## APPLIED RESEARCH

# Deep Learning for Detection of Clinical Operations in Robot-Assisted Percutaneous Renal Access

**BULAT IBRAGIMOV[1,2], JANET ZHEN[2], AND ELIF AYVALI[2]**
[1]Department of Computer Science, University of Copenhagen, 2100 Copenhagen, Denmark
[2]Johnson and Johnson, Redwood City, CA 94065, USA

Corresponding author: Bulat Ibragimov (bulat@di.ku.dk)

**ABSTRACT** Percutaneous nephrolithotomy (PCNL) is the current standard of care for patients with a total renal stone burden > 20 mm. Gaining access to the kidney is a crucial step, as the position of the percutaneous tract can affect the ability to manipulate a nephroscope during the procedure. However, gaining percutaneous access using fluoroscopic guidance has a challenging learning curve, with only a minority of urologists can successfully establish the access. In addition to difficult access, the PCNL carries a risk of bleeding and the need for blood transfusion. Robotic assistance may be a key towards accurate and reliable access. Beyond assisting with renal access, a robotic platform can record data of importance related to the user's activities via sensor-equipped instruments. The analysis of these activities is crucial for understanding what constitutes a successful and safe procedure. In this paper, we harness the powers of machine learning to automatically analyze physicians' activities during robotic-assisted renal access using the Monarch®Platform, Urology. A machine learning framework based on a combination of a 1-dimensional U-net and random forests was developed to find consistent patterns in the sensor data characteristic of needle insertions. This framework retrospectively analyzed data previously obtained from 248 percutaneous renal access procedures. These procedures were performed on 18 human cadaveric models by 17 practicing urologists and one urologist proxy. The framework automatically recognized 94% of all first needle insertions in each procedure and labeled them with an accuracy of 0.81 in terms of the Dice coefficient. The recognition accuracy for secondary insertions was 66%. The automatically detected needle insertions were used to calculate clinical metrics such as tract length, anterior-posterior and cranial-caudal angles of the insertion site, as well as user skills such as trajectory deviation and targeting accuracy.

**INDEX TERMS** Percutaneous nephrolithotomy, machine learning, robotic surgery, human performance analysis.

## I. INTRODUCTION

The urinary stone disease is among the most common acute urinary pathologies and is estimated to affect 10%-15% of US adults [1]. If the disease is discovered at its early stage, small stones could be treated conservatively using a combination of medical expulsive therapy, alpha-blockers, and pain control. The treatment of larger or symptomatic stones is usually performed with Shock Wave Lithotripsy (SWL), Flexible Ureteroscopy (FU), and PCNL. PCNL is one of the treatments of choice for the removal of stones through

small incisions on the patient's skin. The number of PCNLs has been growing by 6% per year, with the fastest growth rates in the US Midwest [2]. The PCNL growth is, however, accompanied by increased complication rates after up to 19% of treatments [2]. Infectious complications and bleeding are the most common intra- and post-operative complications. The predictive factors for such post-treatment complications include large stone size, lengthy PCNL procedure, and high body-mass index (BMI) [3], [4].

Gaining percutaneous access is a delicate task and may be challenging, even for experts. The kidney papillae, the optimal targets for needle insertion, are only a few millimeters in diameter. The presence of arteries and veins surrounding the

The associate editor coordinating the review of this manuscript and approving it for publication was Jingang Jiang ⓘ .

renal pelvis and major calyces in the renal hilum makes any puncture outside the collecting system a potential cause of severe bleeding [5]. Since its introduction in 1976, PCNL in the prone position has been the gold standard for the treatment of large stones [6]. The prone positioning was favored due to the posterior location of the kidneys, the short percutaneous tract to the calyces located at the avascular line of Brodel, and a large puncture area [7]. These advantages are, however, accompanied by the need to reposition the patient after the ureteral catheter placement and complicated anesthesiologic management. As an alternative to prone, supine positioning can overcome the abovementioned challenges with the tradeoff of longer percutaneous tract lengths and potentially increased kidney mobility [8]. Supine patient positioning has been rapidly gaining popularity among physicians, with 20% of procedures done in supine positioning worldwide [9]. There is, therefore, a clinical need for safe and reliable renal access.

There is currently no means to plan a needle insertion with guaranteed success. Physicians primarily rely on fluoroscopy and ultrasound (US) to intra-operatively visualize the anatomy and the percutaneous needle during insertion. Fluoroscopy is the most widely used imaging modality due to the high resolution of the resulting images and low dose exposure [10]. However, both modalities are limited to two-dimensional space, forcing the operator to rely on their professional experience for understanding the true anatomy from the 2D fluoroscopic or US view. To overcome this limitation, physicians who utilize fluoroscopic imaging will often obtain multiple fluoroscopy images or videos at different C-arm angles, which enables confirmation of the alignment of the target calyx and needle insertion trajectory [11]. Although needle alignment at the insertion start is paramount, it does not guarantee overall success, as the needle trajectory needs to be actively corrected to compensate for anatomical motion.

Medical robots are likely to be the solution for improving the accuracy of PCNL procedures. For example, robots could simultaneously record the data from the fluoroscope, track the location of the needle and other instruments, perform automated insertions, etc. There have been several attempts to integrate robots into PCNL procedures [12], [13], [14], [15], [16], [17], [18], [19], [20]. Early studies used kidney phantoms as the targets for needle insertion as phantoms allow precise measurement of the insertion accuracy [13], [14], [16], [18]. At the same time, phantoms are a poor substitute for the human abdominal anatomy. Cadavers and live pigs were utilized as more realistic models [12], [13], [15], [16], [17], [20]. The insertion accuracy was then measured qualitatively as the ratio of successful access to porcine kidneys in one or three needle insertion attempts [15] and quantitatively using three-dimensional computed tomography images acquired at the end of the insertion and fiducial markers placed inside porcine subjects [12], [16], [20]. Two main directions for integrating robots into PCNL procedures, namely remotely controlled and fully automated needle insertions, have been investi-

gated. In both scenarios, the user first identifies the starting point on the subject surface and positions the needle at that point. The needle could then be inserted using the remote controller [12], [13], [16]. Such an approach increases the stability of the insertion tract in contrast to manual insertion and removes the user from the insertion site to protecte them from radiation exposure during intra-operative imaging. Alternatively, a robot could automatically insert the needle after the user sets the starting point and insertion direction are set by the user. The main challenge of such an approach is to ensure that the insertion trajectory stays in agreement with the anatomy, so the researchers develop algorithms for breathing motion compensation via image-based target tracking [14], [15], [17], [18], [20]. Due to significant technical and financial requirements to build robots and conduct robotic experiments, the field of robotic percutaneous procedures remains underinvestigated. The existing studies utilize one phantom or 1-2 pigs. To the best of our knowledge, no use of human cadavers or human patients has been reported. Finally, the studies focus on the assessment of the robotic assistance, whereas data from the robots could be used for other purposes like the assessment of users' skills or treatment outcome prediction.

This study proposes an automated solution for analyzing procedure metrics and quantifying expert performance while establishing renal access in robotic PCNL using a novel surgical robot – Monarch®Platform, Urology (MONARCH). A deep learning-based framework was developed to automatically recognize needle insertion attempts during renal access using the sensor data recorded by the MONARCH. Quantitative features that characterize the user's performance, such as the quality of needle alignment and targeting error, were extracted from automatically recognized insertions and compared to the manually labeled needle insertions.

## II. METHODOLOGY
### A. DATA COLLECTION
Data were obtained from prior recordings of a consecutive series of percutaneous access attempts on a human cadaver model using the MONARCH robotic system. The MONARCH system includes a flexible ureteroscope steered by the user with a video-game-style controller and a manually inserted percutaneous needle. Both the ureteroscope and the needle are equipped with electromagnetic (EM) sensors that report the tip position and orientation of the instruments. The ureteroscope is equipped with an integrated camera that allows the user to navigate inside the renal collecting system under direct vision.

During the simulated procedures, the cadaver was positioned in a modified supine position on the operating table. Saline irrigation was used to distend the kidney for visualization of the anatomy. The amount of fluid used mimicked the ranges used in standard ureteroscopies. The ureteroscope was inserted into the subject's urethra through the bladder, and up the ureter into the kidney collecting system and then docked

on the robotic arms of the MONARCH system. Once the user drove the ureteroscope inside the renal pelvis, they selected a desired calyx for the needle insertion and tagged the papilla using the controller. The user then retracted the ureteroscope so that the papilla of the target calyx was positioned in the center of the ureteroscope camera view (Fig. 1). This location was also tagged using the controller, indicating to the system that the target had been selected.

The user then selected an insertion site on the skin and manually inserted the needle. Selection of the insertion site is purely based on the user's clinical understanding of the optimal trajectory for the needle insertion and their consideration of anatomical structures adjacent to the kidney (such as lung, liver, spleen and colon). The robotic system does not impose any restrictions on the insertion site.

Once the insertion site is selected, the MONARCH system provides visual guidance to the user on how to best align the needle with the target using the guided percutaneous access interface (GPAI). The GPAI provides real-time needle insertion guidance by aligning the needle such that the targeting dot is inside the targeting reticle (seen in Fig. 1). The needle insertion step starts when the aligned needle punctures the subject's skin. The needle insertion step ends when the depth indicator (Fig. 1) turns green, signaling that the needle has reached the target, or turns red, signaling that the insertion should be stopped as the target is missed, and further penetration may result in surrounding tissue damage (Fig. 2). Summarizing, the MONARCH system provided hardware that recorded the relative locations and orientations of the ureteroscope and needle sensors, and visual interface that helps to align the sensors during the insertion. The user performed the needle movements and ureteroscope driving.

Within our analysis, we defined any pauses during needle insertion to still count towards one insertion attempt. Needle retractions before the depth indicator turns green or red do not constitute a new attempt but are considered needle adjustments. Additionally, a new needle insertion attempt does not require the user to completely extract the needle from the cadaveric abdominal tissue. Figure 3 shows a timeframe for an example percutaneous renal access procedure.

## B. NEEDLE INSERTION FEATURES

The MONARCH robotic system records in real-time the location and orientation of the ureteroscope and the needle and the controller commands. At each time point, the ureteroscope is characterized by a vector of positional coordinates $x_{scope}$, and quaternion $q_{scope}$ defined with respect to the electromagnetic field generator (FG). Similarly, the needle at each timepoint is characterized by a vector of positional coordinates $x_{needle}$ and quaternion $q_{needle}$.

Using positions and orientations of the ureteroscope and needle, a number of features were computed for every timestamp of a percutaneous access procedure. First, we computed the relative positions and orientation of the needle against the ureteroscope in 3D. We also computed the linear velocities of the needle and the ureteroscope by fitting local regression

models with Newton's equation of motion as the model function [21]. The linear velocity, for example, can be indicative of ureteroscope retraction and needle insertion.

Several features are computed to capture and quantify the needle alignment. The target alignment error $f_{target}$ estimates if the needle is oriented towards the target, i.e., the 0° angle indicates the perfect alignment. This angle significantly fluctuates at the time of the initial needle alignment. The relative angle between the ureteroscope and the needle is another feature.

Finally, we recorded the depth indicator values of the GPAI and the distance from the needle to the FG. The described-above features characterize the state of the system at every timestamp.

The subject's position on the surgery bed is predefined according to the modified supine PCNL protocol. The MONARCH cart also has a predefined location against the bed. Therefore, the coordinates and orientation of the sensors can be computed with respect to the robotic system and, consequently, the subject. The first angle of interest is the cranial-caudal that measures how a sensor is oriented against the axis passing from the subject's legs to head (Fig. 5a). The second angle is anterior-posterior that measures how a sensor is oriented up or down (Fig. 5b). Similarly, we can represent the ureteroscope orientation with respect to the bed to estimate whether a calyx is anterior or posterior facing as shown in Fig.5d.

## C. U-NET FOR NEEDLE INSERTION DETECTION

A percutaneous renal access procedure starts with the target selection step and ends with either a new target selection or the user manual terminating the procedure with the GPAI. A procedure's features define a 2D array $A$ of size N × M, where N corresponds to the number of timestamps, and M is the number of features per timestamp. The information decoded in these features can be characteristic of specific user activities, including site selection for needle insertion, needle insertion, needle retraction, and secondary insertion. In this study, we are interested in two types of activities: the needle insertion activity, which encompasses both first and secondary insertions, and the no-insertion activity, which covers all the remaining activities. Each timestamp is labeled to belong to insertion/no-insertion resulting in a 1D array $L$ of labels of size N.

The aim is to automatically label percutaneous renal access procedures by learning from a training collection of arrays $A = \{A\}$ and corresponding labels $L = \{L\}$. We implemented a one-dimensional U-Net neural network for automated labeling of $A$ [22]. The encoder part of the U-Net consisted of eight blocks with the feature map sizes of [16, 32, 64, 64, 128, 128, 256, 256], where each block was a sequence of group norm [23], convolution, and ReLU layers. The decoder part of the U-Net has the inverted feature map sizes [256, 128, 128, 64, 64, 32, 16] and the same block composition. All convolution layers have [3 × 1] kernels.

**FIGURE 1.** The guided percutaneous access interface (GPAI) of the Monarch Platform Urology (Auris Health Inc., CA). The interface consists of the ureteroscope camera view (1), target and targeting dot (2) that encodes the alignment between the needle for percutaneous insertion and the target inside the kidney, and depth indicator (3) that turns green when the target is reached or red when the target is missed.
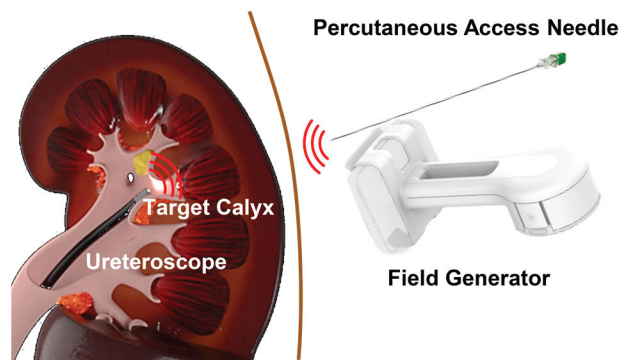


**FIGURE 2.** A schematic illustration of a percutaneous needle access procedure equipment. The flexible ureteroscope equipped with a camera is used to mark the target calyx for the needle insertion. Both ureteroscope and needle are equipped with sensors, which are tracked by the field generator located near the procedure site.

The network loss function is a linear combination of binary cross-entropy and Dice losses weighted with 0.1 and 0.9 coefficients, respectively. The training used an Adam optimizer with an initial learning rate of 0.001 and L2 regularization on the convolution kernels with $\lambda = 0.01$. The network was trained for 100 epochs. The input to the network first passes through average pooling with the kernel $[5 \times 1]$, while the output is upscaled to return to the original input size. This initial downscaling is needed to compensate for feature fluctuations and facilitate the U-Net convergence. The output of the U-Net is a probability vector of length $T$ indicating how likely each timepoint belongs to an insertion.

### D. RANDOM FORESTS FOR PERCUTANEOUS ATTEMPT CLASSIFICATION

The U-Net labels each time point to represent insertion or no-insertion. Such labeling output is insufficient to achieve the overarching aim of estimating the clinical properties of

the insertions and computing the number of needle insertion attempts per procedure. In other words, we need to identify the time points of the skin puncture and end of a needle insertion attempt from the results of U-Net predictions.

Figure 6 illustrates an example where simply thresholding the classification probabilities assigned by U-Net is insufficient. In this example, the user performs an incomplete insertion, then extracts the needle from the subject's body and starts a new insertion attempt. This attempt is then completed. To compute characteristics of the insertion, such as tract length (Fig. 5c), insertion angles, and trajectory deviation, incomplete insertion attempts need to be filtered out (Fig. 6). To classify the result of the U-Net and filter out such scenarios, we first generate candidate insertion regions by thresholding U-Net probabilities and extracting connected regions. Each extracted region will be referred to as a candidate region to be classified as a percutaneous attempt. In Figure 5c, we have three candidate regions where needle insertion activity is taking place.

For each candidate region, we calculate six features. One feature is the duration of the candidate region. Two features are the average target and positional targeting accuracy of the targeting dot on the GPAI, which we expect to be very stable in the middle of an insertion attempt. The distance traveled by the needle and the minimal distance to the target were also included in the feature list. Finally, a binary feature checks if the needle, on average, moves closer to or further away from the target to capture the directionality of the movement.

The regions are classified using an implementation of a random forest classifier [24] from sklearn with standard parameters and the maximum tree depth of 9 splits before a forced leaf. We also augmented the random forest classification with several explicit rules, namely a) the insertion cannot be shorter than $r = 5$ seconds, b) two regions labeled by the
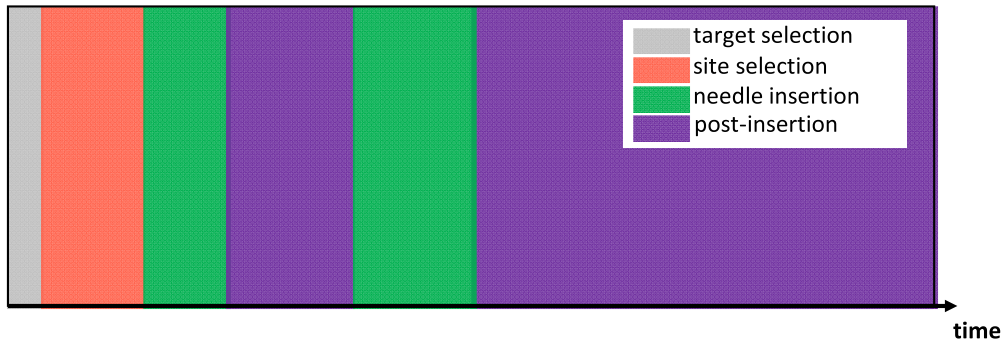
**FIGURE 3.** A schematic illustration of a percutaneous needle access procedure period. The procedure starts from the target selection phase, where the user drives the flexible ureteroscope of the Monarch Platform to select the target, and then retracts the ureteroscope offering a clear view to the selected target location. The next phase is site selection, where the user places the needle to the subject's skin and finds a suitable puncture location while aligning the needle with the target. The next phase is the first needle insertion attempt that goes from the moment when the needle penetrates the skin to the moment the needle reaches or misses the target. Secondary insertions could be executed after the first one if target is missed. Each insertion is followed by a post-insertion phase where the user may perform additional clinical steps such as checking for backflow or inserting a guidewire to the renal collecting system.
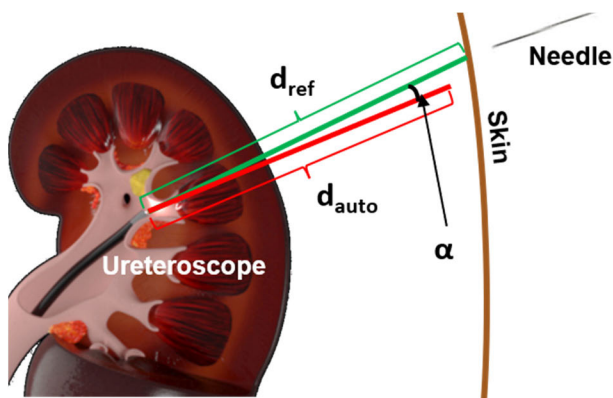


**FIGURE 4.** An example of manually and automatically detected need insertions and the corresponding insertions lengths $d_{ref}$ and $d_{auto}$ and angle between them $\alpha$.

U-Net as insertion that are closer than $p = 2$ seconds from each other are merged into one region, c) the minimal needle track cannot be shorter than $m = 10$ mm, d) the regions are artificially extended to ensure the needle came closer than $k = 5$ mm to the target. These rules are descriptive for both first and secondary insertions. The combination of the U-Net-based timepoint labeling and random forest region classification resulted in the fully automated calculation of needle insertion characteristics from the MONARCH-assisted percutaneous access.

## III. EXPERIMENTS AND RESULTS

### A. DATABASE

The database consisted of recordings from 248 consecutive percutaneous renal access procedures performed on 18 cadavers. The cadavers were both male and female subjects with a body mass index ranging from 19.8 to 39.2. The procedures were conducted by 17 urologists and one urologist proxy. The average experience level of the participants was 3.2, computed on a scale from 1 to 5. A scale of 1 corresponds to a novice user who has not gained percutaneous access in

their practice, and five corresponds to expert urologists who performed hundreds of percutaneous renal access procedures. Only three participants had experience with supine patient positioning, while the majority have been only performing percutaneous access in prone positioning.

Important to note is that within this dataset, not every procedure included a needle insertion attempt nor ended after successful insertion. If a user selected the same or a new target without inserting the needle, the time frame from the previous target selection to the new target selection still constitutes a procedure and was analyzed for potential insertion attempts. The database contains 42 procedures without insertions. At the same time, the user could finish the procedure by abandoning the instruments and the controller without explicitly sending a signal of a procedure end.

The recordings came from more procedures on the right kidneys (57%) than on the left kidneys (43%). For 171 procedures, the experts explicitly determined the target calyx as upper (36%), mid (31%), and lower (33%) pole calyx.

### B. RESULTS

The procedures were analyzed using a needle insertion emulator that displays the 3D location and orientation of the instruments. The engineers inspected the procedure logs to manually label the individual needle insertions by marking their start and end time points. The accuracy of automated labeling for the first insertion in each procedure was 0.81, measured in terms of the Dice coefficient compared against the manual labeling. The framework captured 94% of all first insertions, for which the Dice coefficient was 0.86. The first insertions were used to compute the procedure metrics, which can be subdivided into the site selection metrics and needle insertion metrics. The site selection metrics are determined at moment insertion location is determined, and the insertion starts and includes a) tract length (Fig 5b) relative needle to ureteroscope angle, c) anatomical insertion angle, and d) target orientation (Fig 5). The procedure insertion metrics
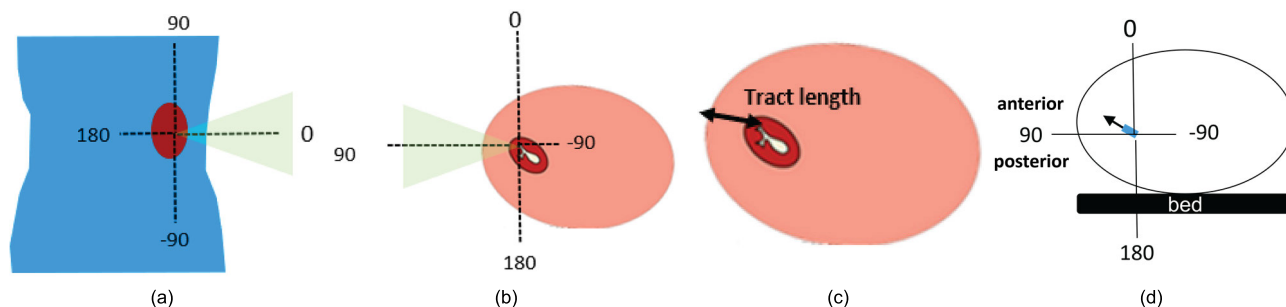
**FIGURE 5.** Illustrations of geometrical features used in the study with the reference to the patient's anatomy. The pink and blue objects correspond to the axial and coronal patient's cross-section, respectively. Red object is a schematic kidney illustration. (a) Caudocranial angle illustration. (b) The anterior-lateral-posterior angle illustration. (c) Tract length for insertion from the patient's skin to the target inside the kidney. (d) The target anterior-lateral-posterior angle illustration.
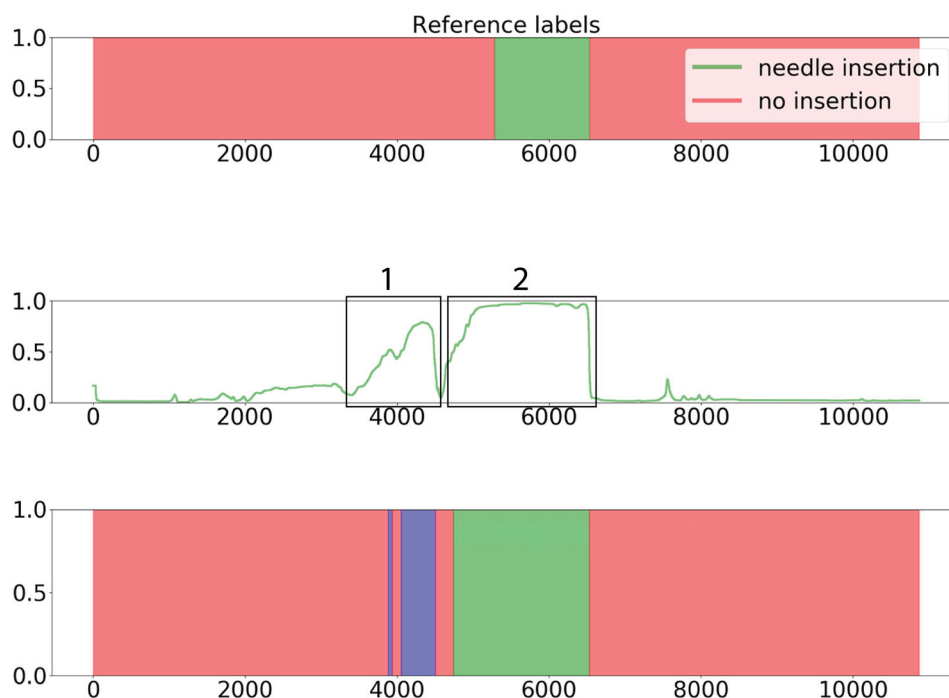


**FIGURE 6.** This example illustrates a common scenario when the user starts (1) inserting the needle but before reaching the target reconsiders and entirely extracts the needle from the subject's body. Then, the user starts (2) a new insertion and completes it following the indication from the guided percutaneous access interface (GUAI) of the Monarch Platform. The initial aborted insertion attempt can be mis-labeled by the U-Net as a needle insertion. The U-Net results are corrected by random-forest-based classification that among other conditions check whether a region labeled by the U-Net as an insertion actually reach the target plane.

include a) minimal distance to the target, b) insertion duration, c) mean positional and angular need deviation during insertion, d) mean and maximal ureteroscope displacement during insertion, i.e., a proxy for anatomical motion. The complete list of metrics and the metric comparison for manually and automatically labeled insertion performed in terms of average ($\pm$ standard deviation), median and 25th and 75th quantiles of the differences are given in Table 1. Outliers in Table 1 are computed using three standard deviations from the average values for each feature. After removing outliers, the average, standard deviation, and quantile values are recomputed to avoid significant skew of the standard deviation values.

We do not compute procedure metrics from secondary insertions but only record their presence. Indeed, the site

selection metrics cannot be estimated from the secondary insertions, while some other metrics, such as secondary insertion duration and angular deviations of the sensors, are not clinically descriptive. A secondary insertion is considered correctly recognized if there is a $\geq 50\%$ overlap between manual and automated labels. In total, 70 secondary needle insertions were distributed among 42 procedures. The recognition accuracy for the secondary insertions was 66%, i.e., 46 of 70 insertions. The recognition accuracy for the procedures that have secondary insertions without necessarily correctly capturing each insertion was 79%, i.e., 33 of 42 procedures.

## IV. DISCUSSION
This paper proposes a novel way of integrating machine learning into percutaneous renal access assessment. In recent

**TABLE 1.** The results of the proposed framework for the automated labeling of needle insertions during percutaneous renal access procedures. The results present the error estimation of quantitative metrics derived from manually and automatically labeled needle insertions. The features are calculated only for the first insertion without considering potential insertion re-attempts. The errors outside the 1.5 interquartile range are considered outliers. The outliers are removed from the reference value range.

| Metric | Units | Reference value range | Errors | | | | |
|--------|-------|-----------------------|--------------------|----------------|--------|----------------|-----------|
| | | | Mean ± st. dev. | 25% quantile | Median | 75% quantile | # outliers |
| Minimal distance from needle to target | mm | [0.06; 7.08] | 2.86 ± 15.18 | 0 | 0 | 0.51 | 5 |
| Angle between needle and ureteroscope orientations | ° | [1.07; 99.72] | 2.30 ± 6.12 | 0.28 | 0.71 | 2.27 | 0 |
| Anatomical insertion angle (anterior-lateral-posterior) | ° | [57.28; 123.8] | 2.19 ± 6.09 | 0.27 | 0.67 | 1.66 | 2 |
| Anatomical insertion angle (caudocranial) | ° | [-41.75; 49.76] | 1.79 ± 3.24 | 0.25 | 0.66 | 1.80 | 0 |
| Target orientation (anterior-lateral-posterior) | ° | [14.94; 172.1] | 0.30 ± 0.85 | 0.02 | 0.07 | 0.17 | 3 |
| Target orientation (caudocranial) | ° | [-95.54; 97.77] | 0.37 ± 1.01 | 0.03 | 0.08 | 0.20 | 1 |
| Insertion tract length | mm | [49.92; 177.6] | 15.86 ± 17.7 | 2.79 | 9.41 | 24.03 | 2 |
| Insertion duration | sec | [12.38; 123.5] | 11.48 ± 14.33 | 2.80 | 7.16 | 14.10 | 3 |
| Mean insertion trajectory deviation | mm | [0.44; 3.34] | 0.78 ± 2.74 | 0.04 | 0.12 | 0.48 | 3 |
| Mean insertion angular deviation | ° | [0.68; 8.32] | 2.21 ± 6.14 | 0.10 | 0.38 | 2.18 | 1 |
| Mean target displacement during insertion | mm | [0.30; 3.88] | 0.47 ± 1.14 | 0.02 | 0.10 | 0.30 | 3 |
| Relative motion between ureteroscope and needle during insertion | ° | [1.44; 88.94] | 8.60 ± 26.45 | 0.44 | 1.32 | 3.64 | 4 |

years, machine learning has dramatically expanded its applicability and was eventually utilized in the prediction of outcomes of staghorn calculi [25]. The researchers were focused on predicting both acute outcomes such as the stone-free status [25], [26], [27], [28], laboratory measurement [26], fever [29], need for transfusion [26], [27] or stenting [27], and late outcomes such as second PCNLs [26] or urosepsis [29]. The studies collected several hundred demographics, clinical and laboratory parameters for each patient who underwent PCNL, applied dimensionality reduction algorithms to select the most information-rich features, and then trained logistic regression [25], k-nearest neighbors [27], random forests [25], support vector machines [25], [27], [28] and neural networks [26] on the resulting features to predict PCNL outcomes. Such models exhibit not only predictive powers but also pinpoint the parameters that are most contributive to the success/failure of PCNL. The authors of such studies concluded that the number of needle insertions needed, BMI, stone size, access calyx pole, kidney comorbidities, procedure duration, and supracostal insertions exhibit the strongest correlation with PCNL outcomes. Except for the stone size and kidney comorbidities, the remaining features are directly associated with or could be estimated during the needle insertion.

The number of needle insertions is predictive of stone-free status for patients with stag horn calculi [25], post-treatment urosepsis [29], [30], and severe hemorrhages [30]. In our study, 79% of all procedures requiring secondary insertions and 66% of individual secondary needle insertion attempts were automatically recognized by the presented framework. Several issues cause the misdetection of secondary insertions. First, some secondary insertions have internal idle states when the user does not move the needle. For such inser-

tions, the framework might capture the insertion but mislabels some of its idle states and therefore does not achieve the required 50%-level overlap with the manual label. Some urologists employ the staccato motion during renal access [31], where the needle is not inserted in one continuous motion but through many small advances and relaxations. During such a motion, the needle could rapidly surpass the target, be retracted, and then continue its movement toward the target. The insertion, therefore, could be erroneously subdivided into two insertion attempts by the labeling scheme. In general, needle insertions rarely consist of monotonous movement toward the target and include small retractions due to tissue resistance, adjustments, and pauses, which makes the recognition of secondary insertions ambiguous even for a human observer. For example, how significant does a needle retraction need to be to constitute an insertion in contrast to needle adjustment? There is no ultimate answer to this question, but empirically, we observed that secondary insertions have an insertion depth of >15mm. To put this threshold into perspective, the mean tract length for the first insertions is 104 mm.

Another key observation is that the U-Net accurately captures most of the needle insertion activity. However, the reference labels only include insertion activity from the start of skin puncture to the end of needle insertion, while incomplete attempts are not analyzed. The incomplete insertion attempts do not serve the end goal of segmenting data from the start of a skin puncture to the insertion end and computing procedure metrics. The incomplete insertion attempts negatively affect the Dice score but do not necessarily compromise the overall framework performance (Fig. 6).

A meta-analysis of the efficacy and safety of percutaneous nephrolithotomy against patients' BMI factor discovered that

the procedure is significantly longer for morbidly obese and obese patients in comparison to patients with normal BMI and overweight patients [32]. Machine learning analysis found BMI to predict stone-free status for patients with stag horn calculi [25] and the need for stenting to manage post-treatment urine leakage [26]. For the subjects with high BMI, the needle needs to travel a longer distance, which correlates with the tract length feature computed in our analysis. The median tract length error was 9.22 mm, which approximately corresponds to a 9% error of the mean tract length of 99.4 mm. The error is, however, skewed due to several large outliers. The median error of 9.22 mm is 1.5 times smaller than the mean error. The errors occur due to the challenges with precise estimation of the skin puncture time. The skin is thick and may be difficult to penetrate, which will require a number of pushes. After penetration, the needle could drift away from the target due to initial misalignment, so the user will decide to pull the needle and realign. In contrast to the tract length, the insertion angles remain stable during the significant part of the insertion, and they are estimated with an error of 1.64° and have very few outliers. The tract length and insertion angles estimate the distance to the kidney from a specific anatomical region on the skin and can approximate the subject's constitution.

Procedure duration is predictive for post-treatment urosepsis and [29], [30] and severe hemorrhages [30]. For outcome prediction, the PCNLs are classified into long, which take more than 90 minutes [30] or 120 minutes [29], and short. The short PCNLs result in more favorable outcomes. The procedure duration is an umbrella characteristic that offers a rough separation between easy and challenging PCNL cases. It must be acknowledged that percutaneous renal access is only a small part of the PCNL, and procedure duration may be associated with the presence of multiple stones, the need for multiple insertions, or other stone removal challenges. Although the proposed framework only targets the access part of the PCNL, the presented methodological and robotic concepts can be extended for recording and automatic extraction of other components that contribute to challenging PCNLs.

The quality of an individual needle insertion attempt has remained non-trivial to analyze without robotic tools and sensors [33]. The precision of needle insertion is encoded in the positional and angular deviation during insertion. These metrics are automatically captured with low errors of 0.49 mm and 1.38° (Table 1). The main source of errors in insertion metric calculation comes from detection outliers, i.e., a relatively few cases when insertions were significantly mislabeled. The main reason for outliers comes from incomplete insertions, when the user starts the insertions, gets dissatisfied, partially retracts the needle, and then continues the insertion. If a complete insertion happens immediately after an incomplete insertion, the algorithm can erroneously stitch these two insertions into one, significantly compromising metric calculation. The future steps will be to improve the recognition and rejection of incomplete insertion attempts. This research opens an avenue toward automated quantitative analysis of percutaneous needle insertions that is critical for evaluating and adopting robotic assistance in percutaneous nephrolithotomy. Moreover, automatic insertion recognition is a prerequisite for understanding which insertion features are essential for successful renal access.

## V. CONCLUSION

This work presents the first attempt at using robotic assistance to record instruments during PCNL and applying deep learning to the automatic labeling of needle insertion activities. The work provides methodology and apparatus for the automated calculation of clinical metrics and characteristic features of user performance during percutaneous renal access.

## REFERENCES

[1] C. D. Scales, A. C. Smith, J. M. Hanley, and C. S. Saigal, "Prevalence of kidney stones in the United States," *Eur. Urol.*, vol. 62, no. 1, pp. 160–165, Jul. 2012, doi: 10.1016/j.eururo.2012.03.052.

[2] H. S. Mirheydar, K. L. Palazzi, I. H. Derweesh, D. C. Chang, and R. L. Sur, "Percutaneous nephrolithotomy use is increasing in the United States: An analysis of trends and complications," *J. Endourol.*, vol. 27, no. 8, pp. 979–983, Aug. 2013, doi: 10.1089/end.2013.0104.

[3] A. Srivastava, K. J. Singh, A. Suri, D. Dubey, A. Kumar, R. Kapoor, A. Mandhani, and S. Jain, "Vascular complications after percutaneous nephrolithotomy: Are there any predictive factors?" *Urology*, vol. 66, no. 1, pp. 38–40, Jul. 2005, doi: 10.1016/j.urology.2005.02.010.

[4] A. Smith, T. D. Averch, K. Shahrour, D. Opondo, F. P. J. Daels, G. Labate, B. Turna, and J. J. M. C. H. de la Rosette, "A nephrolithometric nomogram to predict treatment success of percutaneous nephrolithotomy," *J. Urol.*, vol. 190, no. 1, pp. 149–156, Jul. 2013, doi: 10.1016/j.juro.2013.01.047.

[5] H. Y. Kim, K. W. Lee, and D. S. Lee, "Critical causes in severe bleeding requiring angioembolization after percutaneous nephrolithotomy," *BMC Urol.*, vol. 20, no. 1, p. 22, Mar. 2020, doi: 10.1186/s12894-020-00594-6.

[6] I. Fernström and B. Johansson, "Percutaneous pyelolithotomy: A new extraction technique," *Scandin. J. Urology Nephrol.*, vol. 10, no. 3, pp. 257–259, Jan. 1976, doi: 10.1080/21681805.1976.11882084.

[7] G. R. Sharma, P. N. Maheshwari, A. G. Sharma, R. P. Maheshwari, R. S. Heda, and S. P. Maheshwari, "Fluoroscopy guided percutaneous renal access in prone position," *World J. Clin. Cases WJCC*, vol. 3, no. 3, pp. 245–264, Mar. 2015, doi: 10.12998/wjcc.v3.i3.245.

[8] D. Yuan, Y. Liu, H. Rao, T. Cheng, Z. Sun, Y. Wang, J. Liu, W. Chen, W. Zhong, and J. Zhu, "Supine versus prone position in percutaneous nephrolithotomy for kidney calculi: A meta-analysis," *J. Endourol.*, vol. 30, no. 7, pp. 754–763, Jul. 2016, doi: 10.1089/end.2015.0402.

[9] S. Proietti, M. E. Rodríguez-Socarrás, B. Eisner, V. De Coninck, M. Sofer, G. Saitta, M. Rodriguez-Monsalve, C. D'Orta, P. Bellinzoni, F. Gaboardi, and G. Giusti, "Supine percutaneous nephrolithotomy: Tips and tricks," *Transl. Andrology Urol.*, vol. 8, no. 4, pp. 381–388, Sep. 2019, doi: 10.21037/tau.2019.07.09.

[10] I. Kyriazis, E. Liatsikos, O. Sopilidis, P. Kallidonis, and A. Skolarikos, "European section of urotechnology educational video on fluoroscopic-guided puncture in percutaneous nephrolithotomy: All techniques step by step," *BJU Int.*, vol. 120, no. 5, pp. 739–741, Nov. 2017, doi: 10.1111/bju.13894.

[11] G. Sharma and A. Sharma, "Determining site of skin puncture for percutaneous renal access using fluoroscopy-guided triangulation technique," *J. Endourol.*, vol. 23, no. 2, pp. 193–195, Feb. 2009, doi: 10.1089/end.2008.0170.

[12] T. de Baere, C. Roux, G. Noel, A. Delpla, F. Deschamps, E. Varin, and L. Tselikas, "Robotic assistance for percutaneous needle insertion in the kidney: Preclinical proof on a swine animal model," *Eur. Radiol. Experim.*, vol. 6, no. 1, p. 13, Mar. 2022, doi: 10.1186/s41747-022-00265-1.

[13] T. Komaki, T. Hiraki, T. Kamegawa, T. Matsuno, J. Sakurai, R. Matsuura, T. Yamaguchi, T. Sasaki, T. Mitsuhashi, S. Okamoto, M. Uka, Y. Matsui, T. Iguchi, H. Gobara, and S. Kanazawa, "Robotic CT-guided out-of-plane needle insertion: Comparison of angle accuracy with manual insertion in phantom and measurement of distance accuracy in animals," *Eur. Radiol.*, vol. 30, no. 3, pp. 1342–1349, Mar. 2020, doi: 10.1007/s00330-019-06477-1.

[14] H.-Y. Li, I. Paranawithana, Z. H. Chau, L. Yang, T. S. K. Lim, S. Foong, F. C. Ng, and U.-X. Tan, "Towards to a robotic assisted system for percutaneous nephrolithotomy," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 791–797, doi: 10.1109/IROS.2018.8593689.

[15] E. Ben-David, M. Shochat, I. Roth, I. Nissenbaum, J. Sosna, and S. N. Goldberg, "Evaluation of a CT-guided robotic system for precise percutaneous needle insertion," *J. Vascular Interventional Radiol.*, vol. 29, no. 10, pp. 1440–1446, Oct. 2018, doi: 10.1016/j.jvir.2018.01.002.

[16] T. Hiraki, T. Kamegawa, T. Matsuno, J. Sakurai, Y. Kirita, R. Matsuura, T. Yamaguchi, T. Sasaki, T. Mitsuhashi, T. Komaki, Y. Masaoka, Y. Matsui, H. Fujiwara, T. Iguchi, H. Gobara, and S. Kanazawa, "Robotically driven CT-guided needle insertion: Preliminary results in phantom and animal experiments," *Radiology*, vol. 285, no. 2, pp. 454–461, Nov. 2017, doi: 10.1148/radiol.2017162856.

[17] D. Zhang, Z. Li, K. Chen, J. Xiong, X. Zhang, and L. Wang, "An optical tracker based robot registration and servoing method for ultrasound guided percutaneous renal access," *Biomed. Eng. OnLine*, vol. 12, no. 1, p. 47, 2013, doi: 10.1186/1475-925X-12-47.

[18] B. Challacombe, A. Patriciu, J. Glass, M. Aron, T. Jarrett, F. Kim, P. Pinto, D. Stoianovici, N. Smeeton, R. Tiptaft, L. Kavoussi, and P. Dasgupta, "A randomized controlled trial of human versus robotic and telerobotic access to the kidney as the first step in percutaneous nephrolithotomy," *Comput. Aided Surg.*, vol. 10, no. 3, pp. 165–171, Jan. 2005, doi: 10.3109/10929080500229561.

[19] L.-M. Su, D. Stoianovici, T. W. Jarrett, A. Patriciu, W. W. Roberts, J. A. Cadeddu, S. Ramakumar, S. B. Solomon, and L. R. Kavoussi, "Robotic percutaneous access to the kidney: Comparison with standard manual access," *J. Endourol.*, vol. 16, no. 7, pp. 471–475, Sep. 2002, doi: 10.1089/089277902760367421.

[20] A. Bzostek, S. Schreiner, A. C. Barnes, J. A. Cadeddu, W. W. Roberts, J. H. Anderson, R. H. Taylor, and L. Kavoussi, "An automated system for precise percutaneous access of the renal collecting system," in *CVRMed-MRCAS* (Lecture Notes in Computer Science), J. Troccaz, E. Grimson, and R. Mosges, Eds. Berlin, Germany: Springer, 1997, pp. 299–308, doi: 10.1007/BFb0029249.

[21] F. Sittel, J. Müller, and W. Burgard. (2013). *Computing Velocities and Accelerations from a Pose Time Sequence in Three-dimensional Space*. Accessed: Mar. 30, 2022. [Online]. Available: https://www.semanticscholar.org/paper/Computing-Velocities-and-Accelerations-from-a-Pose-Sittel-M

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (Lecture Notes in Computer Science). N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: SpringerInternational Publishing, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[23] Y. Wu and K. He, "Group normalization," 2018, *arXiv:1803.08494*.

[24] A. Criminisi and J. Shotton, *Decision Forests for Computer Vision and Medical Image Analysis*. London, U.K.: Springer, 2013.

[25] B. M. Z. Hameed, M. Shah, N. Naik, H. S. Khanuja, R. Paul, and B. K. Somani, "Application of artificial intelligence-based classifiers to predict the outcome measures and stone-free status following percutaneous nephrolithotomy for staghorn calculi: Cross-validation of data and estimation of accuracy," *J. Endourol.*, vol. 35, no. 9, pp. 1307–1313, Sep. 2021, doi: 10.1089/end.2020.1136.

[26] A. Aminsharifi, D. Irani, S. Pooyesh, H. Parvin, S. Dehghani, K. Yousofi, E. Fazel, and F. Zibaie, "Artificial neural network system to predict the postoperative outcome of percutaneous nephrolithotomy," *J. Endourol.*, vol. 31, no. 5, pp. 461–467, May 2017, doi: 10.1089/end.2016.0791.

[27] T. Shabaniyan, H. Parsaei, A. Aminsharifi, M. M. Movahedi, A. T. Jahromi, S. Pouyesh, and H. Parvin, "An artificial intelligence-based clinical decision support system for large kidney stone treatment," *Australas. Phys. Eng. Sci. Med.*, vol. 42, no. 3, pp. 771–779, Sep. 2019, doi: 10.1007/s13246-019-00780-3.

[28] A. Aminsharifi, D. Irani, S. Tayebi, T. J. Kafash, T. Shabanian, and H. Parsaei, "Predicting the postoperative outcome of percutaneous nephrolithotomy with machine learning system: Software validation and comparative analysis with Guy's stone score and the CROES nomogram," *J. Endourol.*, vol. 34, no. 6, pp. 692–699, Jun. 2020, doi: 10.1089/end.2019.0475.

[29] S. S. Bansal, P. W. Pawar, A. S. Sawant, A. S. Tamhankar, S. R. Patil, and G. V. Kasat, "Predictive factors for fever and sepsis following percutaneous nephrolithotomy: A review of 580 patients," *Urol. Ann.*, vol. 9, no. 3, pp. 230–233, 2017, doi: 10.4103/UA.UA_166_16.

[30] Y. Wang, F. Jiang, Y. Wang, Y. Hou, H. Zhang, Q. Chen, N. Xu, Z. Lu, J. Hu, J. Lu, X. Wang, Y. Hao, and C. Wang, "Post-percutaneous nephrolithotomy septic shock and severe hemorrhage: A study of risk factors," *Urologia Internationalis*, vol. 88, no. 3, pp. 307–310, 2012, doi: 10.1159/000336164.

[31] S. Park, M. V. Meng, and M. L. Stoller, "Percutaneous nephrolithotomy," in *Urinary Stone Disease: The Practical Guide to Medical and Surgical Management* (Current Clinical Urology), M. L. Stoller and M. V. Meng, Eds. Totowa, NJ, USA: Humana Press, 2007, pp. 621–638, doi: 10.1007/978-1-59259-972-1_32.

[32] Z. Okhunov, M. Helmy, A. Perez-Lansac, A. Menhadji, P. Bucur, S. B. Kolla, J. S. Cho, K. Osann, A. Lusch, and J. Landman, "Interobserver reliability and reproducibility of STONE nephrolithometry for renal calculi," *J. Endourol.*, vol. 27, no. 10, pp. 1303–1306, Oct. 2013, doi: 10.1089/end.2013.0289.

**BULAT IBRAGIMOV** received the B.S. and M.S. degrees in computer science from Kazan Federal University, Tatarstan, in 2010, and the Ph.D. degree from the University of Ljubljana, in 2014.

From 2016 to 2018, he was a Postdoctoral Fellow with the Department of Radiation Oncology, Stanford University. From 2018 to 2019, he was a Senior Research Scientist with Auris Health (Johnson and Johnson). Currently, he is an Associate Professor with the Department of Computer Science, University of Copenhagen, Denmark. His research interests include machine learning in medicine, computer-aided diagnosis, medical image analysis, and human-AI interaction. He received various awards, including the Novo Nordisk Award for Young Data Science Investigator.
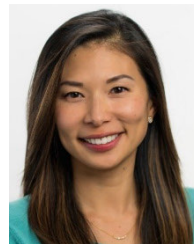
**JANET ZHEN** received the B.S. degree in bioengineering from the University of California San Diego, in 2006, and the M.S. degree in mechanical engineering from the University of California Los Angeles, in 2008.

From 2008 to 2017, she worked in the Aerospace and Defense industry developing computational fluid dynamics models for underwater missile launching systems and supersonic transport vehicles. Driven by a passion for healthcare technology, Janet joined the clinical engineering team at Auris Health in 2017 (later acquired by Johnson & Johnson) to lead the design, development, and validation of the Monarch™ Platform, a groundbreaking surgical robotic system indicated for endourology and bronchoscopy.

**ELIF AYVALI** received the dual B.S. degrees in mechanical engineering and electrical and electronics engineering from Yeditepe University, Istanbul, Turkey, in 2009, and the Ph.D. degree in mechanical engineering from the University of Maryland, College Park, ML, USA, in 2014, focusing on the design and control of shape memory actuated medical robots. She subsequently held a postdoctoral position at the Robotics Institute, Carnegie Mellon University, from 2014 to 2017, where she developed algorithms for complementary situational awareness for robotic surgery. Joining Auris Health, a Johnson and Johnson subsidiary, she led the development of novel capabilities for robot-assisted management of kidney stones. Her primary research interests include the development of robotic systems for safety-critical applications and the verification and validation of machine learning systems.