the scans and the reference scan were pre-processed by thresholding the intensities to values above 100, then downsampled by a factor of 5, and then smoothed with a Gaussian $5 \times 5 \times 5$ filter before calculating the alignment. Fig. 2 shows how one scan is aligned to the reference scan. The purpose of the alignment is to more accurately calculate a probability of the distribution of the locations of the stones.

The locations of all stones in the training set are tracked before and after the alignment. The probability map is calculated by fitting a Gaussian Mixture Model (GMM) in three dimensions with 6 components

to the aligned stone positions. The initial mean for the GMM components are set using k-means. The probability map is then normalized in each slice to get a probability distribution for each slice. Fig. 3 show the locations for all stones before and after alignment. The mean of the probability map in each dimension are shown on the walls and the floor. It can be seen that the alignment reduces the variance of the stone locations and allows for a more accurate probability distribution of the locations of the stones.

3. Results

The dataset consists of 465 clinically acquired unenhanced abdominal CT scans of patients suffering from suspected renal colic. For training the CNN, the scans are randomly divided into 80% training set and 20% for testing after 28 of the scans that contained a too small or too large stone were removed, resulting in 349 training scans and 88 testing scans. The data pre-processing steps consist of selecting volumes-of-interest (VOI) using connected components (CC) and a training set selection method that ranks the CCs.

3.1. Finding connected components

The process of finding connected components (CCs) in one scan

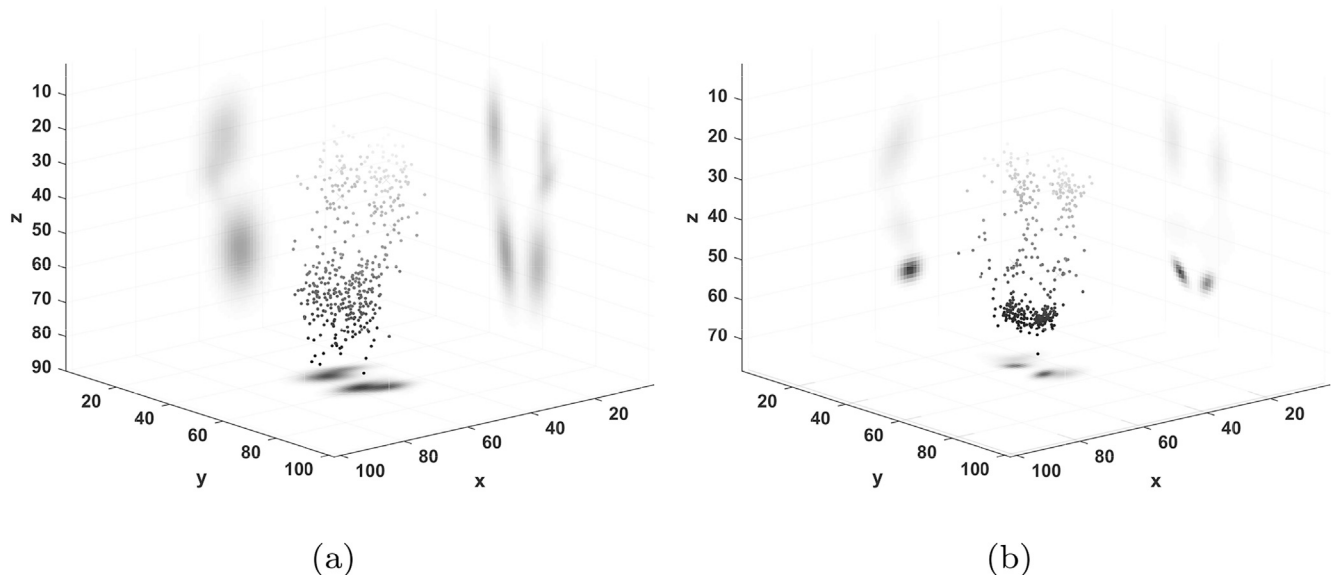


Fig. 3. Stone positions (a) before scan alignment, and (b) after scan alignment. The mean of the probability map in each dimension are shown on the walls and the floor of the plot. The gray color represents the Z-value for easier visualization.

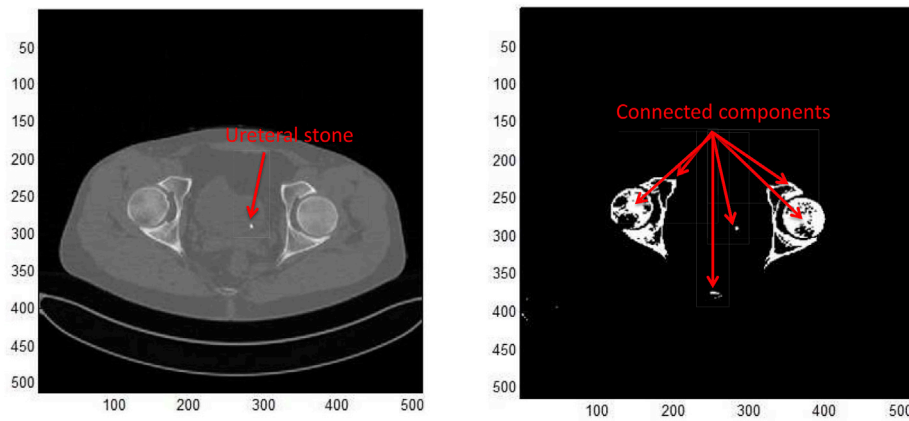


Fig. 4. Process of binarizing and finding connected components. Left image shows an example of an axial 2D section through the pelvis and the lower part of the urinary tract. Right image shows the data after binarizing and the connected components.

consists of first binarizing the volume using the threshold HU_{thresh} and then group all pixels using 6-connectivity in CCs. CCs with a number of voxels below vol_{min} and above vol_{max} were discarded from the analysis.

The binarization threshold (HU_{thresh}) was with no loss of generalization fixed to 250 HU since all common urinary stone types have considerably higher attenuation values [3]. The minimum stone volume (vol_{min}) were chosen to 5 voxels and the maximum stone volume (vol_{max}) were chosen to 500 voxels, corresponding approximately to the volume of a perfect sphere with 2 mm and 9 mm diameter, respectively. The limits were chosen since because virtually all stones smaller than 2 mm pass spontaneously without treatment, while stones with a short axis larger than 9 mm are unlikely to be found in the ureter [17].

The effect of binarizing with the chosen threshold can be seen in Fig. 4 where voxels belonging to soft tissue have been filtered out while urinary stones and skeletal structures remain.

3.2. Training set selection

The remaining number of connected components in each scan after filtering out the ones that are too small or too large can vary between a few hundreds to a few thousands. Each scan contains only one CC containing a ureteral stone and the rest are not ureteral stones. A subset of the non-stones need to be selected to balance the training set. The strategy for selecting the training examples is to select all CCs with a stone and primarily select CCs that resemble a real stone for the non-stone examples. The process of selecting examples for non-stones is done by calculating the euclidean distance between the raw data of a cube of $25 \times 25 \times 25$ around a random pixel in each non-stone CC to the same data for a random real stone and then selecting the top 100 CCs that

have a minimum euclidean distance. Fig. 5(a) show one slice of a 25×25 patch of the raw data for all the stones in the training set and Figure 5(b) show the same data for the non-stones in one scan sorted by the similarity.

In order to increase the number of training examples for the stones, a maximum total of 10 randomized pixels in each CC for the stones are selected, as opposed to only one for the non-stone CCs. Each training example for the stones are then mirrored in the YZ-plane. This results in each scan giving 100 non-stone examples and at most 20 stone examples.

3.3. Evaluating training schemes and hyperparameters

There are a number of design choices and model hyperparameters that need to be set. The first is the choice of input data. In this work we present three training schemes of using 2D, 2.5D, or 3D input data, see Fig. 6. The input size, m , determines the dimensionality of the input data.

The image size m used as input to the CNN is set among $m = [5, 11, 25]$. The filter size n and pooling dimension p are set among $n = [3, 6, 11]$ and $p = [1, 2, 5]$, respectively. The filter size and pooling dimension is set so that the size of the pooling layer is $3 \times 3 \times k$. The number of filters k is set among $k = [20, 50]$ for each of the training schemes.

The attenuation values in the CT scans range between -1024 HU and 3071 HU. In order to decrease the training time of the CNN and avoid getting stuck in early local optima, the values are normalized to values between 0 and 1 by first saturating the upper limit for the attenuation values to 1000 HU and then subtracting each pixel in the scan with a minimum value and dividing by the difference between the maximum and the minimum value according to $I_N = (I - \min(I)) / (\max(I) - \min(I))$.

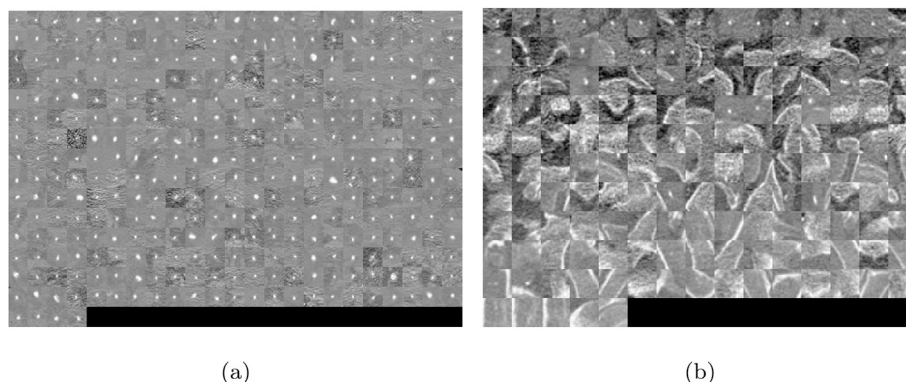


Fig. 5. (a) All stones in the training set. (b) All non-stones in one scan sorted by similarity, calculated with the euclidean distance, to another random real stone.

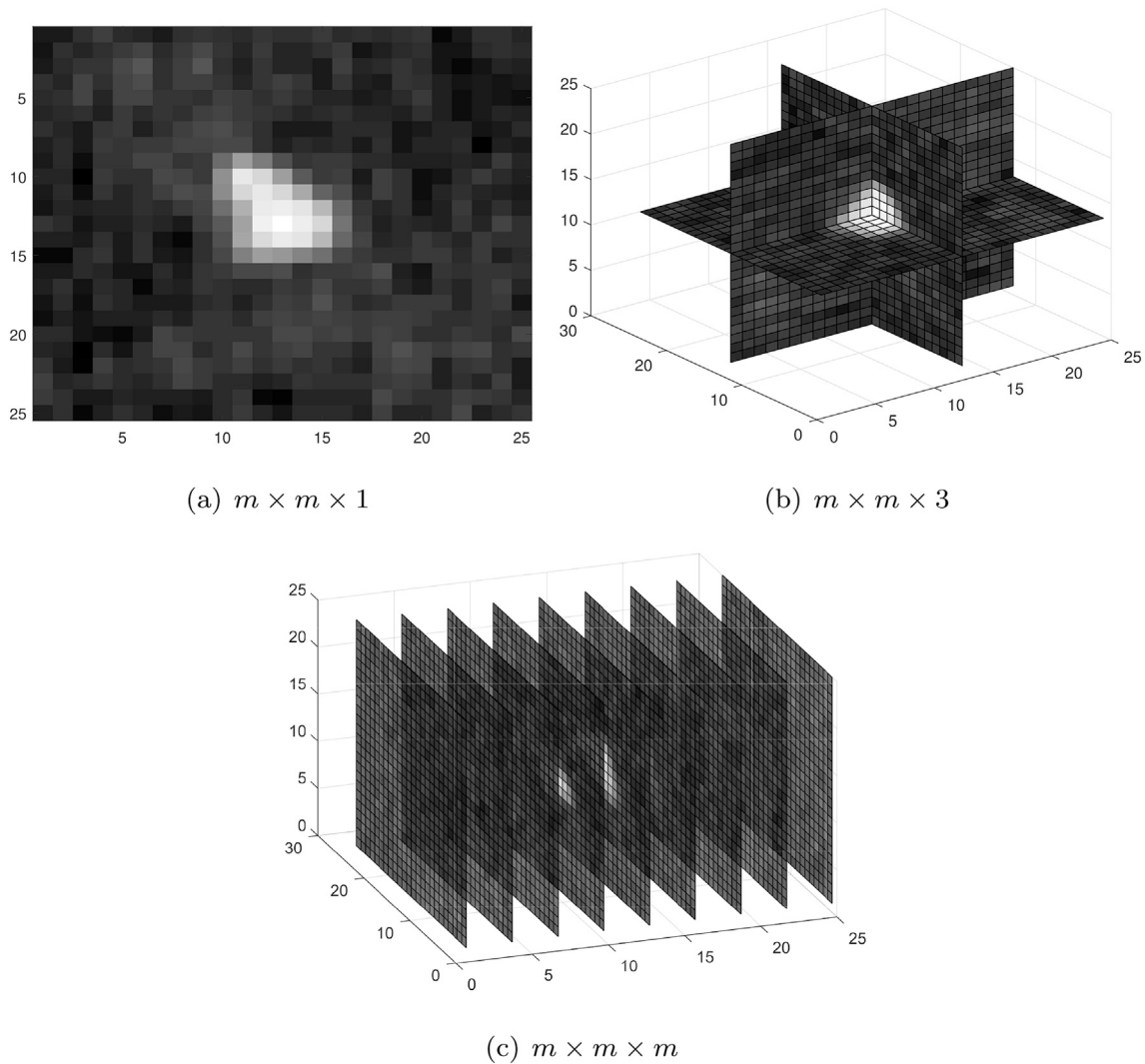


Fig. 6. Three different ways of inputting the data to the algorithm: (a) 2D, (b) 2.5D, and (c) 3D.

$\min(I)$.

The training is done using supervised backpropagation with mini-batch stochastic gradient descent (SGD), early-stopping and momentum. The batch size is selected as 100 and the learning rate is initially set to 0.1 and uses a decaying learning rate if the accuracy on the validation set is not improved over a period of 10 epochs.

Due to the class-imbalance in the training data, the training set is first balanced to have equal class distribution before each training epoch and the F1-score is used instead of classification accuracy for evaluation on the full testing set.

The reported F1-score for the three different training schemes and varying number of hyperparameters on the testing set are shown in Table 1. The highest accuracy with a small CNN model (20 filters) achieves the highest F1-score with the 2.5D scheme and with a small image size of $m = 5$. The highest accuracy with a larger CNN model (50 filters) achieves the highest F1-score with the 2.5D scheme and with a medium image size of $m = 11$. Both the 2.5D and 3D scheme were better than using 2D input. A medium size input image of $m = 11$ achieved a higher F1-score than both the smaller and the larger image sizes for all schemes, meaning that a larger input image size does not improve performance. However, when a large image size of $m = 25$ is used, the performance is increased when the number of filters is increased. This comes at a cost of increased training time. A trade-off between classification performance and training time resulted in the choice of using the 2.5D scheme with image size $m = 11$ and number of filters $k = 20$.

3.4. Classification accuracy

Training was done by randomly divide the original 465 scans into 349 scans for training and 88 scans for testing after removing scans that contain a stone smaller than 5 or larger than 500. From each of the scan in the training set, 100 examples of non-stone were extracted using the method explained in Section 3.2 and 20 examples of stones. The best input method and model architecture from Section 3.3 was selected, namely the training scheme that uses 2.5D input data, image size $m = 11$, filter size $n = 6$, pooling dimension $p = 2$, and number of filters $k = 50$. The model parameters of the CNN was trained using supervised back-propagation and stochastic gradient descent with learning rate 0.1 until

Table 1

F1-score for three training schemes and varying number of filters and image sizes. The highest F1-score for each image size is marked in bold.

	# filters (k)	Image size (m)		
		m = 5	m = 11	m = 25
2D	k = 20	0.688	0.734	0.667
	k = 50	0.626	0.720	0.714
2.5D	k = 20	0.776	0.772	0.719
	k = 50	0.755	0.783	0.735
3D	k = 20	0.701	0.749	0.713
	k = 50	0.723	0.768	0.734

Table 2
Confusion matrix for all CCs in the test set. F1-score 0.35, Recall 0.21, Precision 1.00.

		Predicted	
		Positive	Negative
Actual	Positive	88	0
	Negative	325	18829

convergence.

For testing and evaluation, all pixels in each connected component that are between 5 and 500 pixels large are classified into stone or non-stone and the final prediction is the average over all pixel predictions in that CC. The result was that the model correctly predicted all 88 stones (1 stone in each scan in the test set), 0 false-negatives, 325 false-positives, and a total of 18829 true-negatives, see Table 2 for the confusion matrix. This gives an average of 3.7 ± 4.9 false-positives per scan.

A further analysis of the results can be seen in Fig. 7 were the number of false-positives and true-negatives for each individual CT scan in the test set are shown. The number of true-negatives are mostly constant with

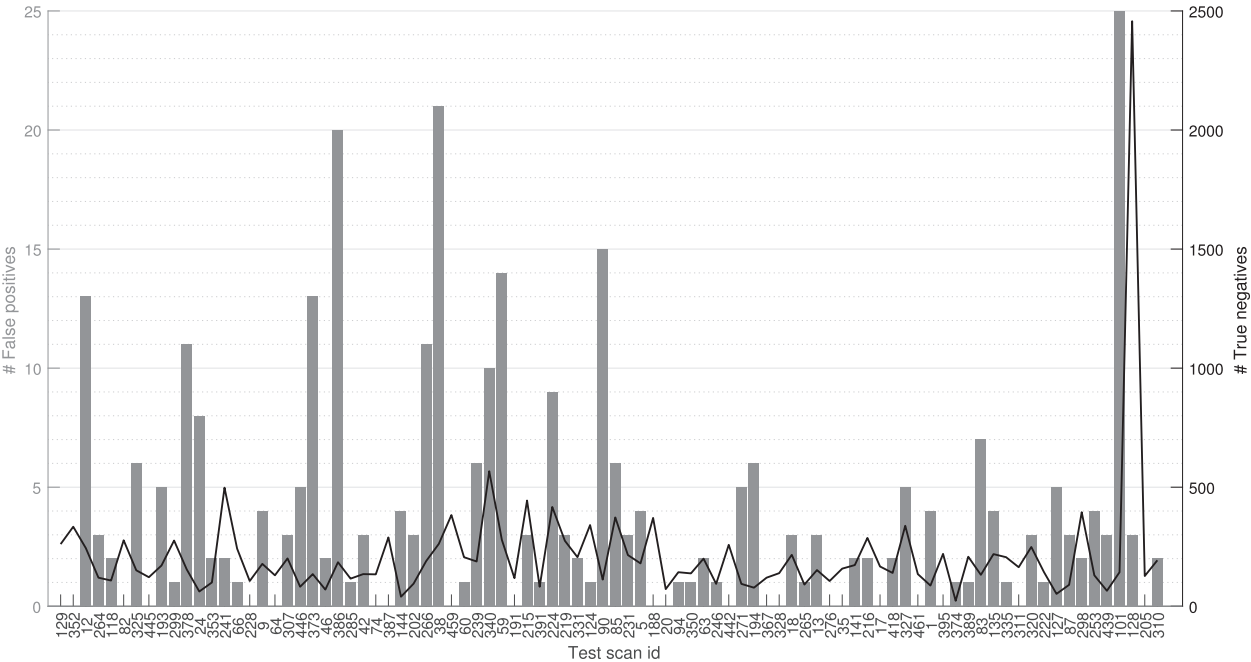


Fig. 7. Number of false-positives (gray bars) and true-negatives (black line) of each scan in the test set.

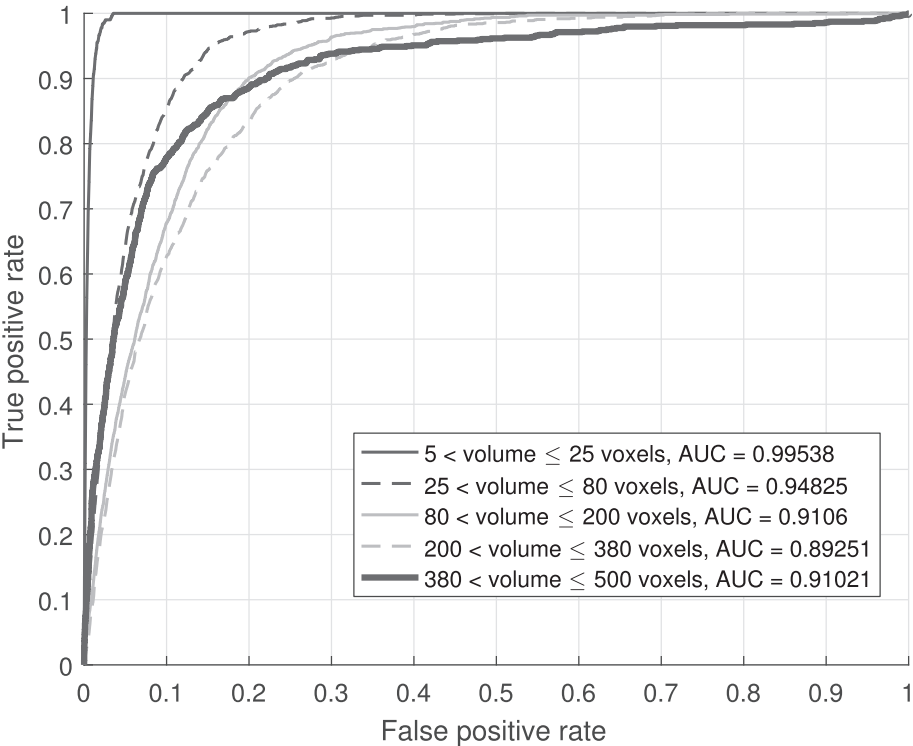


Fig. 8. ROC curve for different sizes of CCs.

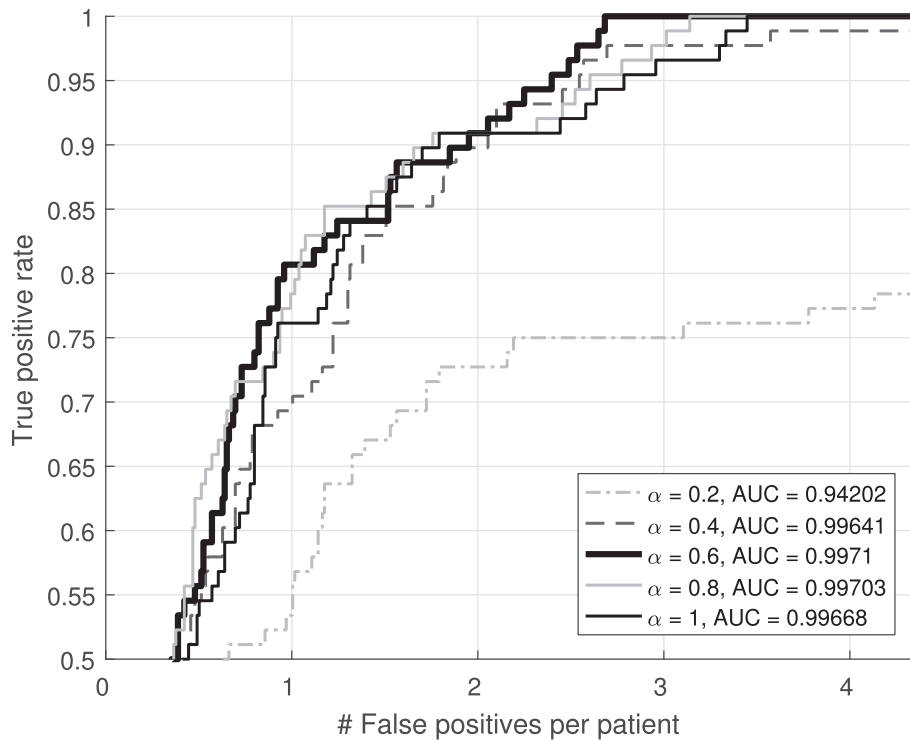


Fig. 9. FROC curve for different trade-offs between classification certainty and stone location probability. The least amount of false positives with a 100% sensitivity is achieved with $\alpha = 0.6$ (black thick line).

an average of 214.0 ± 264.0 per scan, except for scan number 128 that has over 2000 true-negatives. Notice that this large amount of negatives did not have a significant impact on the number of false-positives that was only 3 for this scan. The largest amount of false-positive was detected on scan number 101 with 25 false-positives. A fourth (25%) of the scans have 0 false-positives and 40.9% of the scans have only 1 or 0 false-positives.

Beside the individual scan, the size of the CC also have an influence of the number of false-positives. Fig. 8 shows the ROC curve for different sizes of CCs in the test set. This difference is explained by the use of a small image size of 11×11 pixels and the surrounding information around large stones are not captured. The experiments from Section 3.3 showed that using a larger image size did not improve the overall accuracy since most stones are small. Therefore, we use the anatomical information of the location of the CC to reduce the number of false-positives for large-sized CCs.

3.4.1. Results on using probability map

The probability distribution map from Section 2.2.1 is used to change the output prediction probabilities for the classification from the CNN network. The probability map M gives a probability that the CC at location x is a stone by $P(y = 1|M(x))$ and the CNN model gives a prediction probability that the CC with input data I is a stone by $P(y = 1|I(x))$. The trade-off between how much of these prediction probabilities should influence the final decision is determined by the parameter α by:

$$P(y = 1|I(x), M(x)) = \alpha P(y = 1|I(x)) + (1 - \alpha)P(y = 1|M(x)) \quad (4)$$

where α is between 0 and 1.

The results on a Free Response Operating Characteristic curve (FROC) for different values of α can be seen in Fig. 9. The lowest value for the area-under-the-curve (AUC) is 0.605 and is obtained when $\alpha = 0$, i.e., when only the prediction from the probability map is used and the prediction from the CNN model is ignored. When $\alpha = 1$, i.e., only the prediction from the CNN is used, we get an $AUC = 0.9967$ and on average of

3.69 false-positives per scan with sensitivity 100%, which is the same result reported in Section 3.4. The best trade-off is achieved with $\alpha = 0.6$ which gives an $AUC = 0.9971$ and 2.68 false-positives per scan with sensitivity 100%. The use of anatomical information reduced the amount of false-positives by 27.4%.

4. Discussion and conclusion

In the present study, a Convolutional Neural Network was used for detecting ureteral stones in thin slice CT scans. The main contribution of this work is to approach automatic detection of ureteral stones, by using a representational learning algorithm that is not dependent on the choice of features or segmentation methods, and that takes advantage of the high-resolution CT data.

Without using anatomical information the CNN model achieved a sensitivity of 100% and an average of 3.69 false-positives per patient on a test set of 88 scans. The use of a probability distribution map of stone locations achieved a sensitivity of 100% and an average of 2.68 false-positives per patient.

Convolutional Neural Networks are rapidly becoming a popular method for medical image analysis [18,19]. They have previously been used on a number of medical image modalities, including CT data for tasks such as brain tumor segmentation [16,20], pancreas segmentation [21], and lesion segmentation in livers [22]. To the best of our knowledge, there are no previously published works that uses CNNs on CT scans for ureteral stone detection.

Most previous work for computer aided detection of renal stones use a work-flow of noise reduction, candidate selection, and then a feature-based method for the final classification. The work by Liu et al. uses total variation (TV) flow for image noise reduction and MSER features for finding calculi candidates, and computes a total of 7 texture and shape features to train a Support Vector Machine (SVM) for the task of identifying renal calculi in CT scans [5]. The method was validated on a data set of 192 patients with a false positive rate of 8 per patient and a sensitivity of 69%. The work by Ebrahimi et al. correctly predicted the

presence of a kidney stone in 24 of 29 patients and the absence of a kidney stone in 9 of 10 patients using a method of localization and thresholding [6].

A feature-based approach to differentiate ureteral stones and vascular calcifications has previously been applied by training an Artificial Neural Network (ANN) on a total of 8 shape and texture features on non-contrast CT images [7]. The method was validated on 59 ureter stones and 53 vascular calcifications and resulted in an area under the curve (AUC) of 0.85 for the shape parameters and 0.88 for the texture parameters. The more general problem of detecting the ureteral stones was not addressed in the study.

Our study differentiates from the previous works in several ways. First, the task in the present study is *detection of ureteral stones*, which is a more challenging task for radiologists compared to stones in the kidney, and a more general problem than differentiation between stones and other calcifications. Second, we use the raw pixels on 3D volumes instead of extracted features and we also propose a method for using the anatomical information in the decision process.

In comparison with previous works, the present study represents another step towards an automatic detection algorithm for ureteral stones. However, from a clinical point-of-view, while keeping 100% sensitivity, the false positive rate of 2.7 per CT scan is still not sufficient for replacing the primary diagnostic reading by the radiologist. Radiologists have a reported sensitivity and specificity of approximately 95–100% for ureteral stone CT [3,23]. The promising results in the present study clearly shows the possibilities of CNN for computer aided diagnosis in CT images. With further improvement of the methods, the development of an automatic detection algorithm for ureteral stones ready for clinical use, is within reach. Future work include methods for reducing the amount of pre-processing steps such as finding connected components, binarizing, and selecting training data for class balancing. Another direction for future work is to evaluate the method of using multi-scale CNNs as an option for using anatomical information.

There are some limitations in the current study. The images were acquired at a single radiology department, and more images from different institutions would be needed to improve the generalizability. The smallest stones, that are highly likely to pass spontaneously and the largest stones that are unlikely to be found in the distal ureter were excluded from the study. No scans without ureteral stones were included in the dataset. However, the main outcome measures - sensitivity and false positives rate - are not expected to be altered with negative controls.

In conclusion, the present work describes a CNN method for detecting ureteral stones in thin slice CT. With further development, primary reading of CTs by an automated detection algorithm is feasible.

Conflicts of interest

None declared.

Acknowledgments

The authors would like to acknowledge Antai Llaquet for his contributions. This work has been sponsored by Nyckelfonden (grant OLL-597511) and by Vinnova under the project "Interactive Deep Learning for 3D image analysis".

References

- [1] F.L. Coe, A. Evan, E. Worcester, Kidney stone disease, *J. Clin. Invest.* 115 (10) (2005) 2598–2608.
- [2] O.W. Moe, Kidney stones: pathophysiology and medical management, *The Lancet* 367 (9507) (2006) 333–344.
- [3] A.R. Kambadakone, B.H. Eisner, O.A. Catalano, D.V. Sahani, New and evolving concepts in the imaging and management of urolithiasis: urologists' perspective, *RadioGraphics* 30 (3) (2010) 603–623, <https://doi.org/10.1148/rg.303095146> PMID: 20462984. arXiv, <https://doi.org/10.1148/rg.303095146>, <https://doi.org/10.1148/rg.303095146>.
- [4] K. Doi, Computer-aided diagnosis in medical imaging: historical review, current status and future potential, *Computerized Med. Imag. Graph.* 31 (4–5) (2007) 198–211.
- [5] J. Liu, S. Wang, E.B. Turkbey, M.G. Linguraru, J. Yao, R.M. Summers, Computer-aided detection of renal calculi from noncontrast CT images using TV-flow and MSER features, *Medical Phys.* 42 (1) (2015) 144–153.
- [6] S. Ebrahimi, V.Y. Mariano, Image quality improvement in kidney stone detection on computed tomography images, *J. Image Graphics* 3 (1) (2015) 40–46, <https://doi.org/10.18178/joig.3.1.40-46>.
- [7] H.J. Lee, K.G. Kim, S.I. Hwang, S.H. Kim, S.-S. Byun, S.E. Lee, S.K. Hong, J.Y. Cho, C.G. Seong, Differentiation of urinary stone and vascular calcifications on non-contrast ct images: an initial experience using computer aided diagnosis, *J. Digit. Imag.* 23 (3) (2010) 268–276.
- [8] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324.
- [9] Y. LeCun, B.E. Boser, J.S. Denker, D. Henderson, R.E. Howard, W.E. Hubbard, L.D. Jackel, Handwritten digit recognition with a back-propagation network, in: D.S. Touretzky (Ed.), *Advances in Neural Information Processing Systems 2*, Morgan-Kaufmann, 1990, pp. 396–404.
- [10] K. Suzuki, Pixel-based machine learning in medical imaging, *Int. J. Biomed. Imag.* 2012 (2012) 1 article ID 792079, <https://doi.org/10.1155/2012/792079>.
- [11] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324, <https://doi.org/10.1109/5.726791>.
- [12] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.* 2 (2012) 1097–1105.
- [13] H. Lee, Y. Largman, P. Pham, A.Y. Ng, Unsupervised feature learning for audio classification using convolutional deep belief networks, in: *Proceedings of the 22Nd International Conference on Neural Information Processing Systems*, Curran Associates Inc., USA, 2009, pp. 1096–1104. NIPS'09, <http://dl.acm.org/citation.cfm?id=2984093.2984217>.
- [14] G.W. Taylor, R. Fergus, Y. LeCun, C. Bregler, Convolutional Learning of Spatio-temporal Features, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 140–153, https://doi.org/10.1007/978-3-642-15567-3_11, https://doi.org/10.1007/978-3-642-15567-3_11.
- [15] R. Bellman, *Dynamic Programming*, Courier Corporation, 2013.
- [16] K. Kamnitsas, C. Ledig, V.F. Newcombe, J.P. Simpson, A.D. Kane, D.K. Menon, D. Rueckert, B. Glocker, Efficient multi-scale 3d CNN with fully connected CRF for accurate brain lesion segmentation, *Med. Image Anal.* 36 (2017) 61–78, <http://doi.org/10.1016/j.media.2016.10.004>, <http://www.sciencedirect.com/science/article/pii/S1361841516301839>.
- [17] J. Jendeborg, H. Geijer, M. Alshamari, B. Cierznia, M. Lidén, Size matters: the width and location of a ureteral stone accurately predict the chance of spontaneous passage, *Eur. Radiol.* 27 (11) (2017) 4775–4785. Springer.
- [18] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W.M. van der Laak, B. van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, *Med. Image Anal.* 42 (2017) 60–88, <https://doi.org/10.1016/j.media.2017.07.005>.
- [19] B.J. Erickson, P. Korfiatis, Z. Akkus, T.L. Kline, Machine learning for medical imaging, *RadioGraphics* 37 (2) (2017) 505–515, <https://doi.org/10.1148/rg.2017160130> PMID: 28212054.
- [20] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, H. Larochelle, Brain tumor segmentation with deep neural networks, *Med. Image Anal.* 35 (2017) 18–31, <http://doi.org/10.1016/j.media.2016.05.004>, <http://www.sciencedirect.com/science/article/pii/S1361841516300330>.
- [21] H.R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E.B. Turkbey, R.M. Summers, Deeporgan: multi-level deep convolutional networks for automated pancreas segmentation, in: *Medical Image Computing and Computer-assisted Intervention*, Springer, 2015, pp. 556–564.
- [22] W. Li, F. Jia, Q. Hu, Automatic segmentation of liver tumor in CT images with deep convolutional neural networks, *J. Comput. Commun.* 3 (11) (2015) 146–151, <https://doi.org/10.4236/jcc.2015.311023>.
- [23] S.A. Pfister, A. Deckart, S. Laschke, S. Dellas, U. Otto, C. Buitrago, J. Roth, W. Wiesner, G. Bongartz, T.C. Gasser, Unenhanced helical computed tomography vs intravenous urography in patients with acute flank pain: accuracy and economic impact in a randomized prospective trial, *Eur. Radiol.* 13 (11) (2003) 2513–2520, <https://doi.org/10.1007/s00330-003-1937-1>.