Strategising template-guided needle placement for MR-targeted prostate biopsy

Iani JMB Gayo, Shaheer U. Saeed, Dean C. Barratt,
Matthew J. Clarkson, Yipeng Hu

[1] Department of Medical Physics and Biomedical Engineering,
[2] Wellcome/EPSRC Centre for Interventional and Surgical Sciences,
[3] Centre for Medical Image Computing,
University College London, UK

Abstract. Clinically significant prostate cancer has a better chance to be sampled during ultrasound-guided biopsy procedures, if suspected lesions found in pre-operative magnetic resonance (MR) images are used as targets. However, the diagnostic accuracy of the biopsy procedure is limited by the operator-dependent skills and experience in sampling the targets, a sequential decision making process that involves navigating an ultrasound probe and placing a series of sampling needles for potentially multiple targets. This work aims to learn a reinforcement learning (RL) policy that optimises the actions of continuous positioning of 2D ultrasound views and biopsy needles with respect to a guiding template, such that the MR targets can be sampled efficiently and sufficiently. We first formulate the task as a Markov decision process (MDP) and construct an environment that allows the targeting actions to be performed virtually for individual patients, based on their anatomy and lesions derived from MR images. A patient-specific policy can thus be optimised, before each biopsy procedure, by rewarding positive sampling in the MDP environment. Experiment results from fifty four prostate cancer patients show that the proposed RL-learned policies obtained a mean hit rate of 93% and an average cancer core length of 11 mm, which compared favourably to two alternative baseline strategies designed by humans, without hand-engineered rewards that directly maximise these clinically relevant metrics. Perhaps more interestingly, it is found that the RL agents learned strategies that were adaptive to the lesion size, where spread of the needles was prioritised for smaller lesions. Such a strategy has not been previously reported or commonly adopted in clinical practice, but led to an overall superior targeting performance, achieving higher hit rates (93% vs 76%) and measured cancer core lengths (11.0mm vs 9.8mm) when compared with intuitively designed strategies.

Keywords: Reinforcement Learning · Prostate Cancer · Targeted Biopsy · Planning

1 Introduction

Recent development in multiparametric MR imaging (mpMRI) techniques provides a means of noninasive localisation of suspected prostate cancer [1], which

enables clinicians to target these lesions during the follow-up ultrasound-guided biopsy for further histopathology confirmation. This MR-targeted approach has been shown to reduce both the false positive and false negative detection, compared to previously adopted random biopsy [1, 2], and subsequently motivated research in developing multimodal MR-to-ultrasound image registration [3].

Needle sampling of the MR-identified targets, with or without registration errors, can still be a challenging and arguably overlooked task. Operator expertise was found to be an important predictor in detecting clinically significant prostate cancer [4]. Planning strategies is important for navigating the ultrasound probe, to better observe the targets with respect to imaging, and for manual needle positioning. In transperineal biopsy, the introduction of brachytherapy templates (See Fig.1 for an example) helps the needle deployment - a procedure that is of interest in this study, but choice between 13×13 grid positions remains a subjective decision. For example, a common clinical practice aims at the target centre, but it has been shown to yield an insufficient sampling of the heterogeneous cancer [5] and possibly an inferior diagnostic accuracy in terms of disease-representative grading [6], compared with more spread needle placement. The design of an optimum strategy is further complicated by the need of multiple needles for individual targets, for maximising the hit rate, and the multifoci nature of prostate cancer, which requires repeated sampling of one or more targets.

To the best of our knowledge, there has not been any computer-assisted sampling strategy that takes into account the previous needle deployment(s) or quantitatively optimises patient-and-target-specific needle distribution. In summary, improving the targeting strategy may help reduce the significant false negative rate found in MR-targeted biopsy (reported being as high as 13% [7]), and hence improves the chance of early cancer detection for patients.

Reinforcement learning (RL) has been proposed for medical image analysis tasks [8], such as landmark detection [9], plane finding [10], and for surgical planning such as hysterectomy [11] and orthopaedic operations [12]. It has also been used for needle path planning in minimally-invasive robotic surgery [13], [14]. It is its ability to learn intelligent policies for sequential decision making that provides a potential solution to problems without requiring direct supervision for each action taken, a common constraint in developing machine learning-assisted methods for complex and skill-demanding surgical and intervention applications. This makes RL suitable for finding an optimal targeting strategy, which requires complex decisions for which there is no established best method.

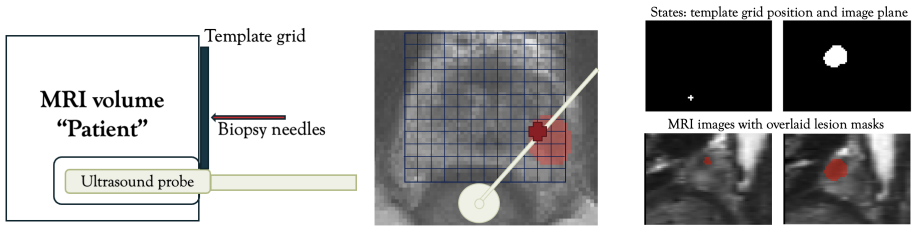
In this study, we investigate the feasibility of using RL to plan patient-specific needle sampling strategies, optimised in pre-operative MR-derived RL environments. We present experimental results based on clinical data from prostate cancer patients and compare the proposed method, using a set of clinically important metrics, to baseline strategies that are designed by human intuition and an interactive targeting performed by two observers. We conclude by reporting a set of interesting observations that demonstrate the benefit of using the proposed RL-learned patient-specific strategies. These indeed adapted effectively to individual procedures and varying targets, for improved final performance of the

sequential target sampling. Consistent hit rates were achieved with less variance for both smaller and larger lesions as a result of the learned adaptive strategies.

2 Method

The agent-environment interactions are modelled as a Markov decision process (MDP), and summarised as a 4-tuple $\langle \mathcal{S}, \mathcal{A}, r, p \rangle$, where $\mathcal{S}$ and $\mathcal{A}$ are the state and action spaces consisting of all possible observed states as input and actions as output for the agent, respectively. $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is the reward function which maps state-action pairs to a real value. The state transition distribution is defined by $p : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \to [0, 1]$ which denotes probability of transitioning to the next state, given the current state-action pair. In this section, we develop an environment for template-guided biopsy sampling of the cancer targets, the MDP components and the policy learning strategy.

2.1 Patient-specific prostate MR-derived biopsy environment



(a) Environment (b) Template grid (c) States and targets

Fig. 1: Simulated biopsy procedure environment. (a) Placement of ultrasound probe and template grid within the MRI volume. (b) Visualisation of ultrasound probe rotation which is always aligned with the chosen template grid position. (c) Examples of (top) states and (bottom) overlaid MR-identified targets

The environment is illustrated in Figure 1a for the targeted biopsy procedures, where virtual biopsy needles are inserted through the perineum via a brachytherapy template grid consisting of 13x13 holes that are 5mm apart. Other needle-based treatments such as cryotherapy, brachytherapy and radiofrequency ablation [15] may also be applicable but are not discussed further in this paper. The position of the transrectal ultrasound probe is approximated within the rectum directly underneath the prostate gland, with a fixed distance to the template grid such that the top of the probe is aligned with the lower side of the template, as shown in the Figure 1a. Both anatomical and pathological structures can be sampled, at any sagittal ultrasound imaging plane given an arbitrary angle, as illustrated in Figure 1b.

We summarised a number of considerations in designing and constructing the adopted biopsy environment as follows. 1) The prostate gland from each MR

volume, the MR-identified targets (as in Figure 1c), and key landmarks such as position of the rectum are all manually segmented from individual patients to construct the biopsy environment. Automated methods for segmenting these regions of interest have been proposed, e.g. [16], [17]. 2) Binary segmentation are provided as observations for the RL agents, as opposed to ultrasound image intensities, which are neither available during planning nor straightforward to synthesize from MR images. We argue that the use of binary representation would be more robust to train the RL agents and the resulting methods are more likely to generalise to different procedures and planning MR images, especially given the existing MR and ultrasound segmentation and registration algorithms described above. 3) Uncertainty in MR-to-ultrasound registration can and should be added to the segmented regions, together with other potential errors in localising these regions during observation such as observer variability in manual segmentation used in this work. We would like to point out that, however, these localisation errors are unlikely to be independent to each other and the dependency of RL model generalisability on how precisely these need to be modelled remain open research questions. Sec. 3 discusses further details adopted in our experiments. 4) In the presented experiments, we focus on targeting the index lesions, those are of largest volumes in each case, to provide first results that show the efficacy of modelling the dynamic biopsy sampling process. However, the described MDP should be directly applicable for and likely to be more effective in cases with multiple lesions.

2.2 The MDP components

State - At a given time point t during the procedure, the agent receives information about its current state $s_t \in \mathcal{S}$: the chosen grid point and the re-sampled 2D image plane obtained by rotating the probe to the current template grid position, as in Figure 1c. This current position is determined by the previous action. This is to test the scenario with least assumptions, where the overall 3D anatomical and pathological information may be corrupted or unreliable due to intra-procedural uncertainties from patient movement and outdated registration.

Actions - The agent takes actions $a_t \in \mathcal{A}$ which modify its position on the template grid. These actions are relative to the current position of the agent (i, j) and are defined as $a_t = (\delta_i, \delta_j)$ such that the new position is given by $(i + \delta_i, j + \delta_j)$, where $\delta_i, \delta_j \in [-15, 15]$. By formulating this relative grid-moving action, we consider the biopsy needle is positioned on the image plane, with an insertion depth that overlaps the needle centre and the centre of the observed 2D target, subject to small predefined positioning errors in each direction. These are commonly adopted practice though not strictly enforced, and we found that increasing the flexibility by independently positioning the ultrasound probe and needle may unnecessarily make the training difficult to converge.

Reward - The reward at the time t is computed based on the reward function $R_t = r(s_t, a_t)$, during training. The agent is rewarded positively if the fired needle obtains samples of the lesion. A penalty is given when chosen grid positions are

outside of the prostate, to avoid hitting surrounding critical structures. From initial experiments, it was found that a greater reward of +5 lead to a faster convergence during training, encouraging the agent to hit the lesions, whilst a penalty of -1 was enough to deter the agent from firing outside the prostate gland. Reward shaping is also introduced to guide the agent towards the lesion, thereby speeding up the learning process. Similar to [9] and [10], a sign function Sgn of the difference between $dist_{t-1}$ and $dist_t$ is computed, where $dist_t$ represents Euclidean distance between target centre and needle trajectory at time t.

$$r = \begin{cases} +5 & \textit{if biopsy needle intersected with target} \\ \text{-1} & \textit{if biopsy needle placed outside prostate} \\ \text{Sgn}(dist_{t-1} - dist_t) & \textit{otherwise} \end{cases} \quad (1)$$

2.3 Policy learning

The navigation and sampling strategy is parameterised by a policy neural network π_θ, with parameters θ, quantifying the probability of performing action a_t given state s_t. Agent's actions then can be sampled from the policy, $a_t \sim \pi_\theta(\cdot|s_t)$. During the policy training, the accumulated reward $Q^{\pi_\theta}(s_t, a_t) = \sum_{k=0}^{T} \gamma^k R_{t+k}$ is maximised, where γ is a discount factor set to 0.9, to obtain the optimal policy π_{θ^*}, $\theta^* = \arg\max_\theta \mathbb{E}_{\pi_\theta}[Q^{\pi_\theta}(s_t, a_t)]$. With continuous actions, policy gradient (PG) and actor-critic (AC) algorithms can thus be adopted for the optimisation.

3 Experiments

Data set - The T2-weighted MR images and their segmentation were acquired from 54 prostate cancer patients. These were obtained as part of clinical trials, PROMIS [1] and SmartTarget [7], where patients underwent ultrasound-guided minimally invasive needle biopsies and focal therapy procedures.

RL algorithm implementation - An agent was trained for each patient individually using the Stable Baselines implementation of PPO [18]. Each agent was trained for 120,000 episodes and a model was selected with the highest average reward after 10 episodes. Each episode was limited to a maximum of 15 time steps, but can terminate early if any 5 needles hit the lesion. At each episode the agent is initialised at random starting positions on the template grid. Based on estimated registration error reported previously [3], random localisation error was added in the observed states, equivalent to a Gaussian noise with a standard deviation of 1.73mm in each of the x, y and z coordinates, or a mean distance error of 3mm. The PPO algorithm [19] was used in reported results, as it guarantees a monotonic reward improvement and stability of training. However, we also report a lack of substantial difference in performance to other tested algorithms, including DDPG [20] and SAC [21]. The policy network was based on ResNet18 [22] architecture, with an additional fully-connected layer for a linear output. An Adam optimiser was used, with a learning rate of 0.0001. It could be of future interest to compare further network architectures and PG/AC algorithms on the proposed RL problem, but is considered beyond the scope of current work.

Biopsy performance metrics - To quantitatively assess target sampling performance, three biopsy-specific metrics are used in this study: hit rate (HR), cancer core length (CCL) and needle area (NA). The HR is the number of needle samples that contains the target, i.e. positive samples, divided by the number of needles fired. Five needles are chosen to represent the maximum number typically used in targeted biopsy [23]. CCL is the total length (in mm) overlapping the target, i.e. the sampled target tissue, in individual needles. CCL $>= 6mm$ often indicates clinical significance [24]. NA estimates the coverage of all fired needles in each episode, defined as the area of an approximating ellipse, $NA = \pi * std_x \times std_y$, where std_x and std_y are standard deviations in the needle navigating x-y plane, defined by the template grid position.

Baseline strategies - Two strategies were compared with the proposed agent, designed to provide an estimate of what clinicians are likely to achieve in practice. For a fair comparison, the same observed targets, the states, and starting positions were used. Student's t-tests were used when comparison is made at a significance level of $\alpha = 0.05$, unless otherwise indicated. The first strategy *(Sweeping strategy)* adopted a simple sweeping of the biopsy needle together with the ultrasound probe, from left to right in a 5mm interval. The target was sampled at the centre of the observed target, i.e. fired, once an image plane is encountered with a lesion. The second strategy *(Scouting strategy)* moves the virtual probe to scout all candidate positions that samples the target, before 5 random ones were selected as fired positions among these candidates. Inter- and intra-operator variance is one of the most important factors that impact the performance of a sampling strategy. For the baseline strategies, additional Gaussian noise was added to the chosen needle/probe positioning with a varying bias and a varying standard deviation (SD). Increasing the SD leads to higher uncertainty in placing the needles, while the bias indicates a targeting strategy that does not aim for the target centre, e.g. for avoiding empty cores or urethra. The experiments were repeated using different combinations of the two variables (each ranging between 0 to 10mm), to test different strategies.

Interactive experiments by two observers - Two human operators, one computer scientist and one biomedical imaging researcher, interacted with a custom-made interface that displays the current template grid position and image plane observed. They were asked to choose where to sample and how far to spread the needles. This is a simple interactive experiment to provide preliminary results that can be compared with the other described strategies.

4 Results

Learned strategy performance - From Table 1, the agent outperforms both baseline strategies in HR and CCL (both p-values$<$0.005), but not in NA, with noticeably smaller variability in HR. Different levels of bias did not lead to significantly different targeting results, while higher SD increased the spread of needles, but reduced CCL and HR. In general, the results between the sweeping

and scouting strategies were not found statistically different in CCL and NA. The scouting strategy resulted in an increased HR (p-value<0.05) compared to sweeping, but does not outperform the agent which still obtains the highest HR.

Table 1: Summary of biopsy performance from the RL agent (top row) and the sweeping and scouting strategies for different bias and SD combinations

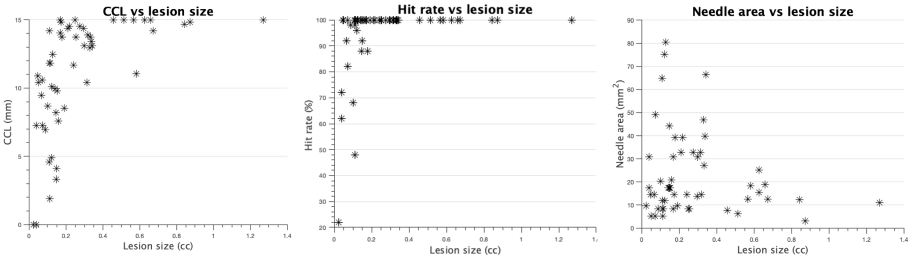
		Baseline 1 (Sweeping)			**Baseline 2 (Scouting)**		
Bias	SD	CCL(mm)	HR(%)	NA(mm^2)	CCL(mm)	HR(%)	NA(mm^2)
Agent		11.13±3.43	93.40 ± 11.44	22.14±18.18	11.13±3.43	93.40 ± 11.44	22.14±18.18
0	0	7.95±3.00	53.36±35.59	23.67±16.57	8.40±2.65	61.10±30.01	31.31±28.53
0	5	7.45±3.19	49.43±34.10	56.81±50.25	8.32±3.04	55.19±30.46	39.84±20.48
0	10	4.89±3.29	30.94±28.71	108.00±86.24	5.67±4.32	43.70±31.99	114.35±71.20
5	0	8.90±4.00	53.69±28.44	36.53±28.91	8.32±3.00	65.19±30.04	36.53±28.91
5	5	7.88±3.52	51.32±32.94	60.44±47.62	7.28±3.17	54.48±31.04	72.08±40.89
5	10	6.06±3.95	41.51±31.58	86.15±63.15	5.74±4.41	44.82±31.02	113.88±79.60
10	0	7.29±3.08	47.80±34.16	27.47±18.38	7.65±3.05	54.81±31.13	27.59±22.95
10	5	7.14±3.38	42.44±28.00	52.78±48.02	6.40±3.64	50.74±29.00	55.45±42.36
10	10	4.74±3.46	28.70±27.22	106.17±115.95	5.98±3.42	47.04±31.22	112.68±95.85

Table 2: Obtained biopsy strategy metrics for the agent and two human observers

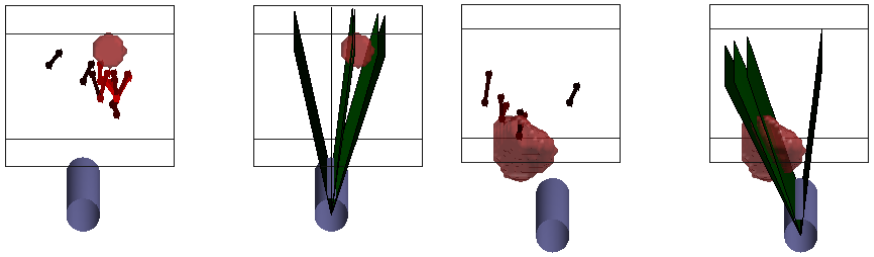
Observers	**CCL (mm)**	**HR(%)**	**NA (mm^2)**
Agent	11.13 ± 3.43	93.40 ± 11.44	22.14 ± 18.18
Observer 1	9.71 ± 3.78	66.30 ± 20.55	42.85 ± 23.36
Observer 2	9.83 ± 3.89	76.30 ± 19.50	64.90 ± 38.84

Interactive experiments by two observers - The agent outperforms both observers in CCL (p-values=0.020 & 0.040), but its NA values are more than double that of the agents, suggesting a potential trade-off between sampling coverage and precision. For HR, the agent outperforms both observers (p-value < 0.001), which demonstrates that the agent could achieve an overall comparable performance as human observers, with a significantly higher CCL.

Learned strategy for varying target sizes - From Figure 2a and Table 3, we observe an interesting behaviour from the learned agent: the smaller the lesions, the larger the spread of the needles. At a volume threshold of 0.4 cc, the mean CCL and NA are statistically different for smaller and larger lesions (p-values=0.002 & 0.040), whilst the difference in HR is not (p-value=0.166). This result may seem counter-intuitive, as one would be cautious in spreading needles for a small target. However, the agent learned to distribute needles more widely for smaller lesions, attempting to maintain the hit rate, given the inevitable presence of target localisation uncertainty described in Sec. 2.1. Visual examples of the learned strategies are shown in Figures 2b and 2c. This learned behaviour is interesting because a) it has not been observed previously, either in literature or in clinical practice. b) it improved the overall targeting performance compared to the target-size-agnostic baseline strategies and c) this could be suggested to urologists and interventional radiologists with or without the proposed RL assistance.



(a) Biopsy metrics CCL, HR and NA vs lesion size



(b) Small lesion size 0.2cc (c) Large lesion size 0.4cc

Fig. 2: (a): CCL, HR and NA as a function of lesion size. (b) and (c): Examples of different sized targets (red), corresponding to the learned policies, represented by the needle sampling positions (red sticks, brighter indicates later time steps) and observed ultrasound images in green. The bounding cube and the cylinder represent the MR prostate volume and probe, respectively.

Table 3: CCL, HR and NA for different lesion sizes using threshold size < 0.4cc

Lesion size	**CCL** (mm)	**HR** $(\%)$	**NA** (mm^2)
Small lesions	10.26 ± 4.19	93.16 ± 16.03	25.26 ± 19.45
Large lesions	14.52 ± 1.18	100.00 ± 0.00	13.02 ± 6.25

5 Discussion and Conclusion

The results show that the developed RL agents are competitive in sampling MR-derived targets, compared with intuitively devised strategies. Higher HR and average CCL were obtained by the agents, which was achieved by reducing the spread of the needles compared to baseline strategies. Furthermore, the learned strategies adapted to patient-specific procedures and varying pathology. The agents learned to achieve similar HR for different sized lesions, by spreading the fired needles more for smaller lesions. Such behaviour has not been observed before, and could be suggested to clinicians for improved targeting performance. Assumptions, such as number of allowed needles, template positioning and uncertainties in localisation/placement, have been made to facilitate the proposed

pre-procedural planning. Some of them may be relaxed for an intra-procedural guidance tool - as a potential extension of this work, the others may require further validation. More importantly, the improved targeting performance provides means in mitigating the cancer under-sampling and help timely diagnosis of a significant number of prostate cancer patients with current MR-targeted biopsy.

Acknowledgement

This work is supported by the EPSRC-funded UCL Centre for Doctoral Training in Intelligent, Integrated Imaging in Healthcare (i4health) [EP/S021930/1], EPSRC [EP/T029404/1], and the Department of Health's NIHR funded Biomedical Research Centre at University College Hospital. This work is also supported by the International Alliance for Cancer Early Detection, an alliance between Cancer Research UK [C28070/A30912; C73666/A31378], Canary Center at Stanford University, the University of Cambridge, OHSU Knight Cancer Institute, University College London and the University of Manchester. This work was also supported by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences [203145Z/16/Z].

References

1. Ahmed, H.U., El-Shater Bosaily, A., Brown, L.C., Gabe, R., Kaplan, R., Parmar, M.K., Collaco-Moraes, Y., Ward, K., Hindley, R.G., Freeman, A., Kirkham, A.P., Oldroyd, R., Parker, C., Emberton, M.: Diagnostic accuracy of multi-parametric MRI and TRultrasound biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. The Lancet. 389, 815–822 (2017). https://doi.org/10.1016/s0140-6736(16)32401-1
2. Simmons, L.A.M., Ahmed, H.U., Moore, C.M., Punwani, S., Freeman, A., Hu, Y., Barratt, D., Charman, S.C., Van der Meulen, J., Emberton, M.: The PICTURE study — Prostate Imaging (multi-parametric MRI and Prostate HistoScanningTM) Compared to Transperineal Ultrasound guided biopsy for significant prostate cancer Risk Evaluation. Contemporary Clinical Trials. 37, 69–83 (2014). https://doi.org/10.1016/j.cct.2013.11.009
3. Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C.M., Emberton, M., Ourselin, S., Noble, J.A., Barratt, D.C., Vercauteren, T.: Weakly-supervised convolutional neural networks for multimodal image registration. Medical Image Analysis. 49, 1–13 (2018). https://doi.org/10.1016/j.media.2018.07.002.
4. Stabile, A., Dell'Oglio, P., Gandaglia, G., Fossati, N., Brembilla, G., Cristel, G., Dehò, F., Scattoni, V., Maga, T., Losa, A., Gaboardi, F., Cardone, G., Esposito, A., De Cobelli, F., Del Maschio, A., Montorsi, F., Briganti, A.: Not All Multiparametric Magnetic Resonance Imaging–targeted Biopsies Are Equal: The Impact of the Type of Approach and Operator Expertise on the Detection of Clinically Significant Prostate Cancer. European Urology Oncology. 1, 120–128 (2018). https://doi.org/10.1016/j.euo.2018.02.002.
5. Calio, B.P., Deshmukh, S., Mitchell, D., Roth, C.G., Calvaresi, A.E., Hookim, K., McCue, P., Trabulsi, E.J., Lallas, C.D.: Spatial distribution of biopsy cores and the

detection of intra-lesion pathologic heterogeneity. Therapeutic Advances in Urology. 11, 1756287219842485 (2019). https://doi.org/10.1177/1756287219842485.
6. Orczyk, C., Hu, Y.P., Gibson, E., El-Shater Bosaily, A., Kirkham, A., Punwani, S., Brown, L., Bonmati, E., Coraco-Moraes, Y., Ward, K. and Kaplan, R., 2017. MP38-07 SHOULD WE AIM FOR THE CENTRE OF AN MRI PROSTATE LESION? CORRELATION BETWEEN MPMRI AND 3-DIMENSIONAL 5MM TRANSPERINEAL PROSTATE MAPPING BIOPSIES FROM THE PROMIS TRIAL. The Journal of Urology, 197(4S), pp.e486-e486
7. Hamid, S., Donaldson, I.A., Hu, Y., Rodell, R., Villarini, B., Bonmati, E., Tranter, P., Punwani, S., Sidhu, H.S., Willis, S., van der Meulen, J., Hawkes, D., McCartan, N., Potyka, I., Williams, N.R., Brew-Graves, C., Freeman, A., Moore, C.M., Barratt, D., Emberton, M.: The SmartTarget Biopsy Trial: A Prospective, Within-person Randomised, Blinded Trial Comparing the Accuracy of Visual-registration and Magnetic Resonance Imaging/Ultrasound Image-fusion Targeted Biopsies for Prostate Cancer Risk Stratification. European Urology. 75, 733–740 (2019). https://doi.org/10.1016/j.eururo.2018.08.007.
8. Zhou, S.K., Le, H.N., Luu, K., Nguyen, H.V., Ayache, N.: Deep reinforcement learning in medical imaging: A literature review. arxiv.org. (2021).
9. Alansary, A., Oktay, O., Li, Y., Folgoc, L.L., Hou, B., Vaillant, G., Kamnitsas, K., Vlontzos, A., Glocker, B., Kainz, B., Rueckert, D.: Evaluating reinforcement learning agents for anatomical landmark detection. Medical Image Analysis. 53, 156–164 (2019). https://doi.org/10.1016/j.media.2019.02.007.
10. Alansary, A., Folgoc, L.L., Vaillant, G., Oktay, O., Li, Y., Bai, W., Passerat-Palmbach, J., Guerrero, R., Kamnitsas, K., Hou, B., McDonagh, S., Glocker, B., Kainz, B., Rueckert, D.: Automatic View Planning with Multi-scale Deep Reinforcement Learning Agents. arXiv:1806.03228 [cs]. (2018).
11. Sato, M., Koga, K., Fujii, T., Osuga, Y.: Can Reinforcement Learning Be Applied to Surgery? IntechOpen (2018).
12. Ackermann, J., Wieland, M., Hoch, A., Ganz, R., Snedeker, J.G., Oswald, M.R., Pollefeys, M., Zingg, P.O., Esfandiari, H., Fürnstahl, P.: A New Approach to Orthopedic Surgery Planning Using Deep Reinforcement Learning and Simulation. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. 12904, 540–549 (2021).
13. Lee, Y., Tan, X., Chng, C.B., Chui, C.K: Simulation of Robot-Assisted Flexible Needle Insertion using Deep Q-Network (2019)
14. Tan, X., Lee, Y., Chng, C.-B., Lim, K.-B., Chui, C.-K.: Robot-assisted flexible needle insertion using universal distributional deep reinforcement learning. International Journal of Computer Assisted Radiology and Surgery. 15, 341–349 (2019). https://doi.org/10.1007/s11548-019-02098-7.
15. Mayo Clinic: Prostate brachytherapy, https://www.mayoclinic.org/tests-procedures/prostate-brachytherapy/about/pac-20384949, last accessed 2020/05/28.
16. Aldoj, N., Biavati, F., Michallek, F., Stober, S., Dewey, M.: Automatic prostate and prostate zones segmentation of magnetic resonance images using DenseNet-like U-net. Scientific Reports. 10, 14315 (2020). https://doi.org/10.1038/s41598-020-71080-0.
17. Dai, Z., Carver, E., Liu, C., Lee, J., Feldman, A., Zong, W., Pantelic, M., Elshaikh, M., Wen, N.: Segmentation of the Prostatic Gland and the Intraprostatic Lesions on Multiparametic Magnetic Resonance Imaging Using Mask Region-Based Convolutional Neural Networks. Advances in Radiation Oncology. 5, 473–481 (2020). https://doi.org/10.1016/j.adro.2020.01.005.

18. StableBaselines: PPO — Stable Baselines3 1.4.1a3 documentation, https://stable-baselines3.readthedocs.io/en/master/modules/ppo.html
19. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal Policy Optimization Algorithms, https://arxiv.org/abs/1707.06347.
20. Lillicrap, T., Hunt, J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., Wierstra, D.: CONTINUOUS CONTROL WITH DEEP REINFORCEMENT LEARNING. (2019).
21. Haarnoja, T., Zhou, A., Abbeel, P., Levine, S.: Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. (2018).
22. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition, https://arxiv.org/abs/1512.03385.
23. Song, G., Ruan, M., Wang, H., Fan, Y., He, Q., Lin, Z., Li, X., Li, P., Wang, X., He, Z., Zhou, L.: How Many Targeted Biopsy Cores are Needed for Clinically Significant Prostate Cancer Detection during Transperineal Magnetic Resonance Imaging Ultrasound Fusion Biopsy? The Journal of Urology. 204, 1202–1208 (2020). https://doi.org/10.1097/JU.0000000000001302.
24. Ahmed, H.U., Hu, Y., Carter, T., Arumainayagam, N., Lecornet, E., Freeman, A., Hawkes, D., Barratt, D.C., Emberton, M.: Characterizing Clinically Significant Prostate Cancer Using Template Prostate Mapping Biopsy. Journal of Urology. 186, 458–464 (2011). https://doi.org/10.1016/j.juro.2011.03.147